

250 Conceptos - Marco A. Alzate - U. Distrital F.J.C.

Prefacio

El nivel de abstracción conceptual que distingue a un magíster/doctor de un ingeniero se verifica principalmente en que el ingeniero tiene la habilidad de seleccionar y utilizar el modelo matemático apropiado para un problema dado (diseño), mientras el magíster/doctor es capaz de elaborar el modelo matemático adecuado para un problema novedoso (investigación). Por eso el magíster/doctor necesita una fundamentación matemática mucho más sólida, pues él debe abstraer de una realidad compleja los aspectos más relevantes asociados con un objetivo particular de estudio, y formularlos en el contexto formal y riguroso de las matemáticas para obtener conclusiones que permitan comprender o controlar el sistema bajo estudio. En redes de comunicaciones, estos modelos matemáticos deben relacionar la capacidad de la red y la demanda de los usuarios con las medidas de desempeño que la red ofrece, de manera que resulten útiles en el diseño, el control y el análisis de las redes. En particular, los modelos matemáticos asociados con las redes de comunicaciones deben representar, fundamentalmente, la incertidumbre de los ingenieros respecto a su objeto de estudio. En efecto, si bien podría pensarse que los componentes de la red (equipos y protocolos) tienen un comportamiento determinístico, ellos existen para satisfacer las demandas de los usuarios, las cuales se presentan en cantidades aleatorias y en instantes de tiempo aleatorios¹.

Considere, por ejemplo, una sesión *ftp* (file transfer protocol). La transferencia de un archivo de cierto tamaño (en bytes) toma cierto tiempo (en segundos), de donde se puede calcular la velocidad de transmisión o tasa de transferencia (en bits por segundo, bps). *ftp* ofrece mediciones de estas tres cantidades al presentar resultados como los siguientes (datos consecutivos reales):

Tamaño del archivo (bytes)	Tiempo de transferencia (segundos)	Tasa de transferencia (bps)
1'352.663	5.2	2'081.020
768.023	3.3	1'861.874
3'114.271	11.9	2'093.627
569344	1.8	2'530.418
8'682.209	34.1	2'036.882
2'118.124	8.0	2'118.124
935.936	3.2	2'339.840

Se puede observar cómo no sólo el tamaño de los archivos y el tiempo de transferencia varían de manera imprevisible, sino que el caudal mismo también cambia entre un archivo y otro a pesar de tratarse de la misma conexión usada de manera consecutiva. Por supuesto, esto no debe sorprendernos: diferentes conexiones se establecen con el servidor *ftp* en instantes de tiempo que no

¹ Algunos protocolos (de acceso múltiple, de control de congestión, de seguridad, etc.) hacen uso de números pseudo-aleatorios, de manera que tampoco son determinísticos.

podemos predeterminar; el número de archivos que se transferirán en cada conexión tampoco es predecible con anterioridad; no se puede conocer de antemano la duración de cada archivo; *ftp* ofrece un servicio confiable, de manera que debe detectar y corregir errores, los cuales se presentan de manera impredecible en los medios de transmisión debido al ruido y a la imperfección de los enlaces; etc. Por supuesto, el magíster/doctor en redes de comunicaciones debe ser capaz de incluir dicha incertidumbre dentro de sus modelos matemáticos, para lo cual se usa el modelado probabilístico de los tiempos entre llegadas y las intensidades de las demandas de los usuarios de una red, así como las pérdidas por errores de transmisión, entre otros eventos inciertos.

En estas notas estudiaremos los fundamentos de la teoría de la probabilidad, las variables aleatorias y los procesos estocásticos, los cuales proporcionan herramientas útiles en el análisis de redes de comunicaciones. El enfoque hacia las redes de comunicaciones se debe a que estas notas constituyen los apuntes de clase de los cursos que he dictado en algunos programas de maestría y doctorado de la Universidad Distrital, que suelen incluir muchos ingenieros electrónicos y de sistemas que inician sus estudios de posgrado con énfasis en teleinformática y temas afines. Sin embargo, los conceptos fundamentales que se estudian constituyen abstracciones matemáticas adecuadas para una gran cantidad de realidades en todas las áreas de las ciencias y las ingenierías. Por eso, éste es un libro de matemática y no de redes de comunicaciones o de ingeniería de teletráfico. Pretende ser una introducción general pero formal a la teoría de las probabilidades, variables aleatorias y procesos estocásticos para estudiantes que inician su formación posgradual, ya sea en teleinformática o en cualquier otra disciplina de la ingeniería. La única particularidad es que, en los ejemplos de aplicación, los dados, las monedas y las bolas de billar han sido cambiados por bits, paquetes de datos y enrutadores. De hecho, el principal objetivo de estas notas no es aprender a usar y desarrollar modelos probabilísticos en redes de comunicaciones; el principal objetivo es migrar desde el enfoque pragmático con que se estudian la probabilidad, las variables aleatorias y los procesos estocásticos en el pregrado hacia la formalidad y la rigurosidad que se requiere en un programa de posgrado en ingeniería. El uso de un tipo específico de ejemplos no es tan significativo pues, después de todo, un estudiante de maestría/doctorado debe tener la capacidad de abstracción necesaria para reconocer las posibles interpretaciones de un ejemplo particular: ¿Cuánto se parece lanzar una moneda para ver si cae cara o sello a recibir un bit para ver si es uno o cero? ¿Cuán diferente es el problema de un juez que tiene que determinar la culpabilidad o inocencia de un acusado a partir de las evidencias, y el problema de un módem que debe determinar si se transmitió un uno o un cero a partir de la señal recibida? Es ese nivel de abstracción y de conceptualización el que se espera de un magister/doctor en ingeniería y, en ese sentido, la escogencia particular de los ejemplos de aplicación no debe ser relevante.

Para facilitar el estudio de los temas del libro, se han enumerado los conceptos, ejemplos y ejercicios de manera consecutiva. En efecto, he notado que, ante el amplio contenido de los cursos que podrían basarse en este libro, resulta conveniente considerar pequeñas unidades conceptuales que puedan irse interconectando, de una en una, hasta formar la intrincada malla conceptual de las probabilidades, las variables aleatorias y los procesos estocásticos. En mi experiencia docente, el uso de conceptos breves y precisos ha permitido a los estudiantes ir construyendo la estructura general del conocimiento gracias a que les facilita recordar conceptos precisos de manera rápida en los momentos en que se

necesitan. Espero que este mismo efecto se logre al subdividir de la misma manera el contenido de estas notas. Por supuesto, la construcción del conocimiento siempre será incompleta si no se hacen muchos ejercicios de aplicación de cada uno de los conceptos enumerados. Aunque iremos añadiendo más ejemplos y ejercicios a estas notas, se recomienda al estudiante complementarlos con muchos ejercicios adicionales de la literatura sugerida.

Por último, quisiera agradecer sinceramente a mis estudiantes de los últimos 26 años, a quienes considero coautores de estas notas.

Marco Alzate, Bogotá, junio de 2016

250 Conceptos - Marco A. Alzate - U. Distrital F.J.C.

I. Introducción

1. Modelado Matemático en Ingeniería

*El modelado matemático consiste en **abstraer** de una realidad compleja los aspectos más relevantes asociados con un **objetivo particular** de estudio, y formularlos en el contexto formal y riguroso de las matemáticas. Para ello, se asocia la realidad bajo estudio con un **sistema**, se identifican los **componentes** que intervienen en dicho sistema, se determinan las **variables descriptivas** de esos componentes, se identifican las **reglas de interacción** entre los componentes y se determina los **parámetros descriptivos** de esas interacciones. Ese sistema así formulado constituye un **modelo matemático** de la realidad compleja que se quiere estudiar. Al operar con dicho modelo matemático, se obtendrán conclusiones que podrían remitirse directamente a la realidad bajo estudio. La utilidad de estas conclusiones para comprender o controlar dicha realidad determinará la **validez** del proceso de modelado matemático. Para evaluar el modelo matemático se pueden usar **técnicas analíticas, experimentales o de simulación**. Estas técnicas no son excluyentes sino que, al contrario, se complementan entre ellas.*

La **ingeniería** es la aplicación creativa de **principios científicos** para diseñar, construir y operar estructuras, máquinas o procesos, de manera que se pueda conocer su comportamiento bajo condiciones específicas de operación, todo con respecto a una función precisa y con criterios estrictos de desempeño, economía y seguridad².

Por principios científicos nos referimos al proceso de generación de conocimiento verificable mostrado en la Figura 1, según el cual observamos el mundo natural (realizamos **experimentos**) de manera que, a partir del **análisis** de las observaciones, obtengamos alguna generalización del comportamiento de la naturaleza. Por supuesto, dicha observación debe obedecer a una pregunta significativa que queremos responder empíricamente, de manera que el análisis de los resultados obedezca a un proceso de razonamiento lógico y coherente. Después de haber acumulado suficiente evidencia empírica, se usa un proceso de **inducción** para construir una **teoría** sobre dicho comportamiento, la cual debe ser **verificable**. Esto es, a partir de la teoría podremos **deducir** algunas **hipótesis** que nos permitirían predecir el comportamiento de la naturaleza ante condiciones particulares. Si estas predicciones resultan satisfechas en experimentos de verificación adecuados, podemos ir ganando confianza en la teoría propuesta. Por supuesto, estos experimentos de verificación deben conducirse con suficiente rigor como para que cualquier persona pueda **replicarlos** independientemente y obtener conclusiones semejantes.

² En particular, la ingeniería electrónica aplica los principios físicos del electromagnetismo y la mecánica cuántica para capturar, almacenar, transmitir y procesar información. Las diferentes áreas de actuación de la ingeniería electrónica (dispositivos de estado sólido, circuitos electrónicos, telecomunicaciones, sistemas de radio, sistemas de control, automática, procesamiento de señales, ingeniería de sistemas, ingeniería de computadores, instrumentación, etc.), algunas de las cuales se han constituido en disciplinas ingenieriles independientes, están todas ellas relacionadas íntimamente con las redes de comunicaciones.

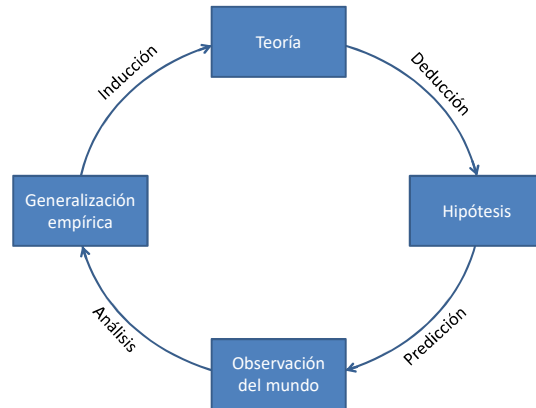


Figura 1. Método científico

Nótese que con este procedimiento se hace imposible demostrar la verdad de una teoría científica: **La ciencia no tiene verdades**, sólo teorías. Si construimos una teoría con una abrumadora evidencia empírica, bastará un solo resultado en contra para desbaratar la teoría. Las teorías científicas son sólo ideas que están evolucionando permanentemente.

En todos los procesos del método científico, desde el planteamiento de la pregunta a resolver y el diseño de experimentos hasta los procesos lógicos de análisis, deducción e inducción, se usa un **modelo de la realidad**. En el caso de los principios científicos que usa la ingeniería, estos modelos son, invariablemente, **modelos matemáticos**. Sidney Harris ha capturado apropiadamente muchas de las características del método científico y de su modelamiento matemático, como muestra la Figura 2.

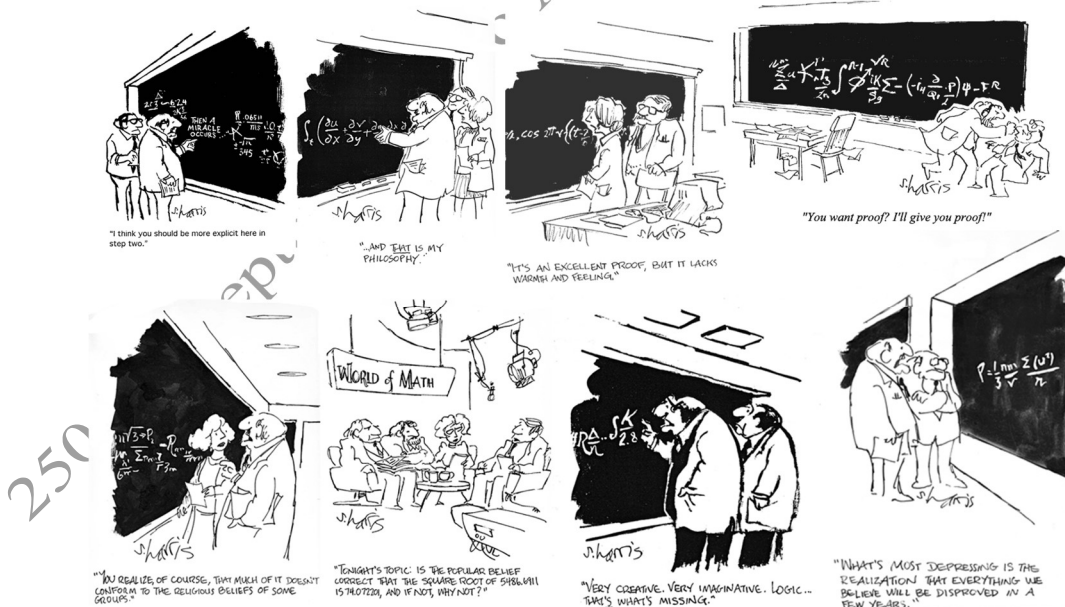


Figura 2. Sobre el método científico (© Sidney Harris, <http://www.sciencecartoonsplus.com>)

Así pues, mientras la ciencia usa el método científico para analizar las observaciones hechas de la naturaleza, proponiendo teorías para comprender y explicar su comportamiento, la ingeniería usa sus resultados para predecir dicho comportamiento de manera que se pueda hacer diseños que permitan

actuar sobre ella misma para facilitar y mejorar las condiciones de vida de la humanidad. Esta consideración suele conducir a comparaciones como las de la **Figura 3** aunque, a medida que la ciencia y la ingeniería se enfocan más hacia el estudio de los sistemas complejos (sistemas que se auto-organizan con base en procesos locales de percepción, aprendizaje, evolución y adaptación), se empieza a difuminar esa diferencia entre ciencia e ingeniería, pues el proceso de comprensión del mundo ya no se distingue del proceso de actuación sobre el mundo, como veremos en la definición 2.

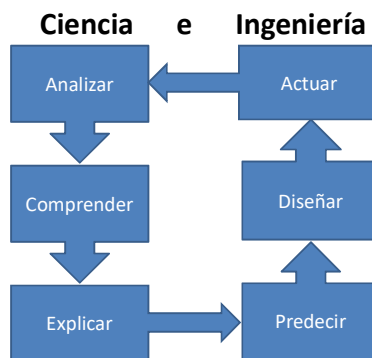


Figura 3. Una comparación simplista pero reveladora entre Ciencia e Ingeniería

Claro, el ingeniero desarrolla sus diseños sobre un modelo matemático antes de implementarlo para actuar en la realidad. **Si no se usa un modelo matemático, no se hace ingeniería.**

Como se mencionó en el resumen de este apartado, el modelado consiste en **abstraer** de una realidad compleja los aspectos más relevantes asociados con un **objetivo particular** de estudio, y formularlos en el contexto formal y riguroso de las matemáticas. Al operar con el modelo así formulado, se obtendrán conclusiones que deberían poderse aplicar a la realidad bajo estudio. La utilidad de estas conclusiones para comprender o controlar el sistema bajo estudio determinará la **validez** del proceso de modelado matemático.

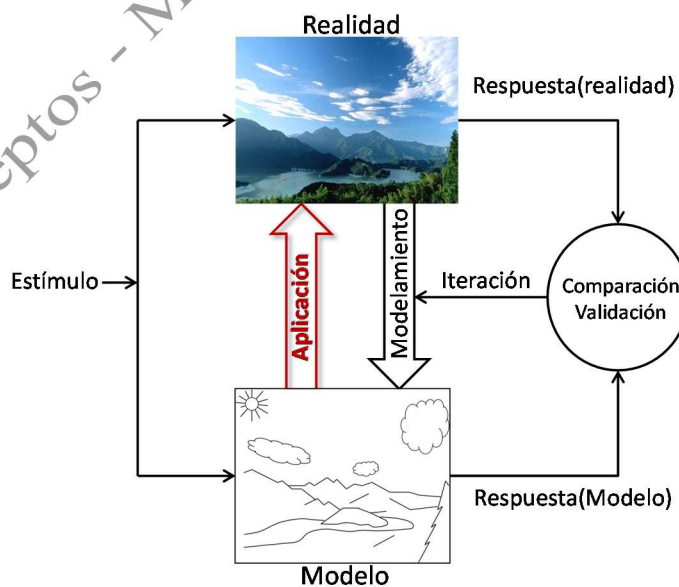


Figura 4. Proceso de modelado matemático

Así pues, el modelado matemático comprende algunas etapas

- Identificación de **componentes**
- Determinación de **variables descriptivas** de esos componentes
- Identificación de **reglas de interacción**
- Determinación de **parámetros** de esas interacciones

Por ejemplo, para describir la traslación de la luna alrededor de la tierra podemos proceder así:

- Identificación de componentes: La luna y la tierra (¡suponemos que ningún otro objeto del universo las perturba!).
- Determinación de variables descriptivas de esos componentes: Posición y velocidad de la tierra y de la luna (¡suponemos que son puntos infinitesimales o, al menos, esferas homogéneas!).
- Identificación de reglas de interacción: Dos objetos se atraen con una fuerza directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia entre ellos (ley universal de la gravedad –Newton–).
- Determinación de parámetros descriptivos de esas interacciones: La constante de proporcionalidad ($G \approx 6.674 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$), la masa de la tierra ($M_T \approx 5.9724 \times 10^{24} \text{ kg}$), la masa de la luna ($M_L \approx 7.346 \times 10^{22} \text{ kg}$), la distancia entre la tierra y la luna ($d \approx 3.84403 \times 10^5 \text{ km}$ en promedio, $3.63026 \times 10^5 \text{ km}$ en el perigeo, $4.05610 \times 10^5 \text{ km}$ en el apogeo) y la velocidad de la luna ($|v| = 1.022 \times 10^3 \text{ ms}^{-1}$ en promedio, $1.082 \times 10^3 \text{ ms}^{-1}$ en el perigeo, $9.7 \times 10^2 \text{ ms}^{-1}$ en el apogeo).

La siguiente figura muestra otro aspecto de la realidad: una pila, un interruptor, una resistencia y un condensador en serie; también se muestra un modelo matemático asociado con esa realidad. Por supuesto, hay muchos aspectos de la realidad que el modelo matemático no captura: La inductancia presente en la resistencia, los efectos de las soldaduras potencialmente defectuosas, la polaridad del condensador electrolítico, los rebotes mecánicos del interruptor, etc. Pero, como se mencionó hace un momento, se trata de construir una representación conceptual de la realidad abstrayendo los aspectos más relevantes asociados con un **objetivo particular**. Si deseamos un modelo para estudiar la carga del condensador, la abstracción alcanzada con el modelo matemático de la Figura 5 es muy apropiada, aunque el mismo modelo no podrá ser útil para un objetivo diferente, como estudiar los efectos térmicos sobre el condensador, por ejemplo.

Otro aspecto fundamental del modelamiento matemático es la generalidad de los modelos. Como muestra la Figura 6, la aplicación de las leyes de Ohm y Kirchhoff al circuito y la aplicación de las leyes de Newton al automóvil (nuevamente, despreciando muchos aspectos de la realidad) conducen a una misma ecuación diferencial lineal ordinaria de primer orden que puede representar un sistema genérico abstracto.

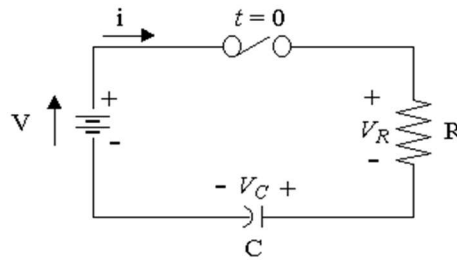
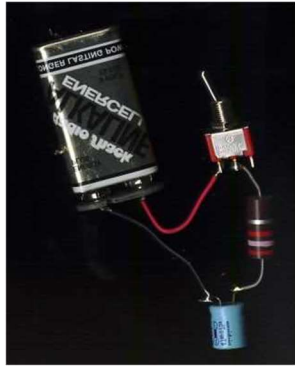


Figura 5. Realidad y modelo matemático

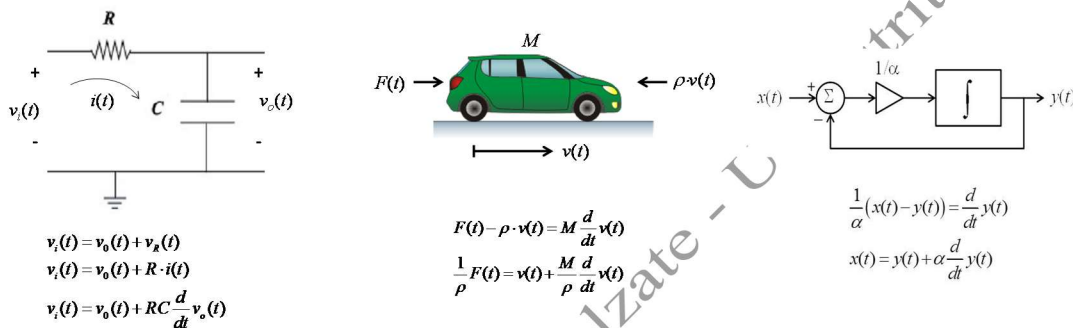


Figura 6. Un modelo matemático es más útil entre más general sea

Lo importante es tener en cuenta que todo modelo matemático en ingeniería es una abstracción del problema de la realidad que queremos modelar, lo cual implica una **simplificación** que limita la validez de los resultados del modelo a las suposiciones que se hayan hecho. Por eso, siempre un modelo matemático particular está asociado necesariamente con un **objetivo** de estudio particular, como muestra la Figura 7. Por ejemplo, mientras mi hijo de cuatro años modelaba un computador como una pantalla, un CD-drive y un joystick, un ingeniero de hardware ve en él unidades de procesamiento, buses, memoria y dispositivos de entrada/salida y un ingeniero de software ve una máquina física, un sistema operativo y un conjunto de aplicaciones. Cada uno de ellos encontrará variables y reglas de interacción para completar su modelo, el cual será válido únicamente para el propósito con que se construyó. Para nosotros, interesados en redes de comunicaciones, un computador es una fuente de tráfico y sus variables descriptivas estarán asociadas con las estadísticas de la generación de paquetes: cuántos paquetes genera en una hora, cuál es la distribución de los tiempos entre llegada, cómo están correlacionados, cuál es el tamaño de los paquetes, etc.

Como se mencionó en el prefacio, en este libro veremos modelos matemáticos que intentan representar redes de comunicaciones. Claro, cada modelo estará asociado con un objetivo particular, lo cual limitará fundamentalmente su rango de aplicabilidad, pero dará luces sobre la dinámica de los fenómenos que ocurren en una red de comunicaciones. Es muy importante que tengamos en cuenta que el modelo no es la realidad y, por lo mismo, las predicciones del modelo no son resultados absolutos del comportamiento que podemos esperar del sistema bajo estudio.

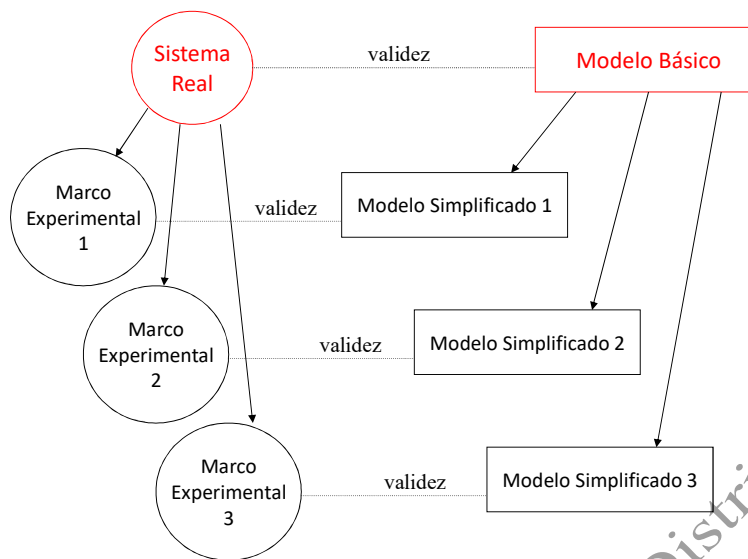


Figura 7. Cualquier modelo matemático es una simplificación ajustada a un marco experimental particular

Lo anterior nos conduce a un punto final igualmente importante: La **evaluación de un modelo matemático**. Por alguna razón que desconozco, siempre que se habla de un modelo matemático las personas (inclusive algunos de los ingenieros, los magisteres y los doctores que estudiarían este libro) piensan en sistemas de ecuaciones. Pero, insistamos, el modelo matemático es apenas una descripción de los componentes, las reglas de interacción y las variables descriptivas de dichos componentes y dichas reglas. Cuando estos elementos se especifican claramente, ya se ha formulado un modelo matemático. Otra cosa es la manera como queramos evaluar dicho modelo. Si la formulación lo permite, el **análisis matemático** (la solución de un sistema de ecuaciones, por ejemplo), es una alternativa deseable –**modelo analítico**–. Pero, generalmente, un modelo que sea capaz de admitir este tipo de soluciones habrá sufrido muchas simplificaciones que lo alejan de la realidad que se quiere modelar. Otra posibilidad es la realización de **experimentos en el laboratorio** con versiones parciales o reducidas del sistema bajo estudio –**modelo experimental**–. Por supuesto, si el ingeniero no tiene un modelo matemático que represente el sistema bajo estudio, le será imposible diseñar experimentos significativos para resolver las preguntas pertinentes del estudio. De todas maneras, los resultados de la experimentación en el laboratorio pueden ser mucho más significativos porque incluirían los efectos de otros componentes que ni siquiera se consideraron explícitamente en el modelo. Una alternativa intermedia es usar la **simulación por computador**, esto es, elaborar un programa que reproduzca las interacciones entre los componentes a medida que transcurre un tiempo virtual en el computador –**modelo de simulación**–. Otra alternativa poco común es experimentar con el mismo sistema que se estudia, pero aún en este caso el ingeniero trabaja con un modelo matemático, así este modelo lo tenga en su mente sólo de manera implícita. Por supuesto, entre más explícito y específico sea el modelo matemático que usa el ingeniero, más significativos serán los resultados de su estudio.

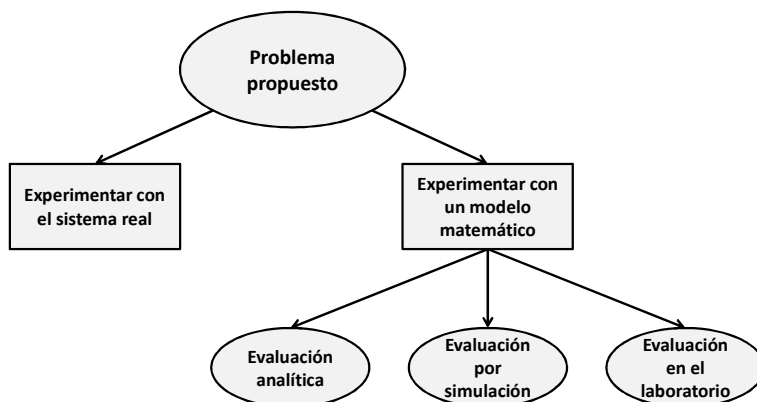


Figura 8. Técnicas de evaluación de un modelo matemático (adaptado de Law&Kelton)

Por ejemplo, considere el modelo del circuito eléctrico de la Figura 6 cuando $RC = 5$ ms, el voltaje de entrada es un escalón de 1 voltio y el condensador estaba inicialmente descargado. El modelo se puede evaluar analíticamente, pues la ecuación diferencial se reduce a $dv_0(t)/dt = 200(1 - v_0(t))$ para $t \geq 0$.

Haciendo $u(t) = 1 - v_0(t)$ podemos separar las variables para obtener $\int \frac{du}{u} = -200 \int dt$, de donde $u(t) = Ce^{-200t}$ o, lo que es lo mismo, $v_0(t) = 1 - Ce^{-200t}$. Como $v_0(0) = 0$, entonces $C=1$, por lo que la respuesta deseada es

$$v_0(t) = 1 - e^{-200t} \quad t \geq 0$$

La segunda alternativa es la simulación: Si consideramos un incremento de tiempo Δt en vez del diferencial dt , y consideramos sólo instantes de tiempo múltiplos de Δt , podríamos aproximar la ecuación diferencial mediante

$$v_i(n\Delta t) \approx v_0(n\Delta t) + RC \frac{v_0(n\Delta t) - v_0((n-1)\Delta t)}{\Delta t}$$

Este es un sistema en tiempo discreto que podemos re-escribir así

$$v_i[n] = v_0[n](1 + \beta) - \beta v_0[n-1], \quad \text{donde } \beta = \frac{RC}{\Delta t}$$

Por ejemplo, con $\Delta t = 0.2$ ms, tendríamos $v_0[n] = (v_i[n] + 25v_0[n-1])/26$ y la simulación se podría realizar, por ejemplo, con el siguiente programa en matlab®:

```

n = -100:400;
vi = (n>=0);
vo = zeros(size(vi));
for m=-100:400
    k = m+101;
    if m<0
        vo(k) = 0;
    else
        vo(k) = (vi(k)+25*vo(k-1))/26;
    end
end
plot(n/5000,vo)
  
```

Listado 1. Simulación de un circuito RC en matlab®

Una tercera alternativa es implementar el circuito de la Figura 5 en el laboratorio y ver la señal del condensador en el osciloscopio. La Figura 9 muestra el resultado del modelo desde los tres tipos de evaluación (tres técnicas de evaluación distintas para un mismo modelo matemático).

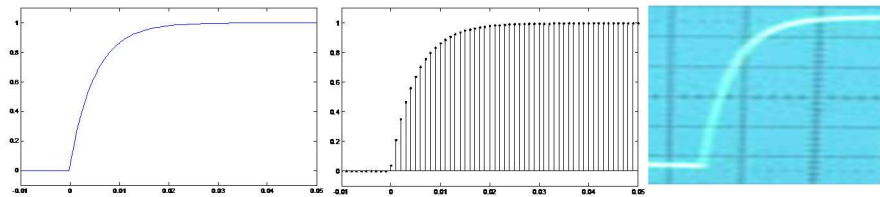


Figura 9. Soluciones analítica, por simulación y experimental de un modelo matemático

En el área de la ingeniería de teletráfico, generalmente se requiere demasiadas simplificaciones para lograr que un modelo matemático se mantenga analíticamente tratable. Por otro lado, resulta demasiado costoso experimentar con el sistema real. En consecuencia, en ingeniería de redes típicamente se usa el método de la simulación por computador para estudiar problemas de diseño y operación de sistemas, especialmente como apoyo a la toma de decisiones. Es importante notar que los dos métodos son complementarios, pues para validar un programa de simulación se suele verificar que, ante escenarios que tengan una solución analítica, ambos métodos de evaluación obtengan los mismos resultados. Esta condición incrementa la confianza de que los resultados del escenario propuesto sean correctos. La validación funciona también en el sentido contrario: Después de haber obtenido un resultado analítico, se puede usar la simulación para verificar los rangos de validez del resultado. En ambos casos, la mejor validación se obtiene al comparar los resultados analíticos o de simulación con resultados experimentales.

Por último, vale la pena aclarar que la simulación es tanto una herramienta como una ciencia. Existe una formalidad muy rigurosa y estricta (la estadística) para asegurar que los estudios de simulación sean significativos y válidos. Sin el cuidado necesario, el uso de la simulación puede conducir a grandes equivocaciones. La simulación de eventos discretos apropiada para la evaluación de modelos matemáticos en ingeniería de teletráfico se revisará con cuidado a partir de la definición 135.

2. Modelos Matemáticos de Sistemas Simples y de Sistemas Complejos

*En el modelo matemático de un **sistema simple** existen pocos componentes, pocas variables descriptivas y pocas reglas de interacción, de manera que el comportamiento global del sistema se puede explicar como el resultado directo de las reglas de interacción entre los componentes. En efecto, tanto los componentes como el sistema entero actúan en una misma escala de tiempo y espacio y resulta fácil encontrar relaciones de causalidad entre las propiedades particulares de los componentes y las propiedades globales del sistema. En el modelo matemático de un **sistema complejo** existen muchos componentes y muchas variables descriptivas cuyas relaciones son no lineales, de manera que, aunque las reglas de interacción sean muy simples, surgen propiedades globales del sistema que no se pueden atribuir directamente a sus componentes sino que son el resultado emergente de las múltiples interacciones no lineales entre ellos. En efecto, los componentes del sistema actúan en escalas de tiempo y espacio diferentes a aquellas en las que se observa el*

comportamiento global del sistema, lo que da lugar a fenómenos de emergencia y auto-organización (formación de patrones, múltiples lazos de realimentación, sincronización, cooperación/competencia, etc.). En ambos casos, los procedimientos de modelado matemático mencionados en el numeral anterior siguen siendo los mismos (componentes, interacciones y variables relacionadas con el objetivo de estudio), pero con sistemas simples se busca una descomposición reduccionista/mecanicista del sistema que permita diseñar el comportamiento deseado, y con sistemas complejos se busca guiar la auto-organización para obtener los comportamientos emergentes que permitan la adaptación autónoma del sistema a los cambios dinámicos del ambiente en que se desenvuelve, mediante aprendizaje o evolución.

Una forma de construir los modelos matemáticos del apartado anterior es considerando que las propiedades y los comportamientos globales de un sistema se pueden deducir o explicar a partir las propiedades y los comportamientos de las partes que lo componen. En este contexto se hace muy apropiado el principio de superposición propio de los sistemas lineales. En la misma línea se encuentra el principio del determinismo, según el cual la incertidumbre se limita a la presencia de errores en las mediciones, en cuyo caso la distribución Gaussiana resulta muy apropiada. Sin embargo, la linealidad y la gaussianidad son, en realidad, poco comunes en la naturaleza.

Históricamente, en ausencia de computadores electrónicos, hasta hace algunas décadas sólo se contaba con métodos analíticos de evaluación de modelos matemáticos, o métodos numéricos calculados manualmente. Los primeros, en general, casi siempre existen para sistemas lineales y gaussianos pero casi nunca existen para sistemas no-lineales y no-gaussianos. Los segundos pueden requerir un número tan exagerado de cálculos que, si no se cuenta con computadores electrónicos, resulta inútil aplicarlos a sistemas con un gran número de componentes cuyas variables descriptivas se relacionan de manera no lineal y no gaussiana. Fue la introducción del computador la que permitió evaluar modelos matemáticos no-lineales y no-gaussianos con muchos componentes, y logró redescubrir fenómenos fascinantes que ya grandes pensadores habían intuido, como el caos, la fractalidad y la emergencia. Con semejante herramienta, la ciencia y la ingeniería pudieron abordar problemas mucho más complejos al permitirse incluir la no-linealidad y la no-gaussianidad en sus modelos matemáticos evaluados numéricamente o mediante simulación. El computador explica en gran medida el vertiginoso desarrollo científico y tecnológico de las últimas décadas.

Entre las herramientas que el computador ayudó a desarrollar para evaluar modelos matemáticos no-lineales y no-gaussianos están los métodos bio-inspirados (redes neuronales artificiales, algoritmos genéticos, inteligencia de enjambre, sistemas inmunes artificiales, sistemas difusos, etc.). Todos ellos han sido muy útiles en la solución de complejos problemas de optimización, regresión o clasificación, fundamentales en el diseño de nuevas soluciones a problemas complejos. La inclusión de estos métodos, entonces, ha permitido extender la aplicabilidad de los modelos matemáticos en ingeniería, especialmente al facilitar la inclusión de requerimientos de robustez (el sistema debe seguir operando satisfactoriamente en un amplio rango de condiciones) y su forma extendida, la adaptabilidad (el sistema debe poder ajustar automáticamente sus parámetros de acuerdo a los cambios en el ambiente). Los grandes avances que estos modelos han permitido alcanzar en materia de comunicaciones,

también han permitido incluir una característica formidable en los diseños de ingeniería: Los sistemas distribuidos, en los que la información local se distribuye para tomar decisiones colectivas en busca de un objetivo global. También estos modelos matemáticos de sistemas robustos, adaptables y distribuidos han contribuido al fascinante mundo tecnológico de hoy.

Así pues, a medida que crece la complejidad de los fenómenos que la ciencia estudia y de los problemas que la ingeniería aborda, crece también la versatilidad de los modelos matemáticos que ellas usan. Hoy, la experiencia ingenieril con sistemas bio-inspirados y la comprensión científica sobre la organización de muchos sistemas naturales, han permitido explorar también otros fenómenos comunes en la naturaleza y otras formas de plantear soluciones de ingeniería a mayores problemas de la humanidad: La auto-organización y la emergencia. Efectivamente, en física, biología, ciencias sociales y ciencias de la computación, entre muchas otras áreas, se ha venido comprendiendo que los comportamientos complejos observados a cierta escala surgen como fenómenos emergentes de procesos de auto-organización a escalas menores, casi siempre mediados por interacciones simples pero no-lineales. Esta comprensión indica que la ingeniería puede enfrentar nuevos problemas básicos de la humanidad, en los ámbitos ambiental, tecnológico, biológico, económico y político, si extiende los nuevos modelos matemáticos que capturan estos fenómenos para lograr diseñar, construir, operar y controlar aparatos, procesos y estructuras que resuelvan problemas de la humanidad, que es el propósito de la ingeniería. Pero estos aparatos, procesos y estructuras serán fundamentalmente diferentes en un sentido básico: Sus partes deberán percibir el mundo, aprender de él, decidir autónoma e inteligentemente comportamientos apropiados, comunicarse entre ellas, al menos en un ámbito local, y, de manera colectiva, permitir la emergencia de patrones funcionales, temporales o espaciales que resuelvan el problema que se trata. A estos nuevos modelos matemáticos en ingeniería se les ha llamado “modelos de inteligencia colectiva”, “modelos de auto-organización guiada”, “modelos de sistemas dinámicos cognitivos”, “modelos de control cooperativo”, “modelos morfogenéticos”, etc. Estos modelos matemáticos ya no buscan que el ingeniero diseñe la solución del problema sino que diseñe el sistema complejo adaptivo que sea capaz de emerger mediante auto-organización la solución del problema, pues el problema mismo cambia dinámicamente en un ambiente complejo. De hecho, el desarrollo de estos modelos matemáticos va difuminando la frontera entre ciencia e ingeniería, pues además de permitir el diseño del sistema de ingeniería, permite proponer explicaciones a los fenómenos emergentes de la naturaleza.

La Figura 10 muestra esquemáticamente el concepto más general de un sistema complejo. Las interacciones no lineales entre sus muchos componentes se presentan en múltiples escalas, de manera que de una escala a otra ocurren fenómenos de emergencia de estructuras y patrones que aparecen por auto-organización, todo como producto de una evolución en el tiempo caracterizada por una alta sensibilidad a las condiciones iniciales. Estas propiedades aparecen en los mercados de valores, las organizaciones políticas, la Internet, las economías nacionales y globales, el clima global, los sistemas ecológicos, etc. En todos estos sistemas existe una gran cantidad de microcomponentes que interactúan de manera tal que la información no se encuentra en los componentes sino en sus interacciones. Por eso, los modelos matemáticos de sistemas complejos involucran herramientas de modelado adicionales a las típicas de los sistemas simples, como se aprecia en la Figura 11. Áreas tan fascinantes como las dinámicas no-lineales y el caos, la realimentación, la evolución, la adaptación,

la interacción entre agentes autónomos mediante teoría de juegos, transiciones de fase, redes, etc. se unen a las técnicas de modelado de sistemas simples (ecuaciones diferenciales y de diferencia lineales, fundamentalmente) para abstraer y conceptualizar los fenómenos propios de los sistemas complejos.

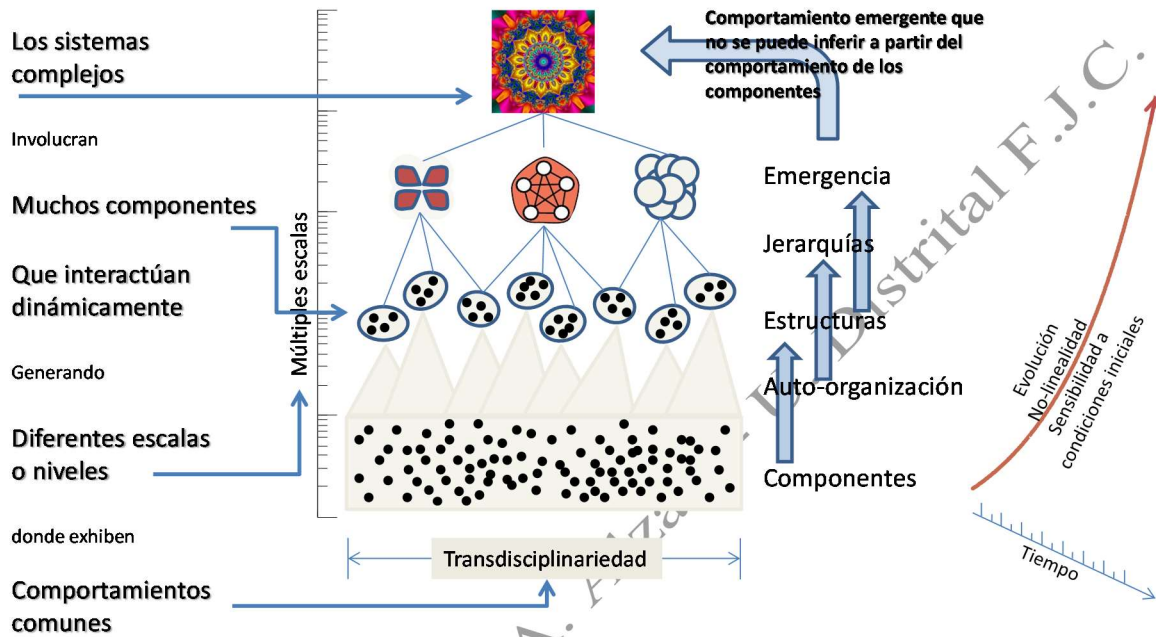


Figura 10. Sistemas Complejos (adaptado de www.art-sciencefactory.com/complexitymap_feb09.html)



Figura 11. Herramientas para el modelado matemático de sistemas complejos (adaptado de Hiroki Sayama, D.Sc., State University of New York, Binghamton)

En lo que respecta a la ingeniería de redes de comunicaciones, el modelado matemático de sistemas complejos busca diseñar sistemas compuestos por agentes cognitivos, esto es, capaces de percibir información del medio ambiente y actuar sobre él de manera que pueda comunicarse directa o indirectamente con otros agentes del sistema, aprender de las señales percibidas incluyendo los efectos de sus propias acciones, y adaptarse así a los cambios ambientales. De esta manera, cada agente cognitivo construye su propio modelo predictivo sobre los aspectos de interés del ambiente y desarrolla sus propias reglas de comportamiento que le permitan actuar sobre el ambiente para llevar a cabo algunas tareas prescritas de manera efectiva y eficiente, ya sea en cooperación o en competencia con otros agentes. En particular, hablando de redes de comunicaciones, ellas están compuestas por un gran número de componentes que interactúan entre ellos simultáneamente con una alta no-linealidad en las dinámicas de los protocolos que lo controlan (potencialmente caótica), fractalidad y multi-fractalidad en las trazas de tráfico que cursa por sus enlaces, topologías físicas y lógicas libres de escala, auto-organización de la demanda de los usuarios al borde de la congestión, y leyes de potencia en la distribución del tamaño de los archivos y la duración de las sesiones que circulan por la red. Todas estas condiciones actuales de las redes de comunicaciones, típicas en todos los sistemas complejos, contrastan con los principios filosóficos de diseño que les dieron origen, los cuales se resumen en una estructura modular jerárquica en la que las diferentes capas de la jerarquía sólo deberían interactuar mediante interfaces precisas y limpias, con protocolos diseñados exclusivamente para la función particular de cada capa. La insuficiencia de estos principios originales, especialmente debida al descubrimiento de interacciones imprevistas entre las capas de la jerarquía funcional, ha conducido a técnicas de diseño “cross-layer” cuyos problemas de optimización distribuida multiobjetivo se resuelven desde la teoría de control de sistemas dinámicos, la teoría de la información y la teoría de juegos, con base en el aprendizaje, la adaptabilidad y la robustez a las condiciones variantes del entorno, utilizando de manera ubicua técnicas de inteligencia computacional como las redes neuronales, la lógica difusa, los algoritmos genéticos y la inteligencia de enjambre, por ejemplo. En la siguiente definición veremos cómo, en ingeniería de teletráfico, se deben usar modelos simples y complejos, siempre considerando la aleatoriedad dentro de ellos.

3. Ingeniería de Teletráfico

*La ingeniería de teletráfico consiste en el uso de modelos matemáticos formales para representar las interacciones que existen entre la **demanda**, la **capacidad** y el **desempeño** de una red de comunicaciones, como muestra la Figura 12. Dichos modelos deben permitir **comprender**, **analizar**, **diseñar** y **controlar** redes de comunicaciones.*

Cada uno de estos tres aspectos puede requerir una gran cantidad de variables para ser descrito adecuadamente, de manera que los símbolos a los que se refiere la figura para representar la demanda (λ), la capacidad (μ) y el desempeño (D) pueden ser arreglos de variables (vectores, posiblemente de dimensión infinita).

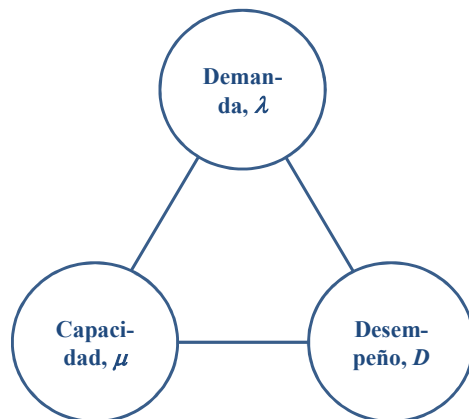


Figura 12. La ingeniería de tráfico estudia las relaciones entre demanda, capacidad y desempeño en una red de comunicaciones

En un ejemplo de extrema sencillez, λ puede ser un escalar que represente la **tasa promedio de llegadas** de paquetes a un enlace de comunicaciones (demanda, en paquetes por segundo), μ puede ser un escalar que represente la **tasa promedio de transmisión** de paquetes que dicho enlace puede alcanzar (capacidad, en paquetes por segundo) y D puede ser un escalar que represente el **retardo promedio** que experimentan los paquetes en dicho enlace, ya sea siendo transmitidos o esperando a ser transmitidos (desempeño, en segundos). Bajo condiciones muy específicas de variabilidad alrededor de esos valores promedios³, las tres variables se relacionan de la siguiente manera:

$$D = \frac{1}{\mu - \lambda} \quad (1)$$

Un resultado así es muy útil porque nos permite determinar el desempeño dadas la capacidad y la demanda (**análisis de redes**, $D=1/(\mu-\lambda)$), o dimensionar la capacidad dada la demanda y el desempeño deseado (**diseño de redes**, $\mu=\lambda+1/D$), o regular la demanda dada la capacidad y el desempeño deseado (**control de redes**, $\lambda=\mu-1/D$). Además, resulta muy significativo el hecho de que, para que dicha expresión tenga sentido, la capacidad μ deba ser superior a la demanda λ o, lo que es lo mismo, que se satisfaga la relación $\rho \equiv \lambda/\mu < 1$, donde ρ , que se conoce como **intensidad de tráfico**, indica la relación entre la demanda (lo que los usuarios exigen, λ) y la capacidad (lo que la red puede ofrecer, μ). En efecto, si la demanda se acerca a la capacidad por la izquierda, el retardo se acerca a infinito; pero si la demanda supera la capacidad, el retardo se haría negativo, indicando que ese caso no está en el rango de validez del análisis que condujo a dicha expresión. La ingeniería de teletráfico busca exactamente este tipo de relaciones que proporcionan comprensión de la dinámica

³ Como veremos más adelante, las llegadas deben formar un proceso de Poisson y los tiempos de transmisión deben ser independientes y estar exponencialmente distribuidos

de las redes de comunicaciones y herramientas matemáticas para análisis, diseño y control de las mismas.

Para determinar relaciones útiles entre la demanda, la capacidad y el desempeño de una red de comunicaciones es necesario estudiar las interacciones dinámicas que se forman cuando los usuarios compiten o colaboran para utilizar los recursos de la red. En efecto, estas redes están conformadas por un conjunto de recursos de capacidad limitada que deben atender las demandas impuestas por los usuarios de la red. La Figura 13 muestra un modelo matemático de una red compuesta por seis enrutadores, cada uno de ellos con dos enlaces de entrada y dos enlaces de salida. Suponemos que cada paquete que llega a un enrutador debe pasar por un proceso de enrutamiento (círculo azul) en el que se decide a cuál de los dos enlaces de salida debe dirigirse el paquete. Como el proceso de decisión toma un tiempo diferente de cero, es posible que un paquete que llegue a un enrutador encuentre el procesador de enrutamiento ocupado con algún otro paquete que llegó con anterioridad, por lo que se dispone también de una memoria (o *buffer*) donde almacenar los paquetes que deben hacer cola para esperar a ser enrutados (rectángulos azules). Los paquetes que salen del proceso de decisión se dirigen a su respectivo enlace de salida (círculos amarillos) pero, nuevamente, como el tiempo de transmisión de un paquete es diferente de cero, es posible que un paquete encuentre el enlace ocupado, por lo que se dispone también de un buffer donde almacenar los paquetes que deben hacer cola para esperar a ser transmitidos (rectángulos amarillos). En este proceso, los paquetes sufren **retardos** (tanto por los tiempos de servicio como por los tiempos de espera en cola) y potenciales **pérdidas** (si, por ejemplo, a su llegada no hay espacio suficiente en el buffer), lo que conduce a una tasa efectiva de paquetes por unidad de tiempo o **caudal**, que puede ser inferior a la demanda impuesta por los usuarios.

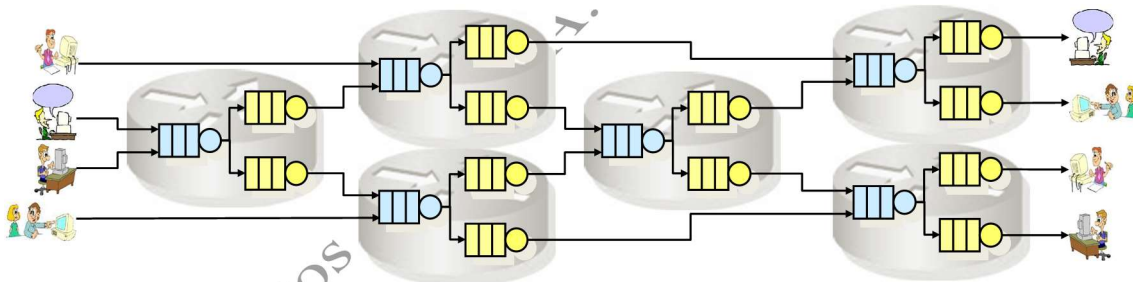


Figura 13. La red de comunicaciones es un conjunto de recursos de capacidad finita que deben ser compartidos por los usuarios

El anterior ejemplo explica por qué los modelos matemáticos más ampliamente utilizados en ingeniería de teletráfico corresponden a la **Teoría de Colas**, que es el área de la investigación operacional que estudia los procesos estocásticos asociados con líneas de espera, donde unos "clientes" hacen fila en espera de un "servicio". De hecho, fue **Agner Erlang** quien fundó la teoría de colas al modelar matemáticamente la red telefónica de Copenhague a comienzos del siglo XX. Posteriormente la teoría de colas se ha aplicado a una gran cantidad de problemas prácticos en flujo de tráfico vehicular, procesos industriales, sistemas operativos, servicio a clientes en diferentes tipos de oficinas, planeación de sistemas de transporte de carga, atención a pacientes en hospitales y centros médicos, etc. Aun así, sigue siendo en el área de las redes de comunicaciones donde mayores avances se han alcanzado en esta disciplina matemática.

En efecto, ha sido en las redes de comunicaciones donde mayor necesidad ha habido de análisis novedosos de sistemas de colas. Con la introducción de los conmutadores automáticos a finales del siglo XIX y su posterior digitalización en la década de 1970, las actividades de análisis de tráfico y optimización de la red para satisfacer unos criterios dados de calidad de servicio (QoS) correspondían al diseño de **redes conmutadas por circuitos** con una sola clase de tráfico. Entonces se desarrolló un método tradicional basado en la **fórmula B de Erlang** que relacionaba la intensidad de tráfico (demanda, A , igual al producto entre el número promedio de llamadas por minuto y la duración promedio de una llamada en minutos), el número de circuitos disponibles (capacidad, N) y la probabilidad de que una llamada entrante no encontrara un circuito libre (desempeño, P_B , igual a la fracción de llamadas perdidas), cuando las llegadas forman un proceso de Poisson y sus duraciones son independientes unas de otras:

$$P_B = \frac{A^N / N!}{\sum_{n=0}^N A^n / n!} \quad (2)$$

Esta es otra relación que obedece a la Figura 12, donde la demanda es A (en vez de λ), la capacidad es N (en vez de μ) y el desempeño es P_B (en vez de D), y que también permite hacer análisis (encontrar P_B dados A y N), diseño (encontrar N dados P_B y A) y control (encontrar A dados P_B y N) de redes.

Mientras se desarrollaba esta infraestructura para las redes telefónicas, en el mundo de la informática se desarrollaban paralelamente las técnicas de transmisión de datos entre computadores, lo cual condujo a las **redes conmutadas por paquetes** con múltiples arquitecturas jerárquicas de protocolos, todas orientadas filosóficamente por el **modelo de referencia OSI**. El propósito fundamental de estas redes era compartir recursos físicos y prestar servicios de transferencia de archivos, correo electrónico y log-in remoto, especialmente. La Figura 13 y la Ecuación (1) se refieren más a este tipo de redes. En la década de 1980, con el desarrollo de las redes públicas de datos tipo **X.25**, las redes telefónicas digitales con **señalización por canal común**, y el desarrollo de la **Internet mediante TCP/IP**, se reconoció la necesidad de integrar la comunicación telefónica con la transmisión de datos por lo que se desarrollaron técnicas híbridas en la llamada **Red Digital de Servicios Integrados (ISDN)**, de donde surgieron tecnologías aún usadas hoy, como **Frame Relay** (relevo de tramas). En la década de 1990 se tuvo la visión de una sola red que integrara no sólo voz y datos sino todo tipo de medios de comunicación (texto, imágenes, video, teleconferencia, etc.), para lo cual se decidió que la conmutación de paquetes era la más apropiada y se desarrolló la red digital de servicios integrados de banda ancha (**B-ISDN**) y su técnica de conmutación, el **Modo de Transferencia Asíncrono (ATM)**. Todos estos desarrollos estuvieron asociados con avances correspondientes en la teoría de colas, incluyendo los conceptos teóricos para ofrecer diferente **calidad de servicio** (QoS –Quality-of-Service-) a diferentes **clases de tráfico** en una misma red conmutada por paquetes. Estos desarrollos incluyen caracterización de tráfico con tiempos entre llegadas y tiempos de servicio correlacionados a corto y a largo plazo, técnicas de control de admisión y control de acceso, tecnologías de conmutación por hardware, conceptos de calidad de servicio con cotas de desempeño entre extremos, el concepto de capacidad equivalente, mecanismos de asignación de recursos de transmisión, gestión activa de colas y de control de congestión mediante realimentación, etc. Sin embargo, mientras se planeaba y se empezaba a implementar la B-ISDN y la tecnología ATM,

apareció la **world-wide-web (WWW)**, una aplicación que disparó a la Internet y su arquitectura TCP/IP como la red donde finalmente se dio la integración fundamental que se buscaba con ATM.

Efectivamente, es al navegar por la web que se evidencian las bondades de tener video, voz, datos, imágenes y audio integrados en una misma aplicación y en una misma red. El fenómeno de la WWW sobre Internet hizo que los conceptos desarrollados para QoS en ATM se aplicaran de alguna manera en redes TCP/IP, con lo que se desarrollaron arquitecturas orientadas a QoS sobre IP tales como servicios integrados (**IntServ**) y servicios diferenciados (**DiffServ**), y protocolos adicionales que buscan acercar conceptualmente IP a ATM, tales como **MPLS, RSVP, RTP, IPv6, QoS-OSPF**, etc. Todos estos mecanismos integran la teoría de colas con **teoría de control** de sistemas dinámicos, en sus formas más avanzadas de control no-lineal, control estocástico, control óptimo, control robusto y control distribuido, simultáneamente, abriendo ampliamente el espectro de los modelos matemáticos aplicables en las redes de comunicaciones y en la ingeniería de teletráfico. Por ejemplo, los usuarios que usan TCP como protocolo de transporte ajustan la tasa de transmisión de acuerdo con la información retroalimentada por los mecanismos de gestión de las colas en los enrutadores (**AQM, Active Queue Management**). Esta información realimentada puede ser en forma de paquetes perdidos, variaciones en el retardo de los paquetes o notificaciones explícitas de congestión. De esta manera, el conjunto usuario/enrutador forma un sistema de control realimentado, como se muestra en la Figura 14. El algoritmo TCP modula el flujo de datos de usuario de manera que, en el instante t , genera $r(t)$ paquetes por segundo. La longitud de la cola en el enrutador, $q(t)$, varía con el flujo $r(t)$ de manera que, en el instante t , se produce una pérdida de paquetes $p(t)$. Debido a los retardos en la red, los usuarios se enteran de estas pérdidas τ segundos tarde y, de acuerdo con ellas, reajusta la tasa $r(t)$ tratando de maximizar alguna función de utilidad.

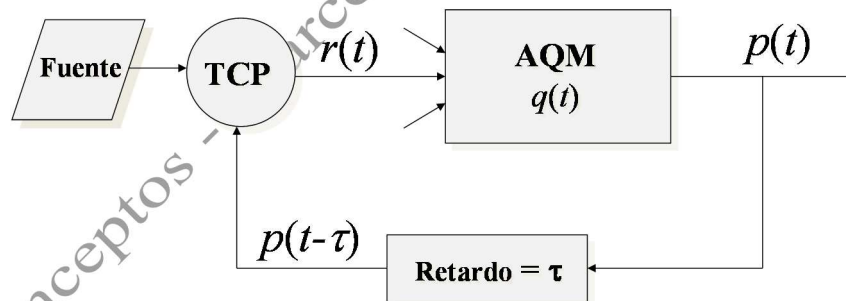


Figura 14. Modelo de control realimentado para el flujo TCP

Cada uno de los mecanismos mostrados en la Figura 14 (TCP, AQM y la dinámica de la cola) determina las tasas de cambio del flujo, la longitud de la cola y la tasa de pérdidas mediante funciones no lineales correspondientes (R , P y Q , respectivamente):

$$r'(t) = R(p(t-\tau)), \quad q'(t) = Q(q(t), r(t)), \quad p'(t) = P(q(t), r(t)) \quad (3)$$

Claramente, ajustar los parámetros de estos mecanismos para maximizar una función de utilidad adecuada es un problema de control óptimo, predictivo, no lineal, distribuido y estocástico, lo cual exige modelos matemáticos adecuados de ingeniería de teletráfico.

Otro desarrollo vertiginoso de las últimas tres décadas se ha dado en el contexto de las **redes inalámbricas**. Junto con la integración de medios de comunicación y la satisfacción de requerimientos de QoS, los usuarios quieren disponibilidad de los servicios de comunicación en todo instante y en todo lugar. Esto sólo es posible si se usan medios inalámbricos de acceso que permitan la movilidad de los usuarios, para lo cual se desarrolló la **telefonía celular**: Una estación base coordina la comunicación en una "célula" o "celda" (una región de cobertura o área de servicio en la que se usa un conjunto particular de canales de radio), permitiendo el acceso inalámbrico de múltiples usuarios móviles a la infraestructura de la red conmutada. A lo largo de cuatro generaciones de tecnología celular se pasó de la transmisión analógica de canales de voz conmutados por circuitos (1G) a una conexión 100% IP entre extremos, sin conmutación de circuitos, donde los paquetes que se intercambian pueden llevar datos, acceso móvil a la www, telefonía IP, televisión (de media o alta resolución), videoconferencia, etc. (4G). Ya se hacen pruebas de una quinta generación con velocidades de acceso de varios Gbps con millones de conexiones por kilómetro cuadrado, lo que traerá a nuestros dispositivos de bolsillo la realidad virtual y la realidad aumentada, además de habilitar la Internet de las cosas (IoT). Con tantas alternativas de acceso, los dispositivos móviles actuales pueden escoger libremente entre acceso mediante **WiFi**, **Bluetooth**, **WiMax**, **LTE** o cualquier tecnología celular, de manera oportunista, para acceder a cualquiera de los servicios mencionados o para conectarse directamente entre ellos sin ninguna infraestructura subyacente, incluyendo redes móviles ad hoc (**MANET**), las cuales se pueden usar para acceso a redes fijas (**redes mesh**), y redes de sensores inalámbricos (**WSN**), incluyendo tecnologías tan útiles como las VANET (Vehicle Ad hoc Network) que ya conducen a la Internet de Vehículos (IoV) para gestión de tráfico urbano.

En este contexto los principios básicos de las arquitecturas funcionales no aplican transparentemente, por lo que se requiere usar técnicas de **diseño cross-layer**. Adicionalmente, la capacidad de adaptabilidad de los sistemas de radio (**SDR** –software defined radio–) hacen posible la implementación de redes de **radio cognitiva**, en donde los equipos cognitivos aprovechan los periodos de tiempo en que las bandas licenciadas no son utilizadas por sus usuarios legítimos para transmitir su propia información. Los terminales cognitivos utilizan propiedades como la **percepción**, el **aprendizaje** y la **adaptación** para llevar a cabo sus funciones, propiedades asociadas con el cerebro humano, lo cual conecta los modelos matemáticos de las redes de comunicaciones con la **inteligencia computacional** y con la teoría de **sistemas dinámicos complejos**. Estos mecanismos permiten resolver cooperativamente los problemas de detección y estimación de señales mediante sistemas **MIMO** (MultiInput-MultiOutput), y permiten la virtualización de los servicios de comunicación móvil sobre los esquemas de computación distribuida "en la nube" (**cloud-computing**). Más aún, con técnicas de comunicación por campo cercano **NFC** (Near Field Communication) como **RFID** (Radio Frequency Identification), que permiten comunicaciones a muy corta distancia para intercambio de datos entre dispositivos, empieza a hacerse realidad la **IoT** (Internet of Things), la cual permite percibir y controlar objetos remotamente a través de la infraestructura de redes, en un paso más de virtualización del mundo real que conducirá rápidamente a sistemas cyber-físicos de objetos inteligentes (smart grids, smart homes, smart cities, etc.).

Todo este desarrollo en las tecnologías de las redes de comunicaciones ha traído grandes retos al modelado matemático para análisis, diseño y control de las mismas, por lo que la ingeniería de teletráfico vive un momento de grandes desafíos en su propósito de encontrar relaciones formales entre la capacidad, la demanda y el desempeño de las redes modernas de comunicaciones. En efecto, con los avances tecnológicos en redes de comunicaciones, avanzan también las técnicas de modelado matemático para ingeniería de teletráfico, partiendo desde la teoría de colas y avanzando hacia conceptos como el ancho de banda equivalente (el cual permite diseñar mecanismos de QoS con base en el estudio de las cotas propuestas por la teoría de grandes desviaciones), la aplicación formal de la teoría de control de sistemas dinámicos (incluyendo conceptos básicos de estabilidad y robustez en control distribuido, robusto, óptimo, no-lineal y estocástico), desarrollos basados en inteligencia computacional (redes neuronales, algoritmos genéticos, lógica difusa, algoritmos de enjambre, sistemas inmunes, etc.) y, en últimas, aplicación formal de la teoría de los sistemas complejos adaptivos (comportamientos emergentes, auto-organización, algoritmos de consenso, fractales, etc.). Algunas propuestas para ingeniería de sistemas complejos con aplicaciones en ingeniería de redes incluyen los sistemas dinámicos cognitivos, la auto-organización guiada y la ingeniería morfogénica, entre otros. Considere por ejemplo un enjambre de robots cuyo movimiento debe ser tal que mantengan un mínimo de conectividad en la red de comunicaciones ad-hoc formada entre ellos, a pesar de que la estructura de la red cambie con el movimiento. Cualquier solución centralizada o jerárquica podría imponer una excesiva sobrecarga a la red en términos de información de control, por lo que se necesitan algoritmos de control descentralizados que relacionen en un modelo matemático distribuido la conectividad inalámbrica local y la movilidad de los robots, de manera que cada robot, haciendo uso únicamente de la información local, tome decisiones individuales y autónomas tales que el conjunto de robot mantengan la conectividad necesaria.

En todos estos esfuerzos recientes sigue siendo necesario modelar la incertidumbre, aun más que en los modelos matemáticos de comienzos del siglo XX. El modelamiento probabilístico que se estudia en este libro es un método muy poderoso para representar la incertidumbre cuantitativamente que ha demostrado su gran utilidad desde hace más de cuatrocientos años y que se complementa maravillosamente con desarrollos más recientes como los sistemas dinámicos caóticos y la lógica difusa, entre otras posibilidades.

II. Conceptos Básicos de Teoría de Probabilidad

4. Experimento Aleatorio

Un experimento es un proceso de observación mediante el cual se selecciona un elemento de un conjunto de posibles resultados. Un experimento aleatorio es aquel en el que el resultado no se puede predecir con anterioridad a la realización misma del experimento.

El modelado matemático consiste en lograr un nivel de abstracción tal que podamos agrupar una gran cantidad de problemas en un solo concepto. Todos sabemos que lanzar una moneda para ver qué lado queda hacia arriba es un experimento aleatorio, cómo también lo fue observar las señales del detector Atlas del LHC (el colisionador de hadrones de 27 km, y 8 mil millones de euros) después de una colisión y detectar si hay en ellas evidencia o no de la existencia del bosón de Higgs⁴. ¿Qué tienen en común estos dos experimentos para poder extraer los elementos de una definición, que aplique también para lanzar un dado y contar los puntos de la cara que queda hacia arriba, para implementar un testbed en el laboratorio y medir el desempeño de un nuevo protocolo de comunicaciones, o para construir y correr un programa de simulación que evalúe el mismo protocolo? Pues bien, en todos ellos hemos llevado a cabo un proceso de observación y, como consecuencia del mismo, hemos seleccionado uno de un conjunto de posibles resultados. Eso es lo que hace el Banco Emisor cuando ajusta las tasas de interés para ver sus efectos en la inflación y lo que hace el protagonista de un libro de probabilidades cuando selecciona una bola de una bolsa para observar su color ¡Así es la capacidad de abstracción del modelado matemático!

Los ejemplos típicos de los cursos de probabilidad incluyen, como acabamos de ver, lanzar una moneda, que equivale a seleccionar un elemento del conjunto $\{cara, sello\}$, lanzar un dado, que equivale a seleccionar un elemento del conjunto $\{1,2,3,4,5,6\}$, o escoger una carta de la baraja de naipes, que equivale a seleccionar un elemento del conjunto $\{(f,n) : f \in \{picas, tréboles, corazones, diamantes\}, n \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}\}$. Otros ejemplos más interesantes para nosotros incluyen medir la tasa de pérdida de paquetes en una conversación VoIP, que equivale a seleccionar un elemento del conjunto $\{x \in \mathbb{Q} \mid 0 \leq x \leq 1\}$, medir el retardo de un paquete de voz en esa misma conversación, que equivale a seleccionar un elemento del conjunto \mathbb{R}^+ (los reales positivos), o verificar el estado de ocupación de un canal de comunicaciones, que equivale a seleccionar un elemento del conjunto $\{libre, ocupado\}$.

En un experimento aleatorio, aunque se mantengan constantes las condiciones bajo las cuales se realizan diferentes instancias del mismo, el resultado no se puede predecir con anterioridad a la realización del experimento. Por ejemplo, como vimos en el prefacio, generalmente no es posible

⁴ En 2012 los experimentos del LHC cruzaron el umbral de los "7 sigmas": Hay una certeza superior al 99.9999999999% de que exista el bosón de Higgs. Si el bosón de Higgs no existe, los resultados de los experimentos se dieron al azar con una probabilidad de 0.000000001%. Note que lanzar una moneda es un experimento que podemos repetir cada vez con idéntica incertidumbre. Si repetimos el experimento "detectar la existencia del bosón de Higgs", será cada vez menos aleatorio en cada nueva repetición.

predecir el caudal, el tamaño del archivo ni el tiempo de transferencia en una transacción *ftp*, lo que indica que transferir un archivo de un servidor a un cliente mediante *ftp* constituye un experimento aleatorio. Igualmente, si desde la ventana de comandos de nuestro PC ejecutamos la instrucción

```
C:>netstat -e 10 > estadisticas.txt
```

y navegamos por Internet durante una hora, generaremos un archivo con algunas estadísticas de la red, incluyendo el número de bytes que se han recibido en períodos de diez segundos. La Figura 15 presenta una gráfica del número de bytes recibidos durante varios períodos en una instancia del experimento. Evidentemente, no estamos en condiciones de predecir cuántos bytes llegarán en el siguiente período, aun cuando podemos afirmar que, por ejemplo, sería muy extraño si llegaran más de 20 Mbytes y, en cambio, sí sería de esperar que llegaran más de 2 Mbytes. De cualquier manera, queda claro que observar el número de bytes recibidos en 10 segundos mientras se navega por Internet constituye un experimento aleatorio.

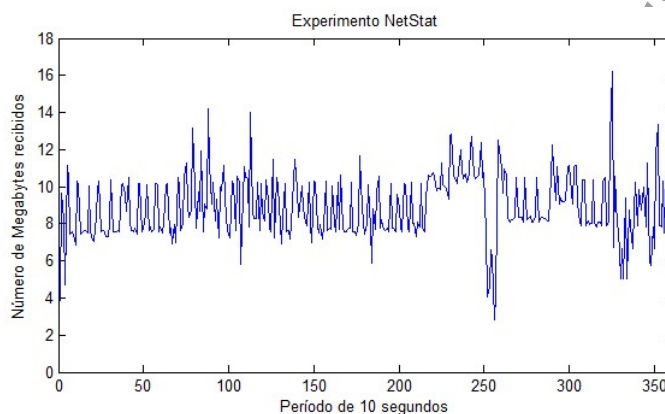


Figura 15. Observar el número de bytes que llegan de la red en un período de 10 segundos constituye un buen ejemplo de lo que es un experimento aleatorio.

¿Por qué no podemos predecir el resultado de un experimento aleatorio? En principio, esta pregunta ha desvelado a muchos científicos de muchas maneras distintas. Las siguientes son tres posibles razones: (1) Desconocemos las leyes naturales que rigen el experimento, (2) conocemos dichas leyes pero son tan complejas que nos es imposible –o resulta indeseable– evaluarlas, (3) existe una indeterminación básica en el universo. La tercera razón es propia de la mecánica cuántica, en la que cada partícula se describe mediante una función de onda cuya magnitud al cuadrado representa la incertidumbre sobre su posición en cada instante. Las primeras dos razones, en cambio, hablan de nuestra ignorancia o nuestra incapacidad, lo que haría de la aleatoriedad un concepto subjetivo, que podría desaparecer con el desarrollo del conocimiento o de la tecnología. Una cuarta fuente de incertidumbre, aunque no de aleatoriedad, es la impredecibilidad de los sistemas dinámicos caóticos. A pesar de ser completamente determinísticos, presentan una sensibilidad a las condiciones iniciales tal que, si queremos predecir algo sobre ellos en un futuro mediano, deberíamos conocer las condiciones iniciales con una precisión absurdamente alta. A pesar del determinismo de estos sistemas, a veces conviene usar modelos probabilísticos para cuantificar la incertidumbre en sus estados futuros. Lo cierto es que, como muestran los experimentos *netstat* o *ftp*, ni siquiera el

más experto ingeniero conocedor de los más íntimos detalles de la implementación de cada protocolo de una red de comunicaciones a todos los niveles de su jerarquía funcional podría predecir los instantes en que cada usuario de la red generará demandas o la magnitud de esas demandas. En consecuencia, aunque una mente privilegiada con infinitos poderes divinos pudiera considerar una red de comunicaciones como un sistema determinístico, a nosotros, pobres mortales, nos toca aceptar nuestra incertidumbre sobre el comportamiento de la red y conformarnos con el hecho de que, al observar la red, estamos llevando a cabo un experimento aleatorio.

5. Espacio Muestral

El espacio muestral de un experimento aleatorio es el conjunto de todos los posibles resultados que podrían observarse en una realización del experimento,

$$\Omega = \{\omega : \omega \text{ es un posible resultado del experimento aleatorio}\}$$

Cuando definimos un experimento como un proceso de observación mediante el cual se selecciona un elemento de un conjunto de posibles resultados, queda claro que, si queremos especificar adecuadamente un experimento, lo primero que debemos describir con precisión es ese conjunto de posibles resultados. En este libro, como es costumbre en la mayoría de textos sobre probabilidades, denotaremos el espacio muestral mediante la letra griega mayúscula Ω (ómega) y sus elementos, de manera genérica, se denotarán mediante la correspondiente letra minúscula ω . Algunos ejemplos que ya se mencionaron en la definición 4 son:

1. Lanzar una moneda y ver qué lado queda hacia arriba: $\Omega = \{\text{cara, sello}\}$.
2. Lanzar un dado y contar los puntos en la cara que queda hacia arriba: $\Omega = \{1,2,3,4,5,6\}$.
3. Escoger una carta de la baraja de naipes: $\Omega = \{\text{picas, tréboles, corazones, diamantes}\} \times \{1,2,3,4,5,6,7,8,9,10,J,Q,K\}$, donde \times representa el producto cartesiano entre los dos conjuntos.
4. Verificar el estado de ocupación de un canal de comunicaciones: $\Omega = \{\text{libre, ocupado}\}$.
5. Medir la fracción de paquetes perdidos durante una hora en una red IP: $\Omega = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$.
6. Medir el retardo experimentado por un paquete de datos mientras transita por una red IP: $\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$.
7. Contar el número de canales PCM libres en un enlace E1: $\Omega = \{0,1,2,\dots,32\}$.
8. Mirar el estado de ocupación de cada uno de los canales en un enlace E1: $\Omega = \{\text{libre, ocupado}\}^{32}$. La potencia indica que se deben ejecutar 32 productos cartesianos del conjunto $\{\text{libre, ocupado}\}$ con síg mismo, con lo que se construye el conjunto de todas las cadenas de 32 símbolos en las que cada símbolo puede tomar uno de los valores *libre* u *ocupado*. Nótese que, aunque el experimento parece sencillo, la *cardinalidad* del espacio muestral (su número de elementos) es mayor a cuatro mil millones ($|\Omega| = 4.294'967.296$).
9. Determinar si un bit, transmitido sobre un canal de comunicaciones, llega correctamente al receptor en el otro extremo del canal: $\Omega = \{\text{si, no}\}$
10. Contar el número de transmisiones que requiere un paquete de datos hasta llegar correctamente a su destino: $\Omega = \{1,2,3,\dots\}$

11. Contar el número de bits recibidos con error en una trama de L bits que llega a través de un canal ruidoso: $\Omega = \{0,1,2,\dots,L\}$
12. Medir durante una hora la fracción de tiempo que un enlace de comunicaciones permanece ocupado: $\Omega = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$.
13. Contar el número de paquetes que llegan a un enrutador de una red de comunicaciones durante un período de una hora: $\Omega = \{0,1,2,\dots\}$.
14. Medir el tiempo que transcurre entre la llegada de dos paquetes consecutivos a un enrutador de una red de comunicaciones: $\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$.
15. etc.

Nótese que, aunque cada experimento aleatorio puede tener solamente un espacio muestral, la ignorancia del modelador respecto a detalles particulares del experimento le puede llevar a considerar espacios muestrales más grandes, lo cual no está mal mientras el espacio muestral supuesto por el modelador, Ω_m , contenga al espacio muestral verdadero, Ω_v . En efecto, en ese caso, simplemente los “posibles resultados” pertenecientes a $\Omega_m \cap \Omega_v^c$ nunca ocurrirán. Por ejemplo, en el experimento 7 supusimos $\Omega = \{0,1,2,\dots,32\}$ porque un enlace E1 tiene 32 canales TDM; sin embargo, un modelador con mayor *conocimiento a priori* podría saber que ese canal E1 hace parte de una red PCM para telefonía, en cuyo caso reduciría su espacio muestral a $\Omega = \{0,1,\dots,30\}$ al considerar que el canal 0 siempre está ocupado con bits de sincronización y el canal 16 siempre está ocupado con bits de señalización. Igualmente, en el experimento 6, algún modelador podría saber cuál es la mínima longitud de los paquetes, L bits, y la máxima capacidad de la ruta, C bps, con lo que podría reducir el espacio muestral a $\Omega = \{x \in \mathbb{R} : x \geq L/C\}$. Cabe anotar que, en muchas ocasiones, un modelador con una gran cantidad de conocimiento a priori que le permita encontrar un conjunto Ω_m muy cercano a Ω_v , puede decidir escoger un espacio muestral aún mayor a Ω_m con el único propósito de simplificar el tratamiento analítico posterior. Por ejemplo, como el número de paquetes que llegan en una hora es un número entero, un modelador podría saber que en el experimento 5 un espacio muestral más cercano al verdadero está contenido en $\{m/n \in \mathbb{Q}^+ : n > 0, m \leq n\}$, donde \mathbb{Q}^+ son los números racionales no negativos. Sin embargo, parece intuitivamente claro que podría ser más fácil considerar el espacio muestral constituido por el intervalo real $[0,1]$.

La Figura 16 muestra un diagrama de Venn que incluye el conjunto Ω compuesto por todos los posibles resultados de todos los posibles experimentos (¿cuán grande es este conjunto?) y, en él, algunos espacios muestrales asignados a un experimento particular, $\Omega_v \subset \Omega_1 \subset \Omega_2 \subset \Omega_3$ y Ω_4 . El verdadero espacio muestral, Ω_v , puede ser un conjunto muy complejo. Ω_1 es el espacio muestral que podría seleccionar un modelador juicioso con una gran cantidad de conocimiento a priori. Ω_2 es el espacio muestral que decidiría seleccionar este mismo modelador para facilitar el análisis posterior. Ω_3 es el espacio muestral seleccionado por otro modelador igualmente juicioso pero que tiene muy poco conocimiento a priori. Por último, Ω_4 es el espacio muestral que seleccionaría un modelador poco juicioso y muy desafortunado, pues no podrá llegar a ningún destino útil por haber empezado parándose sobre arenas movedizas.

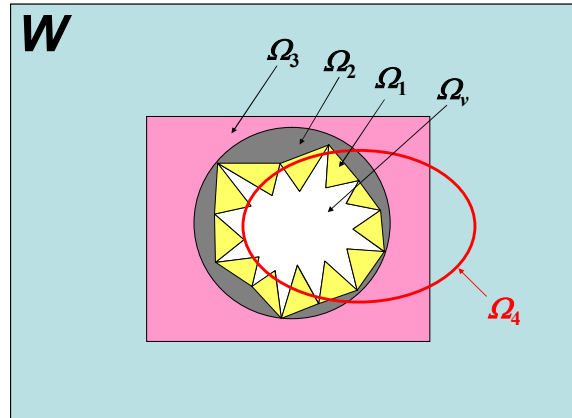


Figura 16. Algunos espacios muestrales asignados a un mismo experimento

6. Evento

Por lo pronto, diremos que un evento es un subconjunto del espacio muestral de un experimento aleatorio. Más adelante (después de la definición 14) restringiremos la definición de evento a un subconjunto "medible", esto es, un subconjunto al que se le pueda asignar una medida de probabilidad.

El evento $A \subset \Omega$ ocurre al realizar una instancia del experimento si el resultado obtenido, ω , pertenece a A , $\omega \in A$.

Supongamos, por ejemplo, que medimos la fracción de paquetes perdidos en una videoconferencia (experimento 5), de manera que los posibles resultados son $\Omega = \{x \in \mathbb{Q} : 0 \leq x \leq 1\}$. Si nos interesa satisfacer un requerimiento de calidad de servicio según el cual no se pueden perder más del 0.1% de los paquetes, deberíamos buscar que una fracción importante de las instancias del experimento correspondieran a elementos del subconjunto $A = \{x \in \Omega : x \leq 0.001\}$. De la misma manera, en cada uno de los experimentos propuestos en la definición 5 podemos definir algunos eventos apropiados:

1. Lanzar una moneda y ver qué lado queda hacia arriba: Los posibles eventos de interés incluyen a los subconjuntos unitarios de Ω , $A = \{cara\}$ y $B = \{sello\}$. No sobra recordar que, mientras *cara* es un posible resultado del experimento, esto es, un elemento de Ω , $\{cara\}$ es un subconjunto unitario de Ω . ¡Y es muy importante reconocer la diferencia! De otro lado, además de los eventos A y B mencionados antes hay otros dos posibles eventos: Ω (el evento cierto) y Φ , (vacío o el evento nulo), pues ellos dos siempre son subconjuntos de Ω .
2. Lanzar un dado y contar los puntos en la cara que queda hacia arriba: En el espacio muestral $\Omega = \{1,2,3,4,5,6\}$ están incluidos conjuntos como $A = \{hay\ más\ de\ tres\ puntos\} = \{4,5,6\}$ y $B = \{hay\ un\ número\ par\ de\ puntos\} = \{2,4,6\}$, así como $A \cap B = \{4,6\}$, por ejemplo.
3. Escoger una carta de la baraja de naipes: En el espacio muestral descrito antes están contenidos, por ejemplo, los eventos $A = \{Una\ figura\ de\ pinta\ roja\} = \{corazones, diamantes\} \times \{J, Q, K\}$ y $B = \{Un\ as\ negro\} = \{(picas, 1), (tréboles, 1)\}$.

4. Verificar el estado de ocupación de un canal de comunicaciones: Como en el ejemplo 1, los posibles eventos de interés son los subconjuntos unitarios $\{\text{libre}\}$ y $\{\text{ocupado}\}$, que son diferentes a los elementos de Ω , libre y ocupado .
5. Medir la fracción de paquetes perdidos durante una hora en una red IP: El espacio muestral es el conjunto de racionales en el intervalo $[0,1]$ de la recta real, donde podemos definir un evento A que dispararía una alarma en el centro de gestión de la red, $A = \{x \in \Omega : 0.1 \leq x\}$: ¡En la última hora se perdió más del 10% de los paquetes!
6. Medir el retardo experimentado por un paquete de voz mientras transita por una red IP dotada con mecanismos VoIP: En este caso, como un paquete que llegue con más de 100 ms (p.ej.) de retardo es descartado en el receptor, un evento de gran interés sería $A = \{x \in \Omega : x > 0.1\} = \{\text{El paquete no alcanza a ser reproducido en el receptor}\}$.
7. Contar el número de canales libres en un enlace E1: Si una videoconferencia requiere 384 kbps, un evento de interés podría ser $A = \{\text{Se puede establecer una videoconferencia}\} = \{6,7,8,\dots,32\}$.
8. Mirar el estado de ocupación de cada uno de los canales en un enlace E1 : $\Omega = \{\text{Libre}, \text{Ocupado}\}^{32}$. Si definimos 33 eventos diferentes $[X_i = \{\text{Hay } i \text{ canales libres}\}, i=0,1,2,\dots,32]$, estaríamos “reconstruyendo” el experimento 7. Sin embargo, mientras “16 canales libres” es un elemento del espacio muestral del experimento 7, en el experimento 8 se trata de un evento (un subconjunto del espacio muestral) compuesto por ¡601'080.390 elementos!
9. Determinar si un bit, transmitido sobre un canal de comunicaciones, llega correctamente al otro lado: $\Omega = \{\text{si}, \text{no}\}$. Como en el experimento 1, no tenemos muchos más eventos que los unitarios $\{\text{si}\}$ y $\{\text{no}\}$, aunque siempre podemos escoger también el evento cierto y el evento Nulo.
10. Contar el número de transmisiones (a través de un canal ruidoso) que requiere un paquete de datos hasta llegar correctamente a su destino: $\Omega = \mathbb{N} = \{1,2,3,\dots\}$. El evento $\{\text{No hay errores de transmisión}\}$ corresponde al subconjunto unitario $\{1\}$.
11. Contar el número de bits con errores en una trama de L bits que se recibe de un canal ruidoso: $\Omega = \{0,1,2,\dots,L\}$. El evento $\{\text{Es necesario retransmitir el paquete}\}$ corresponde al subconjunto $\{1,2,3,\dots,L\} = \{0\}^C$, donde el superíndice C indica complemento en Ω : Será necesario retransmitir el paquete si se daña al menos un bit.
12. Medir durante una hora la fracción de tiempo que un enlace de comunicaciones permanece ocupado: $\Omega = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$. Un posible criterio de gestión de la red podría interpretar el evento $\{x \in \Omega, x \geq 0.8\}$ como $\{\text{Es necesario dispersar el tráfico que cursa sobre el enlace}\}$.
13. Contar el número de paquetes que llegan a un enrutador de una red de comunicaciones durante una hora: $\Omega = \{0,1,2,\dots\}$. Si el enrutador es capaz de atender hasta μ paquetes por hora, un evento de sumo interés para el administrador de la red será $A = \{n \in \Omega : n > \mu\}$, pues la ocurrencia del evento A indica que el enrutador está experimentando congestión.
14. Medir el tiempo que transcurre entre la llegada de dos paquetes consecutivos a un enlace de una red de comunicaciones: $\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$. ¿Cuál sería el evento $A = \{\text{Cuando llega el segundo paquete, el primer paquete ya ha sido transmitido}\}$? La respuesta puede no ser fácil porque depende del estado del enlace (cuántos paquetes había en espera de servicio) cuando llegó el primer paquete, cuya observación constituye otro experimento aleatorio. Sin embargo, si la longitud mínima de los paquetes es L bits y la capacidad del enlace es C bps, sabemos con

seguridad que el evento que nos preguntan está contenido en otro evento mayor, $A \subseteq B = \{x \in \Omega : x > L/C\}$.

En esta definición hemos considerado "evento" como sinónimo de "sub-conjunto de Ω ". Más adelante, cuando definamos la medida de probabilidad (definición 14), distinguiremos entre subconjuntos medibles y subconjuntos no medibles según se les pueda asignar o no una medida de probabilidad. Entonces restringiremos el significado de evento a "subconjunto medible".

7. Frecuencia Relativa

Sea A un subconjunto del espacio muestral de un experimento aleatorio. Si repetimos N veces el experimento y observamos que en N_A de esas repeticiones se obtuvo un elemento de A , decimos que $f_N(A) = N_A/N$ es la frecuencia relativa del subconjunto A en esas N repeticiones del experimento.

Nótese que la notación es muy imprecisa pues $f_N(A)$ no es una función de A subindicada por N . En efecto, en una secuencia diferente de N repeticiones del mismo experimento podríamos obtener un valor distinto de $f_N(A)$. Por ejemplo, considérense las siguientes dos secuencias de 10 lanzadas de un dado:

Secuencia 1 : 2 4 2 6 1 5 3 6 3 3

Secuencia 2 : 6 4 2 5 3 1 5 3 6 4

Si observamos la frecuencia relativa del subconjunto $A = \{\text{el resultado es menor que cuatro}\} = \{1, 2, 3\}$ obtenemos que $f_{10}(A) = 0.6$ en la primera secuencia y $f_{10}(A) = 0.4$ en la segunda secuencia, mientras que, considerando las dos secuencias conjuntamente, obtenemos $f_{20}(A) = 0.5$. Con respecto al subconjunto $B = \{\text{el resultado es un número par}\} = \{2, 4, 6\}$, en cada secuencia individual se obtiene $f_{10}(B) = 0.5$ al igual que en la secuencia conjunta, $f_{20}(B) = 0.5$.

Así pues, calcular la frecuencia relativa de un subconjunto de posibles resultados, A , en N repeticiones de un experimento aleatorio resulta ser **otro experimento aleatorio** (un proceso de observación mediante el cual se selecciona un elemento del conjunto $\{0/N, 1/N, 2/N, \dots, N/N, \}$). Afortunadamente, en muchas ocasiones las frecuencias relativas observadas en diferentes secuencias de experimentos parecen converger a un número muy preciso a medida que el número de repeticiones aumenta en cada secuencia, como se menciona a continuación.

8. Regularidad Estadística

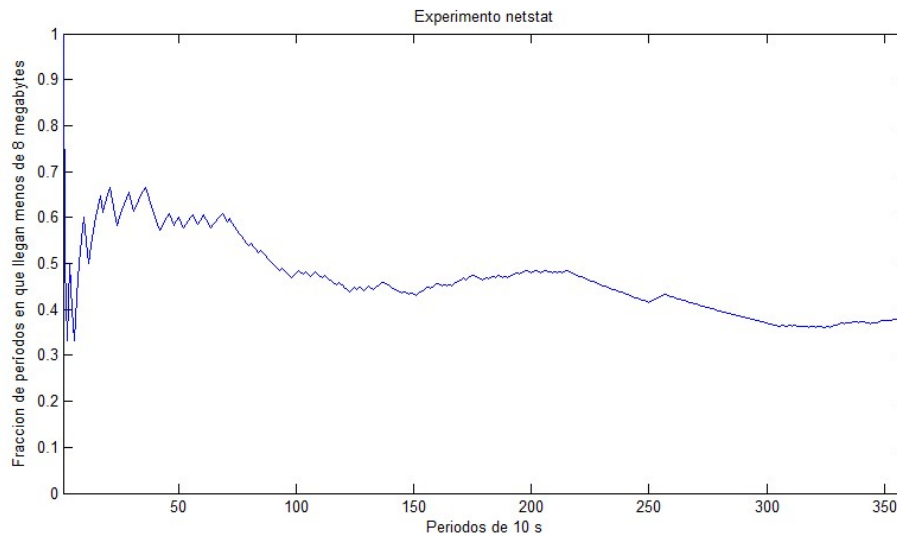
La regularidad estadística es la propiedad que tienen muchos experimentos aleatorios según la cual, al repetir el experimento un gran número de veces bajo condiciones constantes, la frecuencia relativa de los posibles eventos tiende a un valor preciso a medida que aumenta el número de repeticiones.

Aceptar con humildad nuestra incapacidad de predecir el comportamiento de una red de comunicaciones no quiere decir que debamos considerar imposible el diseño de dichas redes con estrictos requerimientos de desempeño. Al contrario, lo que debemos hacer (y lo que han hecho los ingenieros de redes de comunicaciones en los últimos 150 años) es tratar de cuantificar la incertidumbre para así poder usarla como una herramienta a nuestro favor. Afortunadamente, muchos experimentos aleatorios presentan cierta regularidad estadística que facilitan la cuantificación de nuestra incertidumbre.

Considérese, por ejemplo, el experimento netstat de la Figura 15. Supongamos que después de haber observado el número de bytes recibidos durante n períodos de 10 segundos medimos la fracción de períodos en los que llegaron más de 8 Mbytes. Esta fracción es la *frecuencia relativa* del evento $E = \{x \in \mathbb{N} : x < 8000000\}$,

$$f_n(E) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(x_i \in E)$$

donde x_i es el número de bytes recibidos en el i -ésimo período de 10 s y $\mathbb{1}(p)$ es la *función indicadora* de la proposición p , igual a 1 si la proposición p es cierta e igual a 0 si la proposición p es falsa. La Figura 17 muestra una gráfica de $f_n(E)$ vs n , en la que se puede apreciar cómo $f_n(E)$ parece tender a algún valor específico a medida que aumentamos el número de experimentos, n .



250 Figura 17. Aunque no podemos predecir el número de bytes que llegarán en el próximo período de 10 segundos, podemos afirmar que cerca del 38% del tiempo se reciben más de 8 Mbytes.

Es precisamente la regularidad estadística la que nos permite estudiar con rigurosidad los experimentos aleatorios que a diario tiene que realizar un ingeniero a cargo de una red de comunicaciones, pues ella nos permite saber que, *a la larga*, se pueden esperar comportamientos claramente predecibles. Por ejemplo, si la persona que estuvo navegando por la web durante los 60 minutos que duró el experimento netstat de la Figura 15 y de la Figura 17, sigue haciendo el mismo

tipo de consultas durante los siguientes 60 minutos, podríamos afirmar con un alto grado de certeza que “en cerca del 38% de los períodos de 10 s, se espera que lleguen más de 8 Mbytes”.

La teoría de la probabilidades pretende estudiar estas tendencias observadas en las estadísticas que se pueden asociar con un gran número de repeticiones de un experimento aleatorio, pero librándonos de términos imprecisos como “a la larga”, “se espera que”, “cerca de”, etc. Por ejemplo, la teoría de probabilidades querría que dijéramos que “con una confianza del 38%, en el próximo período de 10 s llegarán más de 8 Mbytes”. Así pues, es la regularidad estadística de muchos experimentos aleatorios la que le permite a la teoría de la probabilidad convertirse en una herramienta para cuantificar nuestra incertidumbre.

Es interesante notar que, cuando calculamos estadísticas como la frecuencia relativa de un evento, nos referimos a instancias ya realizadas del experimento aleatorio, sobre las cuales no hay ninguna incertidumbre. La parte más interesante es cuando queremos hacer “inferencia estadística”, esto es, cuando queremos usar esas estadísticas para estudiar la incertidumbre que existe sobre instancias por realizar del experimento aleatorio. Es entonces cuando hablamos de la teoría de probabilidades.

9. Conceptos básicos sobre Conjuntos

*Un conjunto es una colección de elementos. El conjunto de todos los elementos de interés en un contexto dado es el **conjunto universal**, Ω . Si a es un elemento del conjunto A , decimos que $a \in A$ (a “pertenece a” A). En otro caso, decimos que $a \notin A$ (a “no pertenece a” A). Un conjunto está completamente determinado por sus elementos, al punto que dos conjuntos A y B son iguales ($A=B$), si tienen la propiedad que $a \in A \Leftrightarrow a \in B$ (a pertenece a A “si y sólo si” a pertenece a B). La propiedad $a \in A \Rightarrow a \in B$ (“Si” a pertenece a A “entonces” a pertenece a B) define la relación $A \subset B$ (A “es un subconjunto de” B o A “está contenido en” B). En consecuencia, la igualdad entre los conjuntos A y B corresponde a las propiedades $A \subset B$ y $B \subset A$. Siempre es cierto que $A \subset A$. Si $A \subset B$ y $B \subset C$, entonces $A \subset C$. Podemos describir un conjunto especificando la propiedad que determina la pertenencia de un elemento: $A = \{a \in \Omega : P(a)\}$ (“el conjunto A está compuesto por los elementos de Ω que satisfacen la propiedad P ”). Como un conjunto está determinado exclusivamente por sus elementos, existe un único conjunto que no tiene elementos, el **conjunto vacío**, $\phi = \{a \in \Omega : a \neq a\}$. Para cualquier conjunto A , $\phi \subset A$. Si a es un elemento de A ($a \in A$), $\{a\}$ es un **subconjunto unitario** de A ($\{a\} \subset A$). Siempre es cierto que $a \in \{a\}$ aunque sería muy raro que $a \in a$ (Aquí evitaremos conjuntos que pertenezcan a sí mismos para no perdernos en paradojas como la **paradoja de Russell**: Si $A = \{X : X \notin X\}$, ¿ $A \in A$?).*

*La **unión** de dos conjuntos A y B , que se representa mediante $A \cup B$, es el conjunto de los elementos de Ω que pertenecen a A o que pertenecen a B (o que pertenecen a ambos): $A \cup B = \{x \in \Omega : x \in A \vee x \in B\}$. La **intersección** de dos conjuntos A y B , que se representa mediante $A \cap B$, es el conjunto de los elementos de Ω que pertenecen a A y que pertenecen a B : $A \cap B = \{x \in \Omega : x \in A \wedge x \in B\}$. Ambas operaciones son conmutativas y asociativas*

$(A \cup B) = B \cup A$, $A \cap B = B \cap A$, $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$, además de ser distributivas entre ellas $(A \cap (B \cup C)) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. Estas propiedades se extienden a cualquier secuencia de subconjuntos de Ω : $B \cap (\cup_n A_n) = \cup_n (A_n \cap B)$, $B \cup (\cap_n A_n) = \cap_n (A_n \cup B)$, donde $\cup_n A_n = A_1 \cup A_2 \cup A_3 \cup \dots$ y $\cap_n A_n = A_1 \cap A_2 \cap A_3 \cap \dots$. Claramente, $A \subset B \Rightarrow A \cap B = A$ y $A \cup B = B$. Dos conjuntos A y B son **mutuamente excluyentes** si $A \cap B = \phi$.

El **complemento de A** , A^C , está compuesto por los elementos de Ω que no pertenecen a A : $A^C = \{b \in \Omega : b \notin A\}$. Evidentemente, $(A^C)^C = A$, $\phi^C = \Omega$, $\Omega^C = \phi$, $A \cap A^C = \phi$, $A \cup A^C = \Omega$. Las **leyes de DeMorgan** establecen que $(A \cup B)^C = A^C \cap B^C$ y que $(A \cap B)^C = A^C \cup B^C$. Estas leyes se extienden a cualquier secuencia de subconjuntos de Ω : $(\cup_n A_n)^C = \cap_n A_n^C$, $(\cap_n A_n)^C = \cup_n A_n^C$. Con uniones, intersecciones y complementos podemos definir la **diferencia** $A \setminus B = A \cap B^C = \{a \in A : a \notin B\}$ (el conjunto de los elementos de A que no pertenecen a B) y la **diferencia simétrica** $A \Delta B = (A \cap B^C) \cup (A^C \cap B) = \{a \in A \cup B : a \notin (A \cap B)\}$ (el conjunto de los elementos que pertenecen a A ó a B , pero no a ambos).

Otra operación importante entre conjuntos es el **producto cartesiano**, $A \times B = \{(a, b) : a \in A, b \in B\}$. Cada elemento del conjunto $A \times B$ es una **pareja ordenada** en la que el primer componente es un elemento de A y el segundo es un elemento de B , y $A \times B$ es el conjunto de todas esas parejas. Una **función $f: A \rightarrow B$** se construye como un subconjunto del producto cartesiano entre el dominio A y el rango B , donde a cada elemento de A le corresponde uno y sólo uno de los elementos de B , $(f: A \rightarrow B) \subset (A \times B)$.

La **cardinalidad** $|A|$ de un conjunto A es el número de elementos que pertenecen a A . Sólo hay un conjunto con cardinalidad cero, el conjunto vacío ($|\phi| = 0$). Algunos conjuntos pueden tener una cardinalidad infinita contable (como el conjunto de los números racionales) o no contable (como el conjunto de los números reales). A veces existen resultados intuitivos con conjuntos finitos que no se pueden generalizar a conjuntos infinitos.

Un conjunto finito con n elementos distintos se puede ordenar de $n!$ maneras diferentes. Si sólo queremos k de ellos, hay $n!/(n-k)!$ maneras de escogerlos en diferentes órdenes, esto es, hay $n!/(n-k)!$ permutaciones de k elementos entre n . Si cualquier orden en que los escojamos da igual, obtenemos que $k!$ de esas permutaciones son equivalente, de manera de hay $n!/[(n-k)!k!]$ combinaciones de k elementos entre un conjunto de n elementos.

Una secuencia de subconjuntos de Ω , $\{A_n, n=1,2,\dots\}$ es una **partición** de Ω si $A_i \cap A_j = \phi \forall i \neq j$ y $\cup_n A_n = \Omega$. Una secuencia $\{B_n, n=1,2,\dots\}$ es **decreciente** si, $\forall n, B_{n+1} \subseteq B_n$. Una secuencia $\{C_n, n=1,2,\dots\}$ es **creciente** si, $\forall n, C_n \subseteq C_{n+1}$. El **límite de una secuencia decreciente** $\{B_n, n=1,2,\dots\}$ es $\lim_{n \rightarrow \infty} B_n = \cap_{i=1, \dots, \infty} B_i$. El **límite de una secuencia creciente** $\{C_n, n=1,2,\dots\}$ es $\lim_{n \rightarrow \infty} C_n = \cup_{i=1, \dots, \infty} C_i$. Dada cualquier secuencia de conjuntos $\{A_n\}$,

podemos construir una secuencia decreciente $D_n = \bigcap_{i=1 \dots n} A_i$ y una secuencia creciente $C_n = \bigcup_{i=1 \dots n} A_i$. D_n es el conjunto más grande que está contenido en todos los $\{A_i, i=1 \dots n\}$ mientras C_n es el conjunto más pequeño que contiene a todos los $\{A_i, i=1 \dots n\}$. Podemos tomar los respectivos límites $D = \lim_{n \rightarrow \infty} D_n = \bigcap_{i=1 \dots \infty} D_i$ y $C = \lim_{n \rightarrow \infty} C_n = \bigcup_{i=1 \dots \infty} C_i$, donde D es el conjunto más grande contenido en todos los conjuntos de la secuencia y C es el conjunto más pequeño que contiene a todos los conjuntos de la secuencia. Cuando consideramos una secuencia infinita de eventos $A_1, A_2, A_3, \dots \subset \Omega$, podemos construir otras dos secuencias, una decreciente y una creciente así:

$$\left\{ S_m = \sup_{n \geq m} A_n = \bigcup_{n=m}^{\infty} A_n, \quad m=1,2,3,\dots \right\} \quad \left\{ I_m = \inf_{n \geq m} A_n = \bigcap_{n=m}^{\infty} A_n, \quad m=1,2,3,\dots \right\}$$

S_m es el **supremo** de $\{A_m, A_{m+1}, A_{m+2}, \dots\}$ (el conjunto más pequeño que contiene a todos los $A_{n \geq m}$) e I_m es el **ínfimo** de $\{A_m, A_{m+1}, A_{m+2}, \dots\}$ (el conjunto más grande contenido en todos los $A_{n \geq m}$), de manera que $\{S_m, m=1,2,3,\dots\}$ es una secuencia decreciente y $\{I_m, m=1,2,3,\dots\}$ es una secuencia creciente. A medida que m tiende a infinito, podemos construir el límite de cada secuencia,

$$A = \limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k \quad B = \liminf_{n \rightarrow \infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k$$

Cuando ocurre el evento A en un experimento aleatorio con espacio muestral Ω , decimos que ocurrió un número infinito de eventos de la secuencia $\{A_n, n=1,2,\dots\}$. Cuando ocurre el evento B decimos que ocurrieron todos los eventos $\{A_n, n=1,2,\dots\}$, con la posible excepción de un número finito de ellos,

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k = \{ \omega \in \Omega : \forall n \exists k > n : \omega \in A_k \}$$

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k = \{ \omega \in \Omega : \exists n : \forall k > n : \omega \in A_k \}$$

Claramente, $\lim_{n \rightarrow \infty} \inf A_n \subseteq \lim_{n \rightarrow \infty} \sup A_n$. Si también ocurre que $\lim_{n \rightarrow \infty} \sup A_n \subseteq \lim_{n \rightarrow \infty} \inf A_n$, entonces la secuencia de eventos $A_1, A_2, A_3, \dots \subset \Omega$ tiene un límite, el evento L , dado por

$$L = \lim_{n \rightarrow \infty} \sup A_n = \lim_{n \rightarrow \infty} \inf A_n$$

Una manera muy útil de visualizar las relaciones lógicas entre una colección finita de conjuntos es el **diagrama de Venn**. El espacio muestral se representa como un rectángulo y los eventos se presentan como áreas sombreadas dentro de él. Cuando hay demasiados conjuntos por representar en un diagrama de Venn se puede recurrir al mapa de Karnaugh, que representa en una tabla la partición constituida por todos los eventos de interés.

Como podemos inferir de las definiciones 4 a 8, el lenguaje de la teoría de conjuntos es fundamental para la teoría de probabilidades. El concepto de conjunto es una "noción primitiva", es decir, un conjunto no puede definirse en términos de objetos definidos previamente. Algo semejante ocurre

con el concepto de punto, definido en la geometría Euclidiana como "lo que no tiene partes": La definición intuitiva de punto como una ubicación única del espacio que no tiene longitud, área, volumen ni ningún otro atributo dimensional, es el concepto primitivo sobre el cual se construye toda la geometría Euclidiana. Igualmente, la definición intuitiva de conjunto como una "colección de objetos" es el concepto primitivo sobre el cual se construye toda la teoría de conjuntos y, sobre ella, casi toda la matemática. Un niño en pre-escolar entiende fácilmente qué es un conjunto de lápices, así como un niño en primaria entiende qué es el conjunto de los mamíferos y un ingeniero entiende qué es el conjunto de los números reales (o al menos eso creemos). Cantor, Frege y otros matemáticos de finales del siglo XIX supusieron que estos conceptos ingenuos de la teoría de conjuntos eran suficientemente básicos, elementales e intuitivos como para construir toda la matemática sobre ella. Y, hasta cierto punto, Cantor y Frege no se equivocaron: La mayoría de las matemáticas de hoy se basan en la teoría de conjuntos, aunque dicha teoría no es tan elemental, tan básica ni tan intuitiva como se creyó en un principio, pues un uso poco juicioso del concepto de conjunto como "colección de objetos" puede conducir a contradicciones.

Básicamente, Cantor proponía que: (1) Un conjunto es una colección de objetos (los elementos del conjunto) que cumplen con cierta propiedad y, por consiguiente, el conjunto queda definido por la propiedad que determina a sus elementos. (2) Se pueden definir clases de conjuntos: conjuntos cuyos elementos son, a su vez, conjuntos que satisfacen cierta propiedad. (3) Dos conjuntos son iguales si tienen los mismos elementos. Sobre estas tres suposiciones, Frege construyó una rigurosa formalidad basada en sus propios desarrollos de lógica matemática y de lenguajes formales. Cuando, en 1902, Frege había culminado su monumental obra "Las Leyes Fundamentales de la Aritmética", en la que re-escribía toda la aritmética con base en la lógica y la teoría de conjuntos, Bertrand Russell le indicó una grave inconsistencia en su sistema lógico, la paradoja de Russell: ¿El conjunto de los conjuntos que no pertenecen a sí mismos pertenece a sí mismo? Si no, pertenece al tipo de conjuntos que no pertenecen a sí mismos y, por lo tanto, pertenece a sí mismo. Si sí, pertenece al tipo de conjuntos que pertenecen a sí mismos y, por lo tanto, no pertenece a sí mismo. Es decir, este conjunto pertenece a sí mismo sólo si no pertenece a sí mismo. Aunque son raros, hay conjuntos que pertenecen a sí mismos. Por ejemplo, los principios de Cantor permiten definir el conjunto $U = \{a : a=a\}$ o "el conjunto de todo", "el universo", para el cual $U \in U$, obviamente. El conjunto $C = \{a : a \text{ es un concepto matemático}\}$ es, en sí mismo, un concepto matemático, de manera que $C \in C$. El mismo Russell explica su paradoja más claramente: Sólo hay un barbero en la ciudad, quien dice que afeitará sólo a todos aquellos que no se afeiten a sí mismos. ¿Quién afeitará al barbero? Posteriormente Russell reformuló la aritmética sobre una nueva teoría de conjuntos en la que no se permiten los conjuntos que sean elementos de sí mismos pero, igualmente, Gödel encontró una inconsistencia insalvable en la formulación de la aritmética de Russell (el teorema de la incompletitud, que demuestra que en la formulación de Russell de la aritmética –y en cualquier otra formulación que sea recursiva y consistente– hay problemas indecidibles dentro del sistema, como el teorema que dice "este teorema no es demostrable").

En fin, la teoría de conjuntos debió desarrollarse cuidadosamente para evitar el tipo de contradicciones que aparecen fundamentalmente cuando se consideran conjuntos "demasiado grandes" como, por ejemplo, el conjunto de todas las cosas ($U = \{a : a = a\}$). Los matemáticos

Zermelo y Fraenkel redujeron la teoría de conjuntos a un sistema axiomático restringido que no permite paradojas como la de Russell. Existen otras formalizaciones axiomáticas exitosas como la propuesta posteriormente por John von Neumann, Paul Bernays y Kurt Gödel (NBG, una extensión de la de ZF). Ambas teorías se basan en unos principios básicos (axiomas) y unas reglas precisas de obtención de teoremas a partir de axiomas o teoremas previamente demostrados, de manera que no se presente ninguna contradicción (Al menos hasta el día de hoy no se ha encontrado ninguna contradicción... ¡pero aún no se sabe!). Sin embargo, a diferencia de Zermelo y Fraenkel, que siguen usando el conjunto como noción primitiva, Neumann, Bernays y Gödel no usan el concepto de conjunto como noción primitiva sino el concepto de clase: La clase es una "colección de objetos" y "un objeto pertenece a una clase". Un conjunto, en cambio, es un tipo especial de clase que está contenida en alguna otra clase. Así $U = \{a: a=a\}$ no es un conjunto sino una clase y a ella aplican sólo algunos axiomas NBG, pues otros axiomas aplican sólo a las clases que son conjuntos. De esta forma se evitan las paradojas conocidas hasta ahora.

Si evitamos los experimentos aleatorios en los que el espacio muestral mismo sea un posible resultado del experimento (¿el lector puede imaginar un experimento así?), no necesitaremos estudiar los formalismos axiomáticos de teorías de conjuntos. En cambio, nos basaremos en los conceptos intuitivos mencionados al comienzo como resumen de esta definición 9, y de los cuales no daremos más detalles que el ejemplo de operaciones básicas en diagramas de Venn de la Figura 18, que se pueden asociar con las operaciones lógicas entre predicados que involucran la ocurrencia de eventos, como muestra la Figura 19. De hecho, como se mencionó, una diagrama de Venn y un mapa de Karnaugh son equivalentes, como muestra la Figura 20. Las secuencias infinitas y sus límites los tratamos en las definiciones 23, 24 y 25.

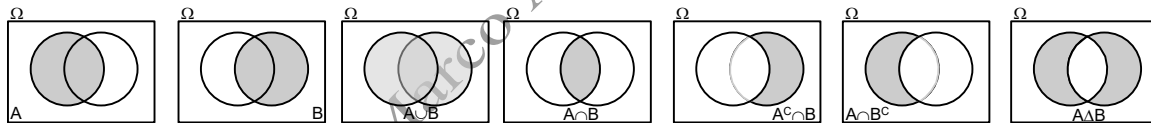


Figura 18. Diagram de Venn de algunas operaciones básicas entre dos conjuntos

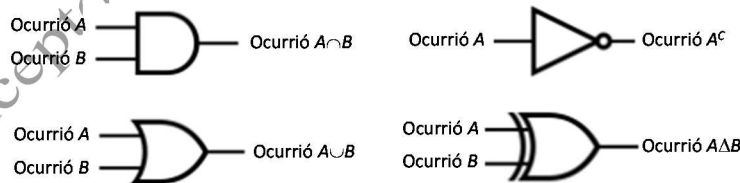


Figura 19. Operaciones entre conjuntos y sentencias lógicas sobre la ocurrencia de eventos

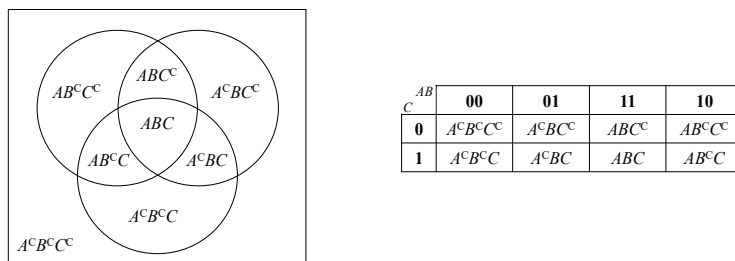


Figura 20. Comparación entre el diagram de Venn de tres subconjuntos y el mapa de Karnaugh de tres proposiciones lógicas (la ocurrencia o no ocurrencia de cada uno de los tres eventos asociados)

10. Conjunto Potencia de $\Omega \{0,1\}^\Omega$

El Conjunto Potencia de Ω es el conjunto de todos los posibles eventos, esto es, la clase de conjuntos conformada por todos los subconjuntos contenidos en Ω ,

$$\{0,1\}^\Omega = \{A : A \subseteq \Omega\}.$$

En aquellos experimentos aleatorios en los que el espacio muestral tiene una cardinalidad finita, es legítimo pensar en enumerar todos los posibles eventos que pueden ocurrir, esto es, todos los posibles subconjuntos de Ω . Para construir esta clase de conjuntos basta con considerar todas las secuencias binarias de $|\Omega|$ bits, donde $|\Omega|$ es la cardinalidad de Ω , de manera que a cada posición en la secuencia le corresponde un elemento de Ω . Así, con cada secuencia construimos un subconjunto conformado por los elementos asociados con un uno en la posición correspondiente de la secuencia. Por ejemplo a la secuencia 0, compuesta por $|\Omega|$ ceros, le corresponde el conjunto vacío, que siempre es un subconjunto de cualquier conjunto; a la secuencia 2^{i-1} , con $i \in \{1, 2, \dots, |\Omega|\}$, compuesta por $|\Omega|-1$ ceros y un uno en la posición i , le corresponde el evento unitario $\{\omega_i\}$; a la secuencia $2^{i-1} + 2^{j-1}$, con $i, j \in \{1, 2, \dots, |\Omega|\}$, $i \neq j$, compuesta por $|\Omega|-2$ ceros y dos unos en las posiciones i y j , le corresponde el evento binario $\{\omega_i, \omega_j\}$; a la secuencia $2^{|\Omega|}-1$, compuesta por $|\Omega|$ unos, le corresponde el espacio muestral mismo que, por definición, es un subconjunto de sí mismo. Debido a esta metodología de construcción, es razonable que al conjunto potencia del espacio muestral Ω se le denote como $\{0,1\}^\Omega$. Más aún, como en $\{0,1\}^\Omega$ hay un conjunto vacío, $|\Omega|$ conjuntos unitarios, $\binom{|\Omega|}{2}$ conjuntos binarios – donde $\binom{m}{k}$ es el número de combinaciones de k elementos escogidos entre m posibles –, $\binom{|\Omega|}{3}$ conjuntos ternarios, etc., la cardinalidad de $\{0,1\}^\Omega$ es $|\{0,1\}^\Omega| = \sum_{n=0}^{|\Omega|} \binom{|\Omega|}{n} = 2^{|\Omega|}$.

En el experimento 4 de la definición 5, por ejemplo, en el que verificamos el estado de ocupación de un canal de comunicaciones, solamente hay dos posibles resultados, por lo que tenemos solamente cuatro posibles eventos:

<i>Libre</i>	<i>Ocupado</i>	<i>Evento</i>
0	0	Φ
0	1	$\{Ocupado\}$
1	0	$\{Libre\}$
1	1	Ω

Pero si viéramos el estado de ocupación de dos canales, considerando cada uno individualmente, tendríamos 16 posibles eventos:

<i>(libre, libre)</i>	<i>(libre, ocupado)</i>	<i>(ocupado, libre)</i>	<i>(ocupado, ocupado)</i>	<i>Evento</i>
0	0	0	0	Φ
0	0	0	1	$\{No\ hay\ canales\ libres\}$
0	0	1	0	$\{(ocupado, libre)\}$

0	0	1	1	{El primer canal está ocupado}
0	1	0	0	{(libre, ocupado)}
0	1	0	1	{El segundo canal está ocupado}
0	1	1	0	{Sólo hay un canal libre}
0	1	1	1	{Al menos un canal está ocupado}
1	0	0	0	{No hay canales ocupados}
1	0	0	1	{Ambos canales están libres o ambos están ocupados}
1	0	1	0	{El segundo canal está libre}
1	0	1	1	{El primer canal está ocupado o el segundo está libre}
1	1	0	0	{El primer canal está libre}
1	1	0	1	{El primer canal está libre o el segundo está ocupado}
1	1	1	0	{Al menos un canal está libre}
1	1	1	1	Ω

Con tres canales tendríamos 256 posibles eventos y con cuatro canales deberíamos considerar 65536 eventos... En el experimento 8, por ejemplo, en el que inocentemente queremos ver el estado de ocupación de cada canal en un enlace E1, tendríamos $2^{4 \cdot 294 \cdot 967 \cdot 296}$ posibles eventos, ¡más de $10^{1 \cdot 292 \cdot 913 \cdot 986}$ eventos (un uno seguido por mil trescientos millones de ceros)! Para hacernos a una idea de la cardinalidad de este conjunto, consideremos que en el universo hay del orden de 2^{265} átomos. Si nos regalan un átomo por cada subconjunto de este espacio muestral, necesitaríamos $2^{4 \cdot 294 \cdot 967 \cdot 031}$ universos como éste. En ese número de universos es altamente probable encontrar uno idéntico al nuestro, excepto porque nuestro amable lector tendría otro color de ojos ¿Cómo nos pudimos meter en un problema tan grande si sólo queríamos monitorear un simple enlace E1?

De hecho, aunque observar los 32 canales de un inofensivo enlace E1 puede generar un número mucho más que astronómicamente grande de posibles eventos, todavía se trata de un conjunto describible (tanto que podemos contar cada uno de los eventos). Pero, ¿podría el lector imaginarse el conjunto potencia del experimento 10? Los más de $10^{1 \cdot 000 \cdot 000 \cdot 000}$ posibles eventos del experimento 8 siguen siendo un número infinitesimalmente pequeño de eventos en comparación con la cardinalidad del espacio muestral del número de transmisiones que se deben hacer para que una trama llegue bien a su destino, pues ese espacio muestral ya tiene un número infinito de elementos, aunque sea "el infinito más pequeño", \aleph_0 , que es la cantidad de números naturales (un conjunto tiene cardinalidad infinita si se puede poner en correspondencia biunívoca con un subconjunto propio de si mismo. Por ejemplo, los naturales son infinitos porque se pueden poner en correspondencia biunívoca con los pares y los reales son infinitos porque se pueden poner en correspondencia biunívoca con el intervalo unitario). Como nos demuestra la paradoja del "Hotel de Hilbert"⁵, $\aleph_0 = \aleph_0 + n = n\aleph_0 = (\aleph_0)^n \forall n \in \mathbb{N}$.

⁵ El único hotel de un pueblo pequeño tiene infinitas habitaciones ($|\mathbb{N}| = \aleph_0$ habitaciones). En el pueblo se juega un campeonato de fútbol en el que cada equipo tiene un número infinito de jugadores. Cuando llega el bus con el primer equipo, se ubica al jugador n en la habitación $n \forall n \in \mathbb{N}$, de manera que todas las habitaciones quedan

Pero, como demostró Georg Cantor, el número de subconjuntos que se pueden construir con los números naturales es mayor que el número de números naturales, $\aleph_0 < \aleph_1 = 2^{\aleph_0}$, donde \aleph_1 es la cantidad de números reales que existen⁶. El conjunto potencia del espacio muestral del experimento 10 tiene tantos eventos como números reales existen!

Más aún, ¿podría el lector imaginar el conjunto potencia del experimento 14? ¡Es el conjunto de todos los subconjuntos que se pueden formar con los números reales no negativos! Si los famélicos 32 canales de un enlace E1 lograrán atemorizarnos de hoy en adelante cada vez que pasemos cerca de la pequeña PBX de la oficina, y temblaremos de terror al tener que contar el número de transmisiones que requiere una trama ¿qué podría hacernos un experimento cuyo espacio muestral sea el conjunto de los números reales? El número de subconjuntos que se pueden formar con los números reales es $\aleph_2 = 2^{\aleph_1} > \aleph_1$. Grandes matemáticos como Bolzano, Cauchy, Weierstrass, Dedekind y Cantor han estudiado estos “monstruos matemáticos”, los “trans-finitos”, algunos de ellos con apreciables consecuencias en su salud mental. Como nos preocupa la salud mental de nuestros lectores, resulta conveniente definir el siguiente concepto, campo- σ de eventos. Esto es, en vez de pretender que se atemorice la próxima vez que vaya a revisar el PBX de la oficina, sólo queremos motivar al lector a seleccionar un conjunto razonablemente pequeño de eventos de interés cada vez que decida modelar un experimento aleatorio (donde “pequeño” puede ser \aleph_0 ó \aleph_1 , pero no \aleph_2 ni ningún otro trans-finito mayor a \aleph_1).

11. Campo- σ de Eventos

Un Campo de Eventos, \mathcal{F} , es una clase de subconjuntos de Ω que satisface los siguientes axiomas: (1) \mathcal{F} es no vacío, (2) si $A \subset \Omega$ es tal que $A \in \mathcal{F}$, $A^c \in \mathcal{F}$, (3) si $A, B \subset \Omega$ son tales que $A, B \in \mathcal{F}$, $A \cup B \in \mathcal{F}$. Un campo- σ de eventos es un campo contablemente aditivo, esto es, que satisface la condición adicional (3^a) si $\{A_n \in \mathcal{F}, n=1, 2, \dots\}$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

ocupadas. Al final se baja el entrenador... ¿dónde ubicar al entrenador? Se le solicita al jugador que está en la habitación n que, por favor, se pase a la habitación $n+1 \forall n \in \mathbb{N}$, con eso todos los jugadores siguen con una habitación pero se desocupa la habitación 1 para el entrenador: $\aleph_0 = \aleph_0 + 1 = \aleph_0 + k \forall k \in \mathbb{N}$. Pero resulta que cada jugador trajo a su mamá... ¿dónde ubicamos a las mamás? Se le solicita al jugador que está en la habitación n que, por favor, se pase a la habitación $2n \forall n \in \mathbb{N}$, con eso todos siguen con una habitación pero se desocupan las habitaciones impares para las mamás: $\aleph_0 = 2\aleph_0 = k\aleph_0 \forall k \in \mathbb{N}$. Pero resulta que el campeonato se juega entre un número infinito de equipos y empiezan a llegar los buses con los demás equipos. ¿dónde ponemos a todos esos jugadores? Le pedimos al jugador que está en la habitación n que se pase a la habitación 2^n y ubicamos al m -ésimo jugador del r -ésimo equipo en la habitación $p(r)^m$, donde $p(r)$ es el r -ésimo número primo mayor a 2: $\aleph_0 = \aleph_0^2 = \aleph_0^k \forall k \in \mathbb{N}$. Nótese que ahora todos los jugadores de todos los equipos tienen su habitación pero, a diferencia de cuando acomodamos sólo al primer equipo y llenamos el hotel, ¡ahora tenemos un número infinito de habitaciones desocupadas (1, 6, 10, 12, 14, 15, 18, 20, 21, 22, 24...)!

⁶ Si podemos contar los puntos en el intervalo unitario $[0,1]$, los tendremos a todos en una correspondencia biunívoca con los naturales en una lista $\{s_1, s_2, s_3, \dots\}$, donde el i -ésimo número es de la forma $s_i = 0.x_{i1}x_{i2}x_{i3}x_{i4} \dots$ siendo x_{ij} el j -ésimo dígito de la expansión binaria de s_i . Pero esa lista es imposible porque en ella hará falta el número $0.y_1y_2y_3y_4 \dots$ con $y_1 \neq x_{11}, y_2 \neq x_{22}, y_3 \neq x_{33}, \dots$. En consecuencia, el número de puntos en el intervalo unitario es mayor que el número de naturales, $\aleph_1 > \aleph_0$. \aleph_1 es el número de números reales, que es igual al número de secuencias infinitas de unos y ceros, que es el mismo número de subconjuntos de los naturales, $\aleph_1 = 2^{\aleph_0}$.

La idea es que más adelante vamos a definir la probabilidad como una función que le asigna una medida real a cada evento de interés. Pero una función no queda bien definida si no especificamos claramente su rango y su dominio. Y, como vimos en la definición 10, no podemos especificar como dominio el conjunto de todos los posibles eventos, pues en muchos casos ese conjunto puede ser monstruoso. Sólo cuando el espacio muestral tiene cardinalidad finita (¡y pequeña!), es posible considerar el conjunto de todos los eventos, el cual es un campo- σ , evidentemente. Pero tampoco podemos seleccionar algunos pocos eventos de interés e ignorar el resto si no le damos una estructura al dominio correspondiente, con el que evitemos llegar rápidamente a inconsistencias. Si nos interesa el evento $A \subseteq \Omega$, ¿cómo no nos podría interesar el evento $A^C = \{\text{No sucede } A\}$? O si nos interesan los eventos A y $B \subseteq \Omega$, ¿cómo no nos podría interesar el evento $A \cup B = \{\text{sucede por lo menos uno de los dos eventos}\}$? Al cerrar el campo de eventos sobre las uniones y los complementos, estamos incluyendo en él todos los eventos asociados con los eventos de interés definidos originalmente, con lo cual podemos asignar medidas de probabilidad a cada evento sin preocuparnos por inconsistencias.

Podemos deducir algunas propiedades adicionales de un campo- σ de eventos a partir de los axiomas que lo definen. Por ejemplo,

1. $\Omega \in \mathcal{F}$.
En efecto, como \mathcal{F} es no-vacío, debe contener al menos un evento A y, por el segundo axioma, también debe contener a A^C . El tercer axioma requiere que la unión de cualquier par de miembros de \mathcal{F} pertenezca también a \mathcal{F} , por lo que $A \cup A^C = \Omega \in \mathcal{F}$.
2. $\Phi \in \mathcal{F}$.
Esta propiedad surge de aplicar el segundo axioma a la propiedad anterior.
3. Si $A \in \mathcal{F}$ y $B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$.
En efecto, por el segundo axioma $A^C \in \mathcal{F}$ y $B^C \in \mathcal{F}$, por lo que el segundo axioma asegura que $A^C \cup B^C \in \mathcal{F}$ y, aplicando nuevamente el segundo axioma, $(A^C \cup B^C)^C = A \cap B \in \mathcal{F}$ (ley de DeMorgan).
4. Similarmente, usando los axiomas 2 y 3^a, podemos decir que si $\{A_n \in \mathcal{F}, n=1,2,\dots\}$, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Así pues, el campo- σ contiene todos los complementos, intersecciones numerables y uniones numerables de cada uno de los conjuntos que lo componen. La virtud de esta construcción es que, con ella, podemos desarrollar todas las funciones lógicas Booleanas sobre los eventos de interés, lo cual nos da la coherencia que necesitamos para expresarnos de manera lógica respecto a los eventos sin salirnos de nuestro campo de eventos, pues estos constituyen una estructura lógica cerrada para la negación, la conjunción y la disyunción. Si nos interesan los eventos $A, B, C \subseteq \Omega$, ¿cómo no nos podría interesar el evento $A \cap B = \{\text{sucedan ambos eventos}\}$? ¿O el evento $A \cap B^C = \{\text{sucede } A \text{ pero no sucede } B\}$? ¿O el evento $(A \cap C^C) \cup (B \cap C \cap A^C) = \{\text{Sucede } A \text{ pero no sucede } C \text{ o suceden } B \text{ y } C \text{ pero no sucede } A\}$? En un campo sigma ocurren todas las operaciones lógicas posibles sobre la ocurrencia o no de cualquier conjunto contable de eventos de interés.

El campo- σ trivial, el más pequeño posible, es $\{\Phi, \Omega\}$. Otro campo sigma trivial, el más grande posible, es el conjunto potencia $\{0,1\}^{\Omega}$. Si sólo nos interesa saber si ocurrió o no un único evento $A \subseteq \Omega$, un campo sigma adecuado es $\{\Phi, A, A^C, \Omega\}$. El conjunto de todas las uniones de conjuntos en una partición contable de Ω es un campo sigma. Si Ω es un conjunto infinito no contable, el conjunto

de subconjuntos de Ω que son contables o cuyo complemento es contable forma un campo- σ que contiene a todos los conjuntos unitarios de Ω , pero que es infinitas veces más pequeño que el conjunto potencia de Ω .

Antes de mostrar algunos ejemplos adicionales, es conveniente incluir una definición adicional.

12. Mínimo Campo- σ de Eventos

Dada una clase de eventos $\mathcal{C} \subseteq \{0,1\}^\Omega$, el mínimo campo- σ de eventos que contiene a \mathcal{C} , $\sigma(\mathcal{C})$, es el campo- σ de menor cardinalidad entre todos los campos- σ que lo contienen.

Como sugeríamos al concluir la definición 10, una vez escogidos el experimento aleatorio y su espacio muestral Ω , lo siguiente por hacer es seleccionar una clase de eventos de interés, \mathcal{C} , y, con esta clase, construir el mínimo campo- σ que contiene a todos los eventos en \mathcal{C} . Este mínimo campo- σ se representa mediante $\sigma(\mathcal{C})$ y se "puede construir" así: Si denotamos \mathcal{H} como el conjunto de campos- σ que contienen a \mathcal{C} , podemos saber que \mathcal{H} no es vacío pues por lo menos $\{0,1\}^\Omega$ es un campo- σ de subconjuntos de Ω que contiene a \mathcal{C} . Definiendo $\sigma(\mathcal{C})$ como la intersección de todos los campos- σ en \mathcal{H} , sabremos que $\sigma(\mathcal{C})$ es el mínimo campo- σ que contiene a \mathcal{C} . En efecto, si \mathcal{F}_1 y \mathcal{F}_2 son dos campos- σ que contienen a \mathcal{C} , $\mathcal{F}_1 \cap \mathcal{F}_2 = \{A \in \mathcal{F}_1 : A \in \mathcal{F}_2\}$ es una clase de eventos que también contiene a \mathcal{C} (pues cada evento en \mathcal{C} está tanto en \mathcal{F}_1 como en \mathcal{F}_2) y que forma un campo- σ porque no es vacío (al menos Ω y \emptyset pertenecen a ambos), si el evento A pertenece a ambos campos, el evento A^c también pertenece a ambos campos, y si los eventos A y B pertenecen a ambos campos, el evento $A \cup B$ también pertenece a ambos campos. Por supuesto, $|\mathcal{F}_1 \cap \mathcal{F}_2| \leq \min(|\mathcal{F}_1|, |\mathcal{F}_2|)$, por lo que la intersección de todos los campos en \mathcal{H} nos da el mínimo campo- σ , $\sigma(\mathcal{C})$. Siendo así, si $\{\mathcal{F}_n, n=1,2,\dots\} \in \mathcal{H}$ es el conjunto de todos los campos- σ que contienen a \mathcal{C} , $\sigma(\mathcal{C})$ se puede definir como $\sigma(\mathcal{C}) = \bigcap_n \mathcal{F}_n$... Sí, debemos admitir que éste no es un procedimiento muy práctico para determinar el mínimo campo- σ que contiene a \mathcal{C} (puede ser que la intersección ni siquiera sea contable), pero al menos así podemos estar seguros que dicho mínimo campo- σ existe.

Como ejemplo, consideremos algunos posibles campos- σ definidos en los 10 primeros experimentos propuestos en las definiciones 5 y 6.

1. La sencillez del espacio muestral del experimento que consiste en lanzar una moneda y ver qué lado queda hacia arriba sugiere que un campo- σ apropiado es el conjunto potencia del espacio muestral. Después de todo, la cardinalidad de dicho campo- σ es solamente 4.
2. En el experimento de lanzar un dado y contar los puntos en la cara que queda hacia arriba incluimos los conjuntos $A = \{\text{hay más de tres puntos}\} = \{4,5,6\}$ y $B = \{\text{hay un número par de puntos}\} = \{2,4,6\}$, con los cuales se puede construir el siguiente campo- σ en $\Omega = \{1,2,3,4,5,6\}$:
 $\mathcal{F} = \{\emptyset, \{5\}, \{4,6\}, \{4,5,6\}, \{2\}, \{2,5\}, \{2,4,6\}, \{2,4,5,6\}, \{1,3\}, \{1,3,5\}, \{1,3,4,6\}, \{1,3,4,5,6\}, \{1,2,3\}, \{1,2,3,5\}, \{1,2,3,4,6\}, \Omega\}$

La cardinalidad de este campo- σ es 16, menor a los 64 eventos del conjunto potencia. Obsérvese que otro campo- σ que también contiene a la clase de eventos $\mathcal{C}=\{A,B\}$ es el siguiente:

$$\mathcal{E} = \{\emptyset, \{6\}, \{5\}, \{5,6\}, \{4\}, \{4,6\}, \{4,5\}, \{4,5,6\}, \{2\}, \{2,6\}, \{2,5\}, \{2,5,6\}, \{2,4\}, \{2,4,6\}, \{2,4,5\}, \{2,4,5,6\}, \{1,3\}, \{1,3,6\}, \{1,3,5\}, \{1,3,5,6\}, \{1,3,4\}, \{1,3,4,6\}, \{1,3,4,5\}, \{1,3,4,5,6\}, \{1,2,3\}, \{1,2,3,6\}, \{1,2,3,5\}, \{1,2,3,5,6\}, \{1,2,3,4\}, \{1,2,3,4,6\}, \{1,2,3,4,5\}, \Omega\}$$

cuya cardinalidad, 32, sigue siendo menor a $|\{0,1\}^\Omega|=64$. Sin embargo $\mathcal{F}=\sigma(\mathcal{C})$ es el mínimo campo- σ que incluye a los eventos de \mathcal{C} . De hecho, nótese que $\mathcal{F}=\mathcal{F}\cap\mathcal{E}$. Para construir un modelo probabilístico de este experimento en donde sólo interesen los eventos en \mathcal{C} , es suficiente con asignar medidas de probabilidad a cada uno de los eventos de \mathcal{F} , y no hace falta asignarle probabilidades a cada evento en \mathcal{E} ni mucho menos a cada evento en $\{0,1\}^\Omega$.

3. En el experimento de escoger una carta de la baraja de naipes definimos los eventos $A = \{\text{Una figura de pinta roja}\} = \{\text{corazones, diamantes}\} \times \{J, Q, K\}$ y $B = \{\text{Un as negro}\} = \{(\text{picas}, 1), (\text{tréboles}, 1)\}$. En este caso, como los eventos son excluyentes (no pueden suceder simultáneamente), el mínimo campo- σ de eventos que incluye a A y B es bastante pequeño: $\mathcal{F}=\{\emptyset, A, B, A\cup B, A^c, B^c, (A\cup B)^c, \Omega\}$. Por supuesto, el conjunto potencia tiene 2^{52} eventos, ¡más de cuatro mil billones (cuatro mil millones de millones)!
4. En el experimento de medir la fracción de paquetes perdidos durante una hora en una red IP teníamos como espacio muestral el intervalo $[0,1]$ de la recta real, donde resulta imposible definir el conjunto potencia. Si definimos una familia de eventos \mathcal{E} compuesto por los intervalos cerrados $\{[0, x], x \leq 1\}$, podríamos considerar el mínimo campo- σ que contiene a \mathcal{E} , $\sigma(\mathcal{E})$. Este conjunto se llama el campo de Borel del intervalo $[0,1]$, $\mathcal{B}([0,1])$ –ver enseguida la definición 13– y, aunque es difícil de describir, sabemos que también contiene todos los intervalos abiertos, semiabiertos, cerrados, puntos aislados y uniones contables de dichos eventos... ¡Todo lo que nos pueda interesar! Claro, hay muchos subconjuntos de $[0,1]$ que no están en $\mathcal{B}([0,1])$ –¡la mayoría!– pero son tan “raros” para nuestros propósitos de modelar la fracción de paquetes perdidos, que no nos interesa incluirlos en nuestro campo- σ de eventos (¡afortunadamente, porque ni siquiera nos queda fácil imaginarlos!).
5. En el experimento de medir el retardo experimentado por un paquete de voz mientras transita por una red VoIP, puede que sólo nos interese el evento $A = \{x \in \Omega : x > 0.1\} = \{\text{El paquete no alcanza a ser reproducido en el receptor}\}$, en cuyo caso el campo- σ de eventos sería elemental: $\mathcal{F}=\{\emptyset, A, A^c, \Omega\}$, a pesar de que el espacio muestral está compuesto por los reales no negativos.
6. Al verificar el estado de ocupación de un canal de comunicaciones los posibles eventos de interés son los subconjuntos unitarios $\{\text{libre}\}$ y $\{\text{ocupado}\}$. En este caso, el conjunto potencia resulta un campo- σ perfecto para trabajar: $\{0,1\}^\Omega = \{\emptyset, \{\text{libre}\}, \{\text{ocupado}\}, \Omega\}$
7. En el experimento de contar el número de canales libres en un enlace E1 podríamos estar interesados en los siguientes dos eventos: $A = \{\text{Se puede establecer una videoconferencia}\} = \{6,7,8,\dots,32\}$ y $B = \{\text{Se puede transmitir video MPEG-4 a por lo menos 768 kbps}\} = \{12,13,14,\dots, 32\}$. En este caso, como el evento A incluye al evento B , $\sigma(\{A,B\})=\{\emptyset, A, B, A^c, B^c, A^c\cup B, A\cap B^c, \Omega\}$.
8. En el experimento de mirar el estado de ocupación de cada uno de los canales en un enlace E1 podemos definir los siguientes 33 eventos $[X_i = \{\text{Hay } i \text{ canales libres}\}, i=0,1,2,\dots,32]$. Como se trata de eventos mutuamente excluyentes, el mínimo campo- σ tendría sólo 2^{33} eventos de interés.

Los eventos de este mínimo campo- σ se podrían asociar, en una correspondencia uno-a-uno, con los eventos del conjunto potencia del experimento 7.

9. Como en los experimentos 1 y 4, el conjunto potencia es un campo- σ apropiado al determinar si un bit transmitido sobre un canal de comunicaciones llega correctamente al lado receptor del canal.
10. Al contar el número de transmisiones (a través de un canal ruidoso) que requiere un paquete de datos hasta llegar correctamente a su destino podemos estar interesados sólo en los eventos $A = \{\text{No hay errores de transmisión}\} = \{1\}$ y $B = \{\text{Mejor desistir de seguir intentándolo}\} = \{16, 17, 18, \dots\}$, con los que se puede construir un pequeño campo- σ con sólo ocho eventos: $\mathcal{F} = \{\Phi, A, B, A \cup B, A^c, B^c, (A \cup B)^c, \Omega\}$.
11. Al contar el número de bits con errores en una trama de L bits que se recibe de un canal ruidoso podríamos estar interesado en el evento en que no se dañó ningún bit, $A = \{0\}$. Un campo- σ apropiado es $\{\Phi, A, A^c, \Omega\}$, donde A^c indica la necesidad de retransmitir la trama.
12. Al medir durante una hora la fracción de tiempo que un enlace de comunicaciones permanece ocupado, dos eventos de interés serían $A = \{x \in \Omega : x \geq 0.8\}$, que indica congestión, y $B = \{x \in \Omega : x \leq 0.2\}$, que indica subutilización. Como A y B son mutuamente excluyentes, el mínimo campo- σ tiene 8 posibles eventos: $\mathcal{F} = \{\Phi, [0.8, 1], [0, 0.2], (0.2, 0.8), (0.2, 1], [0, 0.8), [0.8, 1] \cup [0, 0.2], \Omega\}$.
13. Contar el número de paquetes que llegan a un enrutador de una red de comunicaciones durante una hora. Como mencionamos, el espacio muestral es un subconjunto de los enteros no negativos, $\Omega = \{0, 1, 2, 3, \dots, M\}$, aunque el valor preciso de M puede ser difícil de establecer ya que depende del número de enlaces de entrada, el número de canales en cada enlace, la longitud de los paquetes, etc. Para no correr el riesgo de establecer un valor de M y que lleguen $M+1$ paquetes, una medida segura es suponer que pueden llegar infinitos paquetes, $\Omega = \mathbb{Z}^+$. En este caso, la cardinalidad de Ω es el infinito contable, \aleph_0 . En este caso, el conjunto potencia tiene $\aleph_1 = 2^{\aleph_0}$ eventos y ¡sigue siendo un campo sigma manejable!
14. Medir el tiempo que transcurre entre la llegada de dos paquetes consecutivos a un enlace de una red de comunicaciones. Ahora el espacio muestral es el conjunto de reales no negativos y no vale la pena intentar considerar su conjunto potencia. Nuevamente, si sólo nos interesa saber si transcurrieron menos de 100 ms o no, podemos considerar el campo sigma de eventos $\{\Phi, A, A^c, \Omega\}$, donde $A = \{\text{el tiempo transcurrido es menor a 100 ms}\}$. Sin embargo, es posible que nos interesen, por ejemplo, todos los eventos de la forma $A(x) = \{\omega \in \mathbb{R} : -\infty < \omega \leq x\}$, $x \in \mathbb{R}$. Si es así, necesitamos encontrar el mínimo campo sigma que incluye a los eventos $\{A(x), x \in \mathbb{R}\}$, al que definiremos en el siguiente apartado.

En los 12 primeros ejemplos anteriores nos interesaba un número finito o contable de eventos y resultaba fácil construir con ellos una partición del espacio muestral y considerar como campo- σ todas las uniones de los eventos de la partición⁷. Cuando los eventos de interés son infinitos contables, como en el ejemplo 13, podríamos imaginar un procedimiento parecido. Pero cuando los eventos de

⁷ Si $\{C_i, i \in I\}$ es una partición de Ω , $\mathcal{F} = \{A_J = \cup_{i \in J} C_i, J \subset I\}$ es un campo- σ , como se puede demostrar fácilmente.

interés son infinitos no contables, como en el último ejemplo, debemos considerar campos más abstractos, como muestra la siguiente definición.

13. Campo- σ de Borel de los Reales, $\mathcal{B}(\mathbb{R})$

El campo- σ de Borel de los números reales, $\mathcal{B}(\mathbb{R})$, es el mínimo campo- σ que contiene a todos los intervalos semi-infinitos de la forma $A_x = \{\omega \in \mathbb{R} : -\infty < \omega \leq x\}$, $x \in \mathbb{R}$. Los subconjuntos de \mathbb{R} que pertenecen a $\mathcal{B}(\mathbb{R})$ se denominan “conjuntos de Borel”.

En muchas ocasiones el espacio muestral de nuestros experimentos será el conjunto de los números reales, por lo que se hace muy importante definir un campo sigma de eventos “sencillo” que involucre todos los eventos “razonables” que nos puedan interesar. Como veremos enseguida, el campo de Borel de los números reales incluye a todos los intervalos cerrados, abiertos, semiabiertos, finitos ó semi-infinitos, incluyendo todos los puntos aislados. Cuando nos limitamos a uniones numerables de este tipo de eventos en \mathbb{R} , podemos construir un espacio de probabilidad coherente sobre el cual podremos aplicar toda la lógica booleana sin llegar a inconsistencias.

Veamos qué tipos de eventos se incluyen en $\mathcal{B}(\mathbb{R})$:

1. $(-\infty, x] \in \mathcal{B}(\mathbb{R}) \forall x \in \mathbb{R}$, por definición
2. Aplicando el segundo axioma a los eventos anteriores,
 $(-\infty, x]^c = (x, \infty) \in \mathcal{B}(\mathbb{R}) \forall x \in \mathbb{R}$
3. Como $(-\infty, b]$ y (a, ∞) pertenecen a $\mathcal{B}(\mathbb{R})$,
 $(-\infty, b] \cap (a, \infty) = (a, b] \in \mathcal{B}(\mathbb{R}) \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$.
4. De acuerdo con el punto anterior, $(a - 1/n, a] \in \mathcal{B}(\mathbb{R})$ y, como un campo- σ es cerrado para las intersecciones contables,
$$\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a \right] = [a] \in \mathcal{B}(\mathbb{R}) \forall a \in \mathbb{R}$$
5. De los puntos (3) y (4), $[a] \cup (a, b] = [a, b] \in \mathcal{B}(\mathbb{R}) \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$.
6. $[b]^c \cap (a, b] = (a, b) \in \mathcal{B}(\mathbb{R}) \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$.
7. $[a] \cup (a, b) = [a, b) \in \mathcal{B}(\mathbb{R}) \forall a \in \mathbb{R}, b \in \mathbb{R}, a < b$.

Si todos los intervalos (abiertos, cerrados, semiabiertos, semi-infinitos) y todos los puntos aislados son conjuntos de Borel, al igual que las uniones e intersecciones numerables de dichos subconjuntos, ¿Puede haber algún subconjunto de \mathbb{R} que no sea un conjunto de Borel? Los hay, y son muchos más los subconjuntos de \mathbb{R} que no son de Borel que aquellos que sí son de Borel ¿Qué tal, por ejemplo, el conjunto de los racionales, \mathbb{Q} ? Existe un conjunto infinito de números reales entre cada par de números racionales, por lo que \mathbb{Q} forma un conjunto de puntos aislados, cada uno de los cuales es un conjunto unitario de Borel por el punto 4. Sin embargo, sólo la unión contable de conjuntos de Borel forma un conjunto de Borel... ¡Pero \mathbb{Q} es un conjunto de cardinalidad contable porque, como vimos en la definición 10, $\aleph_0 = \aleph_0^2$. ¿Y los irracionales? Ellos forman el complemento de \mathbb{Q} en \mathbb{R} así que, a pesar de ser la unión incontable de puntos aislados (hay infinitos racionales entre cada par de irracionales), los irracionales siguen siendo un conjunto de Borel debido a la segunda propiedad de

los campos sigma ($\mathcal{B}(\mathbb{R})$ es cerrado para el complemento). Veamos otro subconjunto de \mathbb{R} aún más extraño, el conjunto ternario de Cantor:

Comenzando con el intervalo cerrado $[0,1]$, extraemos de él el segmento central correspondiente al intervalo abierto $(1/3, 2/3)$, dejando los dos intervalos cerrados $[0,1/3]$ y $[2/3,1]$. A cada uno de estos intervalos le extraemos los respectivos segmentos centrales $(1/9, 2/9)$ y $(7/9, 8/9)$, dejando cuatro intervalos cerrados $[0,1/9]$, $[2/9, 3/9]$, $[6/9, 7/9]$ y $[8/9,1]$. A cada uno de estos intervalos le extraemos los respectivos segmentos centrales $(1/27, 2/27)$, $(7/27, 8/27)$, $(19/27, 20/27)$ y $(25/27, 26/27)$, dejando ocho intervalos cerrados. Así seguimos repitiendo el proceso de extracción del tercio de la mitad de cada intervalo cerrado que nos vaya quedando, *Ad Infinitum*, como sugiere la Figura 21. El conjunto que nos queda cuando repetimos la iteración un número infinito de veces es el conjunto de Cantor, \mathcal{C} .

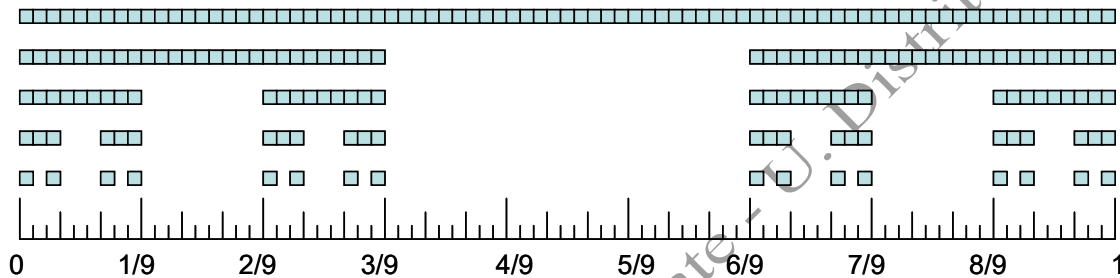


Figura 21. Primeras cuatro iteraciones en la construcción del conjunto de Cantor

¿Es \mathcal{C} un conjunto de Borel? Partiendo de un intervalo de longitud 1, obtenemos 2^1 intervalos de longitud $1/3^1$ en la primera iteración, 2^2 intervalos de longitud $1/3^2$ en la segunda iteración, 2^3 intervalos de longitud $1/3^3$ en la tercera iteración y, en general, 2^n intervalos de longitud $1/3^n$ en la n -ésima iteración. Esto es, después de un número finito de iteraciones seguimos teniendo un número finito de intervalos, que es un conjunto de Borel, pues es la unión de un número finito de intervalos cerrados; pero ¿qué pasa después de un número infinito de iteraciones? ¿Que ya no existe ningún intervalo en \mathcal{C} ! En efecto, en la iteración n , para $n = 1,2,3,\dots$, retiramos 2^{n-1} intervalos, cada uno de longitud 3^{-n} , de manera que la longitud total de los intervalos que extraemos es $\frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \left(\frac{1}{1-\frac{2}{3}}\right) = 1$

. Si de un intervalo de longitud 1 quitamos subintervalos cuya longitud total es 1, nos queda sólo una “nube de polvo” que no contiene ningún intervalo, sólo puntos aislados. Pero cada punto individual de \mathbb{R} es un conjunto unitario de Borel y, si los puntos que quedan en el conjunto de Borel son los extremos de los intervalos que van quedando en cada iteración, tendríamos que \mathcal{C} es la unión contable de puntos aislados... ¡pero no es posible numerar los puntos en el conjunto de Cantor! Existe otra cantidad no numerable de puntos que, sin ser el extremo de ninguno de esos intervalos, jamás se eliminan del conjunto de Cantor, como se puede apreciar para el punto $1/4$ en la Figura 22 : Si no se retiró en la tercera iteración ya jamás será retirado de \mathcal{C} . De hecho, Cantor mostró que hay tantos puntos en \mathcal{C} como en \mathbb{R} , $|\mathcal{C}|=|\mathbb{R}|=\aleph_1$. Basta con notar que el algoritmo de construcción equivale a quitar del intervalo unitario todos los puntos cuya expansión en base 3 incluya algún 1, de manera que los puntos de Cantor son aquellos cuya expansión en base 3 sólo contiene los dígitos 0 y 2 (se entiende que cualquier número que termine en 1 seguido por un número infinito de ceros se escribe terminando

en 0 seguido por un número infinito de 2). En la primera iteración retiramos todos los puntos que tienen un uno en la posición de 3^{-1} ; En la segunda iteración retiramos todos los puntos que tienen un uno en la posición 3^{-2} ; En general, en la n -ésima iteración quitamos todos los puntos que tienen un uno en la posición 3^{-n} . Si nunca retiramos el punto $1/4$ de \mathcal{C} es porque $(1/4)_3 = 0.020202\dots$. Ahora, si expresamos cada punto del intervalo unitario en binario, podemos cambiar cada dígito 1 por 2 e interpretarlo en base 3, con lo cual hemos establecido una relación biunívoca entre los puntos del intervalo unitario y los puntos del conjunto de Cantor:

$$\mathcal{C} \ni x = \sum_{n=1}^{\infty} a_n \cdot 3^{-n} \leftrightarrow y = \sum_{n=1}^{\infty} \frac{a_n}{2} \cdot 2^{-n} \in [0,1], \quad a_n \in \{0,2\}$$

Como Cantor nos enseñó, dos conjuntos tienen la misma cardinalidad si podemos establecer una relación biunívoca entre ellos: ¡hay tantos puntos en \mathcal{C} como en $[0,1]$ y, por consiguiente, como en \mathbb{R} !

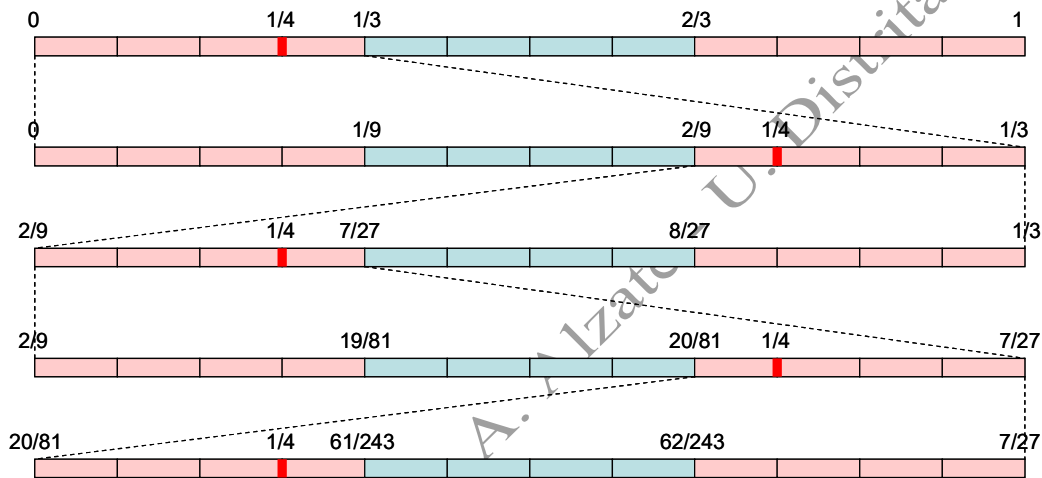


Figura 22. El punto $[1/4]$ pertenece al conjunto de Cantor

En efecto, nótese que si miramos \mathcal{C} en el intervalo $[0, 3^{-n}]$ a través de una lupa que aumente 3^n veces, reconstruiremos el conjunto de Cantor completo. Por eso el conjunto \mathcal{C} constituye un objeto autosemejante (es idéntico a sí mismo en un número infinito de escalas) con dimensión fractal $\ln(2)/\ln(3) \approx 0.6$, a pesar de tener una dimensión topológica 0... \mathcal{C} parece un buen candidato para un subconjunto de \mathbb{R} que no pertenece a $\mathcal{B}(\mathbb{R})$. Después de todo, \mathcal{C} es un conjunto no numerable de puntos aislados cuyo complemento es un conjunto no numerable de intervalos abiertos ¡Pero hasta el conjunto de Cantor es un conjunto de Borel! Basta notar que \mathcal{C} surge de una secuencia contable de intersecciones y uniones de intervalos cerrados:

$$\mathcal{C} = \bigcap_{n=0}^{\infty} \bigcup_{m=0}^{\frac{3^n-1}{2}} \left[\frac{2m}{3^n}, \frac{2m+1}{3^n} \right]$$

Sin embargo, se pueden encontrar subconjuntos del conjunto de cantor que no son Borel. Por ejemplo, evaluando la función de Cantor⁸ en un conjunto de Vitali (ver enseguida) obtenemos un subconjunto

⁸ La función de Cantor $f_c: [0,1] \rightarrow [0,1]$ se construye para cada $x \in [0,1]$ así: (1) expresamos x en base 3, (2) si hay un dígito 1, ponemos cero de él en adelante, (3) cambiamos cada dígito 2 por un dígito 1, (4) interpretamos el resultado como un número binario en $[0,1]$. A la gráfica de f_c se le llama grandilocuentemente “la escalera del diablo”.

del conjunto de Cantor que no es Borel. Vale la pena mencionar que, aunque el conjunto de Cantor parece ser una curiosidad matemática diseñada para mostrar extraños subconjuntos de \mathbb{R} , mediante procedimientos de construcción generalizados semejantes al algoritmo de Cantor, se han desarrollado importantísimos modelos de tráfico en redes de comunicaciones, tales como el modelo wavelet multifractal (MWM).

Bueno, pero si hay menos conjuntos de Borel en \mathbb{R} que subconjuntos de \mathbb{R} , ¿porqué no hemos podido encontrar uno? Mostraremos enseguida una familia infinita contable de conjuntos infinitos no contables de los reales que no son conjuntos de Borel, los conjuntos de Vitali.

Decimos que dos puntos de la recta real, x y y , son equivalentes si su diferencia $x - y$ es un número racional. No hace falta que x y y sean racionales, aunque pueden serlo; es suficiente con que su diferencia sea racional. Por ejemplo, $1/5$ y $1/7$ son equivalentes porque $1/5 - 1/7 = 2/35 \in \mathbb{Q}$, π y $\pi - 1/3$ son equivalentes porque su diferencia es $1/3 \in \mathbb{Q}$, pero π y $\pi/2$ no son equivalentes porque su diferencia es $\pi/2 \notin \mathbb{Q}$. Todos los racionales forman una única clase equivalente y existe otra clase equivalente por cada número irracional en los reales, esto es, existe un número infinito no contable de clases equivalentes (una por cada irracional que no esté separado de otro irracional por una cantidad fraccional), donde cada clase equivalente tiene un número infinito contable de elementos (uno por cada racional). De cada una de esas incontables clases escogemos un elemento en el intervalo $[0, 1]$ y al conjunto resultante le llamamos un “conjunto de Vitali”, V . Ahora, por cada número racional q en el intervalo $[0, 1]$, hacemos una traslación de V sumando q a cada uno de sus elementos. Los conjuntos resultantes son los conjuntos V_q , de manera que ahora tenemos un conjunto contable de conjuntos incontables cuya unión, $\cup_q V_q$, forma el intervalo $[0, 2]$. ¿Cuál es la longitud de cada V_q ? No puede ser cero, porque la longitud del intervalo $[0, 2]$ sería cero. No puede ser diferente de cero, porque la longitud del intervalo $[0, 2]$ sería infinita... Los conjuntos de Vitali no son conjuntos de Borel (ver definición 27).

14. Medida de Probabilidad

Una medida de probabilidad \mathbf{P} asociada a un experimento aleatorio (Ω, \mathcal{F}) es una función $\mathbf{P}: \mathcal{F} \rightarrow \mathbb{R}$ que asigna a cada evento en \mathcal{F} un número real que satisface los siguientes axiomas: (1) $\mathbf{P}(\Omega) = 1$, (2) Si $A \in \mathcal{F}$, $\mathbf{P}(A) \geq 0$, (3) Si $A, B \in \mathcal{F}$ son mutuamente excluyentes ($A \cap B = \emptyset$), $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$. Si \mathcal{F} es un campo- σ infinitamente aditivo, también debe satisfacerse el siguiente axioma adicional: (3^a) Si $\{A_n \in \mathcal{F}, n=1, 2, 3, \dots\}$ es una colección de eventos tal que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$.

Esta definición axiomática, tan útil como es, deja de lado el problema de darle un significado al número que se asigna a cada evento. Lo cierto es que si Kolmogorov estableció esta definición en 1933, en respuesta al reto lanzado por Hilbert en 1900 sobre determinar unas bases formales para la

teoría de la probabilidad, fue inspirado en propiedades fundamentales de los conceptos típicos de probabilidad: (1) Que si repito un experimento un gran número de veces y mido la fracción de experimentos en que sucede el evento A (ver definición 7), la fracción obtenida tiende a $P(A)$ a medida que hago más y más repeticiones. (2) Que si logro describir el espacio muestral como un conjunto de cardinalidad finita en el que ninguno de los eventos unitarios ocurre preferencialmente sobre los otros, la probabilidad de un evento está dada por la cardinalidad del evento sobre la cardinalidad del espacio muestral. (3) Que si consulto a un experto sobre la ocurrencia de un evento en una hipotética realización de un experimento, la probabilidad del evento es el grado de certeza que el experto tiene en que dicho evento ocurra. (4) Que si he acumulado cierta evidencia a favor o en contra de una hipótesis, la probabilidad del evento en el que dicha hipótesis es cierta está dada por el grado de implicación lógica que existe de la evidencia a la hipótesis. (5) etc. Resulta muy afortunado saber que siempre es posible asociar los axiomas propuestos por Kolmogorov a propiedades particulares de la probabilidad en cada una de sus interpretaciones, como describiremos en la definición 15.

En efecto, desde el punto de vista de la formalidad, la matemática no es más que un lenguaje en el que se pueden mezclar unos símbolos de forma adecuada para satisfacer ciertas reglas gramaticales dictadas, por ejemplo, por la lógica, la teoría de conjuntos, etc., de manera que, a partir de unos axiomas, podemos deducir unos teoremas. Pero un modelo matemático es un isomorfismo entre un lenguaje formal de las matemáticas y una simplificación de alguna realidad particular: El isomorfismo induce un significado a las matemáticas. El hecho es que, para cada uno de los diferentes conceptos de probabilidad mencionados en el párrafo anterior, se puede establecer un isomorfismo con la formalidad de Kolmogorov, así como la formalidad de las ecuaciones diferenciales de primer orden con coeficientes constantes pueden asociarse con un circuito RC o con automóvil en movimiento (Figura 6). En un servidor de correo electrónico es posible medir la frecuencia relativa de cada tipo de mensaje spam y aplicar sobre esas medidas la formalidad de Kolmogorov, indicando que la definición formal de probabilidad aplica a la definición (1) de probabilidad. En un juego de cartas es posible contar las cartas para obtener información que permita decidir las acciones que maximicen la probabilidad de ganar, en cuyo caso la formalidad de Kolmogorov aplica a la definición (2) de probabilidad. Si tengo unas evidencias experimentales de las que induzco un nivel de creencia en que ocurra un evento, la formalidad de Kolmogorov debería aplicar a la definición (3) si queremos que esas creencias sean coherentes.

Lo cierto es que esta definición 14 es implacable y precisa, a diferencia de las posibles interpretaciones que le asignemos. Como cualquier formalidad axiomática en matemáticas, los axiomas escogidos por Kolmogorov son la menor cantidad de axiomas que forman un conjunto completo (cualquier proposición que involucre probabilidades puede calificarse como cierta o como falsa), consistente (ninguna proposición podrá ser demostrada como falsa y como verdadera) e independiente (ningún axioma se puede deducir de los demás). Esta formalidad no sólo evita cualquier ambigüedad conceptual sino que enfatiza el pensamiento deductivo con que se debe abordar la teoría de probabilidades: Todos los teoremas se deben derivar de estos tres axiomas (ver definición 17) y es aplicando los axiomas y los teoremas como encontramos las probabilidades de unos eventos en términos de las probabilidades de otros eventos. Por eso aceptamos los axiomas como verdades indiscutibles, algo así como si Kolmogorov los hubiera bajado del monte Sinaí impresos en piedra.

Claro, tenemos la opción de no aceptarlos, en cuyo caso no podemos usar la teoría de probabilidades en nuestros modelos matemáticos de la realidad ingenieril de las redes de comunicaciones y podemos terminar aquí el estudio de este texto. Pero, si optamos por esa opción, nos estaremos privando de una herramienta formidable para nuestro trabajo investigativo y estaremos negándonos a investigar en muchas áreas donde la teoría de las probabilidades es indispensable. Yo, personalmente, creo que la teoría de las probabilidades describe la lógica del pensamiento científico, por lo que no podríamos avanzar muy lejos sin sus axiomas. Sin embargo, para nuestra tranquilidad, enseguida veremos la plausibilidad de los axiomas en las diferentes interpretaciones de la medida de probabilidad.

15. Interpretaciones de la Medida de Probabilidad

Sea un experimento aleatorio (Ω, \mathcal{F}) y un evento $A \in \mathcal{F}$. Una forma de interpretar la probabilidad del evento A es mediante la relación $\mathbf{P}(A) = \lim_{N \rightarrow \infty} f_N(A)$, donde $f_N(A)$ es la frecuencia relativa del evento A en N repeticiones del experimento. Esta interpretación aplica cuando el experimento se puede replicar indefinidamente bajo condiciones idénticas. Otra interpretación muy útil se basa en la relación $\mathbf{P}(A) = |A|/|\Omega|$, donde $|A|$ es la cardinalidad del evento A . Esta interpretación aplica cuando Ω es un conjunto de cardinalidad finita y $P(\{\omega\}) = 1/|\Omega|$ para todo $\omega \in \Omega$. Sin embargo, sea que ninguna, una o ambas interpretaciones apliquen al experimento aleatorio de interés, para nosotros $\mathbf{P}(A)$ es el nivel de certeza que tenemos en que ocurra el evento A cuando ejecutemos el experimento.

Como mencionamos en la definición 7 sobre la regularidad estadística de un experimento aleatorio, si repetimos N veces un experimento con espacio muestral Ω y contamos en cuántas repeticiones ocurrió el evento $A \subset \Omega$, N_A , definimos la frecuencia relativa del evento A en esas N repeticiones como $f_N(A) = N_A/N$. Obsérvese que el proceso de observación de N_A es, en sí mismo, otro experimento aleatorio, de manera que en diferentes conjuntos de N repeticiones podemos obtener diferentes valores de $f_N(A)$. Sin embargo, la regularidad estadística sugiere que, entre más repeticiones hagamos, el valor de $f_N(A)$ tiende a un valor fijo, independientemente del conjunto particular de N repeticiones que seleccionemos.

Por ejemplo, supongamos que deseamos saber cuál es la probabilidad del evento $A = \{\text{a un enrutador llegan más de 1000 bytes en un período de 100 ms}\}$. Para esto medimos la frecuencia relativa de dicho evento en 200 períodos consecutivos y la graficamos en función del número de períodos observados. Si hacemos mediciones durante un minuto, obtendremos tres conjuntos distintos, como muestra la Figura 23, en cada uno de los cuales la frecuencia parece tender a un número cercano a 0.4. Si las condiciones del tráfico permanecen estables durante el minuto de observación y son iguales a las condiciones en el período de 100 ms por cuya probabilidad nos interesamos (que podría ser, por ejemplo, el siguiente período que aún no hemos observado), diríamos que la probabilidad de que

lleguen más de 1000 bytes es “cercana” a 0.4^9 . Pues bien, es fácil ver que los axiomas que definen la probabilidad como una medida de los subconjuntos de Ω contenidos en \mathcal{F} están inspirados en propiedades elementales de la frecuencia relativa. En efecto,

$$(1) f_N(\Omega) = N / N = 1$$

$$(2) \text{ como } N_A \geq 0, f_N(A) \geq 0$$

$$(3) \text{ Si } A \cap B = \Phi, N_{A \cup B} = N_A + N_B, \text{ de manera que } f_N(A \cup B) = f_N(A) + f_N(B).$$

Permítaseme insistir, porque debemos ser cuidadosos con esto, que el límite de la frecuencia relativa es apenas una interpretación de la probabilidad que puede ser útil para los ingenieros de redes de comunicaciones ya que a nosotros nos es posible tomar muchas mediciones con facilidad y en tiempos razonables (medir el retardo de 10000 paquetes, medir el número de errores en 10000 bits transmitidos, medir la condición de “spam” en 10000 mensajes de correo electrónico, etc.). Sin embargo, en muchos casos, el experimento mismo que queremos modelar ni siquiera es repetible, de manera que no tiene sentido considerar esta interpretación. J. M. Keynes, por ejemplo, era economista y cada uno de sus experimentos podía durar décadas; por esa razón, la interpretación frecuentista, que parece objetiva en cuanto a que muestra resultados verificables “a la larga”, lo conduce a expresar su famosa frase: “A la larga, todos estamos muertos”. En estas condiciones, lo mejor es considerar la probabilidad como expresión de simetría o como nivel de confianza. Esta es la visión Bayesiana: La probabilidad es un grado subjetivo de creencia. ¿Qué significa, si no, la frase “mañana en la tarde lloverá con probabilidad 0.7”? Puede que el meteorólogo tenga razones frecuentistas para decir algo así, pero para la persona que lo escucha en televisión es sólo una medida de creencia. O ¿a qué se refiere el adolescente cuando afirma “mi novia me ama con probabilidad 0.7”? Con seguridad no se trata de que el 70% de sus novias lo han amado. Finalmente, como mencionamos en la definición 4, verificar la existencia del bosón de Higgs fue un experimento aleatorio que se realizó en el LHC (en realidad un conjunto muy grande de experimentos aleatorios realizados entre el 10 de septiembre de 2008 y el 4 de julio de 2012), que nunca se repetirá porque ya no será aleatorio.

De todas maneras, desde un punto de vista puramente matemático, la interpretación misma pierde relevancia pues la definición es precisa e implacable: la probabilidad es una función que asigna a cada subconjunto de Ω en \mathcal{F} una medida en \mathbb{R} que satisface tres axiomas básicos. Lo cierto es que, cuando uno está inmerso en un problema de modelado probabilístico, a veces resulta muy útil preguntarse “si yo pudiera repetir este experimento muchas veces, ¿qué esperarías que sucediera a la larga?”, pues la respuesta puede sugerirnos el siguiente paso en el proceso o puede explicarnos un resultado poco intuitivo. De hecho, en este libro echaremos mano de la interpretación frecuentista liberalmente para justificar muchas definiciones o para interpretar muchos resultados.

⁹ De hecho, de acuerdo con la cota de Chebyshev (definición 47(b)), la frecuencia relativa se acerca a la probabilidad a medida que hacemos más y más observaciones, en el sentido de que el evento $B = \{|f_N(A) - \mathbf{P}(A)| > \varepsilon\}$ ocurre con probabilidad menor a $1/(4N\varepsilon^2)$. Más aún, mediante la cota de Chernov (definición 47(c)) se puede demostrar que la aproximación de $f_N(A)$ a $\mathbf{P}(A)$ es exponencial: $\mathbf{P}[\{|f_N(A) - \mathbf{P}(A)| > \varepsilon\}] < \exp(-2N\varepsilon^2)$ –desigualdad de Hoeffding–.

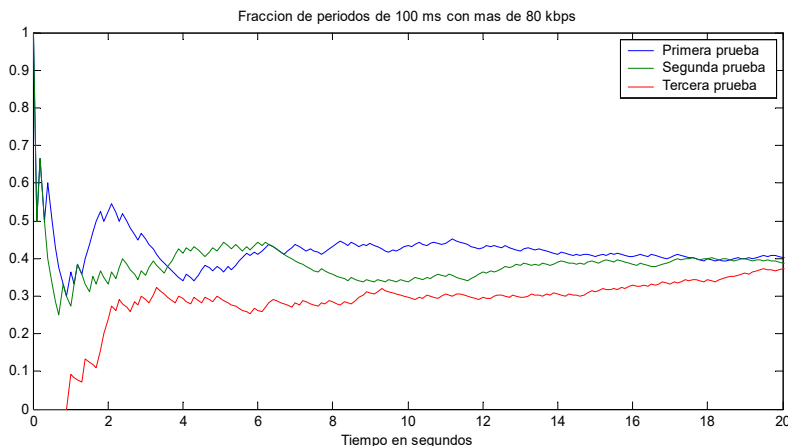


Figura 23. $f_N(A)$ vs N para tres conjuntos distintos de pruebas

Otra interpretación que suele resultar muy útil para estimar las probabilidades de algunos eventos es la interpretación clásica. La probabilidad del evento A es

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Número de resultados favorables al evento } A}{\text{Número total de posibles resultados}}$$

siempre que ninguno de los subconjuntos unitarios de Ω ocurra preferencialmente sobre otros, esto es, siempre que todos los eventos unitarios de Ω sean igualmente probables! Claro, al incluir el concepto a definir dentro de la definición, no estamos definiendo nada. Pero la idea es intuitivamente clara: sacar una carta de la pinta de corazones ocurre con probabilidad $13/52 = 1/4$, pues ninguna carta sale preferencialmente sobre las demás. Obtener un par al lanzar un dado bien balanceado ocurre con probabilidad $3/6=1/2$, pues ningún lado queda hacia arriba preferencialmente sobre los demás. Los axiomas de Kolmogorov aplican igualmente bien para esta interpretación:

- (1) $|\Omega|/|\Omega| = 1$
- (2) como $|A| \geq 0$, $|A|/|\Omega| \geq 0$.
- (3) Si $A \cap B = \emptyset$, $|A \cup B| = |A| + |B|$, de manera que $|A \cup B|/|\Omega| = (|A| + |B|)/|\Omega| = |A|/|\Omega| + |B|/|\Omega|$.

Al hacer uso de esta interpretación, debemos ser muy juiciosos con la premisa de los eventos unitarios "igualmente probables". Si el dado está cargado de manera que $P(\{1\}) = 0.2$ y $P(\{n\})=0.16$ para $n=2,3,4,5,6$, el uso de esta interpretación clásica conducirá a resultados desastrosos. Por eso esta definición sólo aplica a una clase muy pequeña de experimentos aleatorios.

A veces el espacio muestral es infinito no contable, pero aún podemos aplicar la interpretación clásica si usamos alguna medida finita del tamaño de Ω , tal como la longitud o el área:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\text{Medida de la región favorable al evento } A}{\text{Medida total de la región posible}}$$

siempre que todos los subconjuntos de cualquier tamaño dado tengan la misma probabilidad de ocurrir como resultado del experimento:

- (1) $\mu(\Omega)/\mu(\Omega) = 1$
- (2) como $\mu(A) \geq 0$, $\mu(A)/\mu(\Omega) \geq 0$
- (3) Si $A \cap B = \Phi$, $\mu(A \cup B) = \mu(A) + \mu(B)$, de manera que $\mu(A \cup B)/\mu(\Omega) = (\mu(A) + \mu(B))/\mu(\Omega)$
 $= \mu(A)/\mu(\Omega) + \mu(B)/\mu(\Omega)$.

Para nosotros, como ingenieros preocupados por problemas técnicos muy precisos, resulta muy cómodo escoger eclécticamente entre cada una de las interpretaciones la que más nos favorezca o la que mejor nos guíe en el proceso de desarrollar un modelo probabilístico para nuestro problema. Por ejemplo, no hemos sabido de ningún ingeniero de comunicaciones al que le quite el sueño el problema filosófico que implica utilizar un medidor de BER (Bit-Error-Rate) para medir la fracción de bits que se dañan durante su transmisión por un canal de comunicaciones (como en el experimento 9) y después utilizar esa medida como su nivel de confianza en que el próximo bit que transmita se dañe en el canal, aunque así esté mezclando las interpretaciones (1) y (3). Dada la facilidad que tenemos para tomar muchas mediciones, esta interpretación frecuentista de la probabilidad la podemos aprovechar ventajosamente desde la formalidad de Kolmogorov, como se describe en la definición 15.

16. Espacio de Probabilidad

Un espacio de probabilidad es la triplete $(\Omega, \mathcal{F}, \mathbf{P})$ asociada con un experimento aleatorio, donde Ω es el espacio muestral o el conjunto de todos los posibles resultados del experimento, \mathcal{F} es un campo- σ de subconjuntos de Ω construido a partir de una clase de eventos de interés y \mathbf{P} es una función de \mathcal{F} en \mathbb{R} que satisface los axiomas de la definición 14. Como solamente se les puede asignar una medida de probabilidad a los subconjuntos de Ω que pertenecen a \mathcal{F} , a dichos subconjuntos se les denomina "subconjuntos medibles" o "Eventos" (con lo cual cumplimos la promesa de especificar mejor el concepto de evento).

En cualquier caso en que queramos trabajar sobre modelos probabilísticos de una realidad particular, deberemos partir de la descripción explícita del espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ pues, de otra manera, estaremos perdidos: ni siquiera sabremos dónde estamos parados! En efecto, una vez descrito el problema en términos de un espacio de probabilidad, podremos movernos con confianza sobre todos los resultados de la teoría de las probabilidades sabiendo que, mientras seamos consecuentes y rigurosos con ellas, llegaremos a resultados significativos, pues es sobre los axiomas que definen a \mathcal{F} y \mathbf{P} que se construye TODA la teoría de probabilidades.

Hasta este punto, la teoría de probabilidades sería simplemente una rama de la "teoría de las mediciones", que es el área de la matemáticas que estudia las funciones $\mu: \mathcal{H} \rightarrow \mathbb{R}$ que asignan una medida real $\mu(E)$ a cada conjunto E de una colección de conjuntos \mathcal{H} . En teoría de mediciones se estudia formalmente la conveniencia de que \mathcal{H} forme un campo- σ , en cuyo caso $\mu(\cdot)$ es una medida aditivamente contable, como es el caso de las medidas de probabilidad asignadas a los subconjuntos medibles del espacio muestral de un experimento aleatorio. Sin embargo, la definición 21 tratará

sobre la independencia, la cual le dará a la teoría de probabilidades una identidad propia que le permitirá distinguirse de la teoría general de las mediciones.

17. Algunos Resultados Básicos Derivados de los Axiomas de la Probabilidad

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay dos eventos medibles A y $B \in \mathcal{F}$. Entonces (1) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$, (2) $\mathbf{P}(\emptyset) = 0$, (3) $\mathbf{P}(A) \leq 1$, (4) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$, (5) Si $A \subset B$, $\mathbf{P}(A) \leq \mathbf{P}(B)$.

Los anteriores cinco resultados son apenas una muestra mínima de todas las conclusiones que se pueden sacar de los axiomas de la definición 14 pues, como ya se dijo, de los tres axiomas se deriva TODA la teoría de las probabilidades. Sin embargo, como estos cinco resultados se usan cotidianamente cuando se estudian modelos probabilísticos de cualquier sistema, vale la pena tenerlos tan presentes como los mismos axiomas de los que se derivan:

(1) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$

En efecto, como $A \cap A^c = \emptyset$ y $A \cup A^c = \Omega$, los axiomas 1 y 3 conducen a $\mathbf{P}(\Omega) = \mathbf{P}(A^c) + \mathbf{P}(A) = 1$. Restando $\mathbf{P}(A)$ a ambos lados obtenemos el resultado deseado.

(2) $\mathbf{P}(\emptyset) = 0$

En efecto, $A = A \cup \emptyset$, que son eventos mutuamente excluyentes, por lo que podemos aplicar el tercer axioma: $\mathbf{P}(A) = \mathbf{P}(A) + \mathbf{P}(\emptyset)$. Restando $\mathbf{P}(A)$ a ambos lados obtenemos el resultado deseado. Es importante notar que $\mathbf{P}(\emptyset) = 0$ se refiere al “evento imposible” \emptyset . Cualquier otro evento distinto de vacío con probabilidad 0 se conoce como “evento nulo” y, aunque improbable, no es imposible.

(3) $\mathbf{P}(A) \leq 1$

En efecto, como ya demostramos que $\mathbf{P}(A) = 1 - \mathbf{P}(A^c)$, basta con aplicar el segundo axioma en A^c , $\mathbf{P}(A^c) \geq 0$, para obtener el resultado deseado. Como en el caso anterior, es importante notar que $\mathbf{P}(\Omega) = 1$ se refiere al “evento seguro” Ω . Cualquier otro evento distinto de Ω con probabilidad 1 se conoce como “evento casi seguro”, pues es posible que no ocurra.

(4) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

En efecto, podemos expresar $A \cup B$ como la unión de dos eventos mutuamente excluyentes, $A \cup B = A \cup (B \cap A^c)$, de manera que $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c)$. Por otro lado, B también se puede expresar como la unión de dos eventos mutuamente excluyentes, $B = (A \cap B) \cup (B \cap A^c)$, de manera que $\mathbf{P}(B \cap A^c) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$. Remplazando esta expresión de $\mathbf{P}(B \cap A^c)$ en la primera expresión de $\mathbf{P}(A \cup B)$ obtenemos el resultado deseado. Los diagramas de Venn de la Figura 24 representan esta derivación.

(5) Si $A \subset B$, $\mathbf{P}(A) \leq \mathbf{P}(B)$

En efecto, podemos expresar B como la unión de dos eventos mutuamente excluyentes, $B = A \cup (B \cap A^c)$, de manera que $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c)$. Y, como $\mathbf{P}(B \cap A^c) \geq 0$ por el segundo axioma, entonces $\mathbf{P}(B) \geq \mathbf{P}(A)$.

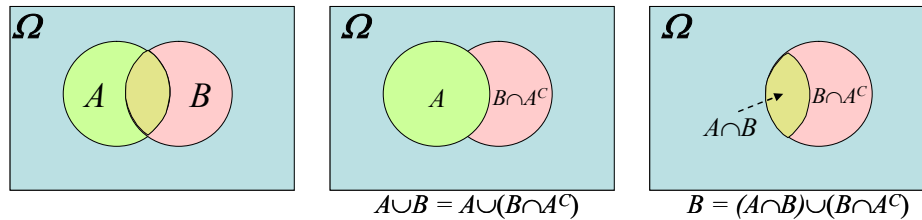


Figura 24. Construcciones para derivar la expresión $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

18. Probabilidad Condicional

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay dos eventos A y $B \in \mathcal{F}$. La probabilidad condicional del evento A dado que se sabe de la ocurrencia del evento B es

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}, \quad \mathbf{P}(B) > 0$$

Esta definición, como tan “sacada de la manga”, en realidad obedece a un concepto muy simple si se le mira desde la interpretación frecuentista de la probabilidad. Supongamos que repetimos N veces el experimento en cuestión y contamos cuántas veces sucedió cada uno de los siguientes eventos: $N_A =$ Número de veces que ocurrió el evento A , $N_B =$ Número de veces que ocurrió el evento B , y $N_{A \cap B} =$ Número de veces que ocurrieron ambos simultáneamente. Ahora consideramos solamente aquellas N_B repeticiones en las que ocurrió el evento B e ignoramos todas las demás. La frecuencia relativa del evento A entre aquellas repeticiones del experimento en que ocurrió B es $f_N(A|B) = N_{A \cap B} / N_B$. Dividiendo el numerador y el denominador por N , obtenemos $f_N(A|B) = f_N(A \cap B) / f_N(B)$, que es una expresión MUY parecida a la definición de probabilidad condicional.

De acuerdo con el anterior resultado, si pudiésemos definir la probabilidad de un evento como el límite de su frecuencia relativa cuando el número de repeticiones tiende a infinito, tendríamos que el condicionamiento sería simplemente una propiedad más de la probabilidad. Pero como la probabilidad es un concepto más abstracto (una función de \mathcal{F} en \mathbb{R} que satisface 3 axiomas), este resultado frecuentista es apenas un motivo de inspiración para la definición propuesta. De todas formas, la definición no nos debe sorprender porque la teoría de probabilidades quiere modelar, precisamente, el comportamiento de ese límite sin obligarnos a gastar un tiempo infinito en hacer un número infinito de repeticiones del experimento.

Volviendo a nuestra definición axiomática, a veces es posible que la condición sea un evento de probabilidad cero, pues una cosa es un evento improbable –con probabilidad cero– y otra cosa es un evento imposible –que no puede ocurrir–. En este caso la probabilidad condicional puede estar indefinida ($0/0$) o puede que, para cualquier secuencia $\{B_n \in \mathcal{F}\}$ tal que $B = \lim_{n \rightarrow \infty} B_n$, exista el límite $P[A|B] = \lim_{n \rightarrow \infty} P[A \cap B_n] / P[B_n]$. Estos límites se mencionarán en las definiciones 23 y 25 pero el caso particular de condicionar en eventos de probabilidad cero se mencionará en el próximo capítulo cuando hablemos de variables aleatorias continuas.

Es legítimo preguntarnos por un nuevo espacio de probabilidad en el que la probabilidad condicional sea una medida de probabilidad válida. Lo primero que notamos, por ejemplo, es que, en el nuevo espacio de probabilidades, el espacio muestral debe ser B , pues nos estamos limitando a estudiar los casos en que tenemos certeza absoluta de que el evento B ocurrió. Pero, ¿cuál sería un nuevo campo de eventos apropiado? Como todos los eventos de interés contenidos en \mathcal{F} se ven reducidos a su intersección con B , es razonable pensar en un campo de eventos como $\mathcal{H} = \{A \cap B : A \in \mathcal{F}\}$. ¿Es éste un campo- σ de subconjuntos de B ? Veamos

- (1) \mathcal{H} es no vacío porque por lo menos $\Phi \in \mathcal{F} \Rightarrow \Phi \cap B = \Phi \in \mathcal{H}$:
 \mathcal{H} es no vacío
- (2) Si $A \in \mathcal{F}$, entonces $A^c (= \Omega \setminus A) \in \mathcal{F}$, de manera que $A \cap B \in \mathcal{F}$ y $A^c \cap B (= B \setminus (A \cap B)) \in \mathcal{F}$, donde $A^c \cap B = B \setminus (A \cap B)$ es el complemento de $A \cap B$ en B :
Si $X \in \mathcal{H}$, entonces $X^c = B \setminus X \in \mathcal{H}$
- (3) Si $A_1 \in \mathcal{F}$, $A_2 \in \mathcal{F}$, entonces $A_1 \cup B \in \mathcal{F}$, $A_2 \cup B \in \mathcal{F}$ y, por consiguiente, $A_1 \cap B \in \mathcal{H}$, $A_2 \cap B \in \mathcal{H}$, entonces $A_1 \cap B \in \mathcal{H}$, $A_2 \cap B \in \mathcal{H}$, $(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B) \in \mathcal{H}$:
Si $X \in \mathcal{H}$, $Y \in \mathcal{H}$, entonces $X \cup Y \in \mathcal{H}$
- (4) Lo mismo se puede verificar para la uniones contables

¿Y será la probabilidad condicional $\mathbf{Q}(\cdot) = \mathbf{P}(\cdot|B)$ una medida válida en (B, \mathcal{H}) ? Veamos:

- (1) $\mathbf{Q}(B) = \mathbf{P}(B|B) = \mathbf{P}(B)/\mathbf{P}(B) = 1$
- (2) $\mathbf{Q}(A) = \mathbf{P}(A|B) \geq 0$
- (3) Si $(A_1 \cap B) \cap (A_2 \cap B) = \Phi$, (en realidad no necesitamos que A_1 y A_2 sean excluyentes, pues basta con que no puedan ocurrir simultáneamente con B), entonces $\mathbf{Q}(A_1 \cup A_2) = \mathbf{P}((A_1 \cup A_2)|B) = \mathbf{P}((A_1 \cup A_2) \cap B) / \mathbf{P}(B) = \mathbf{P}((A_1 \cap B) \cup (A_2 \cap B)) / \mathbf{P}(B) = (\mathbf{P}(A_1 \cap B) + \mathbf{P}(A_2 \cap B)) / \mathbf{P}(B) = \mathbf{P}(A_1|B) + \mathbf{P}(A_2|B) = \mathbf{Q}(A_1) + \mathbf{Q}(A_2)$.
- (4) Lo mismo se puede verificar para uniones contables.

En conclusión, dado el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P}(\cdot))$, podemos construir un nuevo espacio de probabilidad $(B, \mathcal{H}, \mathbf{Q}(\cdot) = \mathbf{P}(\cdot|B))$ condicionando todos los eventos de \mathcal{F} a la ocurrencia del evento B , donde $B \in \mathcal{F}$ y $\mathbf{P}(B) > 0$. Esto es, hemos reducido el espacio original a uno más pequeño.

Esto quiere decir que todo lo que hemos dicho (y diremos) sobre cualquier espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P}(\cdot))$ aplica también al espacio de probabilidad condicionado, $(B, \mathcal{H}, \mathbf{Q}(\cdot) = \mathbf{P}(\cdot|B))$. En particular, como **ya lo demostramos** en la definición 17, (1) $\mathbf{Q}(A^c) = \mathbf{P}(A^c|B) = 1 - \mathbf{P}(A|B) = 1 - \mathbf{Q}(A)$, (2) $\mathbf{Q}(\Phi) = \mathbf{P}(\Phi|B) = 0$, (3) $\mathbf{Q}(A) = \mathbf{P}(A|B) \leq 1$, (4) $\mathbf{Q}(A \cup C) = \mathbf{P}(A \cup C|B) = \mathbf{P}(A|B) + \mathbf{P}(C|B) - \mathbf{P}(A \cap C|B) = \mathbf{Q}(A) + \mathbf{Q}(C) - \mathbf{Q}(A \cap C)$, (5) Si $A \cap B \subset C \cap B$ (en realidad no necesitamos que $A \subset C$, pues basta con que esta condición ocurra dentro de B), $\mathbf{Q}(A) = \mathbf{P}(A|B) \leq \mathbf{P}(C|B) = \mathbf{Q}(C)$.

Considérese, por ejemplo, el experimento 9, en el que transmitimos un bit y vemos si llegó correctamente a su destino en el otro extremo del canal binario. Si consideramos como parte del experimento observar el bit transmitido, nuestro nuevo espacio muestral será $\Omega = \{(0,0), (0,1), (1,0)$,

$(1,1)$ }, donde el resultado (i, j) corresponde a la transmisión del bit i y la recepción del bit j . La probabilidad de que se produzca un error en ese canal es $\mathbf{P}(\{(0,1),(1,0)\})$, que es la probabilidad del evento $ERROR = \{se\ recibe\ un\ bit\ distinto\ al\ bit\ transmitido\}$. Condicionando en el bit transmitido, tenemos dos tipos de error con las siguientes probabilidades

$$\mathbf{P}(\{Recibir\ 0\} \mid \{se\ transmitió\ 1\}) = \mathbf{P}(\{(1,0)\}) / \mathbf{P}(\{(1,0), (1,1)\})$$

$$\mathbf{P}(\{Recibir\ 1\} \mid \{se\ transmitió\ 0\}) = \mathbf{P}(\{(0,1)\}) / \mathbf{P}(\{(0,0), (0,1)\})$$

Dada la simetría que existe en las técnicas de modulación digital, es de esperar que los dos tipos de error tengan la misma probabilidad, en cuyo caso nos encontramos ante un espacio de probabilidad que modela un Canal Binario Simétrico (BSC, binary symmetric channel). Claramente, al utilizar un medidor de BER –*Bit Error Rate*– sobre un canal BSC, estamos tratando de estimar las probabilidades condicionales descritas anteriormente, por lo que el modelo se puede representar como en la Figura 25,

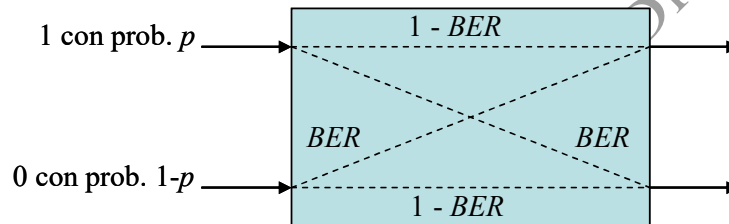


Figura 25. Modelo probabilístico de un canal binario simétrico

donde el espacio de probabilidad que modela el canal es $(\Omega = \{(0,0), (0,1), (1,0), (1,1)\}, \mathcal{F} = \{0,1\}^\Omega, \mathbf{P})$. El conocimiento inicial que tenemos sobre la medida de probabilidad \mathbf{P} en este modelo probabilístico de un canal de comunicaciones es el siguiente:

$$\mathbf{P}[\{(1,0),(1,1)\}] = 1 - \mathbf{P}[\{(0,0),(0,1)\}] = p$$

$$\mathbf{P}[\{(0,0),(1,0)\} \mid \{(1,0),(1,1)\}] = 1 - \mathbf{P}[\{(0,1),(1,1)\} \mid \{(1,0),(1,1)\}] = \dots$$

$$\mathbf{P}[\{(0,1),(1,1)\} \mid \{(0,0),(0,1)\}] = 1 - \mathbf{P}[\{(0,0),(1,0)\} \mid \{(0,0),(0,1)\}] = BER$$

Obsérvese en este ejemplo cómo resulta de fácil “medir” la probabilidad condicional BER. En general, ésta es la gran utilidad de la probabilidad condicional: encontrar la probabilidad de un evento A puede ser muy difícil, pero una vez condicionamos el evento de interés a otro evento B (juiciosamente seleccionado), puede resultar muy fácil encontrar la probabilidad condicional de A dado B . Este truco se repite una y otra vez en el modelado probabilístico de redes de comunicaciones, como tendremos oportunidad de ver en breve. Pero, ¿de qué nos sirve la probabilidad condicional de A dado B si lo que queríamos encontrar era la probabilidad de A ? El siguiente teorema explica dónde reside la utilidad del “truco”.

19. Teorema de la Probabilidad Total

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay un evento $A \in \mathcal{F}$ y una secuencia de eventos en \mathcal{F} , $\{B_k\}, k=1,2,\dots$ que forman una partición de Ω . Entonces

$$\mathbf{P}(A) = \sum_k \mathbf{P}(B_k) \mathbf{P}(A | B_k)$$

Esta relación es más fácil de ver si consideramos la partición más pequeña, constituida por B y B^c , como muestra la Figura 26. En efecto, con esta partición podemos expresar el evento A como la unión de dos eventos mutuamente excluyentes, $A = (A \cap B) \cup (A \cap B^c)$, de manera que $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^c)$. Pero, por la definición misma de la probabilidad condicional, $\mathbf{P}(A \cap B) = \mathbf{P}(B) \mathbf{P}(A|B)$ y $\mathbf{P}(A \cap B^c) = \mathbf{P}(B^c) \mathbf{P}(A|B^c)$, de manera que $\mathbf{P}(A) = \mathbf{P}(B) \mathbf{P}(A|B) + \mathbf{P}(B^c) \mathbf{P}(A|B^c)$. La generalización a particiones más numerosas (incluyendo aquellas contablemente infinitas) es inmediata.

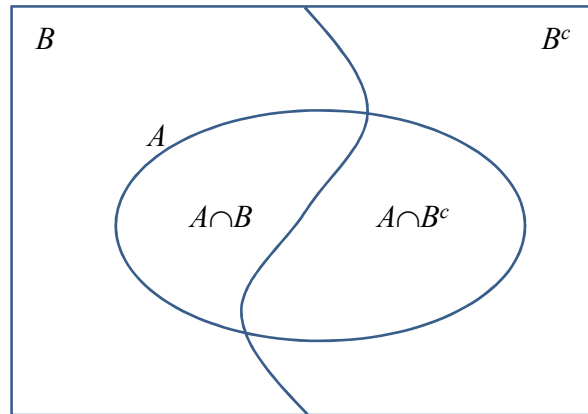


Figura 26. Diagrama de Venn para verificar el teorema de la probabilidad total

En el modelo del canal simétrico binario de la Figura 25, por ejemplo, ¿cuál será la probabilidad de recibir un cero? Podemos condicionar en el bit transmitido, ya que las probabilidades condicionadas en el bit transmitido son conocidas:

$$\begin{aligned} \mathbf{P}(\{Rx 0\}) &= \mathbf{P}(\{Tx 0\}) \mathbf{P}(\{Rx 0\} | \{Tx 0\}) + \mathbf{P}(\{Tx 1\}) \mathbf{P}(\{Rx 0\} | \{Tx 1\}) \\ &= (1-p)(1-BER) + pBER = 1-BER-p+2pBER \end{aligned}$$

Claro, la probabilidad de recibir un uno debe ser uno menos la probabilidad de recibir un cero, lo cual puede ser verificado mediante la probabilidad total:

$$\begin{aligned} \mathbf{P}(\{Rx 1\}) &= \mathbf{P}(\{Tx 0\}) \mathbf{P}(\{Rx 1\} | \{Tx 0\}) + \mathbf{P}(\{Tx 1\}) \mathbf{P}(\{Rx 1\} | \{Tx 1\}) \\ &= (1-p)BER + p(1-BER) = BER + p - 2pBER \end{aligned}$$

Observese que si $BER=0$ ó si $BER=1$, no existiría ninguna duda en el receptor sobre el bit transmitido, pues el bit recibido tendrá toda la información necesaria para identificar al primero sin equivocaciones. Cualquier otro valor de BER genera incertidumbre en el receptor, especialmente en el caso extremo en que $BER = 0.5$, pues en este caso obtenemos que $\mathbf{P}(\{Rx 1\}) = \mathbf{P}(\{Rx 0\}) = 0.5$, independientemente de p , de manera que podemos ahorrarnos el canal y hacer que en el receptor se lance una moneda equilibrada por cada bit transmitido.

Una pregunta de mucho interés para el módem receptor es la siguiente: Dado que recibí cierto símbolo a la salida del canal, ¿cuáles son las probabilidades del respectivo símbolo a la entrada del canal? La siguiente regla es muy útil para este tipo de preguntas.

20. Regla de Bayes

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay un evento $A \in \mathcal{F}$ y una secuencia de eventos $\{B_k\}$, $k=1,2,\dots$ que forman una partición de Ω y que también pertenece a \mathcal{F} . Entonces

$$\mathbf{P}(B_k | A) = \frac{\mathbf{P}(B_k)\mathbf{P}(A | B_k)}{\sum_j \mathbf{P}(B_j)\mathbf{P}(A | B_j)}$$

Esta regla surge directamente de la definición de la probabilidad condicional, pues $\mathbf{P}(A \cap B_k) = \mathbf{P}(B_k) \mathbf{P}(A | B_k)$ y $\mathbf{P}(A) = \sum_j \mathbf{P}(B_j)\mathbf{P}(A | B_j)$, de manera que la regla de Bayes sólo es una forma de expresar la definición $\mathbf{P}(B_k | A) = \mathbf{P}(A \cap B_k) / \mathbf{P}(A)$.

Considérese nuevamente el canal binario simétrico (BSC) donde el transmisor se caracteriza por transmitir 1 con probabilidad p y transmitir 0 con probabilidad $1 - p$ y el canal se caracteriza por una tasa de error por bit BER, como muestra la Figura 25. Si el destino recibe 0, ¿cuál es la probabilidad de que la fuente haya enviado 1?

$$\begin{aligned} P(\{Tx1\} | \{Rx0\}) &= \frac{P(\{Tx1\})P(\{Rx0\} | \{Tx1\})}{P(\{Tx1\})P(\{Rx0\} | \{Tx1\}) + P(\{Tx0\})P(\{Rx0\} | \{Tx0\})} \\ &= \frac{p \cdot BER}{p \cdot BER + (1 - p)(1 - BER)} \end{aligned}$$

La regla de Bayes se considera el fundamento del método científico y, en general, de la racionalidad: Yo parto de unas creencias iniciales (las probabilidades a priori $P(B_k)$ que hablan de los posibles estados de la naturaleza, los cuales no puedo medir directamente) y los relaciono con observaciones concretas de la realidad, el evento A (que es un hecho, una realidad objetiva), para lograr unas nuevas creencias (las probabilidades a posteriori $P(B_k|A)$), mejoradas por la consideración de hechos objetivos. La racionalidad consiste, precisamente, en estar dispuesto a cambiar mis creencias para ajustarlas a los hechos observados. Cómo a los posibles estados de la naturaleza (las hipótesis) se asigna unos niveles de creencia (las probabilidades a priori), para ajustar esas mismas creencias (las probabilidades a posteriori), algunos filósofos y teóricos de la estadística hablan de la subjetividad que implica la regla de Bayes como algo negativo. Sin embargo, la regla de Bayes da claridad a la toma de decisiones bajo condiciones de escasa información y resultados inciertos. Por ejemplo, la existencia del bosón de Higgs era incierta antes de 2012 pero, dados los resultados de los experimentos del LHC, la probabilidad de la existencia del bosón de Higgs es superior a 0.9999999999999999. Este es uno de los resultados científicos que se dan en términos de la regla de Bayes. Pero igual puede decir el adolescente que no sabe si su compañera de salón está enamorada

de él o no, y ajusta su nivel de creencia de acuerdo con las miradas que recibe de ella. O un médico que no sabe si un paciente tiene cierta enfermedad o no y ajusta su nivel de creencia de acuerdo con las pruebas de laboratorio que le ordena. O un juez que no sabe si un acusado es culpable o no y ajusta su nivel de creencia de acuerdo con las evidencias presentadas por la fiscalía y la defensa. Así, la regla de Bayes es el fundamento de los métodos de investigación en casi todas las ciencias y, por supuesto, en ingeniería. Por ejemplo, el usuario de una red puede tener una creencia a priori del ancho de banda disponible para él. Pero si ajusta su creencia de acuerdo con datos experimentales tales como el retardo, las pérdidas y las variaciones de retardo (jitter), obtendrá un estimado más realista del ancho de banda disponible. Todos los anteriores son ejemplos de lo que en ciencias y estadística se conoce indistintamente como prueba de hipótesis, teoría de la detección, toma de decisiones, técnicas de clasificación, etc., y, en todas ellas, la regla de Bayes es la herramienta fundamental.

21. Eventos Independientes

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay dos eventos A y $B \in \mathcal{F}$. A y B son independientes si y sólo si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ o, equivalentemente, si $\mathbf{P}(A|B) = \mathbf{P}(A)$ y $\mathbf{P}(B|A) = \mathbf{P}(B)$.

Tres eventos medibles A , B y C son independientes si se cumplen las siguientes cuatro condiciones: (1) $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, (2) $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$, (3) $\mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C)$, y (4) $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$.

En general, los eventos medibles $\{A_n, n=1,2,\dots\}$ forman una secuencia de eventos

independientes si
$$\mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i) \quad \forall I \subset \{1,2,\dots\}$$

Nuevamente, ésta es simplemente una definición. Pero es una definición muy poderosa, pues nos dice que si A y B son independientes, nuestra incertidumbre respecto a la ocurrencia de A no cambia cuando nos informan que B ocurrió. En términos de nuestra interpretación frecuentista, supongamos que hacemos N repeticiones del experimento en las que observamos que A ocurrió N_A veces, B ocurrió N_B veces, y ambos ocurrieron simultáneamente $N_{A \cap B}$ veces. Supongamos que la frecuencia relativa del evento A en N repeticiones del experimento, N_A/N , tiende al mismo valor al que tiende la frecuencia relativa del evento A en las N_B veces que ocurrió el evento B , $N_{A \cap B}/N_B$, a medida que N tiende a infinito. Siendo así, para estimar $\mathbf{P}(A)$ como el límite de la frecuencia relativa, nos daría igual si nos fijamos en todas las N repeticiones o sólo en aquellas N_B repeticiones en que ocurrió B , pues la restricción del espacio muestral de Ω a B no altera la frecuencia relativa de A .

A veces puede ser fácil identificar si dos eventos son o no son independientes. Por ejemplo sean $A = \{\text{me va a ir bien en mi matrimonio}\}$ y $B = \{\text{Mi prometida y yo tenemos el mismo nivel de educación}\}$. Nadie duda que $\mathbf{P}(A|B) > \mathbf{P}(A)$ y que $\mathbf{P}(A|B^c) < \mathbf{P}(A)$, de manera que A y B no son independientes. Sin embargo, si definimos C como el evento $\{\text{Yo soy Tauro y mi novia es Libra}\}$, resulta sorprendente la

cantidad de personas que creen que $\mathbf{P}(A|C) \in \{0,1\}$ independientemente de B . Yo, personalmente, creo que A y C son eventos independientes, de manera que $\mathbf{P}(A|C) = \mathbf{P}(A)$.

En nuestro mundo de las redes de telecomunicaciones, en muchas ocasiones debemos admitir que ciertos eventos no son independientes, aunque preferimos suponer independencia para mantener el análisis matemático tratable. Por ejemplo, muchos resultados útiles suponen que la presencia de errores de transmisión en una trama es independiente de la presencia de errores en la trama inmediatamente anterior. Tal vez en enlaces satelitales o de fibra óptica se pueda argumentar la validez de esa suposición, pero no en enlaces terrestres de radio o de cobre donde los errores se pueden deber, por ejemplo, a la ignición eléctrica de un motor de combustión o a la operación cercana de un horno de microondas. Igualmente, al modelar el tráfico sobre una red, muchas veces preferimos suponer que el tiempo entre la llegada del paquete $n-1$ y la del paquete n es independiente del tiempo entre la llegada del paquete n y la del paquete $n+1$. Seguramente, si se trata del punto de acceso a la red de un gran número de usuarios, esta suposición de independencia se pueda justificar. Pero si se trata de paquetes de un mismo flujo o si los paquetes ya han sido sometidos a interacciones debidas a los protocolos de la red, es muy difícil aceptar que sus tiempos entre llegadas puedan ser independientes. Sin embargo, tan poderoso es el concepto de independencia que, aún en estos casos, a veces preferimos suponer independencia con la esperanza de que los resultados obtenidos al final del análisis no estén muy alejados de la realidad.

Por alguna razón muy común (que no he logrado detectar!), muchos estudiantes neófitos de teoría de probabilidad suelen equiparar la independencia de dos eventos con la exclusión mutua entre ellos. Si A y B son mutuamente excluyentes y por lo menos uno de ellos tiene probabilidad mayor que cero, resulta imposible que sean independientes porque $\mathbf{P}(A|B) = 0$ y $\mathbf{P}(B|A) = 0$, de manera que sólo podrían ser independientes si ambos eventos son nulos. De la misma manera, si dos eventos son independientes, resulta imposible que sean mutuamente excluyentes, a menos que ambos sean eventos nulos. Considérese, por ejemplo, el experimento de seleccionar un punto de un rectángulo unitario como el de la figura 9, descrito mediante el siguiente espacio de probabilidad

$$(\Omega = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 1\}, \mathcal{F} = \mathcal{B}(\mathbb{R}^2) \cap \Omega, \mathbf{P}(A) = \text{Área}(A) \forall A \in \mathcal{F}).^{10}$$

Sean $A = \{(x, y) \in \Omega : x < 0.5\}$ y $B = \{(x, y) \in \Omega : y < 0.5\}$. Claramente $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(A^c) = \mathbf{P}(B^c) = 1/2$. A y B no son mutuamente excluyentes porque $A \cap B = \{(x, y) \in \Omega : x < 0.5, y < 0.5\} \neq \emptyset$. Pero A y B sí son independientes porque el área de $A \cap B$ es $1/4$, que es la mitad del área de B , de manera que $\mathbf{P}(A|B) = 1/2 = \mathbf{P}(A)$ o, mejor aún, $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = 1/4$. De manera semejante, A y A^c son mutuamente excluyentes, por lo que $\mathbf{P}(A|A^c) = 0 < \mathbf{P}(A) = 0.5$: A y A^c no pueden ser independientes.

¹⁰ Por definición, la “intersección” entre un conjunto y una clase de conjuntos, $A \cap \mathcal{H}$, se refiere a la restricción de la clase \mathcal{H} al evento A , esto es, $A \cap \mathcal{H} = \{A \cap H : H \in \mathcal{H}\}$ es una nueva clase (reducida) de eventos.

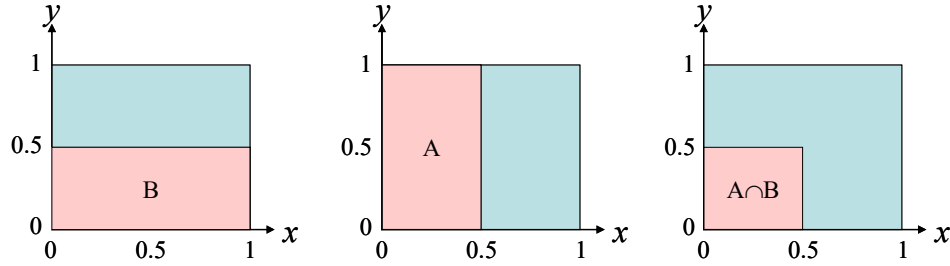


Figura 27. Distinción entre independencia y exclusión mutua

Es importante notar que, si A y B son independientes, también lo son A^c y B . En efecto, como B es la unión de dos eventos mutuamente excluyentes, $(A \cap B) \cup (A^c \cap B)$, encontramos que $\mathbf{P}(A^c \cap B) = \mathbf{P}(B) - \mathbf{P}(A \cap B) = \mathbf{P}(B) - \mathbf{P}(A)\mathbf{P}(B) = (1 - \mathbf{P}(A))\mathbf{P}(B) = \mathbf{P}(A^c)\mathbf{P}(B)$. Por la misma razón, si A y B son independientes, también lo son A y B^c , así como A^c y B^c .

Tres eventos pueden ser independientes por pares, sin necesidad de que sean tres eventos independientes. Considere, por ejemplo, una fuente de información que es capaz de generar tres símbolos $\{a, b, c\}$ con los cuales puede construir nueve mensajes, $\Omega = \{abc, acb, bac, bca, cab, cba, aaa, bbb, ccc\}$, cada uno con probabilidad $1/9$. Sea $A_k = \{\text{el } k\text{-ésimo símbolo del mensaje es "a"}\}$, $k=1,2,3$, de manera que $\mathbf{P}(A_1) = \mathbf{P}(\{abc, acb, aaa\}) = 1/3$, $\mathbf{P}(A_2) = \mathbf{P}(\{bac, cab, aaa\}) = 1/3$ y $\mathbf{P}(A_3) = \mathbf{P}(\{bca, cba, aaa\}) = 1/3$. Claramente, A_1, A_2 y A_3 son independientes por pares porque $\mathbf{P}(A_i \cap A_j) = \mathbf{P}(\{aaa\}) = 1/9 = \mathbf{P}(A_i)\mathbf{P}(A_j)$ si $i \neq j$. Sin embargo no son tres eventos independientes porque $\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(\{aaa\}) = 1/9 \neq \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3) = 1/27$.

El caso contrario también puede ocurrir: $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ pero $\mathbf{P}(A \cap B) \neq \mathbf{P}(A)\mathbf{P}(B)$. Un ejemplo trivial pero definitivo puede ser el siguiente: Se mide el tiempo de transferencia ftp de un archivo de 100 kbytes y se definen los siguientes eventos $A = \{\text{menos de 1 segundo}\}$, $B = \{\text{menos de 100 ms}\}$ y $C = \{0 \text{ segundos}\}$. Claramente, $C \subset B \subset A$ de manera que $0 = \mathbf{P}(C) < \mathbf{P}(B) < \mathbf{P}(A) < 1$, por lo que $\mathbf{P}(A \cap B) = \mathbf{P}(B) > \mathbf{P}(A)\mathbf{P}(B)$, por lo cual A y B no son independientes, pero $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = 0$, de manera que la probabilidad de la intersección de los tres eventos es igual al producto de las tres probabilidades individuales.

Como ejemplo del poder del concepto de independencia considere la red de la Figura 28 en la que cada enlace falla con probabilidad p , independientemente de los otros enlaces. ¿Cuál es la probabilidad de que exista una ruta desde A hasta C ? Sean $E_i = \{\text{El enlace } e_i \text{ está bueno}\}$, $i=1,2,3,4,5$, y $R = \{\text{Existe una ruta entre } A \text{ y } C\}$. Considerando la partición del espacio muestral dada por E_5 y E_5^c , podemos aplicar el teorema de la probabilidad total así:

$$\mathbf{P}(R) = \mathbf{P}(E_5)\mathbf{P}(R | E_5) + \mathbf{P}(E_5^c)\mathbf{P}(R | E_5^c)$$

Donde $\mathbf{P}(E_5) = 1 - p$, $\mathbf{P}(R | E_5) = 1$, y $\mathbf{P}(E_5^c) = p$, de manera que

$$\mathbf{P}(R) = 1 - p + p \mathbf{P}(R | E_5^c)$$

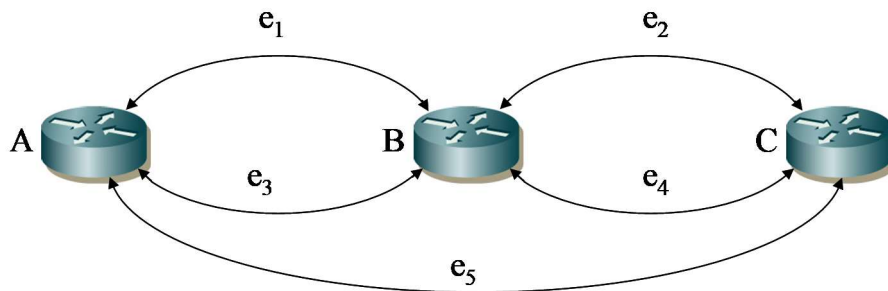


Figura 28. Red para análisis de confiabilidad

Si e_5 está dañado, A y C sólo se podrán comunicar a través de B , para lo cual se requiere que de A se pueda llegar a B ($E_1 \cup E_3$) y que de B se pueda llegar a C ($E_2 \cup E_4$):

$$\mathbf{P}(R | E_5^c) = \mathbf{P}((E_1 \cup E_3) \cap (E_2 \cup E_4))$$

Aquí es cuando la suposición de independencia facilita enormemente el problema pues, si E_1 y E_3 son independientes de E_2 y E_4 , $\mathbf{P}(R | E_5^c)$ será el producto de las dos probabilidades que, por simetría, son idénticas:

$$\mathbf{P}(R | E_5^c) = \mathbf{P}(E_1 \cup E_3)^2$$

Pero $E_1 \cup E_3 = (E_1^c \cap E_3^c)^c$, por lo que podemos aplicar nuevamente la independencia de E_1 y E_3 : $\mathbf{P}(E_1 \cup E_3) = 1 - \mathbf{P}(E_1^c \cap E_3^c) = 1 - \mathbf{P}(E_1^c)\mathbf{P}(E_3^c) = 1 - p^2$. Reemplazando,

$$\mathbf{P}(R) = 1 - p + p(1 - p^2)^2$$

Si los enlaces no fallaran independientemente unos de otros, la solución del problema sería enormemente compleja.

Dos eventos pueden no ser independientes, a menos que se condicionen a un tercer evento: $\mathbf{P}(A \cap B) \neq \mathbf{P}(A)\mathbf{P}(B)$ pero $\mathbf{P}(A \cap B | C) = \mathbf{P}(A | C)\mathbf{P}(B | C)$, en cuyo caso se dice que A y B son condicionalmente independientes. Considérese, por ejemplo, un enlace en tiempo discreto donde la unidad de tiempo es el tiempo de transmisión de un paquete. Sea $X[n]$ = Número de paquetes transmitidos hasta el instante n , con $X[0] = 0$. Definamos los siguientes eventos:

$A_2 = \{X[2] = 1\} = \{01, 10\}$, donde 01 indica 0 paquetes transmitidos en el primer slot y 1 paquete transmitido en el segundo slot.

$A_3 = \{X[3] = 2\} = \{011, 101, 110\}$

$A_4 = \{X[4] = 2\} = \{0011, 0101, 0110, 1001, 1010, 1100\}$.

Si en cada slot se transmite un paquete con probabilidad p independientemente de los slots vecinos, tenemos

$$\mathbf{P}(A_2) = 2p(1-p), \quad \mathbf{P}(A_3) = 3p^2(1-p), \quad \mathbf{P}(A_4) = 6p^2(1-p)^2$$

Obsérvese que $\mathbf{P}(A_2 \cap A_4) = \mathbf{P}(\{0101, 0110, 1001, 1010\}) = 4p^2(1-p)^2 \neq \mathbf{P}(A_2)\mathbf{P}(A_4) = 12p^3(1-p)^3$, de manera que A_2 y A_4 no son independientes. Sin embargo consideremos las siguientes probabilidades:

$$\mathbf{P}(A_2 | A_3) = \mathbf{P}(\{011, 101\}) / \mathbf{P}(\{011, 101, 110\}) = 2p^2(1-p) / 3p^2(1-p) = 2/3$$

$$\mathbf{P}(A_4 | A_3) = \mathbf{P}(\{0110, 1010, 110\}) / \mathbf{P}(\{011, 101, 110\}) = 3p^2(1-p)^2 / 3p^2(1-p) = 1 - p$$

$$\mathbf{P}(A_2 \cap A_4 | A_3) = \mathbf{P}(\{0110, 1010\}) / \mathbf{P}(\{011, 101, 110\}) = 2p^2(1-p)^2 / 3p^2(1-p) = 2(1-p)/3$$

Claramente, $\mathbf{P}(A_2 \cap A_4 | A_3) = \mathbf{P}(A_2 | A_3) \mathbf{P}(A_4 | A_3)$, de manera que A_2 y A_4 son condicionalmente independientes dado A_3 .

El anterior ejemplo es una Cadena de Markov (ver definición 123), cuya principal característica es que, aunque el futuro depende del pasado, el futuro resulta condicionalmente independiente del pasado cuando se conoce el presente. Esta propiedad me parece un principio importante para aplicar a una vida positiva: Todo mi futuro depende solamente de quién soy yo en este momento, independientemente de cómo llegué a ser lo que soy. Mi futuro sólo dependerá de mi pasado si yo no sé quién soy en este momento.

22. Experimentos Compuestos

Sean $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ y $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ dos espacios de probabilidad que corresponden a dos experimentos aleatorios. Si realizamos los dos experimentos y vemos el par de resultados como un único resultado de un experimento aleatorio compuesto, tendremos un nuevo espacio de probabilidad conjunto en el que el espacio muestral es el producto cartesiano $\Omega_1 \times \Omega_2$, correspondiente a los pares ordenados $\{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$. Los eventos de interés, en general, pertenecerán al producto cartesiano $\mathcal{F}_1 \times \mathcal{F}_2$, correspondiente a los conjuntos de pares ordenados $\{(\omega_1, \omega_2) \in A_1 \times A_2\}, \forall A_1 \in \mathcal{F}_1, \forall A_2 \in \mathcal{F}_2$ ¹¹. Sin embargo, $\mathcal{F}_1 \times \mathcal{F}_2$ no necesariamente forma un campo sigma de subconjuntos de $\Omega_1 \times \Omega_2$, por lo que será necesario cerrarlo para los complementos y las uniones contables. Así, un campo sigma apropiado para el experimento compuesto es $\sigma(\mathcal{F}_1 \times \mathcal{F}_2)$, el mínimo campo sigma de subconjuntos de $\Omega_1 \times \Omega_2$ que incluye los eventos que pertenecen a $\mathcal{F}_1 \times \mathcal{F}_2$, que son los pares (ω_1, ω_2) en los que ω_1 pertenece a algún evento A_1 en \mathcal{F}_1 y ω_2 pertenece a algún evento A_2 en \mathcal{F}_2 .

Por ejemplo, sean $\Omega_1 = \{a, b\}$ con $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$ y $\Omega_2 = \{1, 2, 3\}$ con $\mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3\}, \{1, 2, 3\}\}$. El espacio muestral del experimento compuesto es $\Omega_1 \times \Omega_2 = \{a1, a2, a3, b1, b2, b3\}$ y el campo sigma del experimento compuesto debe incluir a los conjuntos $\mathcal{F}_1 \times \mathcal{F}_2 = \{\emptyset, \{a3\}, \{b3\}, \{a1, a2\}, \{a3, b3\}, \{b1, b2\}, \{a1, a2, a3\}, \{a1, a2, b3\}, \{a3, b1, b2\}, \{b1, b2, b3\}, \{a1, a2, a3, b3\}, \{a1, a2, b1, b2\}, \{a3, b1, b2, b3\}, \{a1, a2, a3, b1, b2\}, \{a1, a2, b1, b2, b3\}\}$. Evidentemente, éste no es un campo sigma de subconjuntos de $\Omega_1 \times \Omega_2$ aunque \mathcal{F}_1 y \mathcal{F}_2 sean campos sigma de subconjuntos de Ω_1 y Ω_2 , respectivamente. Para completar un campo sigma apropiado para el experimento compuesto, debemos cerrar $\mathcal{F}_1 \times \mathcal{F}_2$ con los complementos y las uniones, lo que implica añadir seis eventos compuestos adicionales:

$$\sigma(\mathcal{F}_1 \times \mathcal{F}_2) = \{\emptyset, \{a3\}, \{b3\}, \{a1, a2\}, \{a3, b3\}, \{b1, b2\}, \{a1, a2, a3\}, \{a1, a2, b3\}, \{a3, b1, b2\}, \{b1, b2, b3\}, \{a1, a2, a3, b3\}, \{a1, a2, b1, b2\}, \{a3, b1, b2, b3\}, \{a1, a2, a3, b1, b2\}, \{a1, a2, b1, b2, b3\}, \{a1, a2, a3, b1, b2, b3\}\}$$

Nótese que $|\mathcal{F}_1 \times \mathcal{F}_2| = 10$, $|\sigma(\mathcal{F}_1 \times \mathcal{F}_2)| = 16$ y $|\{0, 1\}^{\Omega_1 \times \Omega_2}| = 64$.

¹¹ Donde $\phi \times A = \phi \forall A \subset \Omega$

Falta considerar la medida de probabilidad en este nuevo espacio medible $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$, que podría ser tan simple como $P(A_1, A_2) = P_1(A_1)P_2(A_2)$ si los dos experimentos que componen el experimento compuesto son independientes¹². En general, puede haber alguna dependencia entre ellos, de manera que $P(A_1, A_2) = P_1(A_1)P_{21}(A_2|A_1) = P_2(A_2)P_{12}(A_1|A_2)$, donde las probabilidades condicionales $P_{21}(A_2|A_1)$ y $P_{12}(A_1|A_2)$ involucrarían las dependencias que puedan ocurrir entre los dos experimentos (esto es, considerando la naturaleza de los experimentos, en cuyo caso las probabilidades condicionales hablan sobre los procesos de observación correspondientes).

Sigamos considerando el ejemplo anterior añadiendo medidas de probabilidad. El primer espacio de probabilidad es $(\Omega_1 = \{a, b\}, \mathcal{F}_1 = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}, \mathbf{P}_1 = \{P_1(\emptyset) = 0, P_1(\{a\}) = p, P_1(\{b\}) = 1 - p, P_1(\{a, b\}) = 1\})$, que corresponde a escoger una caja entre dos cajas, una marcada con la letra a y otra marcada con la letra b , y lees la letra marcada en la caja escogida. El segundo espacio de probabilidad es $(\Omega_2 = \{1, 2, 3\}, \mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3\}, \{1, 2, 3\}\}, \mathbf{P}_2 = \{P_2(\emptyset) = 0, P_2(\{1, 2\}) = q, P_2(\{3\}) = 1 - q, P_2(\{1, 2, 3\}) = 1\})$, que corresponde a escoger una bola de billar de una bolsa oscura y leer el número marcado en la bola, donde algunas bolas están marcadas con el número 1, otras con el número 2 y las demás con el número 3. Sabemos que el espacio de probabilidad asociado con el experimento compuesto es $(\Omega_3 = \Omega_1 \times \Omega_2 = \{a1, a2, a3, b1, b2, b3\}, \mathcal{F}_3 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2) = \{\emptyset, \{a3\}, \{b3\}, \{a1, a2\}, \{a3, b3\}, \{b1, b2\}, \{a1, a2, a3\}, \{a1, a2, b3\}, \{a3, b1, b2\}, \{b1, b2, b3\}, \{a1, a2, a3, b3\}, \{a1, a2, b1, b2\}, \{a3, b1, b2, b3\}, \{a1, a2, a3, b1, b2\}, \{a1, a2, b1, b2, b3\}, \{a1, a2, a3, b1, b2, b3\}\}, \mathbf{P}_3)$, pero ¿cómo es la medida de probabilidad \mathbf{P}_3 ? Depende de la naturaleza de la composición entre los experimentos. Si la bola se escoge de una bolsa oscura que contiene todas las bolas y la caja se escoge de una bolsa oscura que contiene las dos cajas, el resultado del experimento 1 no tiene nada que ver con el resultado del experimento 2, de manera que los experimentos son independientes:

$$\begin{aligned} P_3(\emptyset) = 0, P_3(\{a3\}) = p(1-q), \quad P_3(\{b3\}) = (1-p)(1-q), \quad P_3(\{a1, a2\}) = pq, \quad P_3(\{a3, b3\}) = 1-q, \\ P_3(\{b1, b2\}) = (1-p)q, \quad P_3(\{a1, a2, a3\}) = p, \quad P_3(\{a1, a2, b3\}) = 1+2pq-p-q, \quad P_3(\{a3, b1, b2\}) = p+q-2pq, \\ P_3(\{b1, b2, b3\}) = 1-p, \quad P_3(\{a1, a2, a3, b3\}) = 1-q(1-p), \quad P_3(\{a1, a2, b1, b2\}) = q, \quad P_3(\{a3, b1, b2, b3\}) = 1-pq, \\ P_3(\{a1, a2, a3, b1, b2\}) = p+q-pq, \quad P_3(\{a1, a2, b1, b2, b3\}) = 1-p(1-q), \quad P_3(\{a1, a2, a3, b1, b2, b3\}) = 1. \end{aligned}$$

Pero si el experimento compuesto es tal que la bola se obtiene de la caja seleccionada, la probabilidad del experimento compuesto dependerá de la distribución de las bolas en cada una de las dos cajas. Si la fracción de bolas marcadas con 3 en la caja a es r y la fracción de bolas marcadas con 3 en la caja b es s , el experimento sólo tendrá sentido si $q = 1 - pr - (1-p)s$, lo cual demuestra la dependencia de los dos experimentos.

Es fácil generalizar esta idea a una secuencia de experimentos, $\{(\Omega_n, \mathcal{F}_n, \mathbf{P}_n), n = 1, 2, \dots, N\}$, donde N puede ser cualquier número natural o, inclusive, infinito. Por ejemplo, los intentos repetidos de un experimento generarán nuevos espacios de probabilidad en los que

¹² Aquí debemos ser cuidadosos. El concepto de independencia que invocamos aquí no es exactamente el mismo de la definición 21 porque los eventos no están definidos en el mismo espacio de probabilidad. Por eso necesitamos esta definición 22. Los eventos

- el espacio muestral para n repeticiones es el n -ésimo producto cartesiano del espacio muestral del experimento individual, $\Omega^{(n)} = \Omega \times \Omega \times \dots \times \Omega$;
- el campo de eventos medibles será el mínimo campo- σ que incluya los n -ésimos productos cartesianos del campo de eventos del experimento individual, $\mathcal{F}^{(n)} = \sigma(\mathcal{F} \times \mathcal{F} \times \dots \times \mathcal{F})$;
- la medida de probabilidad de un evento medible en el nuevo espacio dependerá de las posibles dependencias entre repeticiones del experimento. Si son independientes, por ejemplo, la probabilidad de un evento medible en el nuevo espacio será el producto de las medidas de probabilidad de los eventos respectivos en cada repetición individual.

La tercera característica del nuevo espacio de probabilidad es la razón por la que en teoría de probabilidades se le da tanto énfasis y tanta importancia al concepto de independencia. Varios autores coinciden que es la independencia la que hace de la teoría de la probabilidad un área de la matemática aparte de la teoría de la medida.

El ejemplo más típico es el de la repetición de experimentos de Bernoulli. En cada repetición el campo de eventos es $\mathcal{F} = \{\emptyset, A, A^C, \Omega\}$ para algún $A \subset \Omega$, como en el caso de lanzar una moneda ($A = \{\text{cara}\}$), ver el estado de ocupación de un enlace ($A = \{\text{ocupado}\}$) o medir el retardo de un paquete y ver si supera 100 ms ($A = \{\omega \in \Omega \mid \omega > 0.1\}$). Cuando hacemos dos repeticiones, el espacio muestral es Ω^2 y los eventos de interés pertenecen a \mathcal{F}^2 , y, en general, cuando hacemos n repeticiones, el espacio muestral es Ω^n y los eventos de interés pertenecen a \mathcal{F}^n . Aquí hay dos interpretaciones importantes para el experimento compuesto: En una interpretación estamos viendo estadísticas de un único experimento desde un punto de vista frecuentista, de manera que si en las n repeticiones obtenemos n_A veces el evento A , diríamos que la probabilidad de A es cercana a n_A/n . Pero ahora hacemos una segunda interpretación: No se trata de n veces un experimento de Bernoulli sino de un único nuevo experimento cuyo único proceso de observación consiste en repetir n veces un experimento de Bernoulli. Un posible resultado es que en las primeras n_A repeticiones obtengamos A y en las restantes $n - n_A$ obtengamos A^C , cuya probabilidad es $p^{n_A}(1-p)^{n-n_A}$ si las repeticiones son independientes. Ahora las preguntas sobre el experimento (único) se refieren a cuáles repeticiones obtuvieron A y cuáles obtuvieron A^C . Por ejemplo, una pregunta válida en ese único experimento es cuántas de las n repeticiones fueron A , en cuyo caso obtendríamos la frecuencia relativa asociada con la interpretación anterior del experimento compuesto en donde hacemos n experimentos diferentes.

Como otro ejemplo consideremos una “regla de parada” tal como lanzar una moneda repetidamente hasta que aparezca una cara y, entonces, detenerse. El espacio muestral más sencillo sería $\cup_{n=0,1,2,\dots} \{\{s\}^n, c\}$ que es el conjunto de secuencias de cero o más sellos seguidos por una cara. Por supuesto, surge la inquietud si debemos incluir el caso $n = \infty$. Si consideramos que la probabilidad de que eventualmente aparezca una cara es 1, tal vez sería útil eliminar el caso $n = \infty$ de nuestro espacio muestral pues, después de todo, ¿cómo podríamos llegar a observar ese resultado? Si la probabilidad de cara es mayor a cero, no importa cuán pequeña sea, eventualmente ocurrirá una cara con probabilidad 1. Las siguientes definiciones aclararán esta idea fundamental de la ocurrencia “eventual” de un evento.

23. Continuidad de la Medida de Probabilidad

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay una secuencia creciente de eventos $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ y una secuencia decreciente de eventos $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ contenidos en \mathcal{F} . Entonces

$$P\left[\lim_{i \rightarrow \infty} A_i\right] = \lim_{i \rightarrow \infty} P[A_i] \quad \text{y} \quad P\left[\lim_{i \rightarrow \infty} B_i\right] = \lim_{i \rightarrow \infty} P[B_i]$$

Para una secuencia creciente de eventos sabemos por la definición 9 que el límite de la secuencia es la unión de todos los eventos

$$\lim_{i \rightarrow \infty} A_i = \bigcup_{n=1}^{\infty} A_n$$

la cual se puede expresar como unión de eventos mutuamente excluyentes;

$$\lim_{i \rightarrow \infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup (A_4 \setminus A_3) \cup \dots$$

para aplicar el tercer axioma de la definición 14:

$$P\left[\lim_{i \rightarrow \infty} A_i\right] = P[A_1] + P[A_2 \setminus A_1] + P[A_3 \setminus A_2] + P[A_4 \setminus A_3] + \dots$$

$$P\left[\lim_{i \rightarrow \infty} A_i\right] = P[A_1] + \sum_{n=2}^{\infty} P[A_n \setminus A_{n-1}] = P[A_1] + \lim_{i \rightarrow \infty} \sum_{n=2}^i P[A_n \setminus A_{n-1}]$$

$$P\left[\lim_{i \rightarrow \infty} A_i\right] = P[A_1] + \lim_{i \rightarrow \infty} \sum_{n=2}^i (P[A_n] - P[A_{n-1}]) = P[A_1] + \lim_{i \rightarrow \infty} (P[A_i] - P[A_1])$$

$$P\left[\lim_{i \rightarrow \infty} A_i\right] = \lim_{i \rightarrow \infty} P[A_i]$$

Igualmente, el límite de una secuencia decreciente de eventos es la intersección de todos los eventos (definición 9)

$$\lim_{i \rightarrow \infty} B_i = \bigcap_{n=1}^{\infty} B_n$$

de manera que, según la ley de DeMorgan, la secuencia de complementos es una secuencia creciente:

$$\left(\lim_{i \rightarrow \infty} B_i\right)^C = \left(\bigcap_{n=1}^{\infty} B_n\right)^C = \bigcup_{n=1}^{\infty} B_n^C = \lim_{i \rightarrow \infty} B_i^C$$

a la que aplica el resultado anterior:

$$P\left[\lim_{i \rightarrow \infty} B_i^C\right] = \lim_{i \rightarrow \infty} P[B_i^C]$$

esto es,

$$1 - P\left[\lim_{i \rightarrow \infty} B_i\right] = 1 - \lim_{i \rightarrow \infty} P[B_i]$$

o, lo que es lo mismo,

$$P\left[\lim_{i \rightarrow \infty} B_i\right] = \lim_{i \rightarrow \infty} P[B_i]$$

Consideremos, por ejemplo, una secuencia de lanzamientos de una moneda equilibrada. ¿Con qué probabilidad obtendremos eventualmente una cara? Si definimos el evento A_n como el evento en el

que ha habido al menos una cara en las primeras n lanzadas habremos creado una secuencia creciente de eventos, ya que $A_1 \subset A_2 \subset A_3 \subset \dots$. Claramente, el evento "obtendremos eventualmente una cara" corresponde al evento $\lim_{n \rightarrow \infty} A_n$. Por la continuidad de la medida de probabilidad sabemos que $P[\lim_{n \rightarrow \infty} A_n] = \lim_{n \rightarrow \infty} P[A_n]$. La probabilidad de A_1 es $1/2$, la probabilidad de A_2 es $3/4$, la probabilidad de A_3 es $7/8$ y, en general, la probabilidad de A_n es $(2^n - 1)/2^n$. Esto es, $\lim_{n \rightarrow \infty} P[A_n] = 1$, por lo que $P[\lim_{n \rightarrow \infty} A_n] = 1$: Eventualmente obtendremos una cara con probabilidad 1. Claro, como la probabilidad de cara es tan alta, es de esperar que veamos una cara con bastante frecuencia. Pero si apostamos al baloto en todas sus rifas, también es cierto que eventualmente lo ganaremos pues cada vez que lo juguemos ganaremos con probabilidad $p = (1/16)(5/43)(4/42)(3/41)(2/40)(1/39) = 1/15'401.568 > 0$, de manera que el evento {Ganaremos al menos una vez en n intentos} tiene probabilidad $1 - (1-p)^n \rightarrow 1$. El problema es que esta vez no veremos que dicho evento ocurra con mucha frecuencia pues, si participamos en todas las apuestas, ganaremos en promedio una vez cada ciento cincuenta mil años¹³. Sin embargo, es interesante notar que si, podemos seguir jugando infinitas veces al baloto, lo ganaremos un número infinito de veces, como lo demuestra el segundo lema de Borel-Cantelli que veremos en la definición 25, después de la siguiente breve introducción.

24. Límite del supremo y límite del ínfimo

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay una secuencia de eventos $\{A_1, A_2, A_3, \dots\}$. Definimos el límite del supremo y el límite del ínfimo de la secuencia mediante las siguientes relaciones:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k = \{\omega \in \Omega : \forall n \exists k > n : \omega \in A_k\}$$

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k = \{\omega \in \Omega : \exists n : \forall k > n : \omega \in A_k\}$$

Cuando ocurre el evento $\limsup_{n \rightarrow \infty} A_n$ decimos que ocurrió un número infinito de eventos de la secuencia $\{A_n, n=1,2,\dots\}$, por lo que el límite del supremo también se conoce como $\{A_n, \text{i.o.}\}$ (infinitely often). Cuando ocurre el evento $\liminf_{n \rightarrow \infty} A_n$ decimos que ocurrieron todos los eventos de la secuencia $\{A_n, n=1,2,\dots\}$, con la posible excepción de un número finito de ellos, por lo que el límite del ínfimo también se conoce como $\{A_n, \text{e.a.}\}$ (eventually always).

Claramente, $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$. Si también ocurre que $\lim_{n \rightarrow \infty} \sup A_n \subseteq \lim_{n \rightarrow \infty} \inf A_n$, entonces la secuencia de eventos A_1, A_2, A_3, \dots tiene como límite el evento L dado por

$$L = \lim_{n \rightarrow \infty} \sup A_n = \lim_{n \rightarrow \infty} \inf A_n$$

Para empezar, consideremos primero una secuencia de números reales x_1, x_2, x_3, \dots . El supremo de la secuencia $\{x_m, m \geq n\}$ es la cota superior más pequeña de $\{x_m, m \geq n\}$ (por eso al supremo también se le conoce como LUB, *least upper bound*). Dicho de otra manera, el supremo de $\{x_m, m \geq n\}$ es el

¹³ Ver definición 43(b)

mínimo $x \in \mathbb{R}$ tal que $x \geq x_m \forall m \geq n$. Nótese que el supremo no necesariamente pertenece a la secuencia $\{x_m, m \geq n\}$. Si así fuera, al supremo le diríamos también el máximo, pues en este caso coinciden. En otras palabras, el máximo debe hacer parte de la secuencia mientras que el supremo no está necesariamente contenido en la secuencia. Por ejemplo, $\sup_{n \geq 0} (1 - e^{-n}) = 1$ aunque $\max_{n \geq 0} (1 - e^{-n})$ no existe porque la secuencia tiende a 1 sin llegar jamás a ese valor. Sin embargo $\sup_{n \geq 0} ((-1)^n) = \max_{n \geq 0} ((-1)^n) = 1$. De manera análoga, el ínfimo de una secuencia de números es la máxima cota inferior o GLB, *greatest lower bound*: el ínfimo de $\{x_m, m \geq n\}$ es el máximo $x \in \mathbb{R}$ tal que $x \leq x_m \forall m \geq n$. Para secuencias finitas, el ínfimo es el mismo mínimo como el supremo es el mismo máximo. Pero en secuencias infinitas pueden darse los casos en que el ínfimo y el mínimo coincidan, o que sólo exista el ínfimo o que no exista ninguno de ellos. Por ejemplo, el ínfimo de la secuencia $X = \{1/n, n \in \mathbb{N}\}$ es 0, pero no es el mínimo de X porque $0 \notin X$. El supremo de X es 1, que es a su vez el máximo de X porque $1 \in X$.

Por supuesto, $s_n = \sup\{x_m, m \geq n\}$ es una secuencia no creciente en n mientras que $i_n = \inf\{x_m, m \geq n\}$ es una secuencia no decreciente en n , como muestra la Figura 29.

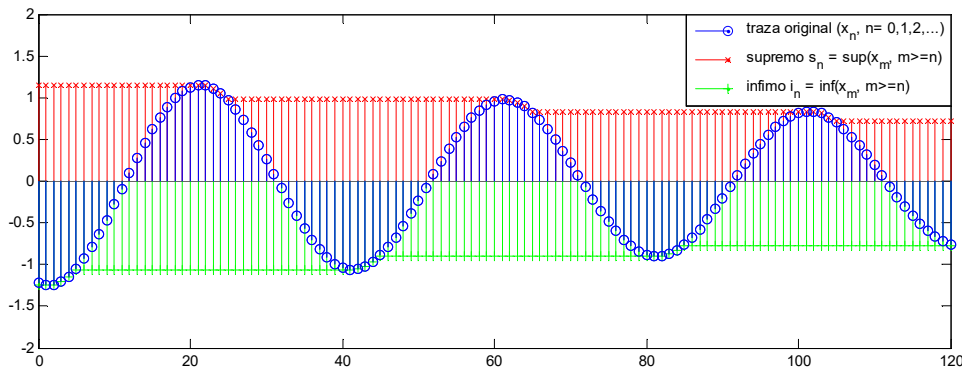


Figura 29. Supremo e ínfimo de una secuencia de números reales

Consideremos ahora el límite de la secuencia decreciente $s_n = \sup\{x_m, m \geq n\}$ cuando n tiende a infinito, conocida como *lim sup*,

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\sup_{m \geq n} x_m \right)$$

Si l_s es el límite del supremo de $\{x_n\}$, para cualquier $\varepsilon > 0$ existe un natural N tal que $x_n < l_s + \varepsilon$ para todo $n > N$, y no hay un número menor que l_s que satisfaga esta misma propiedad. En otras palabras, cualquier número mayor que l_s eventualmente se convertirá en una cota superior de la secuencia $\{x_n\}$, de manera que sólo un número finito de elementos de la secuencia pueden ser mayores a $l_s + \varepsilon$.

Igualmente se puede hablar del límite de la secuencia creciente $i_n = \inf\{x_m, m \geq n\}$ cuando n tiende a infinito, o *lim inf*,

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\inf_{m \geq n} x_m \right)$$

Si l_i es el límite del ínfimo de $\{x_n\}$, para cualquier $\varepsilon > 0$ existe un natural N tal que $x_n > l_i - \varepsilon$ para todo $n > N$, y no hay un número mayor que l_i que satisfaga esta misma propiedad. En otras palabras, cualquier número menor que l_i eventualmente se convertirá en una cota inferior de la secuencia $\{x_n\}$, de manera que sólo un número finito de elementos de la secuencia pueden ser menores a $l_i - \varepsilon$. En la Figura 29, el límite de la señal roja cuando n tiende a infinito es el $\lim sup$ de la secuencia $\{x_n\}$ y el límite de la señal verde cuando n tiende a infinito es el $\lim inf$ de la secuencia $\{x_n\}$. Si los dos límites coinciden, la secuencia converge al valor común de los límites del supremo y del ínfimo, esto es, si existe $L = \lim_{n \rightarrow \infty} x_n$, entonces $L = \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n$, como muestra la Figura 30.

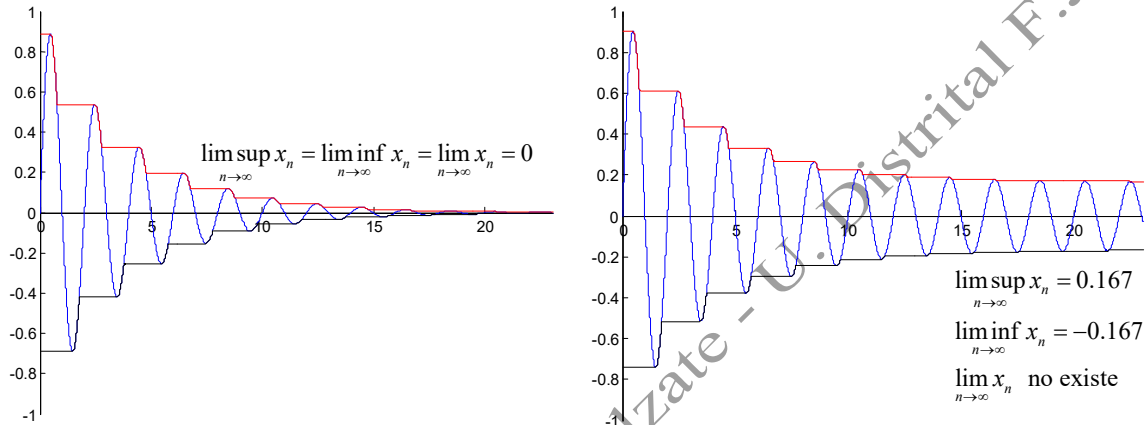


Figura 30. Los límites del supremo y del ínfimo pueden ser diferentes. Si son iguales, corresponden al límite de la secuencia (se cambió stem por plot para mejor visualización).

Permítame insistir en algunas propiedades evidentes y muy interesantes del limsup y el liminf de una secuencia de números son las siguientes:

- $\liminf_{n \rightarrow \infty} \{x_n\} \leq \limsup_{n \rightarrow \infty} \{x_n\}$
- Si $\limsup_{n \rightarrow \infty} \{x_n\} < \infty$, $\forall \varepsilon > 0 \exists N$ tal que $\forall m > N x_m < \varepsilon + \limsup_{n \rightarrow \infty} \{x_n\}$ (decimos que, $\forall \varepsilon > 0$, $x_m < \varepsilon + \limsup_{n \rightarrow \infty} \{x_n\}$ “para todos los m suficientemente grandes” o *eventually always* (e.a.))
- Si $\limsup_{n \rightarrow \infty} \{x_n\} < \infty$, $\forall \varepsilon > 0 \forall N \exists m > N$ tal que $x_m > -\varepsilon + \limsup_{n \rightarrow \infty} \{x_n\}$ (decimos que, $\forall \varepsilon > 0$, $x_m > -\varepsilon + \limsup_{n \rightarrow \infty} \{x_n\}$ “para un número infinito de subíndices m ” o *infinitely often* (i.o.))
- Si $\liminf_{n \rightarrow \infty} \{x_n\} < \infty$, $\forall \varepsilon > 0 \exists N$ tal que $\forall m > N x_m > -\varepsilon + \liminf_{n \rightarrow \infty} \{x_n\}$ (decimos que, $\forall \varepsilon > 0$, $x_m > -\varepsilon + \liminf_{n \rightarrow \infty} \{x_n\}$ “para todos los m suficientemente grandes” o *eventually always* (e.a.))
- Si $\liminf_{n \rightarrow \infty} \{x_n\} < \infty$, $\forall \varepsilon > 0 \forall N \exists m > N$ tal que $x_m < \varepsilon + \liminf_{n \rightarrow \infty} \{x_n\}$ (decimos que, $\forall \varepsilon > 0$, $x_m < \varepsilon + \liminf_{n \rightarrow \infty} \{x_n\}$ “para un número infinito de subíndices m ” o *infinitely often* (i.o.))

Estos conceptos de limsup y liminf para secuencias de números reales son una introducción perfecta para entender los mismos conceptos asociados con una secuencia de eventos $\{A_1, A_2, A_3, \dots\}$. En este caso podemos estar interesados en saber cuántos de los eventos A_n ocurren en una realización del experimento aleatorio. El supremo y el ínfimo de la secuencia son dos nuevas secuencias, la primera decreciente y la segunda creciente, definidas como

$$\sup A_{k \geq n} = \bigcup_{k=n}^{\infty} A_k \qquad \inf A_{k \geq n} = \bigcap_{k=n}^{\infty} A_k$$

(ver definición 9). El evento $\sup\{A_{k \geq n}\}$ ocurre cuando ocurre al menos uno de los $\{A_k, k \geq n\}$. El evento $\inf\{A_{k \geq n}\}$ ocurre cuando ocurren todos los $\{A_k, k \geq n\}$. El límite del supremo de la secuencia de eventos es el evento en el que ocurre un número infinito de eventos de la secuencia. Igualmente, el límite del ínfimo es el evento en que ocurren todos los eventos de la secuencia con excepción de un número finito de ellos:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k \qquad \liminf_{n \rightarrow \infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k$$

En efecto, $\omega \in \limsup_{n \rightarrow \infty} A_n$ quiere decir que $\omega \in \cup_{k=1.. \infty} A_k$, y $\omega \in \cup_{k=2.. \infty} A_k$, y $\omega \in \cup_{k=3.. \infty} A_k$ y, en general, $\omega \in \cup_{k=m.. \infty} A_k$ para todo $m \geq 1$. Esto implica que debe existir algún $i \geq 1$ tal que $\omega \in A_i$. Pero también es cierto que $\omega \in \cup_{k=(i+1).. \infty} A_k$, por lo que debe existir algún $j > i$ tal que $\omega \in A_j$. Pero, como $\omega \in \cup_{k=(j+1).. \infty} A_k$, debe existir algún $l > j$ tal que $\omega \in A_l$. Continuando de esta manera es fácil ver que existe una secuencia infinita $i < j < l < \dots$ de eventos en $\{A_n, n=1,2,3,\dots\}$ a los que ω pertenece, es decir, cuando ocurre $\limsup_{n \rightarrow \infty} A_n$ ocurre un número infinito de eventos en $\{A_n, n=1,2,3,\dots\}$. Por eso al evento $\limsup_{n \rightarrow \infty} A_n$ también se le conoce como $\{A_n, \text{i.o.}\}$, donde i.o. representa la expresión en inglés *infinitely often*.

De otro lado, $\omega \in \liminf_{n \rightarrow \infty} A_n$ quiere decir que $\omega \in \cap_{k=1.. \infty} A_k$, o que $\omega \in \cap_{k=2.. \infty} A_k$, o que $\omega \in \cap_{k=3.. \infty} A_k$ o, en general, que $\omega \in \cap_{k=m.. \infty} A_k$ para algún m , de donde $\omega \in A_i$ para todo $i \geq m$. En consecuencia, ω pertenece a todos los A_n con la posible excepción de un número finito de ellos. Por eso al evento $\liminf_{n \rightarrow \infty} A_n$ también se le conoce como $\{A_n, \text{e.a.}\}$, donde e.a. representa la expresión en inglés *eventually always*.

Sea, por ejemplo, el espacio de probabilidad $([0,1], \mathcal{B}([0,1]), P(A)=\text{longitud}(A))$. Definamos en él las siguientes secuencias de eventos: $\{A_n=[1/n,1], n=1,2,\dots\}$, $\{B_n=[(n-1)/4n,(3n+1)/4n], n=1,2,\dots\}$, $\{C_n=[(n-1)/4n,(3n-1)/4n], n=1,2,\dots\}$. Claramente, $P[A_n]=1-1/n \rightarrow 1$, $P[B_n]=(n+1)/2n \rightarrow 1/2$, $P[C_n]=1/2$ para todo n . En efecto, $\{A_n\}$ es una secuencia creciente con $\lim_{n \rightarrow \infty} A_n = \cup_{n=1.. \infty} A_n = (0,1]$, de manera que $P[\lim_{n \rightarrow \infty} A_n]=\lim_{n \rightarrow \infty} P[A_n]$. $\{B_n\}$, en cambio, es una secuencia decreciente con $\lim_{n \rightarrow \infty} B_n = \cap_{n=1.. \infty} B_n = [1/4,3/4]$, de manera que $P[\lim_{n \rightarrow \infty} B_n]=\lim_{n \rightarrow \infty} P[B_n]$. Hasta aquí sólo hemos verificado la continuidad de la probabilidad. Sin embargo, $\{C_n\}$ no es una secuencia creciente ni decreciente. Podemos construir la secuencia $S_n = \cup_{m=n.. \infty} C_m = [(n-1)/4n,3/4]$ cuya probabilidad es $P[S_n]=(2n+1)/4n \rightarrow 1/2$. En efecto, $\limsup_{n \rightarrow \infty} C_n = \lim_{n \rightarrow \infty} S_n = [1/4,3/4]$, cuya probabilidad es $1/2$. También podemos construir la secuencia $I_n = \cap_{m=n.. \infty} C_m = [1/4,(3n-1)/4n]$ cuya probabilidad es $P[I_n]=(2n-1)/4n \rightarrow 1/2$. En efecto, $\liminf_{n \rightarrow \infty} C_n = \lim_{n \rightarrow \infty} I_n = [1/4,3/4]$, cuya probabilidad es $1/2$. En este caso, $\lim_{n \rightarrow \infty} C_n = \liminf_{n \rightarrow \infty} C_n = \limsup_{n \rightarrow \infty} C_n = [1/4, 3/4]$, y se cumple nuevamente que $P[\lim_{n \rightarrow \infty} C_n]=\lim_{n \rightarrow \infty} P[C_n]$.

Las secuencias de eventos $\{A_1, A_2, A_3,\dots\}$ y los eventos correspondientes $\limsup A_n = \{A_n, \text{i.o.}\}$ y $\liminf A_n = \{A_n, \text{e.a.}\}$ son de gran importancia en redes de telecomunicaciones. Por ejemplo, si A_n es el evento en el que el buffer de un nodo de comunicaciones se encuentra vacío en el instante n , el evento $\{A_n, \text{i.o.}\}$ indica que el buffer volverá a estar desocupado una y otra vez, indicando que no hay

congestión. El evento $\{A_n^C, \text{e.a.}\}$, en cambio, indica que desde algún momento el buffer dejará de estar desocupado, lo que implica que la congestión será inevitable.

25. Lemas de Borel-Cantelli

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que hay una secuencia de eventos A_1, A_2, A_3, \dots . Sea $\{A_n, \text{i.o.}\}$ el evento en el que un número infinito de los A_n ocurre,

$$\{A_n, \text{i.o.}\} = \limsup_{k \rightarrow \infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

El primer lema de Borel-Cantelli establece que

$$P(\{A_n, \text{i.o.}\})=0 \text{ si } \sum_n P(A_n) < \infty$$

El segundo lema de Borel-Cantelli establece que

$$P(\{A_n, \text{i.o.}\})=1 \text{ si } \sum_n P(A_n)=\infty \text{ y los } A_n \text{ son eventos independientes}$$

(En el primer lema de Borel-Cantelli, los A_n no necesariamente deben ser independientes)

Para demostrar el primer lema de Borel/Cantelli debemos recordar primero que si $s_n = \sum_{k=1..n} x_k$ converge a $s < \infty$ cuando n tiende a infinito, entonces x_n converge a cero cuando n tiende a infinito. En efecto, $x_n = s_n - s_{n-1}$ de manera que $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} s_n - \lim_{n \rightarrow \infty} s_{n-1} = s - s = 0$. Siendo así, la suposición $\sum_n P(A_n) < \infty$ implica que $\lim_{n \rightarrow \infty} P(A_n) = 0$. Ahora, como $A = \limsup_{n \rightarrow \infty} A_n \subseteq \bigcup_{m=n.. \infty} A_m$ para todo n , entonces, para cualquier n ,

$$P(A) \leq P\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} P(A_m) \xrightarrow{n \rightarrow \infty} 0 \text{ cuando } \sum_{n=1}^{\infty} P(A_n) < \infty$$

Para demostrar el segundo lema consideremos el complemento de A ,

$$A^C = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^C$$

Por la continuidad de la medida de probabilidad (definición 23),

$$P\left[\lim_{N \rightarrow \infty} \bigcap_{m=n}^N A_m^C\right] = \lim_{N \rightarrow \infty} P\left[\bigcap_{m=n}^N A_m^C\right] = P\left[\bigcap_{m=n}^{\infty} A_m^C\right]$$

y, como suponemos independencia,

$$P\left[\lim_{N \rightarrow \infty} \bigcap_{m=n}^N A_m^C\right] = \prod_{m=n}^{\infty} (1 - P(A_m))$$

pero es bien sabido que para cualquier $x \geq 0$, $1 - x \leq e^{-x}$, por lo que

$$P\left[\lim_{N \rightarrow \infty} \bigcap_{m=n}^N A_m^C\right] \leq \prod_{m=n}^{\infty} \exp(-P(A_m)) = \exp\left(-\sum_{m=n}^{\infty} P(A_m)\right) = 0$$

donde la igualdad a cero se debe a que estamos suponiendo que $\sum_{n=1.. \infty} P(A_n) = \infty$. Entonces

$$P[A^C] = \lim_{n \rightarrow \infty} P\left[\bigcap_{m=n}^{\infty} A_m^C\right] = 0 \Rightarrow P[A] = 1$$

Estos lemas son muy poderosos, pues la ocurrencia de $\{A_n, \text{i.o.}\}$ se puede establecer con probabilidad 1 o con probabilidad 0 con sólo ver la convergencia de una secuencia de números. El ejemplo básico es una secuencia infinita de lanzadas de una una moneda bien balanceada: Obtendremos sello un

número infinito de veces con probabilidad 1 porque las lanzadas son independientes y $\sum_n P(\text{Sello en } n) = \sum_n 1/2^n = \infty$. Aún si la moneda se va deteriorando de manera que la probabilidad de sello en la n -ésima lanzada es $1/(n+1)$, obtendremos sello un número infinito de veces con probabilidad 1 por la divergencia de la serie armónica, aunque cada vez los sellos estarán más espaciados. Sin embargo, si la moneda se deteriora tanto con cada lanzada que la probabilidad de sello en la n -ésima lanzada es $1/(n+1)^2$, finalmente dejaremos de ver sellos porque $\sum_n 1/(n+1)^2 = \pi^2/6 - 1 < \infty$.

En próximos capítulos, los lemas de Borel-Cantelli serán muy útiles al determinar la convergencia de secuencias de variables aleatorias, al determinar la recurrencia de los estados de una cadena de Markov, etc. En machine learning los lemmas de Borel-Cantelli son muy útiles para determinar si, eventualmente, los algoritmos de aprendizaje convergirán a un error tolerable. En redes de comunicaciones es muy importante saber que algunos eventos se repetirán indefinidamente (se produce una transmisión sin errores), o que se producirá eventualmente un evento (se transmite correctamente un paquete específico), etc.

26. Modelo Probabilístico

Cuando representamos el comportamiento de un sistema físico mediante un experimento aleatorio, al espacio de probabilidad correspondiente se le denomina Modelo Probabilístico.

Como ingenieros de redes de telecomunicaciones, diariamente nos enfrentamos a problemas tecnológicos particulares caracterizados por nuestra incertidumbre sobre los resultados de las mediciones que no nos es posible observar, ya sea porque son mediciones en el futuro (“Diseña una red en el que el máximo retardo de un paquete sea menor a 50 ms el 99% de las veces”) o porque no tenemos acceso directo a ellas (“Determine el ancho de banda disponible a lo largo de una ruta para el flujo entre dos usuarios dados”). Esta situación se presenta de manera mucho más explícita cuando nuestra actividad profesional está asociada con la investigación y el desarrollo, como se espera que suceda con los estudiantes de postgrado que estudian este libro. En estos casos, se hará necesario especificar el problema en términos de un experimento aleatorio al que se le debe asociar un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ adecuadamente seleccionado. Como este espacio de probabilidad, aunque descrito con toda la formalidad de la teoría de probabilidades, ya no se refiere a un concepto completamente abstracto e inmaterial del mundo de las matemáticas teóricas sino a un modelo matemático de un sistema real, cuyos resultados deben ser aplicables nuevamente a la realidad de donde provino, se le denomina Modelo Probabilístico.

El objetivo del modelado de sistemas, en general, es representar una realidad altamente compleja mediante un modelo matemático que sea lo más sencillo posible pero que capture los aspectos más relevantes que afectan el desempeño del sistema en estudio, de manera que la complejidad no imposibilite el análisis del modelo y que dicho análisis nos permita comprender mejor el comportamiento del sistema real. En el caso del modelado probabilístico, se trata de reformular la pregunta que nos queremos responder sobre el sistema real en términos de algunas características de

un experimento aleatorio. Una vez se ha hecho explícito el experimento, debemos especificar un espacio de probabilidad correspondiente, que se convertirá en un Modelo Probabilístico del Sistema Real.

Ya discutimos en qué consiste el proceso de determinar el espacio muestral y el conjunto de eventos de interés con los que debemos construir el campo- σ de eventos medibles. Sin embargo especificar la medida de probabilidad es un proceso algo más elaborado. El espacio de probabilidad exige que determinemos $\mathbf{P}(A)$ para todo $A \in \mathcal{F}$. Lo que podemos hacer en un primer paso es determinar la probabilidad de algunos eventos en \mathcal{F} mediante algún proceso inductivo que tenga en cuenta la naturaleza precisa del “proceso de observación” de la definición 4. En este primer paso podemos usar mediciones de frecuencias relativas (en cuyo caso debemos ser muy juiciosos con el diseño estadístico de los experimentos) o, en el peor de los casos, elucubrar mediante argumentos razonables sobre independencias, simetrías, uniformidades u otras propiedades que podamos derivar del conocimiento inicial que tengamos de la naturaleza del experimento aleatorio. A partir de las probabilidades de estos eventos iniciales, en un segundo paso podemos calcular las probabilidades de otros eventos mediante el uso juicioso de las herramientas con que dotamos al lector en este libro, tales como los axiomas de la definición 14, los resultados de la definición 17, el teorema de la probabilidad total, la regla de Bayes, la composición de experimentos, etc. Mientras el primer paso es un proceso inductivo asociado con la naturaleza del experimento aleatorio, el segundo paso es un proceso deductivo asociado con la rigurosidad matemática de la teoría de probabilidades. En un tercer paso, debemos inferir las implicaciones que los resultados deducidos en el paso dos tengan sobre la realidad compleja que estamos modelando. Este es otro proceso inductivo que le da validez al modelo probabilístico si las implicaciones son correctas y útiles (o lo invalida si las implicaciones son incorrectas o inútiles).

Desafortunadamente, este libro no es sobre modelado probabilístico de redes de telecomunicaciones (cómo seleccionar un modelo para una realidad tecnológica compleja dada) sino sobre cómo analizar un modelo probabilístico dado. De hecho, obsérvese que a partir de la definición 16 (espacio de probabilidad) todas las definiciones subsecuentes (17 a 25) siempre han empezado con la frase “Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad en el que ...”. Para nosotros, en este libro, el modelo siempre va a estar dado y simplemente nos dedicaremos a obtener herramientas para el proceso deductivo del segundo paso en el modelado probabilístico. Más aún, para ahorrarnos tener que reescribir esa frase en todas las definiciones que siguen, en el siguiente capítulo definiremos la variable aleatoria con el propósito de dejar explícito un modelo probabilístico particular $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{H}(\cdot))$ que será el que usemos casi siempre de ahí en adelante. En este libro apenas vamos a llenar la caja de herramientas del lector con un buen conjunto de aparatos útiles, entre los cuales hemos puesto ya los tres axiomas, cinco de las innumerables conclusiones que se pueden sacar de ellos, el teorema de la probabilidad total y la regla de Bayes: diríamos que ya pusimos en la caja el serrucho y el martillo. En los siguientes capítulos colocaremos en ella algunas herramientas más sofisticadas. Sin embargo, sólo la experiencia le permitirá al lector aprender a “construir muebles” con las herramientas de su caja a partir de “la madera” proporcionada por las redes de comunicaciones. Lo más que podemos hacer (y esa es la razón de este libro) es escoger ejemplos del mundo de las redes de comunicaciones, pues pienso que el aprendizaje de las técnicas de modelado sólo es posible mediante el estudio de ejemplos y la sorprendente capacidad de generalización que exhibe la inteligencia humana. Claro, como nuestros

ejemplos son puramente pedagógicos, invitamos al lector a que se afilie a IEEE, ACM y tantas otras sociedades técnicas que publican excelentes revistas en las que cada artículo es un ejemplo realista de un proceso de modelado probabilístico llevado a cabo por uno o varios expertos.

27. ¿Por qué es necesario definir un espacio de probabilidad (Ω, \mathcal{F}, P) ?

No es posible definir una medida de probabilidad sobre todos los subconjuntos del espacio muestral $[0,1]$ de manera que satisfaga los tres axiomas.

Supongamos un experimento aleatorio en el que el espacio muestral Ω es el intervalo unitario $[0,1]$, de manera que la probabilidad de que el resultado del experimento caiga en cualquier intervalo de Ω es igual a la longitud del intervalo: $P[[0,1/2]] = 1/2$, $P[(1/4,1/2]] = 1/4$, $P[(1/8,1/3) \cup [2/3, 8/9]] = 31/72$, $P[\{1/2\}] = 0$, $P[\{1/2\} \cup \{3/4\}] = 0$, $P[\{1/2\} \cup (2/3, 8/9)] = 2/9$, etc.

Si A_1, A_2, A_3, \dots son subconjuntos disyuntos de $[0,1]$, sabemos que $P[A_1 \cup A_2 \cup A_3 \cup \dots] = P[A_1] + P[A_2] + P[A_3] + \dots$. Pero es imposible extender esta propiedad a una suma incontable de eventos disyuntos pues, si la extendiéramos, deberíamos esperar que:

$$P([0,1]) = \sum_{x \in [0,1]} P[\{x\}]$$

lo cual es evidentemente falso: mientras la expresión de la izquierda es igual a uno, la expresión de la derecha es igual cero¹⁴. Esta es la razón por la cual debemos restringirnos a considerar únicamente operaciones contables y es por eso que necesitamos definir la probabilidad sobre un campo sigma, esto es, cerrado para los complementos y las uniones contables.

Consideremos la misma medida de probabilidad en el mismo espacio muestral (el intervalo unitario $[0,1]$, donde la probabilidad de cualquier intervalo es igual a la longitud del intervalo). El resultado fundamental de esta definición²⁷ (que no es posible definir una medida de probabilidad para todos los subconjuntos del espacio muestral $[0,1]$ que satisfaga los tres axiomas) se puede demostrar recurriendo a los conjuntos de Vitali que, como mencionamos, no son conjuntos de Borel. Supongamos que se puede definir $P[A]$ para cualquier $A \subset [0,1]$ y derivemos una contradicción de dicha suposición. Para esto definamos una relación de equivalencia¹⁵ en $[0,1]$ de la siguiente manera: $x \sim y$ si y sólo si la diferencia $y-x$ es racional. Esta relación de equivalencia genera una partición de Ω . Sea C un conjunto formado con uno y sólo un elemento de cada clase de equivalencia¹⁶ que no

¹⁴ La suma sobre un conjunto incontable $|A| = \aleph_1$, $x = \sum_{\alpha \in A} x_\alpha$, se define como el supremo de $\sum_{i \in I} x_i$, entre todos los subconjuntos contables I contenidos en A . Así es como sabemos que la suma no contable de ceros es igual a cero.

¹⁵ Una relación en un conjunto Ω es una función booleana en $\Omega \times \Omega$ (dados $x, y \in \Omega$ sólo puede ocurrir que x esté relacionada con y ($x \sim y$) o no). Una relación es de equivalencia si (1) es reflexiva ($x \sim x \forall x \in \Omega$), (2) es simétrica ($x \sim y \Leftrightarrow y \sim x$) y (3) es transitiva ($x \sim z$ si $x \sim y$ y $y \sim z$). Dada una relación de equivalencia \sim y un elemento $x \in \Omega$, la clase equivalente de x es el conjunto de todos los y tales que $x \sim y$. Claramente, dos clases equivalentes sólo pueden ser idénticas o disyuntas. Dada una relación de equivalencia \sim , el conjunto de clases de equivalencia forman una partición de Ω .

¹⁶ La posibilidad de escoger uno y sólo un elemento de cada subconjunto se conoce como el "axioma de la selección".

Parece obvio si tenemos un número finito de conjuntos finitos (tome exactamente un músico de cada una de las bandas de rock en Bogotá). Pero cuando tenemos infinitos conjuntos, cada uno de ellos de cardinalidad infinita, el axioma de selección no es tan obvio y debe proponerse, precisamente, como un axioma.

contenga 0 (si escogimos cero, lo podemos cambiar, por ejemplo, por $1/2$, que es equivalente a cero). Sea $C \oplus r$ el desplazamiento circular de C por r :

$$C \oplus r = \{c+r : c \in C, c+r < 1\} \cup \{c+r-1 : c \in C, c+r > 1\}$$

Claramente, $P[A \oplus r] = P(A)$ $0 \leq r \leq 1$. Además, cada punto en $(0,1]$ pertenece a la unión $\cup_{r \in [0,1) \cap \mathbb{Q}} (C \oplus r)$ de desplazamientos circulares de C . Más aún, como C contiene exactamente un elemento de cada clase equivalente, todos los conjuntos $C \oplus r$ son disyuntos para $r \in [0,1) \cap \mathbb{Q}$. De esta manera, el tercer axioma exige que

$$P[(0,1]] = \sum_{r \in [0,1) \cap \mathbb{Q}} P[C \oplus r]$$

Pero $P[C \oplus r] = P[C]$, de donde

$$1 = P[(0,1]] = \sum_{r \in [0,1) \cap \mathbb{Q}} P[C]$$

Esta es la contradicción que estábamos buscando: Al sumar un número contable infinito de veces el mismo número no negativo sólo podemos obtener 0 si $P[C]=0$ u obtener ∞ si $P[C]>0$, pero nunca podremos obtener 1. En conclusión, si las probabilidades han de satisfacer propiedades razonables, no podemos definir las para todos los posibles subconjuntos de $[0,1]$ sino que debemos restringirnos a algunos conjuntos medibles.

28. Sobre los Conceptos de Aleatoriedad y Probabilidad

El concepto de aleatoriedad presenta muchas dificultades intuitivas, que aún son materia de controversia entre filósofos y que pueden confundir a los estudiantes neófitos. En el análisis de modelos probabilísticos debemos usar con precaución la intuición, sólo como una guía que siempre debe ser corroborada por el formalismo axiomático de Kolmogorov, pues en muchas ocasiones la intuición puede fallar drásticamente. De todas maneras, dada la naturaleza de la mayoría de experimentos que se refieren a redes de comunicaciones, en los que casi siempre están involucrados o un gran número de usuarios, o un gran número de paquetes, o un gran número de bits, etc., la intuición basada en la interpretación de la probabilidad como frecuencia relativa suele sugerir caminos acertados en el proceso hacia el objetivo del modelado probabilístico en redes de comunicaciones. Después de todo, como dijo Laplace, “En el fondo, la teoría de las probabilidades no es más que sentido común reducido al cálculo”.

Todos los seres humanos se encuentran con el azar en cada momento de sus vidas: en el noticiero recibimos estadísticas económicas, encuestas políticas y predicciones del clima; desde niños jugamos con dados, cartas y monedas; hacemos filas en los bancos y supermercados; padecemos trancones en las avenidas; al decidir dónde invertir nuestros ahorros debemos evaluar el riesgo y ponderarlo respecto a los posibles rendimientos de cada una de las opciones; como usuarios de las redes modernas de comunicaciones sufrimos tiempos excesivos de respuesta, falta de disponibilidad en los recursos de la red, ruidos y recortes en las señales que recibimos, etc. Sin embargo, a pesar de nuestra vasta experiencia con el azar, el concepto de aleatoriedad sigue estando muy alejado del común de la gente

e, inclusive, de muchas personas muy bien preparadas en distintas profesiones (incluyendo las ciencias y la ingeniería: Me consta!). No es sólo por la necesidad de entender conceptos tan abstractos como los asociados con la teoría de medidas (definiciones 10, 11, 12, 13, 23, 25 y 27) sino por las múltiples alternativas de interpretación de la medida de probabilidad.

Y es que el concepto de aleatoriedad se va alcanzando gradual y muy lentamente. Por ejemplo, cuando mi hija de cinco años y mi hijo de cuatro años deseaban resolver alguna disputa de manera “justa”, recurrían a un juego infantil muy tradicional en Colombia: Uno de ellos canta “Pi NU no, pin DOS, pin TRES, pin CUA tro, pin CIN co, pin SEIS, pin SIE te, PIno, CHItto, SErás, TÚ” mientras que con el dedo índice se señala a sí mismo o al hermanito, alternando la dirección con cada sílaba en mayúsculas. María Alejandra iniciaba señalándose ella misma, mientras que Juan Diego parecía escoger *al azar* si iniciaba señalándose él o señalando a la hermanita. Aunque María Alejandra fue una fervorosa defensora del método de conciliación, Juan Diego empezó a perder la confianza en él porque, extrañamente, sólo ganaba la mitad de las veces en que él mismo contaba.

La humanidad misma parece haber seguido ese mismo proceso gradual y lento que sigue cada ser humano individualmente, pues sólo hasta el siglo XVI se empezó a formalizar un concepto que, hasta entonces, era sólo el mecanismo de expresión de las voluntades divinas. ¡Y qué útil resultaba ser el vocero de los dioses cuando se usaban dados no balanceados! Lo cierto es que los seres humanos (y la humanidad entera en su conjunto) primero aprendemos mediante la intuición y, después, sobre esa base, empezamos a formalizar conceptos. Cuando la intuición es correcta, ese proceso es formidable porque ayuda profundamente en la comprensión de temas difíciles. Yo mismo, como profesor de Procesamiento Digital de Señales, Sistemas de Comunicación, Redes de Comunicaciones y Control de Sistemas Dinámicos, me preocupo por presentar los conceptos de manera que los estudiantes primero los capturen de manera intuitiva antes de aprenderlos desde las formalidades matemáticas o tecnológicas. Pero jamás intento hacer eso como profesor de Probabilidades, Variables Aleatorias y Procesos estocásticos porque, en los problemas asociados a estos temas, la intuición suele fallar miserablemente! Por eso he dejado esta discusión para el final de este capítulo, cuando ya hemos visto la formulación axiomática de Kolmogorov, pues en este caso es mucho mejor presentar la formalidad antes que la intuición. Ya en el siglo XVIII DeMoivre lo mencionó: Los problemas que tienen que ver con el azar suelen parecer fácilmente solucionables mediante el sentido común, cuando en realidad casi nunca es así.

Veamos algunos ejemplos:

(1) Un presentador de un concurso de televisión le ofrece que escoja una de tres puertas sabiendo que sólo una de ellas conduce a un gran premio mientras que las otras dos sólo llevan a pequeños premios de consolación. Llamemos a a la puerta que usted escoge. Una vez usted escogió su puerta, el presentador le revela una de las otras dos puertas que conducía a un premio de consolación. Llamemos b a la puerta revelada por el presentador y c a la otra puerta. Ahora el presentador le pregunta: “¿Desea quedarse con a o prefiere cambiarse a c ?”. ¿Cuál sería la mejor estrategia en este juego? He aquí el razonamiento que hace la gran mayoría de personas:

Al principio cada puerta tenía una probabilidad $1/3$ de conducir al premio mayor, independientemente de la que yo escogiera. Sin embargo, una vez el presentador me revela que b no tiene el premio, me quedan sólo dos puertas, a y c . Como sólo una de las dos conduce al premio, la probabilidad de que cada una de ellas conduzca al premio es $1/2$. El hecho de que yo haya escogido antes la puerta a no cambia el nuevo hecho de que ahora tengo dos puertas, una de las cuales conduce al premio grande y otra al premio de consolación. Luego da igual si me quedo con a o si me cambio a c , pues en ambos casos ganaré o perderé con probabilidad $1/2$.

Deténgase usted, señor lector, a pensar un poco en el análisis anterior antes de seguir leyendo. Ahora sí, he aquí el análisis (correcto) que hacen muy pocas personas y que, seguramente, fue el que usted hizo:

Sean los eventos $A = \{a \text{ tiene el premio}\}$, $B = \{b \text{ tiene el premio}\}$ y $C = \{c \text{ tiene el premio}\}$. En un principio, $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/3$. Si el presentador me hubiera informado de la ocurrencia del evento B^C antes de que yo escogiera la puerta a , tendría el caso del análisis anterior, $\mathbf{P}(A|B^C_{\text{antes}}) = \mathbf{P}(C|B^C_{\text{antes}}) = 1/2$. Sin embargo el protocolo del juego me permitió escoger primero, cuando todavía $\mathbf{P}(A)$ era un tercio. Un hecho cierto en ese momento era que al menos una de las otras dos puertas debía conducir a un premio de consolación, de manera que cuando el presentador me revela que era b la que conducía al premio de consolación no me dio ninguna información nueva: La probabilidad de A cuando yo la escogí era de $1/3$, independientemente de que se me revelara B^C después de mi escogencia, $\mathbf{P}(A|B^C_{\text{después}}) = \mathbf{P}(A) = 1/3$. Una vez informado de la ocurrencia del evento B^C , la única alternativa al evento A es el evento C , de manera que $\mathbf{P}(C|B^C_{\text{después}}) = 1 - \mathbf{P}(A|B^C_{\text{después}}) = 1 - \mathbf{P}(A) = 2/3$. Es mejor estrategia cambiarse a la puerta c .

Nótese que la solución correcta no concuerda con la solución intuitiva porque hay un asunto de protocolo que pasa inadvertido. Si el presentador le hubiera revelado una puerta mala antes de que usted hubiera escogido la puerta a , él tendría dos posibles puertas para escoger. Cuando él lo deja escoger primero, usted le reduce sus posibilidades de dos a una, con probabilidad $1/3$. Un gran amigo mío lo pone de la siguiente manera para ver la bondad intuitiva de la solución correcta: Suponga que no son tres sino mil puertas, de las cuales una sola conduce al premio. Usted escoge una, el presentador le revela 998 que no tenían premio... ¿Se quedaría con la que escogió primero? Es “casi seguro” que el premio está en la otra puerta!

(2) Supongamos que cada nuevo bebé que llega al mundo es niño o niña con probabilidad $1/2$, independientemente del resto de la humanidad y, en particular, independientemente de sus hermanitos y hermanitas. Bajo esta suposición consideremos los siguientes problemas:

Se encuentra con una amiga a quien no veía desde hacía diez años y sostienen la siguiente conversación:

Usted: “¡Hola! ¿tienes hijos?”

Amiga: “Sí. Tengo dos”

Usted: “¿alguna niña?”

Amiga: “Sí”

Usted: “¡Adiós!”

¿Cuál es la probabilidad de que su amiga tenga dos niñas?

Al otro día se encuentra otra vez con su amiga y ve que lleva a una niña de la mano:

Usted: “¡Hola! ¿Esta niña tan preciosa es hija tuya?”

Amiga: “Sí”

Usted: “¡Adiós!”

¿Cuál es la probabilidad de que su amiga tenga dos niñas?

A diferencia del ejemplo anterior, en este caso distintas personas hacen diferentes razonamientos. Sin embargo casi todo el mundo da por hecho que la pregunta es la misma y, por lo tanto, exige la misma respuesta. Después de todo, dicen, desde el primer día yo ya sabía que mi amiga tenía por lo menos una niña, así que en el segundo día no obtengo ninguna información nueva al ver exactamente eso: una niña hija de mi amiga, que yo ya sabía que existía! Como verla o no verla no hace ninguna diferencia, la pregunta (y la respuesta) no cambia de un día para otro.

Deténgase usted, señores lector, a pensar un poco en el análisis anterior antes de seguir leyendo. Ahora sí, he aquí el análisis (correcto) que hacen muy pocas personas y que, seguramente, fue el que usted hizo:

Cada hijo de mi amiga puede ser una niña (f) o un niño (m). El primer día supe que mi amiga tenía dos hijos, de manera que el espacio muestral del experimento consistente en observar el género de cada uno de los hijos de mi amiga es $\Omega = \{(f,f), (f,m), (m,f), (m,m)\}$, donde cada evento unitario tiene probabilidad $1/4$. Pero también supe que el evento $\{(m,m)\}$ no ocurrió. Luego la pregunta que me hago el primer día es $\mathbf{P}(\{(f,f)\} \mid \{(f,f), (f,m), (m,f)\}) = \mathbf{P}(\{(f,f)\}) / (1 - \mathbf{P}(\{(m,m)\})) = 1/3$. Sin embargo el segundo día me estoy preguntando por la probabilidad de que el otro hijo (el que no estoy viendo) sea una niña, lo cual ocurre con probabilidad $1/2$. ¿La probabilidad de que mi amiga tenga dos niñas aumentó de $1/3$ a $1/2$ solamente porque pude observar a una niña de mi amiga, siendo que yo ya sabía que tenía por lo menos una niña? No. Es solamente que la pregunta que me estoy haciendo es distinta, ¡a pesar de que la formulación parece idéntica!

Los dos ejemplos anteriores son paradojas que confunden inclusive a algunos expertos en probabilidades (tal vez usted mismo haya releído cada caso varias veces y haya tomado papel y lápiz para ver con detalle qué es lo que está pasando). Hay otros muchos ejemplos que, aunque no confundían a los expertos, si resultan paradójicos para el común de la gente, como el siguiente:

(3) Ante un juez presentan a una persona muy honorable, x , acusada de haber cometido un crimen. La reputación de x la precede, de manera que el evento $X = \{x \text{ cometió el crimen}\}$ tiene una probabilidad muy bajita: $\mathbf{P}(X) = 0.001$. La fiscalía presenta dos testigos muy confiables, y y z , cada uno de los cuales dice la verdad con probabilidad 0.9 . Más aún, estos testigos son independientes pues no se conocen entre sí y observaron los hechos desde sitios distintos. Durante el juicio se presentan los eventos $Y = \{y \text{ dice que } x \text{ cometió el crimen}\}$ y $Z = \{z \text{ dice que } x \text{ cometió el crimen}\}$. Según la más antigua tradición legislativa y judicial (Moisés, Hamurabi, Nemqueteba) y la opinión

de la mayoría de personas, x es condenado porque el testimonio de dos testigos independientes muy confiables es evidencia suficiente. ¿Se hizo justicia?

Veamos cómo se afecta la probabilidad de X cuando la condicionamos a los eventos Y y Z :

$$P(X|Y \cap Z) = \frac{P(X)P(Y \cap Z|X)}{P(X)P(Y \cap Z|X) + P(X^c)P(Y \cap Z|X^c)} = \frac{P(X)P(Y|X)P(Z|X)}{P(X)P(Y|X)P(Z|X) + P(X^c)P(Y|X^c)P(Z|X^c)}$$

$$P(X|Y \cap Z) = \frac{(0.001)(0.9)(0.9)}{(0.001)(0.9)(0.9) + (0.999)(0.1)(0.1)} = \frac{3}{40}$$

Ciertamente los testigos aumentan la probabilidad de X 75 veces (desde 0.001 hasta 0.075), pero condenar a alguien cuando la probabilidad de que haya cometido el crimen es menos de 0.1 me parece muy injusto!

Lo que pasa, como lo anunció DeMoivre, es que la intuición suele fallar estrepitosamente en asuntos de aleatoriedad. ¿Cuántas veces lo ha abordado un voceador de lotería en la calle anunciándole un dato valiosísimo: “Hace cinco semanas que el número 3 no sale en la lotería y aquí le tengo un billetico terminado en 3”. Él espera convencerlo de que compre el billete porque el hecho de que el 3 no haya salido cinco veces seguidas aumenta la probabilidad de que salga la próxima vez. De hecho haga una encuesta entre sus amigos no ingenieros ni científicos y observará que muchos de ellos creen que, si se ha lanzado cinco veces una moneda bien equilibrada y en las cinco ocasiones se ha obtenido cara, aumentan las probabilidades de que en la próxima lanzada se obtenga sello pues, después de todo, el sello debería aparecer en la mitad de las lanzadas¹⁷.

Las paradojas y las concepciones populares y erróneas sólo indican que el concepto de aleatoriedad parece obvio, cuando en realidad no lo es. De hecho, su misma interpretación ha enfrentado a importantes científicos. Ya 500 años antes de Cristo Leucipo había manifestado que nada ocurre por azar sino que todo obedece a la razón y a la necesidad. Así pues, lo que denominamos azar se refiere a los efectos de causas escondidas que están fuera de nuestro conocimiento o fuera de nuestro control, como sostuvo Demócrito, discípulo de Leucipo. Los sofistas reafirmaron este concepto en contra de Epicuro, quien sostenía que si todos los eventos tenían una causa conocible, el hombre carecería de libre albedrío... El cristianismo ayudó a afianzar el concepto sofista, pues el resultado final de todos los experimentos debía obedecer, necesariamente, a la voluntad de Dios: En un universo en el que todo está sometido a la voluntad de Dios (y a su plan de salvación) sólo nuestra ignorancia puede abrirle espacios al azar. Con la aparición de la mecánica Newtoniana pocos años después, se terminó de consolidar la visión determinista: La voluntad de Dios se manifestaba en leyes del movimiento

¹⁷ Existe un muy reconocido locutor y comentarista deportivo en Colombia que debe ser muy bien formado porque le dicen “doctor”. A él le escuché decir en una ocasión lo siguiente: “La historia muestra que el equipo A le ha ganado al equipo B en el 70% de los encuentros que han disputado. Sin embargo, en los últimos tres meses se han enfrentado cinco veces y en todas ellas ha ganado B . Luego, si la teoría de las probabilidades no falla, A debería ganar en el próximo partido”. Claro, A perdió el partido de esa tarde contra B y no por una falla de la teoría de las probabilidades sino porque, al menos desde hacía tres meses, A parecía ser un equipo de “troncos”. Si el locutor doctor quiso ser optimista, debió suponer que cada partido es independiente de los demás, en cuyo caso la probabilidad de que A ganara esa tarde sería 0.7. Sin embargo, parece mucho más correcto pensar que la probabilidad de que A le ganara esa tarde a B dado que A llevaba cinco partidos seguidos perdiendo contra B era un número muy cercano a cero.

que eran asequibles para el hombre a través de las matemáticas. Según Laplace, sólo necesitamos conocer la masa, la posición y la velocidad de cada partícula del Universo en un instante dado para predecir con precisión su destino último y su pasado más remoto. Siendo el comportamiento del universo tan determinístico y predecible, ¿cómo pudo, entonces, desarrollarse tan profunda y aceleradamente la teoría de las probabilidades durante este período? Porque si bien el mundo estaba sometido al determinismo de la voluntad inmutable de Dios, nuestra capacidad de observación era limitada: El hombre comete errores y la probabilidad nos permite cuantificar el error. Ya Galileo y Tycho Brahe formularon proposiciones fascinantes sobre el error en las mediciones astronómicas; es inevitable, es simétrico y entre más pequeño sea más probable es. Estas ideas sugieren tomar muchas mediciones y promediarlas, de donde surgen las leyes de los grandes números. Thomas Simpson fue el primero en introducir la teoría de los juegos de azar, cuando en 1756 dijo que si cada fuente de error se comportaba como un dado, el error total se debe comportar como la suma de muchos dados. En 1808 Gauss usó la famosa campana $\exp(-x^2/2)/\sqrt{2}$, justificado por el teorema del límite central, propuesto por Laplace 2 años después. Tendremos oportunidad en este libro de divertirnos un buen rato con los tres conceptos: la campana gaussiana, la ley de los grandes números y el teorema del límite central.

Si bien el determinismo se remonta a los atomistas como Leucipo y Demócrito, es precisamente el estudio de las partículas subatómicas el que reivindica a Epicuro: La naturaleza puede ser inherentemente aleatoria, ¿puede ser que haya una indeterminación básica en el universo! En 1900 Max Planck explicó (exitosamente) por qué los cuerpos calientes no irradiaban en todas las posibles frecuencias, diciendo que la radiación se daba en “cuantos” de energía. Lo que sorprendió a Planck es que esta idea resultó más que un truco matemático cuando muchos físicos empezaron a encontrar más comportamientos cuánticos en las partículas subatómicas. En 1926 Heisenberg fue el primero en advertir que, siendo así, jamás nos sería dado conocer el estado del universo como lo proponía Laplace, pues para observar una pequeña partícula debemos iluminarla con, por lo menos, un cuanto de luz, alterando irremediamente el estado que queríamos observar. Esto conduce al principio de incertidumbre como una de las leyes básicas de la naturaleza: si conocemos la posición exacta de una partícula no podemos saber nada sobre su velocidad y viceversa. Schrödinger, de hecho, describe las partículas mediante una ecuación de onda con la que evalúa *la probabilidad* de que una partícula se encuentre, en un instante de tiempo particular, en un punto dado del espacio! Albert Einstein, quien contribuyó notablemente al desarrollo de la mecánica cuántica con su estudio de la radiación de cuerpo negro, nunca se sintió cómodo con esta idea pues él era abiertamente determinista. Una conversación entre Einstein y Max Born pudo haber ocurrido hace 2500 años entre Demócrito y Epicuro: Einstein (o Demócrito): “Dios no juega a los dados”. Born (o Epicuro): “Y ¿quiénes somos nosotros para decidir a qué puede o no puede jugar Dios?”¹⁸

¿Dónde está, pues, la aleatoriedad de una secuencia de resultados obtenidos al repetir muchas veces un experimento? Si creemos que la naturaleza sufre de una indeterminación básica, la aleatoriedad de la secuencia está en el experimento mismo que la generó; si creemos que las leyes que rigen el

¹⁸ En una carta a Max Born Einstein escribió "La mecánica cuántica es realmente imponente. Pero una voz interior me dice que aún no es la buena. La teoría dice mucho, pero no nos aproxima realmente al secreto del 'viejo'. Yo, en cualquier caso, estoy convencido de que Él no tira dados".

experimento son conocidas pero muy complejas y difíciles de evaluar, la aleatoriedad de la secuencia está en nuestra incapacidad computacional para calcular el siguiente resultado; y si creemos que las leyes existen pero no las conocemos, la aleatoriedad de la secuencia está en nuestra ignorancia. En cualquiera de los tres casos, la aleatoriedad se caracteriza por nuestra incapacidad para predecir el resultado del siguiente experimento, como propusimos en la primera definición. En los dos últimos casos, sin embargo, la aleatoriedad se vuelve un asunto subjetivo: alguien que esté mejor capacitado que nosotros para predecir el resultado del siguiente experimento encontrará que la secuencia es menos aleatoria de lo que nosotros creemos. Consideremos, por ejemplo, las siguientes secuencias de números:

1	2	3	4	5	6	7	8	9	10	...
2	4	6	8	10	12	14	16	18	20	...
4	7	2	6	4	7	2	6	4	7	...
3	1	4	1	5	9	2	6	5	3	...

En cada una de ellas es fácil predecir el siguiente número, ¿cierto? 11 para la primera secuencia, porque los números parecen ir de uno en uno; 22 para la segunda secuencia porque los números parecen ir de dos en dos; 2 para la tercera secuencia porque los números parecen tener un período de longitud 4; ¿identifica usted fácilmente el siguiente número de la cuarta secuencia? Parece una secuencia aleatoria, hasta que reconocemos en ella la expansión decimal de π : el siguiente número es 5. Se diría que las cuatro secuencias anteriores son completamente determinísticas, aunque un estudiante de primaria que no haya visto trigonometría podría considerar que la cuarta secuencia es aleatoria ¿Qué tal la siguiente secuencia?

7	5	7	9	6	3	9	2	7	...
---	---	---	---	---	---	---	---	---	-----

Difícil adivinar el siguiente número, ¿cierto? Pero se trata de una secuencia completamente determinística. Iniciando con $Z_0 = 7182$, hago $Z_{i+1} = [\text{floor}(Z_i^2/100)]_4$, donde $\text{floor}(x)$ es la parte entera de x , y $[x]_4$ es el número compuesto por los cuatro dígitos menos significativos de x (las unidades, decenas, centenas y unidades de mil en x). El i -ésimo número de la secuencia corresponde a las unidades de mil en Z_i . La siguiente tabla muestra cómo se construye la secuencia completa. Nótese que a partir del último dígito en la secuencia mostrada, continúa una cadena infinita de ceros. Sin embargo, hasta donde se mostró, parecía una secuencia completamente aleatoria, aunque no lo sería para quien conozca el algoritmo y el valor de Z_0 . ¿Dónde está, entonces, la aleatoriedad?

i	Z_i	Z_i^2
0	7182	51581124
1	5811	33767721
2	7677	58936329
3	9363	87665769
4	6657	44315649
5	3156	09960336
6	9603	92217609

7 2176 04734976
8 7349 54007801

Nadie duda que las lanzadas consecutivas de un dado generan una secuencia aleatoria porque la única manera de conocer el siguiente número de la secuencia es lanzando el dado una vez más. Pero, si usamos un algoritmo para calcular π con cualquier precisión deseada y nos muestran su expansión decimal a partir del dígito 100, ¿no consideraríamos la secuencia como aleatoria? En 1888, Venn verificó que los primeros 707 dígitos de la expansión decimal de π satisfacen criterios importantes de aleatoriedad: cada dígito aparece en la secuencia aproximadamente el mismo número de veces que los demás dígitos, sin ninguna estructura aparente. 101 años después, Gregory y David Chudnovsky verificaron el mismo comportamiento para más de mil millones de dígitos en la expansión de π . ¿Qué más podría uno esperar de una secuencia aleatoria? Pero sabemos, por supuesto, que la expansión decimal de π es completamente determinística! (ver definición 135(d)).

Como la teoría de la probabilidad no se puede construir sobre la subjetividad que da la ignorancia, Kolmogorov mismo, en su esfuerzo por formalizar la teoría, llegó a un concepto muy interesante de aleatoriedad: Si la complejidad de una secuencia está dada por la longitud del programa de computador más pequeño capaz de generarla (en un modelo computacional particular, tal como una máquina de Turing), una secuencia es aleatoria cuando su complejidad es máxima, esto es, cuando el único algoritmo que la puede generar es el algoritmo que la menciona, elemento por elemento. Por supuesto, π resulta poco complejo, pues el algoritmo es sencillo: basta contar cuántos diámetros caben en una circunferencia¹⁹, para lo cual el computador puede evaluar iterativamente una serie de potencias. Pero una secuencia de lanzadas de un dado es aleatoria, porque necesitamos lanzar los dados para poder especificar la secuencia. Esta definición pone el concepto de aleatoriedad en términos formales, muy al estilo de Kolmogorov, pues la aleatoriedad de una secuencia ya no depende de la apreciación del observador sino que es una medida objetivamente cuantificable (más adelante tendremos oportunidad de introducirnos brevemente en la teoría de la complejidad y la teoría de la información que sustentan este concepto).

Pero, igualmente, queda casi sin resolver la pregunta de cómo asignar las probabilidades a los eventos del campo- σ escogido para analizar un modelo probabilístico de una realidad compleja. Fundamentalmente, se trata de evaluar nuestra confianza en que el evento suceda cuando realicemos el experimento, para lo cual podemos valer nos de la frecuencia relativa observada en experimentos anteriores o en razonamientos plausibles sobre la naturaleza del experimento, tales como las simetrías proporcionadas por eventos equiprobables (ver la definición 26). En los capítulos siguientes formularemos diferentes modelos para muchos casos típicos de experimentos que surgen una y otra vez en el mundo de las redes de comunicaciones, los cuales podremos usar para generar hipótesis sobre las probabilidades de algunos eventos y, con nuestras herramientas, encontrar las probabilidades de otros eventos o algunas estadísticas de interés.

¹⁹ Es fácil calcular rápidamente millones de dígitos de π mediante, por ejemplo, el algoritmo de Chudnovsky:

$$\frac{1}{12\pi} = \sum_{n=0}^{\infty} (-1)^n \frac{(6n)! (3591409 + 545140134 \cdot n)}{(3n)!(n!)^3 \cdot 640320^{(3n+3/2)}}$$

Por último, vale la pena mencionar la teoría de las posibilidades como una formalidad reciente para tratar con cantidades inciertas, alternativa (y a veces complementaria) a la teoría de las probabilidades. Sobre los eventos pertenecientes a un campo aditivo \mathcal{F} de subconjuntos de un espacio muestral Ω , se define la posibilidad del evento $A \in \mathcal{F}$ como una función $\mathbf{Po} : \mathcal{F} \rightarrow \mathbb{R}$ que satisface los siguientes postulados: $\mathbf{Po}(\emptyset) = 0$, $\mathbf{Po}(\Omega) = 1$, $\mathbf{Po}(A \cup B) = \max(\mathbf{Po}(A), \mathbf{Po}(B))$ si A y B son conjuntos disyuntos en \mathcal{F} . Este último postulado se puede extender igualmente a campos- σ infinitamente aditivos, como en la definición 14. Esta medida de posibilidad está más asociada con la función de membresía de un elemento en un conjunto difuso, esto es, un conjunto al que se puede pertenecer con cierto grado de pertenencia en el rango $[0, 1]$, en cuyo caso la posibilidad de un evento es el máximo entre las funciones de membresía de sus miembros. Como un ejemplo revelador de la diferencia entre los dos conceptos, considere que lleva muchos días perdido en el desierto y encuentra dos botellas llenas de un líquido de apariencia deliciosa. La etiqueta en una botella dice que la probabilidad de que su contenido sea potable es 0.9, mientras la etiqueta en la otra botella dice que la posibilidad de que sea potable es 0.9. ¿Cuál líquido consumiría usted? Es muy probable que el contenido de la primera botella sea agua pura, aunque, en el peor de los casos, la primera botella podría contener ácido sulfúrico o cianuro. En cambio puede estar seguro que la segunda botella no contiene ni agua pura (porque su grado de membresía en el conjunto de los líquidos potables sería 1) ni ácido sulfúrico (porque su grado de membresía en el conjunto de los líquidos potables sería cero) sino, talvez, gaseosa, en cuyo caso su consumo podría tener algún efecto negativo en su salud, aunque insignificante ante la alternativa de morir de sed.

De cualquier manera, la incertidumbre es un aspecto ineludible en la gran mayoría de situaciones que enfrenta el ser humano. Como mencionamos en la definición 14, hay muchas razones para considerar la incertidumbre en los modelos matemáticos de la realidad que nos interesa. En muchas ocasiones, por ejemplo, no conocemos el verdadero estado del sistema de interés porque tenemos sólo una observación parcial del mismo. Un médico, por ejemplo, puede observar los síntomas de su paciente, pero casi nunca podrá observar su enfermedad y, ciertamente, nunca podrá observar su prognosis, aunque debe tomar decisiones importantes respecto a la enfermedad y su prognosis para recomendar un tratamiento. Por otro lado, las observaciones de aquellos aspectos observables de nuestro sistema suelen ser ruidosas y distorsionadas, por lo que debemos considerar, además, el error de nuestras observaciones. Adicionalmente, el estado del sistema ni siquiera tiene una relación determinística con nuestras observaciones, de manera que, aunque ellas fueran completas y exactas, puede existir aún alguna incertidumbre con respecto al estado real del sistema. Continuando con el ejemplo anterior, un conjunto de síntomas puede asociarse con diferentes enfermedades, aunque la experiencia puede indicar que algunas enfermedades son más probables que otras dado el conjunto de síntomas. La incertidumbre debida a estos aspectos generadores de incertidumbre (dificultad para observar el estado real del mundo, dificultad para medir con precisión los aspectos medibles del mundo y dificultad para modelar con precisión la dependencia entre lo observado y lo inobservable) podría reducirse con los avances en el conocimiento científico y en el desarrollo tecnológico. Sin embargo, aún nisiquiera es claro que el universo sea realmente determinístico cuando se estudia con un nivel suficiente de detalle. En el dominio actual de la mecánica cuántica, ciertamente no lo es; pero tampoco

lo es con respecto a nuestra comprensión actual de muchos otros aspectos del mundo (sin mencionar, por ejemplo, el caos).

Esta presencia inevitable de la incertidumbre sobre el estado real del mundo nos exige considerar todos los posibles estados, aunque nuestras observaciones no nos permitan eliminar muchas de esas posibilidades. Y, en estas condiciones, resulta muy conveniente argumentar no sólo sobre los estados posibles, sino sobre las probabilidades de cada uno de esos posibles estados. De esta manera, las decisiones que tomemos (por ejemplo, decisiones de ingeniería, de tratamiento médico, de planeación económica, etc.) será mucho más significativas y mucho más acertadas. En este contexto, la inutilidad de la intuición en la teoría de las probabilidades a la que se refería DeMoivre se puede matizar con otras tres ideas íntimamente relacionadas. La primera es del gran biólogo Stephen Jay Gould: “el desconocimiento de la teoría de las probabilidades puede ser el mayor impedimento para el conocimiento científico”. La segunda es de Pierre Simon Laplace: “En el fondo, la teoría de las probabilidades no es más que sentido común reducido al cálculo”. La tercera es de Louis Pasteur : “La suerte sólo favorece a las mentes preparadas”.

250 Conceptos - Marco A. Alzate - U. Distrital F.J.C.

III. Conceptos Básicos de Variables Aleatorias

29. Variable Aleatoria

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, una variable aleatoria (va) es una función $X: \Omega \rightarrow \mathbb{R}$ tal que, $\forall x \in \mathbb{R}$, el evento $A(x)$ definido como $\{\omega \in \Omega : X(\omega) \leq x\}$ es un evento medible ($A(x) \in \mathcal{F}$)

La variable aleatoria le asigna a cada elemento del espacio muestral un número real, de manera tal que las imágenes de los eventos en \mathcal{F} resultan ser conjuntos de Borel en \mathbb{R} , como sugiere la Figura 31.

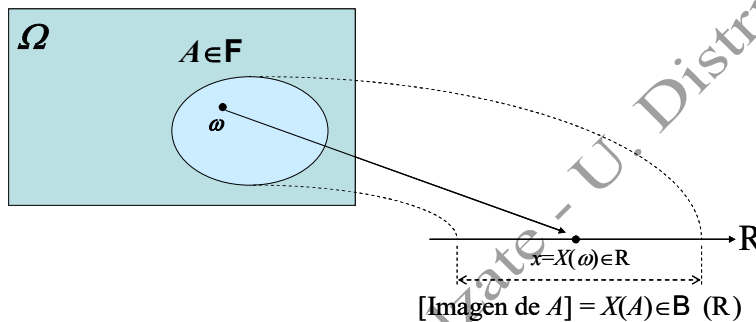


Figura 31. Concepto de Variable Aleatoria

Lo primero que podemos ver es que una variable aleatoria no es una variable sino una función; y no es aleatoria sino determinística: a cada $\omega \in \Omega$ le corresponde uno y sólo un valor real, $X(\omega)$. El nombre (aparentemente inapropiado) de variable aleatoria se debe a razones históricas, pero se convierte en un buen truco mnemotécnico: una variable aleatoria no es una variable sino una función y no es aleatoria sino determinística.

Considere el ejemplo 4 en el que observamos la ocupación de un canal de comunicaciones. El espacio de probabilidad del experimento está dado por el espacio muestral $\Omega = \{libre, ocupado\}$, el campo- σ de eventos $\mathcal{F} = \{0,1\}^{\Omega} = \{\Phi, \{libre\}, \{ocupado\}, \Omega\}$, y la medida de probabilidad $\mathbf{P}: \mathcal{F} \rightarrow \mathbb{R}$ dada por $\mathbf{P}(\Phi) = 0$, $\mathbf{P}(\{libre\}) = 1-p$, $\mathbf{P}(\{ocupado\}) = p$ y $\mathbf{P}(\Omega) = 1$. Sobre este espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ podemos definir una función determinística $X: \Omega \rightarrow \mathbb{R}$ dada por los valores $X(libre) = 0$ y $X(ocupado) = 1$. ¿Es esta función una variable aleatoria? Claramente,

$$A(x) = \{\omega \in \Omega : X(\omega) \leq x\} = \Phi \in \mathcal{F} \quad \forall x < 0$$

$$A(x) = \{\omega \in \Omega : X(\omega) \leq x\} = \{libre\} \in \mathcal{F} \quad \forall 0 \leq x < 1$$

$$A(x) = \{\omega \in \Omega : X(\omega) \leq x\} = \Omega \in \mathcal{F} \quad \forall x \geq 1$$

Como $A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$, X es una variable aleatoria. Es más, podemos graficar la probabilidad de $A(x)$ para cada $x \in \mathbb{R}$, pues $\mathbf{P}(A(x)) = \mathbf{P}(\Phi) = 0$ si $x < 0$, $\mathbf{P}(A(x)) = \mathbf{P}(\{libre\}) = 1-p$ si $0 \leq x < 1$, y $\mathbf{P}(A(x)) = \mathbf{P}(\Omega) = 1$ si $x \geq 1$. La Figura 32 muestra esta probabilidad $\mathbf{P}(A(x))$ como función de x .

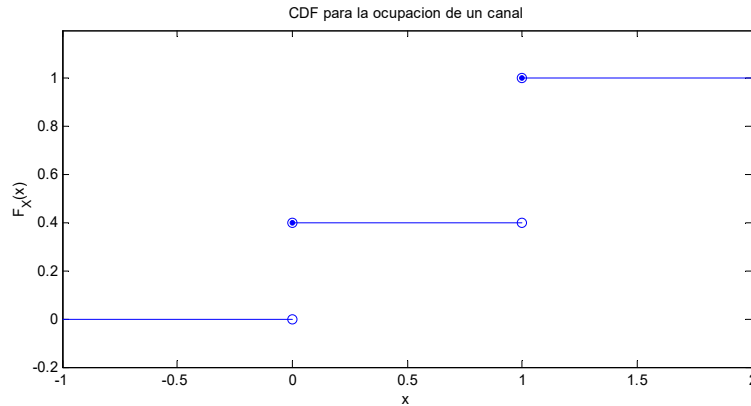


Figura 32. $P(\{\omega \in \Omega : X(\omega) \leq x\})$ en el ejemplo de la ocupación de un canal con $X =$ Número de canales ocupados, cuando $P(\{\text{ocupado}\})=0.6$

Claro, no cualquier función $X: \Omega \rightarrow \mathbb{R}$ es una *va*. Por ejemplo, consideremos el espacio de probabilidad $(\Omega = \{a_1, a_2, a_3, a_4\}, \mathcal{F} = \{\emptyset, \{a_1\}, \{a_2\}, \{a_1, a_2\}, \{a_3, a_4\}, \{a_1, a_3, a_4\}, \{a_2, a_3, a_4\}, \Omega\}, \mathbf{P} = \{0, p, q, p+q, 1-p-q, 1-q, 1-p, 1\})$. Este es un espacio de probabilidad bien definido pues \mathcal{F} es una clase de subconjuntos de Ω cerrado para el complemento y para la unión, y \mathbf{P} es una medida de probabilidad de \mathcal{F} en \mathbb{R} que satisface los tres axiomas. Definamos la función $X(a_i) = i, i=1,2,3,4$. ¿Es X una *va*? ¡Esta vez no! En efecto, para $x < 1, A(x) = \emptyset \in \mathcal{F}$ con $\mathbf{P}(A(x)) = 0$; para $1 \leq x < 2, A(x) = \{a_1\} \in \mathcal{F}$ con $\mathbf{P}(A(x)) = p$; para $2 \leq x < 3, A(x) = \{a_1, a_2\} \in \mathcal{F}$ con $\mathbf{P}(A(x)) = p+q$; para $x \geq 4, A(x) = \Omega \in \mathcal{F}$ con $\mathbf{P}(A(x)) = 1$; pero para $3 \leq x < 4, A(x) = \{a_1, a_2, a_3\} \notin \mathcal{F} \dots$ Y la definición de *va* exige que $A(x)$ debe ser medible para todo $x \in \mathbb{R}$. Pero es fácil verificar que, por ejemplo, $Y(a_i) = \lceil i/2 \rceil$ sí es una variable aleatoria en este espacio de probabilidad ($\lceil z \rceil$ es el techo de z , esto es, el mínimo entero mayor o igual a z). Efectivamente, $Y(a_1) = Y(a_2) = 1, Y(a_3) = Y(a_4) = 2$, de manera que para $y < 1, \mathbf{P}(A(y)) = \mathbf{P}(\emptyset) = 0$; para $1 \leq y < 2, \mathbf{P}(A(y)) = \mathbf{P}(\{a_1, a_2\}) = p+q$; y para $2 \leq y, \mathbf{P}(A(y)) = \mathbf{P}(\Omega) = 1$.

En general, si $|\Omega|$ es finita o contable y $\mathcal{F} = \{0,1\}^\Omega$, cualquier función de Ω en \mathbb{R} es una *va*. Pero si $|\Omega|$ es infinita incontable, podemos convertir cualquier función $X: \Omega \rightarrow \mathbb{R}$ en una *va* si construimos un espacio de probabilidad para ella con el mínimo campo- σ que contiene los eventos $A(x) = \{\omega \in \Omega : X(\omega) \leq x\}, \forall x \in \mathbb{R}$, y asignamos alguna medida de probabilidad a dichos eventos. De esta manera nos aseguramos que la imagen de los eventos en el espacio de probabilidad sean conjuntos de Borel en \mathbb{R} .

Observe que en muchos casos el espacio muestral mismo está contenido en (o es igual a) el conjunto de los números reales, de manera que $X(\omega) = \omega$ es una *va* perfectamente válida, como lo podría ser cualquier otra función de \mathbb{R} en \mathbb{R} . Este es el caso de los ejemplos 2, 5, 6, 7, 10, 11, 12, 13 y 14 de la definición 5. En el caso del ejemplo 4, como acabamos de ver, la función bivaluada $X(\text{libre})=0, X(\text{ocupado})=1$ es una variable aleatoria. En el experimento 9, ver si un bit se recibe con error o no, la *va* dada por $X(\text{si})=1, X(\text{no})=0$, tiene un comportamiento muy parecido al caso del ejemplo 4.

En el caso del ejemplo 8 (verificar el estado de ocupación de cada canal de una trama E1) podríamos definir una va en ese experimento asociando cada uno de los elementos del espacio muestral con el número binario de 32 bits conformado de la siguiente manera: asignamos un cero a cada canal libre y un uno a cada canal ocupado y decimos que cada dígito representa la potencia de dos asociada con la posición del canal en la trama, $X(\omega) = \sum_{i=0}^{31} 2^i 1(\text{el } i\text{-ésimo canal en } \omega \text{ está ocupado})$ ²⁰. Siendo así,

el rango de la va será el conjunto de los números enteros desde 0 hasta $2^{32}-1 = 4.294'967.295$. Entre muchas otras variables aleatorias que podríamos imaginar en este mismo experimento se puede mencionar el ancho de banda libre en la trama, que es un múltiplo de 64 kbps:

$$Y(\omega) = 64000 \sum_{i=0}^{31} 1(\text{el } i\text{-ésimo canal en } \omega \text{ está libre}).$$

30. Función de Distribución de Probabilidad Acumulativa, CDF

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se define una variable aleatoria $X: \Omega \rightarrow \mathbb{R}$. La Función acumulativa de distribución de probabilidad de X es la función $F_X: \mathbb{R} \rightarrow \mathbb{R}$ definida como $F_X(x) = \mathbf{P}(\{\omega \in \Omega: X(\omega) \leq x\})$, $\forall x \in \mathbb{R}$. Le diremos la CDF por la sigla en inglés de Cumulative Distribution Function.

Obsérvese de dónde surge la importancia de que los eventos $A(x) = \{\omega \in \Omega: X(\omega) \leq x\}$ sean medibles en el espacio de probabilidad en que se define la función (va) X : Si no fuera así no se podría definir la CDF de X (al menos no con dominio en todos los reales).

No hay manera de destacar suficientemente la importancia de esta función en lo que resta de nuestro estudio en este libro. Para empezar, obsérvese que, mientras $\mathbf{P}(\cdot)$ es una medida de conjuntos (una función de \mathcal{F} en \mathbb{R}) y $X(\cdot)$ es una función de Ω en \mathbb{R} , $F_X(\cdot)$ es, por primera vez en este libro, un función de los reales en los reales. Con teoría de conjuntos fueron pocas las herramientas que pudimos guardar en nuestra caja de herramientas: tres axiomas, algunas propiedades derivadas de ellos, probabilidad total y regla de Bayes. Pero ahora, con funciones de \mathbb{R} en \mathbb{R} , podemos echar mano del análisis real para atiborrar nuestra caja de herramientas.

En el ejemplo 14, el tiempo entre llegadas puede tomar cualquier valor no negativo y menor que infinito, de manera que una posible CDF (de hecho, una CDF típica en estos casos) es la que se muestra en la Figura 33, donde x está medida en ms.

²⁰ Recordemos que $1(s)$ es la función indicadora de la sentencia s , igual a 1 si la sentencia s es cierta e igual a 0 si la sentencia s es falsa, como se dijo en la definición 8.

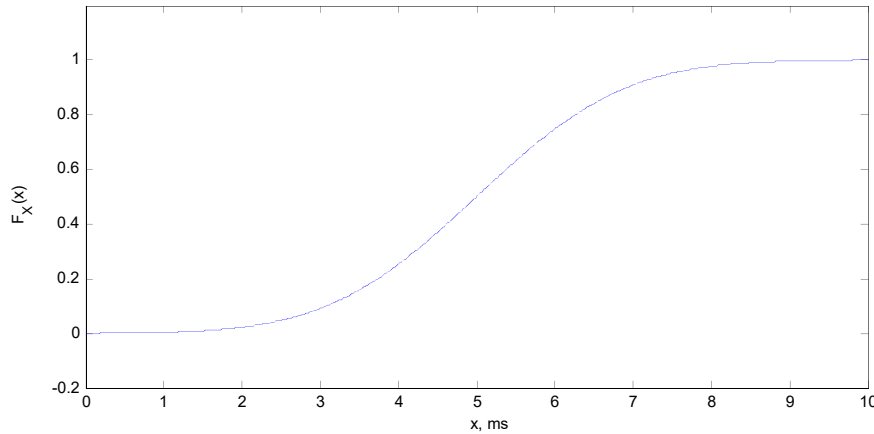


Figura 33. CDF para el tiempo entre llegadas de paquetes consecutivos

31. Propiedades de la CDF

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se define una variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ con CDF $F_X(\cdot)$. Entonces,

- (a) La CDF es no-negativa: $F_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
- (b) La CDF es no-decreciente: si $x_1 < x_2$ entonces $F_X(x_1) \leq F_X(x_2)$
- (c) La CDF es acotada: $F_X(-\infty) = 0, F_X(\infty) = 1$.
- (d) La CDF es continua por la derecha: $F_X(x^+) = F_X(x)$.

En efecto, como la CDF $F_X(x)$ es una medida de probabilidad de un evento indicado por el número real x , las propiedades anteriores son las formas que toman algunas propiedades de la medida de probabilidad. Para cada número real x definamos $A(x)$ como el evento $\{\omega \in \Omega : X(\omega) \leq x\}$, medible en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$.

- (a) El segundo axioma de la definición 14 exige que $\mathbf{P}(A(x)) \geq 0$, de donde surge la no-negatividad de $F_X(x)$.
- (b) Si $x_1 < x_2$, $A(x_2) = A(x_1) \cup \{\omega \in \Omega : x_1 < X(\omega) \leq x_2\}$, de manera que $A(x_1) \subseteq A(x_2)$ y, de acuerdo con el quinto resultado de la definición 17, $\mathbf{P}(A(x_1)) \leq \mathbf{P}(A(x_2))$, por lo que $F_X(x)$ debe ser no-decreciente.
- (c) Como $A(-\infty) = \emptyset$, el segundo resultado de la definición 17 exige que $F_X(-\infty) = 0$. Y como $A(\infty) = \Omega$, el primer axioma de las probabilidades exige que $F_X(\infty) = 1$. Dado que $F_X(x)$ es no decreciente, estos resultados implican que $F_X(x)$ está acotada en el rango $[0, 1]$.
- (d) Para cualquier $n \in \mathbb{N}$, $A(x+1/n) = A(x) \cup \{\omega \in \Omega : x < X(\omega) \leq x+1/n\} = A(x) \cup B_n(x)$, donde definimos $B_n(x)$ como $\{\omega \in \Omega : x < X(\omega) \leq x+1/n\}$. Como $A(x)$ y $B_n(x)$ son mutuamente excluyentes, aplica el tercer axioma de la definición 14, $\mathbf{P}(A(x+1/n)) = \mathbf{P}(A(x)) + \mathbf{P}(B_n(x))$, de donde $\mathbf{P}(B_n(x)) = F_X(x+1/n) - F_X(x)$. A medida que n tiende a infinito, la cota superior en el intervalo que define a $B_n(x)$ tiende a x , pero x está por fuera del intervalo por la cota inferior, que es abierta en x , por lo

que $B_n(x)$ tiende a Φ . Formalmente, $\lim_{n \rightarrow \infty} B_n(x) = \bigcap_{k=1}^{\infty} \left\{ \omega \in \Omega : x < X(\omega) \leq x + \frac{1}{k} \right\} = \Phi$, de manera que $\lim_{n \rightarrow \infty} \mathbf{P}(B_n(x)) = \mathbf{P}(\Phi) = 0$. En consecuencia, $\lim_{n \rightarrow \infty} F_X(x + \frac{1}{n}) = F_X(x)$, que es la definición de continuidad por la derecha, $F_X(x^+) = F_X(x)$.

Obsérvese que las funciones de la Figura 32 y la Figura 33 satisfacen las cuatro propiedades anteriores. Veamos otros dos ejemplos sencillos en los que definimos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ y una función determinística $X: \Omega \rightarrow \mathbb{R}$, de manera que podemos deducir su CDF, $F_X(x)$, a partir de la medida de probabilidad \mathbf{P} , y verificar las propiedades anteriores.

Primero consideremos el estado de ocupación de dos enlaces, donde cada uno está ocupado con probabilidad p independientemente del otro. El espacio muestral de este experimento aleatorio es

$$\Omega = \{(libre, libre), (libre, ocupado), (ocupado, libre), (ocupado, ocupado)\}$$

Y la probabilidad de cada subconjunto unitario se puede calcular fácilmente por la suposición de independencia:

$$\mathbf{P}(\{(l, l)\}) = (1-p)^2, \quad \mathbf{P}(\{(l, o)\}) = \mathbf{P}(\{(o, l)\}) = p(1-p), \quad \mathbf{P}(\{(o, o)\}) = p^2$$

Con esta asignación de probabilidad podemos considerar todos los subconjuntos de Ω en nuestro campo- σ de eventos y calcular fácilmente sus probabilidades sumando las probabilidades de sus respectivos subconjuntos unitarios, pues son mutuamente excluyentes:

(l, l)	(l, o)	(o, l)	(o, o)	$A \subseteq \Omega$	$\mathbf{P}(A)$
0	0	0	0	Φ	0
0	0	0	1	$\{(o, o)\}$	p^2
0	0	1	0	$\{(o, l)\}$	$p(1-p)$
0	0	1	1	$\{(o, l), (o, o)\}$	$p(1-p) + p^2 = p$
0	1	0	0	$\{(l, o)\}$	$p(1-p)$
0	1	0	1	$\{(l, o), (o, o)\}$	$p(1-p) + p^2 = p$
0	1	1	0	$\{(l, o), (o, l)\}$	$2p(1-p)$
0	1	1	1	$\{(l, o), (o, l), (o, o)\}$	$2p(1-p) + p^2 = 1 - (1-p)^2$
1	0	0	0	$\{(l, l)\}$	$(1-p)^2$
1	0	0	1	$\{(l, l), (o, o)\}$	$p^2 + (1-p)^2 = 1 - 2p(1-p)$
1	0	1	0	$\{(l, l), (o, l)\}$	$p(1-p) + (1-p)^2 = 1-p$
1	0	1	1	$\{(l, l), (o, l), (o, o)\}$	$p(1-p) + p^2 + (1-p)^2 = 1 - p(1-p)$

1	1	0	0	$\{(l,l),(l,o)\}$	$p(1-p)+(1-p)^2 = 1-p$
1	1	0	1	$\{(l,l),(l,o),(o,o)\}$	$p(1-p)+p^2+(1-p)^2 = 1-p(1-p)$
1	1	1	0	$\{(l,l),(l,o),(o,l)\}$	$2p(1-p)+(1-p)^2 = 1-p^2$
1	1	1	1	Ω	1

Teniendo completamente especificado nuestro espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, podemos definir la función determinística $X: \Omega \rightarrow \mathbb{R}$ que asigna a cada elemento del espacio muestral el número de canales ocupados:

$$X((l,l)) = 0, \quad X((l,o)) = X((o,l)) = 1, \quad X((o,o)) = 2$$

Entonces podemos definir los eventos $A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \quad \forall x \in \mathbb{R}$ y sus respectivas probabilidades, las cuales corresponden a la CDF, $F_X(\cdot)$

$$A(x) = \begin{cases} \Phi & x < 0 \\ \{(l,l)\} & 0 \leq x < 1 \\ \{(l,l),(l,o),(o,l)\} & 1 \leq x < 2 \\ \Omega & 2 \leq x \end{cases} \Rightarrow F_X(x) = P(A(x)) = \begin{cases} 0 & x < 0 \\ (1-p)^2 & 0 \leq x < 1 \\ 1-p^2 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

Como muestra la Figura 34, esta función satisface las cuatro propiedades de una CDF: no-negativa, no-decreciente, acotada y continua por la derecha.

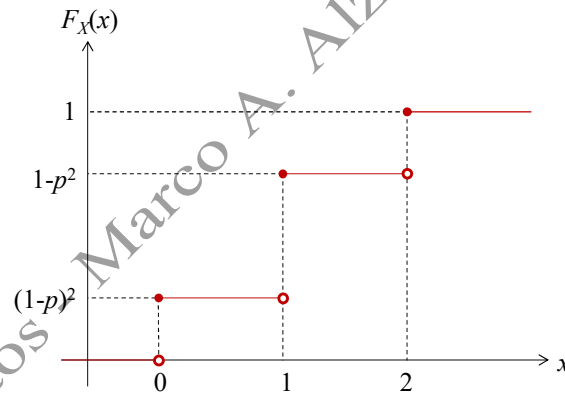


Figura 34. CDF del número de enlaces ocupados entre dos enlaces cuando cada uno se encuentra ocupado con probabilidad p , independientemente del otro

En el segundo ejemplo escogeremos un punto al azar del cuadrado unitario,

$$\Omega = \{(a,b) \in \mathbb{R}^2 : a \in [0,1], b \in [0,1]\}$$

Por supuesto, en este experimento no podemos escoger el conjunto potencia de Ω como dominio de la probabilidad, pues sería un dominio demasiado grande. Por eso escogemos como campo- σ de eventos el campo de Borel de \mathbb{R}^2 , restringido al cuadrado unitario (ver la definición 55 y la nota de pie de página de la definición 21). En ese espacio muestral y con ese campo de eventos, una medida de probabilidad adecuada para cada subconjunto medible es su área, como muestra la Figura 35.

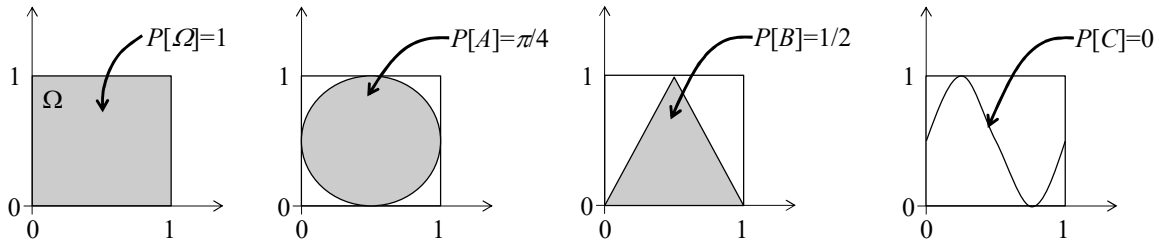


Figura 35. Probabilidad de algunos conjuntos de Borel en el cuadrado unitario de \mathbb{R}^2

Ahora que tenemos la especificación completa del espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, podemos definir una función determinística $X: \Omega \rightarrow \mathbb{R}$. En este caso, escogemos la siguiente función: $X(a,b) = \min(a,b)$. ¿Cuáles son los conjuntos $A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \forall x \in \mathbb{R}$, y cuáles son sus probabilidades? La Figura 36(a) muestra estos conjuntos: todos los puntos (a,b) del cuadrado unitario tales que $\min(a,b) \leq x$. Su probabilidad se puede expresar como la suma de las áreas de dos subconjuntos mutuamente excluyentes: Uno de área x y otro de área $x(1-x)$ para un área total de $x(2-x)$. Evidentemente, si $x < 0$, $A(x) = \emptyset$ y, si $x > 1$, $A(x) = \Omega$. Por consiguiente, la función de distribución acumulativa de esta variable aleatoria es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x(2-x) & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Como se muestra en la Figura 36(b).

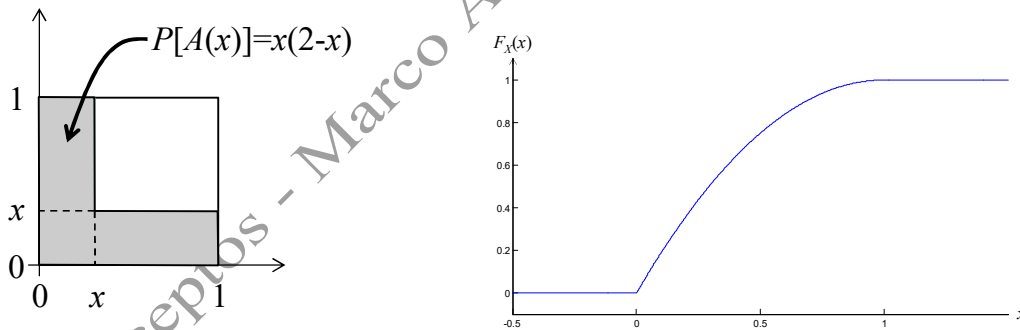


Figura 36. $A(x) = \{(a,b) \in \Omega : \min(a,b) \leq x\}$, $F_X(x) = P[A(x)] = x(2-x)$

Nuevamente, esta función satisface las cuatro propiedades de una CDF: es no-negativa, no-decreciente, acotada y continua por la derecha.

Una de las razones por las que la CDF es un concepto tan fundamentalmente importante es que cualquier función de \mathbb{R} en \mathbb{R} que satisfaga las anteriores cuatro propiedades es una CDF válida en el espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(\cdot))$. Esto es, no es estrictamente necesario considerar un espacio de probabilidad sobre el cual podamos definir una ν para la cual construiríamos la respectiva CDF de acuerdo con las probabilidades de los eventos medibles en el espacio original, como hicimos en los dos ejemplos anteriores. Podemos tomar el camino inverso: Considerar una CDF y definir con

ella una *va* apropiada en el espacio $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(\cdot))$. Siendo así, para especificar completamente una variable aleatoria basta con describir su *CDF*: Decir qué valores toma y cómo se distribuye la probabilidad sobre esos valores. ¡No hace falta definir ningún otro espacio de probabilidad subyacente!

Por ejemplo, considere la función $g(x) = (1 - e^{-\lambda x})u(x)$, donde $u(x)$ es el escalón unitario que vale 1 si $x \geq 0$ y vale 0 en otro caso, y λ es un número real positivo. Esta es una función no-negativa, no-decreciente, acotada y continua. Por lo tanto, podemos suponer la existencia de una *va* X que toma valores en los reales no negativos y asignarle la *CDF* $F_X(x) = g(x)$, con lo que construiríamos un espacio de probabilidad formalmente definido. Si establecemos la hipótesis de que dicho espacio modela el tiempo de vida útil de los componentes de una red, por ejemplo, podríamos construir y evaluar así un modelo probabilístico de confiabilidad. A las variables aleatorias con $F_X(x) = 1 - e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$, se les conoce como variables aleatorias exponenciales, como se describe en la definición 44.

Como de ahora en adelante vamos a trabajar casi exclusivamente con variables aleatorias, vamos a despreocuparnos desde ahora por la definición explícita de un espacio de probabilidad, pues tácitamente dejaremos que dicho espacio sea $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(\cdot))$. Tanto es así, que de ahora en adelante nos tomaremos muchas libertades en la notación. Por ejemplo, en vez de hablar de $\mathbf{P}(\{\omega \in \Omega : X(\omega) \in B\})$, donde B es un conjunto de Borel, diremos solamente $\mathbf{P}(X \in B)$. Por supuesto, formalmente este es un error gramatical que podría confundirse con un garrafal error conceptual porque las probabilidades no se asignan a sentencias lógicas sino a subconjuntos medibles de Ω . Pero como ya no necesitamos hacer referencia a un espacio muestral subyacente, es simplemente nuestra convención para referirnos a la probabilidad del evento medible B en el espacio muestral \mathbb{R}^{21} . Es importante insistir en lo que decíamos en el primer capítulo: Si no tenemos perfectamente definido un espacio de probabilidad para nuestro modelo, no sabremos dónde estamos parados. Sólo estamos diciendo que, mientras nuestro modelo probabilístico se base en una variable aleatoria, el correspondiente espacio de probabilidad puede dejarse implícitamente definido. Por esta razón, otra libertad en la notación será la de cambiar la frase “Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se define una *va* $X: \Omega \rightarrow \mathbb{R}$ con *CDF* $F_X(x)$ ” por la frase “Sea $F_X(x)$ la *CDF* de alguna *va* X ” (a menos, claro, que necesitamos referirnos explícitamente al espacio de probabilidad subyacente).

32. Probabilidad de Algunos Subconjuntos de \mathbb{R}

Sea $F_X(\cdot)$ la *CDF* de alguna *va* X . Por simplicidad, denotemos $\mathbf{P}(X \in B)$ como $\mathbf{P}(B)$ para cualquier $B \in \mathcal{B}(\mathbb{R})$. Entonces

- (a) $\mathbf{P}((-\infty, a]) = F_X(a) \quad \forall a \in \mathbb{R}$
- (b) $\mathbf{P}((a, \infty)) = 1 - F_X(a) \quad \forall a \in \mathbb{R}$
- (c) $\mathbf{P}((a, b]) = F_X(b) - F_X(a) \quad \forall a, b \in \mathbb{R}, a < b$
- (d) $\mathbf{P}([a]) = F_X(a^+) - F_X(a^-) \quad \forall a \in \mathbb{R}$

²¹ Ahora un evento medible es, simplemente, un conjunto de Borel en los reales, $B \in \mathcal{B}(\mathbb{R})$.

- (e) $\mathbf{P}((-\infty, a)) = F_X(a) - \mathbf{P}([a]) = F_X(a^-) \forall a \in \mathbb{R}$
 (f) $\mathbf{P}([a, \infty)) = 1 - F_X(a) + \mathbf{P}([a]) = 1 - F_X(a^-) \forall a \in \mathbb{R}$
 (g) $\mathbf{P}((a, b)) = F_X(b) - F_X(a) - \mathbf{P}([b]) = F_X(b^-) - F_X(a) \forall a, b \in \mathbb{R}, a < b$
 (h) $\mathbf{P}([a, b]) = F_X(b) - F_X(a) + \mathbf{P}([a]) = F_X(b) - F_X(a^-) \forall a, b \in \mathbb{R}, a < b$
 (i) $\mathbf{P}([a, b)) = (F_X(b) - \mathbf{P}([b]) - (F_X(a) - \mathbf{P}([a]))) = F_X(b^-) - F_X(a^-) \forall a, b \in \mathbb{R}, a < b$

Como de costumbre, estas propiedades surgen de los tres axiomas de la probabilidad, como mostraremos a continuación:

- (a) Esta es la definición 30
 (b) Este es el resultado 1 de la definición 17, aplicado a (a)
 (c) $(-\infty, b] = (a, b] \cup (-\infty, a]$ son dos eventos disyuntos, por lo que aplica el tercer axioma: $F_X(b) = \mathbf{P}((a, b]) + F_X(a)$. Restando $F_X(a)$ a ambos lados se obtiene el resultado.
 (d) De acuerdo con el resultado anterior, $\mathbf{P}((a-1/n, a]) = F_X(a) - F_X(a - 1/n)$ para todo entero n mayor o igual a 1. En el límite cuando n tiende a infinito, el evento $(a-1/n, a]$ tiende a $\lim_{n \rightarrow \infty} (a - 1/n, a] = \bigcap_{j=1}^{\infty} (a - 1/j, a] = [a]$, mientras que $F_X(a-1/n)$ tiende a $F_X(a^-)$. Por la propiedad (d) de la definición 31, $F_X(a) = F_X(a^+)$. Poniendo los tres resultados juntos obtenemos $\mathbf{P}([a]) = F_X(a^+) - F_X(a^-) \forall a \in \mathbb{R}$.
 (e) $(-\infty, a) = (-\infty, a) \cup [a]$ son dos subconjuntos mutuamente excluyentes, por lo que aplica el segundo axioma de la definición 14: $F_X(a) = \mathbf{P}([a]) + \mathbf{P}((-\infty, a))$. Restando $\mathbf{P}([a])$ a ambos lados obtenemos el resultado.
 (f) $[a, \infty) = (-\infty, a)^C$, donde el superíndice C se refiere al complemento respecto al conjunto de los reales. Aplicando el primer resultado de la definición 17 al resultado (e) anterior, $\mathbf{P}([a, \infty)) = 1 - F_X(a) + \mathbf{P}([a]) \forall a \in \mathbb{R}$.
 (g) Como $(a, b) = (a, b) \cup [b]$, $F_X(b) - F_X(a) = \mathbf{P}((a, b)) + \mathbf{P}([b])$. Restando $\mathbf{P}([b])$ a ambos lados se obtiene el resultado.
 (h) Como $[a, b] = (a, b) \cup [a]$, $\mathbf{P}([a, b]) = F_X(b) - F_X(a) + \mathbf{P}([a])$.
 (i) Como $[a, b) = [a, b) \cup [b]$, podemos aplicar el tercer axioma y el resultado (h) anterior, $\mathbf{P}([a, b)) = F_X(b) - F_X(a) + \mathbf{P}([a]) = \mathbf{P}([a, b]) + \mathbf{P}([b])$. Restando $\mathbf{P}([b])$ obtenemos $\mathbf{P}([a, b)) = (F_X(b) - \mathbf{P}([b]) - (F_X(a) - \mathbf{P}([a])))$

Cada uno de estos resultados tiene interpretaciones importantes. En particular, consideremos el punto (d): Si un punto individual x_0 de \mathbb{R} tiene una probabilidad diferente de cero, la *CDF* de la correspondiente *va* debe tener una discontinuidad en ese punto, de manera que $F_X(x_0) = F_X(x_0^-) + \mathbf{P}([x_0])$. Por otro lado, si la *CDF* de una variable aleatoria X es una función continua, la probabilidad de cada punto individual es cero, pues la continuidad significa que $F_X(x^+) = F_X(x^-)$ para todo x . Este efecto lo podemos ver con claridad en la Figura 37, donde graficamos una *CDF* con dos puntos de discontinuidad y seleccionamos cuatro subconjuntos en el eje horizontal para los cuales graficamos sus respectivas probabilidades en el eje vertical.

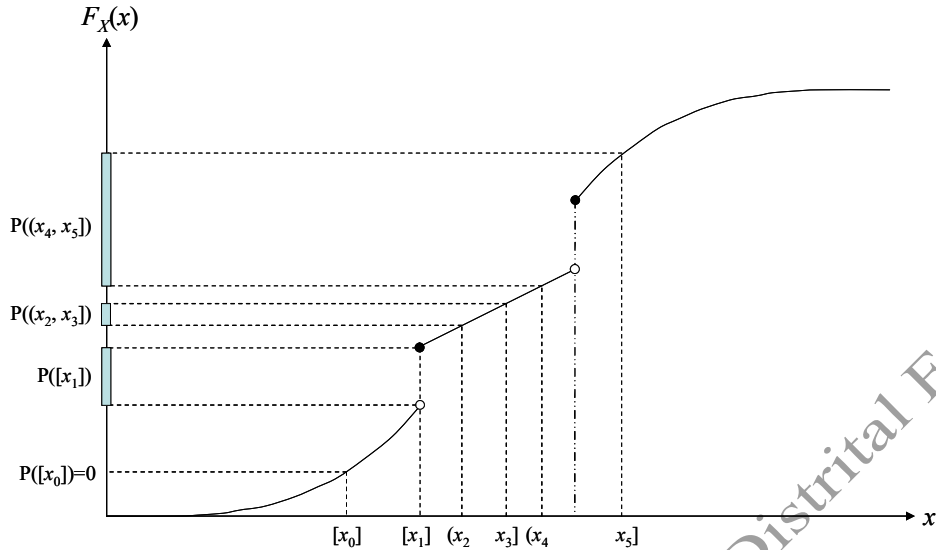


Figura 37. Probabilidad de algunos intervalos

Claramente, x_0 es un punto en el que $F_X(\cdot)$ es continua y, por lo tanto, como subconjunto unitario de \mathbb{R} , tiene una probabilidad igual a cero. A diferencia de x_0 , x_1 es un punto de discontinuidad, donde la discontinuidad corresponde a un salto de longitud $\mathbf{P}([x_1])$; el evento unitario $[x_1]$ puede suceder con probabilidad mayor que cero. Obsérvese que x_0 también puede suceder, a pesar de que su probabilidad es cero! De hecho, todos los valores en el rango mostrado en la figura pueden suceder, aunque sólo dos de ellos con probabilidad diferente de cero. De muchas maneras, nuestra vida está construida a partir de eventos que, aunque tenían probabilidad cero, ocurrieron para hacer de nosotros lo que somos hoy: casi todo lo que ocurre a nuestro alrededor ocurre a pesar de tener probabilidad cero. Se diría que cada uno de nosotros es un milagro! Este es un aspecto importante por considerar con las regiones en que la *CDF* de una *va* es continua. Consideremos, por ejemplo, el intervalo $(x_2, x_3]$: Cada punto individual de ese intervalo tiene probabilidad cero, aunque la probabilidad de que la *va* tome algún valor dentro de ese intervalo es $\mathbf{P}((x_2, x_3]) = F_X(x_3) - F_X(x_2) > 0$. De acuerdo con la figura, esta probabilidad es pequeña comparada con la probabilidad de que la *va* tome un valor en el intervalo $(x_4, x_5]$, el cual contiene un punto de discontinuidad (llamémosle x_a), de manera que $\mathbf{P}((x_4, x_5]) = \mathbf{P}((x_4, x_a)) + \mathbf{P}((x_a, x_5]) + \mathbf{P}([x_a])$. Todos los demás puntos del intervalo tienen probabilidad cero aunque, en conjunto, tienen una probabilidad mayor que la de $[x_a]$ (ver definición 27).

Como acabamos de ver, en general, una *CDF* puede tener puntos de discontinuidad, regiones monótonamente crecientes y regiones donde toma un valor constante. Sin embargo, a veces resulta conveniente describir estas *CDF* generales como la combinación convexa de dos *CDF*s, una continua en todo el rango \mathbb{R} y otra que es constante en intervalos delimitados por un número contable de discontinuidades. Por ejemplo, si $F_1(x)$ toma una forma semejante a la de la Figura 33 y $F_2(x)$ toma una forma semejante a la de la Figura 32, la combinación convexa $F_X(x) = \alpha F_1(x) + (1-\alpha)F_2(x)$, $0 \leq \alpha \leq 1$, tomaría una forma semejante a la de la Figura 37. El hecho de que pueda haber puntos de discontinuidad con probabilidad mayor que cero y puntos de continuidad con probabilidad igual a cero motiva la definición de variables aleatorias continuas y discretas aunque, estrictamente, no es

una definición necesaria. Se dice que X es una variable aleatoria continua si $F_X(x)$ es una función continua para todo $x \in \mathbb{R}$. Se dice que X es una variable aleatoria discreta si la imagen de Ω es un subconjunto contable de \mathbb{R} , en cuyo caso la CDF toma la forma de una suma acumulada de escalones, $F_X(x) = \sum_k p_k u(x - x_k)$, donde $u(x)$ es el escalón unitario que vale 0 si $x < 0$ y vale 1 si $x \geq 0$, de manera que $p_k = F_X(x^+) - F_X(x^-) = P[X=x_k]$. En otro caso, se dice que X es una variable aleatoria híbrida o mixta (Figura 38).

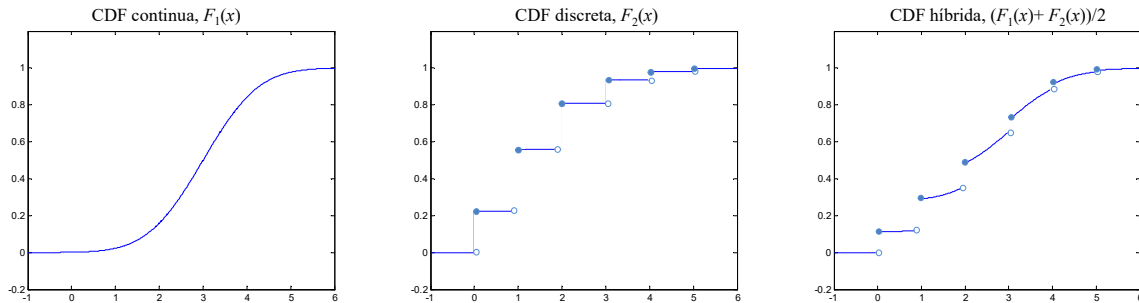


Figura 38. Funciones de distribución acumulativa continua, discreta e híbrida

Volvamos a los primeros seis ejemplos de la definición 6:

1. Lanzar una moneda y ver qué lado queda hacia arriba. El espacio de probabilidad de este experimento es $(\Omega = \{cara, sello\}, \mathcal{F} = \{0,1\}^\Omega, \mathbf{P}(\{cara\})=\mathbf{P}(\{sello\})=0.5)$, de donde podemos definir la variable aleatoria discreta X dada por $X(cara)=0$ y $X(sello)=1$, cuya CDF se grafica en la Figura 39(a).
2. Lanzar un dado y contar los puntos en la cara que queda hacia arriba: $\Omega = \{1,2,3,4,5,6\}$. En este caso $X(\omega)=\omega$ es una variable aleatoria discreta en el que cada posible valor ocurre con probabilidad $1/6$, como se muestra en la Figura 39(b).
3. Escoger una carta de la baraja de naipes: El espacio muestral de este experimento es $\Omega = \{(f,n) : f \in \{picas, tréboles, corazones, diamantes\}, n \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}\}$. Sea Y una va que asigna a cada palo un número entero así: $Y(picas) = 0, Y(tréboles)=1, Y(corazones)=2, Y(diamantes)=3$. Esta es una variable aleatoria discreta en la que cada posible valor ocurre con probabilidad $1/4$. Sea Z otra va que asigna a cada figura un número entero así: $Z(n) = n-1$ si $n \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}, Z(J)=10, Z(Q)=11$ y $Z(K)=12$. Esta es otra variable aleatoria discreta en la que cada posible valor ocurre con probabilidad $1/13$. La variable aleatoria $X(f,n) = 13Y(f) + Z(n)$ toma valores en el rango de números enteros $[0, 51]$, donde 0 le corresponde al as de picas y 51 le corresponde al rey de diamantes. Cada posible valor en el rango de X ocurre con probabilidad $1/52$, de manera que X es una variable aleatoria discreta cuya CDF es como se muestra en la Figura 39(c).
4. Medir la fracción de paquetes perdidos durante una hora en una red IP: $\Omega = \{\omega \in \mathbb{R} : 0 \leq \omega \leq 1\}$. Nuevamente, $X(\omega)=\omega$ es una va adecuada para la cual quisiéramos que el valor $X=0$ ocurriera con una probabilidad significativa. Por consiguiente se trata de una va mixta cuya CDF tiene un punto de discontinuidad en el origen, como muestra la Figura 39(d). La forma particular de esta CDF

- puede ser diferente, dependiendo de las condiciones particulares de la red. En el caso que se muestra, se trata de la fracción de pérdidas en un simple enrutador que conecta una pequeña red local con Internet. La probabilidad de que no hayan pérdidas durante una hora es 0.8 y la probabilidad de que se pierdan menos de un cuarto de los paquetes es, para efectos prácticos, uno.
5. Medir el retardo experimentado por un paquete de datos mientras transita por una red IP. Como el espacio es $\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$, una *va* perfectamente válida es $X(\omega) = \omega$. El quinto capítulo desarrollaremos varios modelos probabilísticos para este experimento, uno de los cuales conduce a la *CDF* mostrada en la Figura 39(e). Se trata de una variable continua cuya distribución muestra que el 50% de los paquetes tardan menos de 100 ms y el otro 50% tarda entre 100 y 200 ms.
 6. Verificar el estado de ocupación de un canal de comunicaciones: $\Omega = \{\text{libre}, \text{ocupado}\}$. Aquí, la variable definida en el ejemplo 1 resulta válida. La Figura 39(f) muestra la *CDF* cuando la probabilidad de ocupación es 0.8.

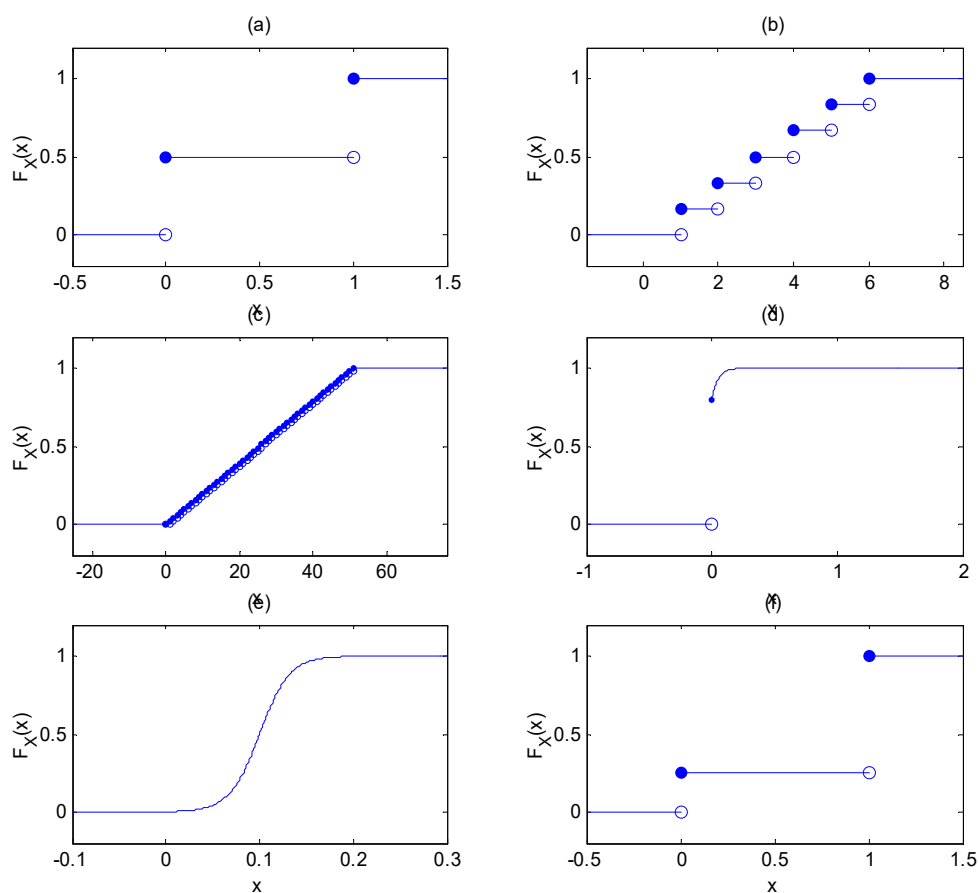


Figura 39. Función de Distribución Acumulativa (CDF) de las variables aleatorias mencionadas en los primeros seis ejemplos de la definición 6

33. Función de Densidad de Probabilidad, *pdf*, y Función de Distribución de Probabilidad, *pmf*

Sea $F_X(\cdot)$ la CDF de alguna va X . La **función de densidad de probabilidad** de X (*pdf* por "probability density function"), $f_X(x)$, se define como la derivada de $F_X(x)$, esto es,

$$f_X(x) = \frac{d}{dx} F_X(x), \forall x \in \mathbb{R}$$

Si X toma sólo un conjunto contable de posibles valores $\{x_1, x_2, x_3, \dots\}$, la CDF toma la forma de una suma acumulada de escalones, $F_X(x) = \sum_k p_k u(x - x_k)$, donde $u(x)$ es el escalón unitario que vale 0 si $x < 0$ y vale 1 si $x \geq 0$ y $p_k = P[X=x_k]$ es la **función de distribución de probabilidad** de X , (*pmf* por "probability mass function"). En este caso la *pdf* toma la forma $f_X(x) = \sum_{k=1}^{\infty} p_k \delta(x - x_k)$, donde $p_k = P(X=x_k)$, y $\delta(x)$ es el impulso de Dirac

Después de haber pasado por la situación extraña para un ingeniero de tener que manipular una función que va de una clase de subconjuntos a los reales, $P:\mathcal{F} \rightarrow \mathbb{R}$, logramos una forma menos rara de función que va del conjunto muestral a los reales, $X:\Omega \rightarrow \mathbb{R}$. Estas funciones eran raras porque a X habría que evaluarla sobre el color de una bola de billar o sobre la figura de una carta de naipes, mientras que a P habría que evaluarla sobre subconjuntos de objetos como esos. Con la CDF tenemos por primera vez en este libro una función que va de \mathbb{R} en \mathbb{R} y, por supuesto, no podía pasar mucho tiempo antes de que un ingeniero sintiera la necesidad de derivarla (para tranquilidad de nuestros lectores, pronto hallaremos oportunidad hasta de aplicar transformadas de Fourier ☺).

Por ejemplo, la *pdf* de una variable aleatoria exponencial, como la definida en el ejemplo de la definición 31, es $f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} (1 - e^{-\lambda x}) = \lambda e^{-\lambda x}, x \geq 0$.

Recordemos que para los puntos $x \in \mathbb{R}$ en los que la CDF $F_X(x)$ es continua, la probabilidad $P(X=x)$ era cero. Sin embargo, sabemos por la definición 32(c) que la probabilidad de que X caiga en un pequeño intervalo $(x, x+\Delta x]$ es $P(x < X \leq x+\Delta x) = F_X(x+\Delta x) - F_X(x)$. Así pues, la *pdf* se puede considerar como el límite de $P(x < X \leq x+\Delta x) / \Delta x$ cuando Δx tiende a cero, lo cual justifica su nombre como *densidad* de probabilidad:

$$f_X(x)\Delta x \approx P(x < X \leq x+\Delta x)$$

Esto es, si bien la va X toma el valor x con probabilidad cero, $f_X(x)\Delta x$ nos dice cuál es la probabilidad de un intervalo muy pequeño cercano a x , que resulta un valor proporcional a la longitud del intervalo (si Δx es suficientemente pequeño), con $f_X(x)$ como factor de proporcionalidad, según muestra la Figura 40. Por esta razón, si $F_X(x)$ es continua en $x=x_0$, la probabilidad de que X tome el valor x_0 es cero, aunque x_0 es un valor posible si $f_X(x)$ es mayor a cero en $x=x_0$; sin embargo, si $F_X(x)$ es constante en un intervalo alrededor de x_0 , de manera que $f_X(x)$ es cero en $x=x_0$, entonces x_0 no sólo es improbable sino que también es imposible.

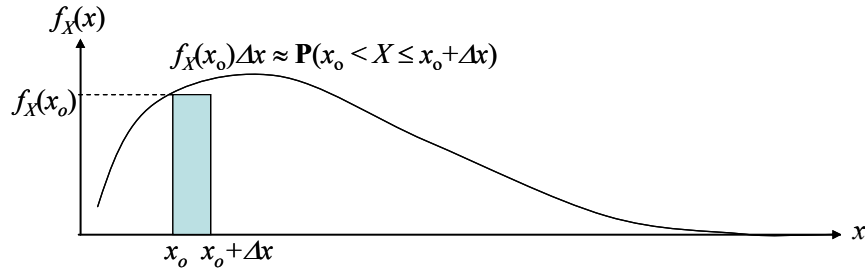


Figura 40. Interpretación de la *pdf*

Es de anotar que la interpretación anterior también tiene una aplicación práctica, pues sugiere una técnica de estimación de la *pdf* de una variable aleatoria mediante el cálculo de la frecuencia relativa de intervalos pequeños en una larga secuencia de muestras de la variable aleatoria. Al dividir cada término de la frecuencia relativa por la longitud del intervalo correspondiente, tendremos un estimado de la *pdf*.

La definición de la *pdf* como la derivada de la *CDF* puede ser muy general en cuanto puede aplicarse a cualquier tipo de variable aleatoria si aceptamos que la *pdf* puede tener discontinuidades (cuando la *CDF* es continua pero no derivable) y singularidades (cuando la *CDF* $F_X(x)$ tiene discontinuidades). En particular, si X es una *va* discreta (que toma sus valores en el conjunto contable $\{x_1, x_2, x_3, \dots\}$), su derivada será cero en todo punto excepto en los de discontinuidad, en los cuales la derivada se hace singular. En consecuencia, la *pdf* de una variable aleatoria discreta es un tren de impulsos de Dirac²²,

$$f_X(x) = \sum_{k=1}^{\infty} \mathbf{P}(X = x_k) \delta(x - x_k)$$

en el que el área debajo de cada impulso corresponde a la respectiva función de distribución de probabilidad, o *pmf*, $p_k = \mathbf{P}(X = x_k) = F_X(x_k^+) - F_X(x_k^-)$. En este libro hablaremos en general de la *pdf* y, sólo cuando sea estrictamente necesario o conveniente, particularizaremos para la *pmf*. Por ejemplo, cuando verificábamos el estado de ocupación de un canal de comunicaciones y definíamos $X(\text{libre})=0$ y $X(\text{ocupado})=1$, obteníamos la *CDF* mostrada en la Figura 39(f) si la probabilidad del evento $\{\text{ocupado}\}$ fuese 0.8. Evidentemente, se trata de una *va* discreta con la *pmf* mostrada en la Figura 41.

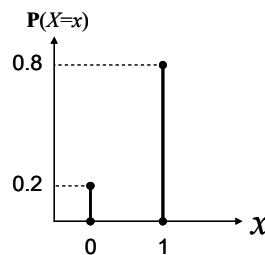


Figura 41. *pmf* de la *va* generada por la ocupación de un canal, cuya *CDF* aparece en la Figura 39(f)

²² Recordemos que el impulso de Dirac $\delta(x)$ vale cero en cualquier valor $x \in \mathbb{R}$, excepto en $x=0$, y que

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

34. Propiedades de la pdf y la pmf

Sea $f_X(\cdot)$ la pdf de alguna va X . Entonces

$$(a) f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$$

$$(b) F_X(x) = \int_{-\infty}^x f_X(a) da$$

$$(c) \int_{-\infty}^{\infty} f_X(a) da = 1$$

de manera que, para variables discretas, las anteriores propiedades se pueden reescribir en términos de la pmf así:

$$(a') p_k \geq 0$$

$$(b') F_X(x) = \sum_{k: x_k \leq x} p_k$$

$$(c') \sum_k p_k = 1$$

La primera propiedad se debe a que la CDF es no decreciente. La segunda propiedad es, simplemente, el teorema fundamental del cálculo. Y la tercera propiedad, que surge de evaluar la segunda en el punto $x=\infty$, es simplemente el primer axioma de las probabilidades definido en 14(a). La propiedad (a') es el segundo axioma de las probabilidades, y las propiedades (b') y (c') surgen de (b) y (c) evaluando la integral

$$\int_a^b f_X(x) dx = \int_a^b \sum_k p_k \delta(x - x_k) dx = \sum_k p_k \int_a^b \delta(x - x_k) dx = \sum_{k: a < x_k \leq b} p_k$$

Obsérvese que, a la luz de las anteriores propiedades y la interpretación sugerida por la Figura 40, podemos evaluar varias probabilidades en términos de la pdf (o la pmf) así:

$$(1) P(a < X \leq b) = \int_a^{b^-} f_X(x) dx \quad \left(P(a < X \leq b) = \sum_{k: a < x_k \leq b} p_k \right)$$

$$(2) P(a \leq X \leq b) = \int_{a^-}^{b^-} f_X(x) dx \quad \left(P(a \leq X \leq b) = \sum_{k: a \leq x_k \leq b} p_k \right)$$

$$(3) P(a \leq X < b) = \int_a^{b^-} f_X(x) dx \quad \left(P(a \leq X < b) = \sum_{k: a \leq x_k < b} p_k \right)$$

$$(4) P(a < X < b) = \int_{a^+}^{b^-} f_X(x) dx \quad \left(P(a < X < b) = \sum_{k: a < x_k < b} p_k \right)$$

donde las diferencias sutiles en los límites de la integral se refieren a la necesidad de incluir o excluir posibles impulsos de Dirac. Por supuesto, si X es una va continua (esto es, si la CDF es continua), los

cuatro intervalos contemplados en la columna izquierda tienen la misma probabilidad, pues en ese caso la probabilidad de cada punto individual es cero.

Nuevamente, como de la *pdf* (o la *pmf*) de una variable aleatoria se puede obtener la correspondiente *CDF*, cualquier función $g: \mathbb{R} \rightarrow \mathbb{R}$ que satisfaga las propiedades anteriores es la *pdf* de alguna variable aleatoria.

Como se nota en las expresiones anteriores, por brevedad basta con referirse solamente la *pdf*, pues todo lo que digamos de ella se extiende inmediatamente a la *pmf* en el caso de variables discretas, a través de la expresión $\int_a^b f_X(x)dx = \sum_{k:a < x_k \leq b} p_k$. Más aún, como la necesidad de distinguir entre las propiedades (a) y (a'), (b) y (b'), y (c) y (c') obedecen al uso de la integral de Riemann en las expresiones de probabilidad basadas en la *pdf*, simplificaremos la terminología y la notación si usamos la integral de Lebesgue (o, para este caso, su forma menos general de Riemann-Stieltjes) para evitar referirnos separadamente a la *pdf* o a la *pmf*. En efecto, usando la integral de Lebesgue, la probabilidad del evento $X \in A$ para algún conjunto de Borel A se expresa como $\int_A dF_X(x)$, tanto para variables continuas como para variables discretas o mixtas, donde

$$\mathbf{P}(X \in A) = \int_A dF_X(x) = \begin{cases} \int_A f_X(x)dx & \text{si } X \text{ es continua} \\ \sum_{k:x_k \in A} p_k & \text{si } X \text{ es discreta} \end{cases}$$

Para los lectores poco familiarizados con la teoría de mediciones o con el análisis real, baste pensar que el término $\mathbf{P}(X \in A) = \int_A dF_X(x)$ es, simplemente, una notación sencilla para referirse indistintamente a cualquiera de las dos expresiones $\int_A f_X(x)dx$ o $\sum_{k:x_k \in A} p_k$, según corresponda. Por supuesto, en muchas ocasiones será necesario hacer la distinción correspondiente, en cuyo caso volveremos a la sumatoria basada en la *pmf* o a la integral de Riemann basada en la *pdf*, que corresponden a la respectiva integral de Lebesgue en cada caso. Sin embargo, haremos una brevísima mención de esta notación, pues es la más común en la buena literatura técnica sobre redes de comunicaciones.

35. Integrales de Riemann e integrales de Riemann-Stieltjes

Sea $g: \mathbb{R} \rightarrow \mathbb{R}$ una función acotada definida en el intervalo $(a, b]$. Subdividamos dicho intervalo mediante algunos valores crecientes de x así $\{x_i\} = \{a = x_0 < x_1 < x_2 < \dots < x_n = b\}$. Para cada una de estas subdivisiones del intervalo podemos definir las siguientes dos sumas:

$$S = \sum_{i=1}^n M_i \cdot (x_i - x_{i-1}) \qquad s = \sum_{i=1}^n m_i \cdot (x_i - x_{i-1})$$

donde

$$M_i = \sup_{x_{i-1} < x \leq x_i} g(x) \qquad m_i = \inf_{x_{i-1} < x \leq x_i} g(x)$$

La integral superior de Riemann de g en $[a,b]$ es el ínfimo de los S sobre todas las posibles subdivisiones $\{x_i\}$, mientras que la integral inferior de Riemann de g en $[a,b]$ es el supremo de los s sobre todas las posibles subdivisiones $\{x_i\}$. La integral superior siempre será mayor o igual a la integral inferior pero, si las dos son iguales, el valor común es la integral de Riemann de g en $[a,b]$, que se denota así:

$$\int_a^b g(x)dx = \inf_{\{x_i\}} \sum_{i=1}^n \left[(x_i - x_{i-1}) \cdot \sup_{x_{i-1} < x \leq x_i} g(x) \right] = \sup_{\{x_i\}} \sum_{i=1}^n \left[(x_i - x_{i-1}) \cdot \inf_{x_{i-1} < x \leq x_i} g(x) \right]$$

En la integral de Riemann-Stieltjes hacemos la misma partición del mismo intervalo, pero en vez de considerar la longitud del subintervalo $(x_i - x_{i-1})$, consideramos el incremento de una función integradora (en nuestro caso, una función de distribución acumulativa):

$$\int_a^b g(x)dF_X(x) = \inf_{\{x_i\}} \sum_{i=1}^n \left[(F(x_i) - F(x_{i-1})) \cdot \sup_{x_{i-1} < x \leq x_i} g(x) \right] = \sup_{\{x_i\}} \sum_{i=1}^n \left[(F(x_i) - F(x_{i-1})) \cdot \inf_{x_{i-1} < x \leq x_i} g(x) \right]$$

La notación que se usa para referirnos a la integral de Riemann o a la integral de Riemann-Stieltjes es bastante clara:

$\int_a^b g(x)dx$	Integral de Riemann : El diferencial dx se refiere al límite de la longitud de cada intervalo, $\Delta x_i = x_i - x_{i-1} \rightarrow dx$
$\int_a^b g(x)dF_X(x)$	Integral de Riemann-Stieltjes : El diferencial $dF_X(x)$ se refiere al límite del incremento de la función integradora que, en el caso de una CDF, corresponde a la probabilidad de cada intervalo, $\Delta F_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \rightarrow dF_X(x)$

Nótese que con la integral de Riemann-Stieltjes, el concepto de integral supera la noción geométrica de "área bajo la curva", apropiado para la integral de Riemann. Ahora estamos considerando otra medida de los intervalos diferente a su longitud. Por otro lado, si describimos el diferencial $dF_X(x)$ mediante la *pdf*, $dF_X(x) = f_X(x)dx$, la integral Riemann-Stieltjes de $g(x)$ con respecto a la función integradora $F_X(x)$ equivale a la integral de Riemann de la función $g(x)f_X(x)$.

Si consideramos que la longitud de un intervalo es una medida del intervalo y que la probabilidad de que una variable aleatoria caiga en un intervalo es otra medida del intervalo, tanto la integral de Riemann como la de Riemann-Stieltjes son integrales con respecto a una medida particular de intervalos.

De todas maneras, el lector desinteresado puede simplemente considerar un asunto de notación:

$$\int_A g(x)dF_X(x) = \begin{cases} \int_A g(x)f_X(x)dx & \text{si } X \text{ es continua} \\ \sum_{k:x_k \in A} g(x_k)p_k & \text{si } X \text{ es discreta} \\ \alpha \int_A g(x)f_1(x)dx + (1-\alpha) \sum_{k:x_k \in A} g(x_k)p_k & \text{si } X \text{ es mixta} \end{cases}$$

En el tercer caso, nos referimos a una variable aleatoria mixta

$$F_X(x) = \alpha F_1(x) + (1-\alpha)F_2(x)$$

donde $F_1(x)$ es la CDF de una variable aleatoria continua con $pdf f_1(x)$, $F_2(x)$ es la CDF de una variable aleatoria discreta con $pmf p_k$ y $0 \leq \alpha \leq 1$ (Figura 38).

36. Valor Esperado de una Variable Aleatoria

Sea $F_X(\cdot)$ la CDF de alguna va X . El Valor esperado de X se define como $E[X] = \int_{\mathbb{R}} x dF_X(x)$. Al valor esperado también se le conoce como media, esperanza o primer momento de X .

Supongamos que obtenemos N calificaciones parciales en un curso de procesos estocásticos, $\{X_1, X_2, \dots, X_N\}$, cada una de ellas en el rango $\{0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$. Como cada calificación parcial puede tener, de alguna manera, algún componente aleatorio, el profesor querría tomar muchas muestras. Pero como, de todas maneras, al final del semestre el departamento le exige

un solo número, el profesor entrega el promedio, $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$. Esto puede ser injusto, porque debe

haber circunstancias distintas entre alguien que obtiene tres en todas las notas parciales y alguien que obtiene cinco en el 60% de ellas y cero en el 40% restante, aunque ambos obtienen un mismo promedio de 3.0. Sin embargo, como toca representar toda la secuencia $\{X_1, X_2, \dots, X_N\}$ mediante un solo número, casi nadie duda que el promedio \bar{X} es la mejor selección posible, especialmente si N es un número grande. Lo ideal sería presentar “la calificación final” como un histograma con la frecuencia relativa de cada posible valor de las calificaciones parciales, pero la administración de semejante proceso de calificación sería muy costosa para el departamento.

¿Cómo se relaciona ese número mágico \bar{X} , el promedio, con la distribución de la va X ? Consideremos la suma que se usa para el promedio y recalculémosla usando la asociatividad de la suma, así

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{N} \sum_{k=1}^{11} N_k x_k = \sum_{k=1}^{11} \frac{N_k}{N} x_k$$

donde x_k es el k -ésimo posible valor de X (en este caso $x_k = (k-1)/2$ para $k=1,2,\dots,11$), y N_k es el número de veces que se obtuvo la calificación x_k entre las N calificaciones parciales. La máxima justicia de esa calificación final se obtendría cuando el número de calificaciones parciales tendiera a infinito, en cuyo caso, de acuerdo con la definición 15,

$$\lim_{N \rightarrow \infty} \bar{X} = \sum_{k=1}^{11} x_k \left(\lim_{N \rightarrow \infty} \frac{N_k}{N} \right) = \sum_{k=1}^{11} x_k \mathbf{P}(X = x_k) = \sum_{k=1}^{11} x_k p_k \equiv \int_{\mathbb{R}} x dF_X(x)$$

Si X fuese una variable aleatoria continua de la que tomamos N muestras, bastaría con discretizar el rango de posibles valores en M subintervalos de longitud Δx , de manera que

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{N} \sum_{k=1}^M N_k x_k = \sum_{k=1}^M \frac{N_k}{N} x_k$$

donde N_k es el número de muestras X_i que caen en el k -ésimo intervalo y x_k es algún punto dentro del k -ésimo intervalo que satisface la igualdad de la suma (el cual existe por el teorema del valor medio). Si hacemos que el número de muestras N tienda a infinito, la relación N_k/N tiende a la probabilidad de que X caiga en el k -ésimo intervalo (según nuestra pragmática interpretación frecuentista de la definición 15) que, de acuerdo con la definición 33, equivale aproximadamente a $f_X(x_k)\Delta x$, si Δx es suficientemente pequeño:

$$\lim_{N \rightarrow \infty} \bar{X} = \sum_{k=1}^M x_k \lim_{N \rightarrow \infty} \frac{N_k}{N} \approx \sum_{k=1}^M x_k f_X(x_k) \Delta x$$

Ahora sólo basta con considerar el límite en el que Δx tiende a cero (en cuyo caso M debe tender a infinito) para que la aproximación sea exacta:

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \bar{X} = \int_{\mathbb{R}} x f_X(x) dx \equiv \int_{\mathbb{R}} x dF_X(x)$$

$\Delta x \rightarrow 0$

Así pues, el valor esperado no es más que una generalización del promedio numérico cuando consideramos un número infinito de muestras de la variable aleatoria. Más aún, de acuerdo con el ejemplo de las calificaciones, si el promedio es la estadística más sencilla que mejor resume la secuencia total de calificaciones, el valor esperado es la estadística más sencilla que mejor resume la distribución de una variable aleatoria.

Obsérvense también, en la interpretación anterior, las formas particulares que toma la integral de Lebesgue cuando se aplica a variables continuas y discretas separadamente, en cuyo caso utilizamos explícitamente la *pmf* o la *pdf* en vez de un diferencial general de la *CDF*:

$$E[X] = \int_{\mathbb{R}} x dF_X(x) \equiv \begin{cases} \sum_k x_k p_k & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Podemos derivar otra expresión muy útil para $E[X]$ cuando $X \geq 0$:

$$E[X] = \int_0^{\infty} x dF_X(x) = \int_0^{\infty} dF_X(x) \int_0^x du = \int_{x=0}^{\infty} \int_{u=0}^x dF_X(x) du$$

La Figura 42 muestra cómo podemos realizar el cambio del orden de integración:

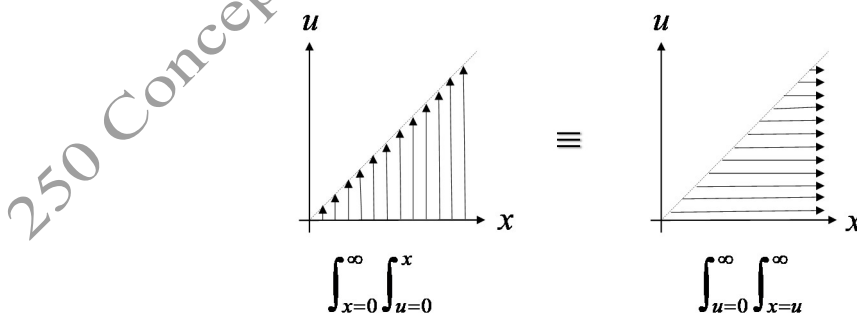


Figura 42. Cambio del orden de integración para $E[X]$, $X \geq 0$

Entonces

$$E[X] = \int_{x=0}^{\infty} \int_{u=0}^x dF_X(x) du = \int_{u=0}^{\infty} \left(\int_{x=u}^{\infty} dF_X(x) \right) du = \int_0^{\infty} P[X > u] du$$

De donde podemos concluir que

$$\text{Si } X \geq 0, E[X] = \int_0^{\infty} (1 - F_X(x)) dx$$

Que es la expresión útil a la que nos referíamos.

Una segunda estadística importante para resumir la distribución de una *va* es la varianza pero, para definirla, debemos conocer el valor esperado de una función de una variable aleatoria, lo cual requiere tres definiciones previas.

37. Función de una Variable Aleatoria

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se define una variable aleatoria $X: \Omega \rightarrow \mathbb{R}$. Sea $g: \mathbb{R} \rightarrow \mathbb{R}$ una función de los reales en los reales. Sea $Y: \Omega \rightarrow \mathbb{R}$ una función del espacio muestral en los reales tal que a cada $\omega \in \Omega$ le asigna la cantidad real $Y(\omega) = g(X(\omega))$. Si $\forall y \in \mathbb{R}$, el evento $B(y)$ definido como $\{\omega \in \Omega : Y(\omega) \leq y\}$ es un evento medible (esto es, si $B(y) \in \mathcal{F}$), entonces Y es una nueva variable aleatoria.

La relación entre las funciones $X(\omega)$ y $Y(\omega) = g(X(\omega))$ se muestran en la Figura 43. Bajo ambas transformaciones, las imágenes inversas de cualquier conjunto de Borel deben corresponder a eventos medibles en \mathcal{F} . Esto es, si $B \in \mathcal{B}(\mathbb{R})$, entonces $X^{-1}(B) \in \mathcal{F}$ y $Y^{-1}(B) \in \mathcal{F}$. Como X está definida en el espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(\cdot))$, cualquier función $g(\cdot)$ que transforme conjuntos de Borel en conjuntos de Borel generará una variable aleatoria válida $Y = g(X)$. Basta con verificar que la imagen inversa de cualquier intervalo $(-\infty, y]$, $y \in \mathbb{R}$, sea un conjunto de Borel, $B_y = g^{-1}((-\infty, y]) \in \mathcal{B}(\mathbb{R})$ porque, siendo así, $F_Y(y) = P[X \in B_y]$ y la variable aleatoria Y queda completamente definida²³.

²³ g es una función de los reales en los reales, $g: \mathbb{R} \rightarrow \mathbb{R}$, de manera que el argumento de g debe ser un número real. Sin embargo, podemos hablar de la imagen de un subconjunto $A \subset \mathbb{R}$ bajo la transformación g , $B = g(A) = \{y = g(x) : x \in A\} \subset \mathbb{R}$, así como de la imagen inversa de un conjunto $B \subset \mathbb{R}$, $A = g^{-1}(B) = \{x = g^{-1}(y) : y \in B\} \subset \mathbb{R}$. Para que $Y = g(X)$ sea una variable aleatoria se necesita que $B \in \mathcal{B}(\mathbb{R}) \Leftrightarrow A \in \mathcal{B}(\mathbb{R})$.

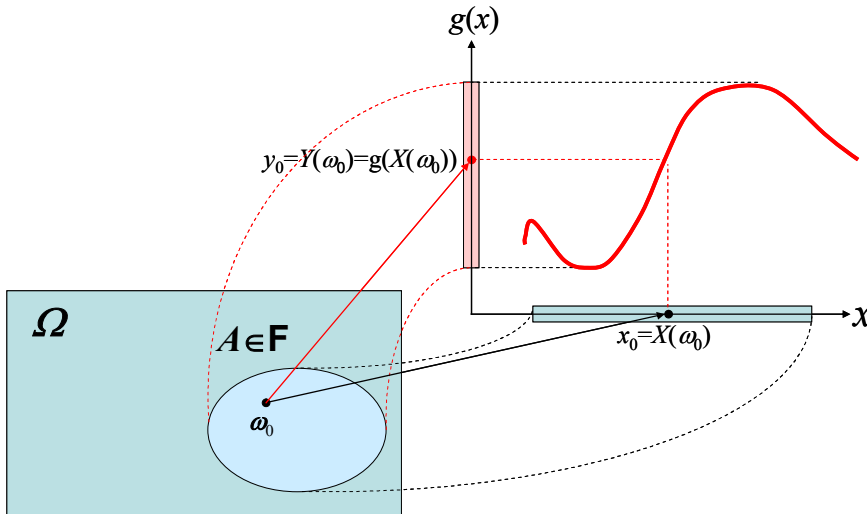


Figura 43. Concepto de Función de una Variable Aleatoria

Por ejemplo, si L es una variable aleatoria que representa la longitud en bits de un paquete de datos que se transmite por un enlace de capacidad C bps, el tiempo de transmisión del paquete será una nueva variable aleatoria dada por $T = (L+h)/C$, donde h es la longitud total de los encabezados que se le añaden al paquete en capas inferiores de la pila de protocolos. Definiendo $g(l) = (l+h)/C$, tenemos que $g^{-1}(t) = Ct-h$, de manera que $g^{-1}((-\infty, t]) = (-\infty, Ct-h]$ y, en consecuencia, g transforma conjuntos de Borel en conjuntos de Borel: $T=g(L)$ es una variable aleatoria válida.

Así como es difícil imaginar subconjuntos de los reales que no sean conjuntos de Borel, es difícil imaginar funciones $g: \mathbb{R} \rightarrow \mathbb{R}$ que no definan una nueva variable aleatoria. Por ejemplo, podría parecer que $g(x) = \sin(1/x)$ no definiría una variable aleatoria por su extraño comportamiento cerca de $x=0$, como muestra la Figura 44. Pero la verdad es que, para cualquier $y_0 \in \mathbb{R}$, el evento $\{\omega \in \Omega : Y(\omega) \leq y_0\}$ corresponde a una unión contable de conjuntos medibles definidos por X mediante intervalos, $\cup_{n \in \mathbb{Z}} \{\omega \in \Omega : 1/(2n\pi + \theta_2) \leq X(\omega) \leq 1/(2n\pi + \theta_1)\}$, donde $[\theta_1, \theta_2]$ es el primer intervalo con $\theta_1 \geq 0$ en el que $\sin(\alpha) \leq y_0$ si $\alpha \in [\theta_1, \theta_2]$.

$Y(\omega) = \sin(1/X(\omega))$ es una variable aleatoria si $X(\omega)$ es una variable aleatoria

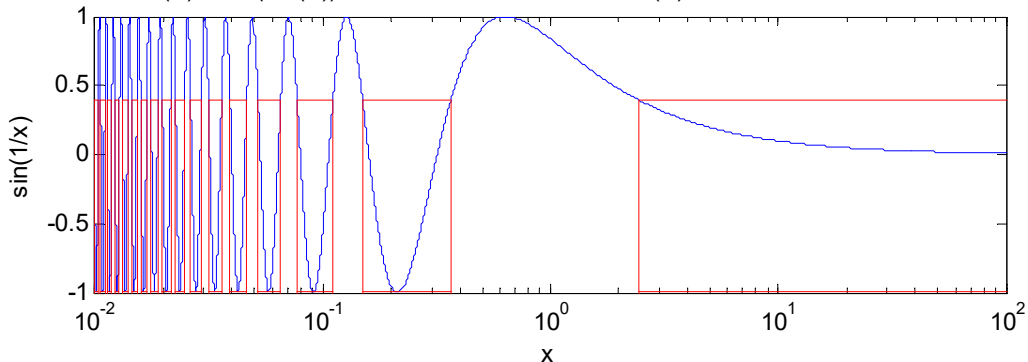


Figura 44. $Y(\omega)=g(X(\omega))$ es una variable aleatoria si $g: \mathbb{R} \rightarrow \mathbb{R}$ transforma conjuntos de Borel en conjuntos de Borel

Para otro ejemplo interesante, consideremos la secuencia "tienda de campaña",

$$x_{n+1} = \begin{cases} 3x_n & x_n < 1/2 \\ 3(1-x_n) & x_n \geq 1/2 \end{cases}$$

y definamos la función $g:[0,1] \rightarrow \mathbb{R}$ así: $g(x_0) = 1$ si la secuencia $\{x_0, x_1, x_2, \dots\}$ permanece siempre dentro del intervalo unitario ($x_n \in [0,1] \forall n \in \mathbb{N}$) y $g(x_0) = 0$ en otro caso. Es fácil ver que, si x_0 inicia en el intervalo $(1/3, 2/3)$, ya en el primer paso x_1 habrá excedido el intervalo unitario, como muestra la gráfica de la izquierda de la Figura 45, de manera que $g(x_0)=0$ para $x_0 \in (1/3, 2/3)$. De la misma manera, si $x_0 \in (1/9, 2/9) \cup (7/9, 8/9)$, ya en el segundo paso x_2 habrá excedido el intervalo unitario, como muestra la gráfica central de la Figura 45, de manera que $g(x_0)=0$ para $x_0 \in (1/9, 2/9) \cup (7/9, 8/9)$. La gráfica de la derecha de la Figura 45 muestra lo que ocurre en el tercer paso: En los intervalos $(1/27, 2/27)$, $(7/27, 8/27)$, $(19/27, 20/27)$ y $(25/27, 26/27)$, $g(x_0)=0$. Siguiendo de esta manera, queda claro que $g(x_0)$ vale uno en el conjunto de Cantor y vale cero fuera del conjunto de Cantor, esto es, $g(x)$ es la función indicadora del conjunto de Cantor.... ¿Es $Y=g(X)$ una v.a. si X está uniformemente distribuida en $[0,1]$?... En adelante supondremos que la expresión $Y=g(X)$ siempre definirá una nueva variable aleatoria Y cuando X sea una variable aleatoria.

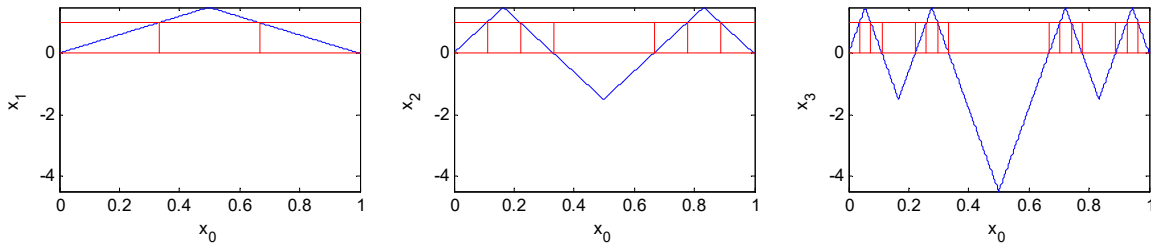


Figura 45. Funcion tienda de campaña $g:[0,1] \rightarrow \mathbb{R}$

38. pdf de una Función de una Variable Aleatoria

Sea $F_X(\cdot)$ la CDF de alguna va X y sea Y otra va definida mediante $Y=g(X)$, donde g es una función de los reales en los reales. Entonces la CDF de Y , $F_Y(y)$, satisface

$$dF_Y(y) = \sum_{i=1}^n dF_X(x_i)$$

donde $\{x_1, x_2, \dots, x_n\}$ son las raíces de la ecuación $y=g(x)$.

Si X es continua con pdf $f_X(\cdot)$ y g es una función diferenciable en todo punto, la pdf de Y está dada por

$$f_Y(y) = \sum_{i=1}^n f_X(x_i) \frac{1}{|g'(x_i)|}$$

donde $g'(x)$ es la derivada de $g(x)$. Si X es una va discreta, la pmf de Y está dada por

$$P(Y = y_j) = \sum_{i: y_j = g(x_i)} P(X = x_i)$$

En efecto, considérese el ejemplo mostrado en la Figura 46 en la que, para el punto y seleccionado, existen tres raíces de la ecuación $y=g(x)$, pues $g(x_1)=g(x_2)=g(x_3)=y$. Por el tercer axioma de la definición 14,

$$P(y < Y \leq y + \Delta y) = P(x_1 < X \leq x_1 + \Delta x_1) + P(x_2 + \Delta x_2 < X \leq x_2) + P(x_3 < X \leq x_3 + \Delta x_3)$$

donde todos los incrementos son positivos con la excepción de Δx_2 , que es menor que cero. A medida que Δy se hace más y más pequeña, obtenemos la expresión original de la definición, $dF_Y(y) = \sum_{i=1,2,3} dF_X(x_i)$. Si X es discreta, esta suma se interpreta como $P(Y=y) = P(X=x_1) + P(X=x_2) + P(X=x_3)$. Ahora veamos la interpretación de la suma si X es continua.

Si Δy es suficientemente pequeño, $dF_Y(y) = \sum_{i=1,2,3} dF_X(x_i)$ se puede reescribir de la siguiente manera:

$$f_Y(y)\Delta y \approx f_X(x_1)\Delta x_1 + f_X(x_2)|\Delta x_2| + f_X(x_3)\Delta x_3$$

donde la aproximación se hace exacta a medida que Δy tiende a cero. En términos generales, si existen n raíces, tenemos

$$f_Y(y) \approx \frac{1}{\Delta y} \sum_{i=1}^n f_X(x_i) |\Delta x_i| = \sum_{i=1}^n \frac{f_X(x_i)}{\Delta y / |\Delta x_i|}$$

donde, en el límite cuando Δy tiende a cero, obtenemos exactitud en la igualdad:

$$f_Y(y) = \lim_{\Delta y \rightarrow 0} \sum_{i=1}^n \frac{f_X(x_i)}{|\Delta y / \Delta x_i|} = \sum_{i=1}^n f_X(x_i) \frac{1}{|g'(x_i)|}$$

Nótese que si la ecuación $y=g(x)$ no tiene raíces, $dF_Y(y)=0$, como muestra la Figura 47(a). Igualmente, si las raíces forman un continuo, la Y puede tener un componente discreto aunque X sea continua, como muestra la Figura 47(b).

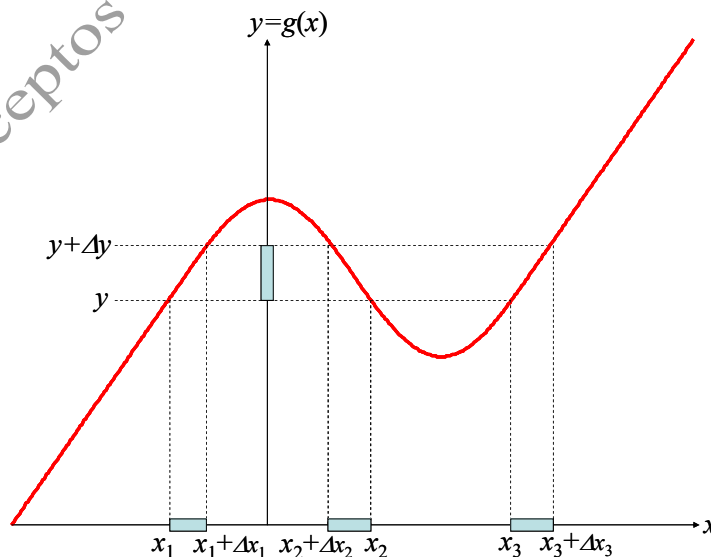


Figura 46. Construcción para encontrar $f_Y(y)$ cuando $Y=g(X)$

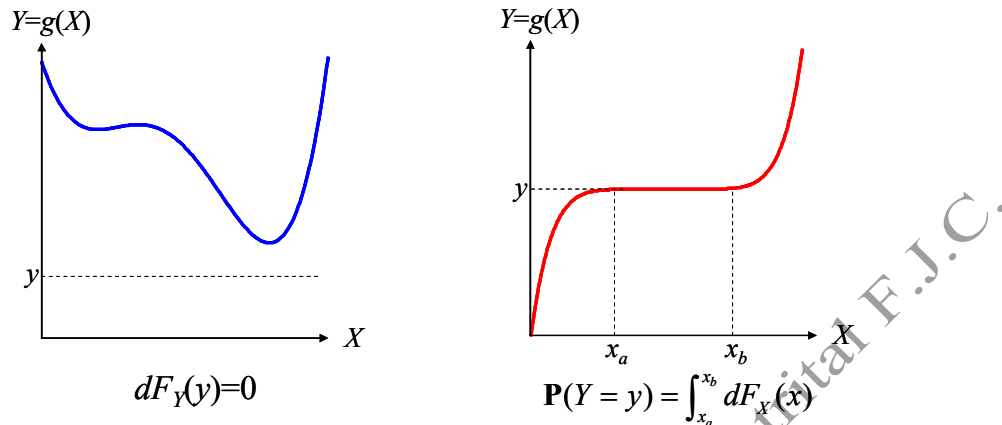


Figura 47. Casos en que $y=g(x)$ no tiene raíces (azul) y en que las raíces de $y=g(x)$ forman un continuo (rojo)

A manera de ejemplo, considérese la transmisión de un archivo desde un servidor ftp a través de un enlace de C bps. Si la longitud del archivo, L , tiene una pdf $f_L(l)$, ¿cuál será la pdf $f_T(t)$ del tiempo de transferencia, T ? Digamos que $T = L/C + t_0$, donde t_0 es el tiempo de establecimiento de la conexión ftp. Si definimos $g(l) \equiv l/C + t_0$ obtenemos que la única raíz de $t=g(l)$ es $l = C(t - t_0)$. En este caso la derivada de $g(l)$ es constante, $g'(l) = 1/C$. Consecuentemente, $f_T(t) = C f_L(C(t - t_0))$.

Como un segundo ejemplo, considérese la eficiencia en la transmisión de un paquete cuya longitud es una variable aleatoria L con pdf $f_L(l)$, cuando se le añade un encabezado de h bits: $E = g(L) = L/(L+h)$. La única raíz de $e=g(l)$ es $l = h e/(1-e)$ y la derivada de $g(l)$ es $g'(l) = h/(h+l)^2$. En consecuencia, $f_E(e) = \frac{h}{(1-e)^2} f_L\left(\frac{h e}{1-e}\right)$. La Figura 48 muestra las respectivas distribuciones de

L y E cuando $h = 192$ bits y L tiene una distribución exponencial $f_L(l) = \exp(-l/1024)/1024$.

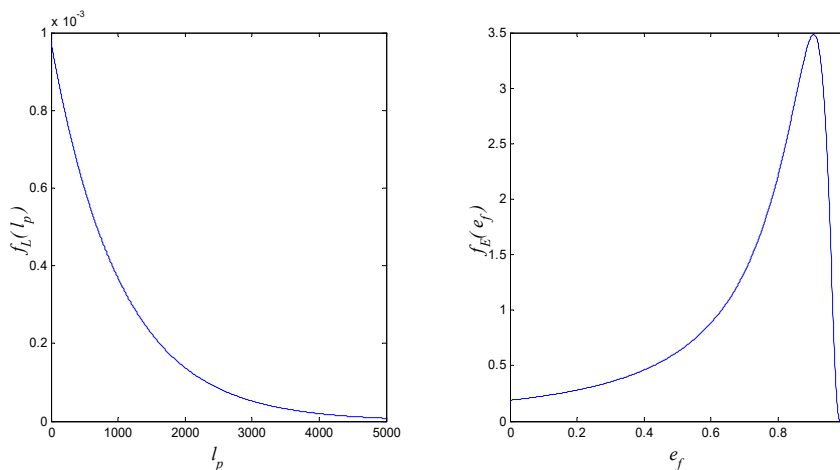


Figura 48. Funciones de densidad de probabilidad de la longitud de un paquete (a) y de la eficiencia en la transmisión cuando se añaden 192 bits de encabezado (b)

39. Valor Esperado de una Función de una Variable Aleatoria

Sea $F_X(\cdot)$ la CDF de alguna va X y sea Y otra va definida mediante $Y=g(X)$, donde g es una función de los reales en los reales. Entonces el valor esperado de Y está dado por

$$E[Y] = \int_{\mathbb{R}} g(x) dF_X(x)$$

En efecto, si por simplicidad suponemos que $g(\cdot)$ es una función monótonamente creciente, de la definición 38 sabemos que $dF_X(x) = dF_Y(g(x))$, por lo que

$$E[Y] = \int_{\mathbb{R}} y dF_Y(y) = \int_{\mathbb{R}} g(x) dF_X(x) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{si } X \text{ es continua} \\ \sum_k g(x_k) p_k & \text{si } X \text{ es discreta} \end{cases}$$

Para los valores de y en los que $y=g(x)$ tenga varias raíces, la expresión es la misma por asociatividad.

En el ejemplo de la transmisión de un archivo desde un servidor ftp a través de un enlace de C bps cuando la longitud del archivo, L , tiene una pdf $f_L(l)$, ¿cuál será el valor esperado del tiempo de transferencia $T = L/C + t_0$? Acabamos de ver que $f_T(t) = C f_L(C(t - t_0))$, de donde podemos verificar que $E[T] = \int_0^{\infty} (t_0 + l/C) f_L(l) dl = t_0 + E[L]/C$.

40. Varianza de una Variable Aleatoria

Sea X una va con valor esperado $E[X]$. La varianza de X , $V[X]$, se define como $V[X] = E[(X - E[X])^2]$. La desviación estándar de X , σ_X , se define mediante la relación $V[X] = \sigma_X^2$.

Supongamos que mandamos medir a uno de nuestros técnicos más brillantes una variable aleatoria X . El técnico es brillante pero perezoso y tramposo, por lo que decide inventarse algún número a y decir que ése fue el valor que midió. Como la equivocación será $X-a$, él quisiera escoger a de manera que la diferencia $X-a$ sea lo más cercana a cero posible. Para conseguir esto, el brillante técnico querría minimizar $(X-a)^2$ pero, como ésta es una función de una variable aleatoria, decide escoger el valor de a que minimiza $E[(X-a)^2]$. Por supuesto, la manera simple de encontrar el valor apropiado de a es observando la derivada de $E[(X-a)^2]$ respecto a a :

$$\begin{aligned} \frac{d}{da} E[(X-a)^2] &= E\left[\frac{d}{da}(X-a)^2\right] = E[2(a-X)] = 2 \int_{\mathbb{R}} (a-x) dF_X(x) \\ &= 2a \int_{\mathbb{R}} dF_X(x) - 2 \int_{\mathbb{R}} x dF_X(x) = 2(a - E[X]) \end{aligned}$$

Debido a la convexidad de la función $g(a)=E[(X-a)^2]$, el único valor extremo corresponde a un mínimo, así que basta con igualar la anterior derivada a cero para obtener el valor de a que minimiza el error cuadrado promedio (MSE –Mean Square Error–), $a=E[X]$. Así pues, cuando reemplazamos una variable aleatoria por su valor esperado minimizamos el MSE, el cual es, precisamente, la varianza de X , $V[X] = E[(X - E[X])^2]$. Correspondientemente, la desviación estándar σ_X es una medida de qué tan dispersos están los valores observados de X respecto a su valor medio, $E[X]$.

41. Propiedades del Valor Esperado y la Varianza de una Variable Aleatoria

Sea X una variable aleatoria con valor esperado $E[X]$ y varianza $V[X]$ y c una constante entre los reales. Entonces,

$$(a) E[X+c] = E[X]+c$$

$$(b) E[cX] = c E[X]$$

$$(c) V[X+c] = V[X]$$

$$(d) V[cX] = c^2 V[X]$$

$$(e) V[X] = E[X^2] - E[X]^2$$

Estas propiedades son muy fáciles de verificar:

$$(a) E[X+c] = \int_{\mathbb{R}} (x+c)dF_X(x) = \int_{\mathbb{R}} xdF_X(x) + c \int_{\mathbb{R}} dF_X(x) = E[X]+c$$

$$(b) E[cX] = \int_{\mathbb{R}} cxdF_X(x) = c \int_{\mathbb{R}} xdF_X(x) = cE[X]$$

$$(c) V[X+c] = E[(X+c) - (E[X]+c)]^2 = E[(X-E[X])^2] = V[X]$$

$$(d) V[cX] = E[(cX - cE[X])^2] = E[c^2(X-E[X])^2] = c^2E[(X-E[X])^2] = c^2V[X]$$

$$(e) V[X] = E[(X-E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - E[X]^2$$

donde la demostración de la propiedad (c) hace uso de la propiedad (a), la demostración de la propiedad (d) hace uso de la propiedad (b) y la demostración de la propiedad (e) hace uso de las propiedades (a) y (b). Estas propiedades se usarán con tanta cotidianidad que, finalmente, deberán ser recordadas como conceptos fundamentales de las variables aleatorias.

42. Momentos de una Variable Aleatoria

El n -ésimo momento de una variable aleatoria X es $E[X^n]$. El n -ésimo momento central es $E[(X-E[X])^n]$.

De acuerdo con lo anterior, el valor esperado es el primer momento y la varianza es el segundo momento central. El *skewness* es una cantidad muy útil relacionada con el tercer momento central, $S[X] = E[(X-E[X])^3]/V[X]^{3/2}$, que mide la simetría de la *pdf* de X alrededor de su valor medio (si $S[X]=0$, la *pdf* de X es simétrica alrededor de $E[X]$; si $S[X]<0$, la *pdf* “se recuesta” hacia la izquierda; y si $S[X]>0$, la *pdf* “se recuesta” hacia la derecha). El *kurtosis* es otra cantidad relacionada con el cuarto momento central, $K[X] = E[(X-E[X])^4]/V[X]^2 - 3$, que mide qué tan plana o puntuda es la *pdf* de X (entre más negativo es $K[X]$, la *pdf* de X tiende a ser más plana; entre más positivo, la *pdf* tiende a ser más puntuda. La referencia $K[X]=0$ corresponde a la distribución gaussiana, descrita en las definiciones 44 y 45. La Figura 49 muestra las características de una *pdf*, representadas por los cuatro primeros momentos de X .

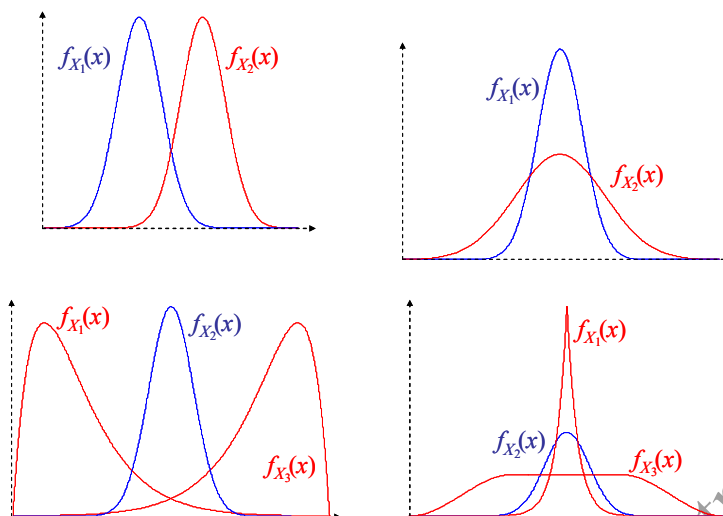


Figura 49. Efecto de los cuatro primeros momentos en la pdf de una va. En la parte superior izquierda, $E[X_1] < E[X_2]$. En la parte superior derecha, $V[X_1] < V[X_2]$. En la parte inferior izquierda, $S[X_1] < S[X_2] = 0 < S[X_3]$. Y en la parte inferior derecha, $K[X_1] > K[X_2] = 0 > K[X_3]$

43. Algunas Variables Aleatorias Discretas

- (a) Una variable aleatoria de **Bernoulli** con parámetro $p \in [0,1]$ toma dos posibles valores, $X \in \{0,1\}$, con $\mathbf{P}[X=1]=1-\mathbf{P}[X=0]=p$. Su valor esperado es p y su varianza es $p(1-p)$.
- (b) Una variable aleatoria geométrica con parámetro $p \in [0,1]$ toma valores enteros positivos, $X \in \{1,2,3,\dots\}$, de manera que $\mathbf{P}[X=k]=p^{k-1}(1-p)$. Su valor esperado es $1/(1-p)$ y su varianza es $p/(1-p)^2$.
- (c) Una variable aleatoria **binomial** con parámetros (n,p) , donde n es un entero positivo y p un real en el intervalo $[0,1]$, toma valores enteros no negativos en el rango $\{0,1,2,\dots,n\}$, de manera que $\mathbf{P}[X=k]=\binom{n}{k}p^k(1-p)^{n-k}$. Su valor esperado es np y su varianza es $np(1-p)$.
- (d) Una variable aleatoria de **Poisson** con parámetro $\rho > 0$ toma valores enteros no negativos, $X \in \{0,1,2,\dots\}$, de manera que $\mathbf{P}[X=k]=\frac{\rho^k}{k!}e^{-\rho}$. Tanto su valor esperado como su varianza son iguales a ρ .
- (e) Una variable aleatoria **uniforme discreta** con parámetros (m,n) , donde m y n son enteros tales que $m \leq n$, toma valores en el rango de números enteros $\{m, m+1, m+2, \dots, n-1, n\}$, de manera que $\mathbf{P}[X=k]=1/(n-m+1)$ si k está en el rango mencionado. Su valor esperado es $(m+n)/2$ y su varianza es $(n-m)(n-m+2)/12$.

A continuación damos algunos ejemplos de modelos probabilísticos en redes de comunicaciones basados en las anteriores variables aleatorias, y demostramos los resultados obtenidos respecto a la

media y la varianza de cada una de ellas. Todos ellos son de gran importancia práctica en telecomunicaciones, en especial el modelo de tráfico Poisson.

- (a) Muchos fenómenos aleatorios en el estudio de redes de comunicaciones pueden modelarse mediante variables aleatorias de Bernoulli, como ya se ha mencionado previamente.

Sea $X=1$ si un enlace se encuentra ocupado y $X=0$ si el mismo enlace se encuentra desocupado. Entonces X es una variable aleatoria de Bernoulli, donde el parámetro p es la utilización del enlace.

Sea $X=1$ si un bit transmitido sobre un enlace de radio punto-a-punto llega con error al otro extremo del enlace, y $X=0$ si el bit llega correctamente. Entonces X es una variable aleatoria de Bernoulli, donde el parámetro p es la tasa de error del canal, BER (Bit Error Rate).

El siguiente es el primero de una serie de modelos de tráfico que estudiaremos en este libro. Hay un enlace por el que se transmiten celdas ATM (Asynchronous Transfer Mode), donde el tiempo se discretiza en unidades correspondientes al tiempo de transmisión de una celda. En cada unidad de tiempo puede llegar una celda con probabilidad p o no llegar ninguna celda con probabilidad $1-p$. Dada una unidad particular de tiempo, sea $X=1$ si llega una celda en esa unidad y $X=0$ si no llega ninguna celda. Entonces X es una variable aleatoria de Bernoulli, donde p es la tasa de llegadas, en celdas/unidad de tiempo.

En cualquiera de los tres casos tenemos que $E[X] = 1 \cdot p + 0 \cdot (1-p) = p$, $E[X^2] = 1 \cdot p + 0 \cdot (1-p) = p$ y $V[X] = E[X^2] - E[X]^2 = p - p^2 = p(1-p)$, lo cual cobra mucho sentido a la luz de los ejemplos propuestos. En el modelo de tráfico, por ejemplo, si $p=0$, el promedio es 0 con varianza 0 pues en ninguna unidad de tiempo llegan paquetes; si $p=1$, el promedio es 1 con varianza 0 pues cada unidad de tiempo trae un paquete. La máxima varianza se da con $p = 1/2$, que corresponde a la máxima incertidumbre sobre la llegada de paquetes: si p es menor que $1/2$, tenemos mayor certeza de que no llegará un paquete y, si p es mayor que $1/2$, tenemos mayor certeza de que llegará un paquete. La figura 2.13 muestra la distribución, el promedio y la varianza del número de celdas que llegan por unidad de tiempo en función de la tasa de llegadas, p .

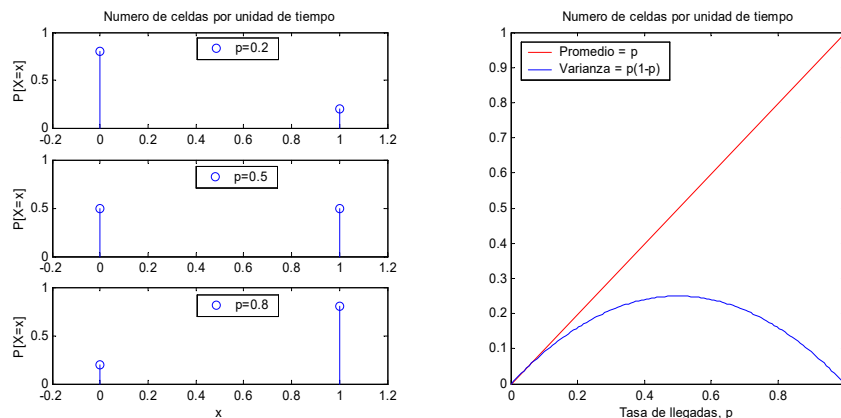


Figura 2.13 Distribución, promedio y varianza del número de celdas que llegan por unidad de tiempo bajo un modelo de Tráfico de Bernoulli

(b) La variable aleatoria geométrica surge de repetir un experimento de Bernoulli hasta que se obtenga uno de los resultados deseados, asegurando que cada experimento sea independiente de los demás. De hecho, existen cuatro formas posibles de definir una distribución geométrica, dependiendo si el experimento se repite hasta obtener un cero o hasta obtener un uno y, en cada caso, si el experimento exitoso se cuenta o no:

- Repetir hasta obtener 1 y no contar el 1: $\mathbf{P}[X=k]=p(1-p)^k, \quad k=0,1,2,\dots$
- Repetir hasta obtener 1 y contar el 1: $\mathbf{P}[X=k]=p(1-p)^{k-1}, \quad k=1,2,3,\dots$
- Repetir hasta obtener 0 y no contar el 0: $\mathbf{P}[X=k]=p^k(1-p), \quad k=0,1,2,\dots$
- Repetir hasta obtener 0 y contar el 0: $\mathbf{P}[X=k]=p^{k-1}(1-p), \quad k=1,2,3,\dots$

En la definición 43(b) se escogió el cuarto caso, que puede corresponder al siguiente ejemplo: Al transmitir un paquete por un enlace no-confiable se produce un error detectable con probabilidad p . El paquete se retransmite cuantas veces sea necesario hasta que llegue sin errores detectables al otro extremo del enlace, y se cuenta el número de transmisiones que se requieren, X . La probabilidad de tener que hacer una sola transmisión es la misma probabilidad de que el paquete llegue sin errores, $\mathbf{P}[X=1] = \mathbf{P}[\text{Primera transmisión exitosa}] = 1-p$. Será necesario hacer dos transmisiones si hay un error en la primera transmisión y la segunda resulta exitosa, lo cual ocurre con probabilidad $\mathbf{P}[X=2] = \mathbf{P}[\text{Primera transmisión con error}]\mathbf{P}[\text{Segunda transmisión exitosa} | \text{Primera con error}]$. Si la presencia de errores es independiente de una transmisión a otra, la anterior probabilidad condicional es igual a la correspondiente probabilidad incondicional, $\mathbf{P}[X=2] = \mathbf{P}[\text{Primera transmisión con error}]\mathbf{P}[\text{Segunda transmisión exitosa}] = p(1-p)$. En general, será necesario hacer k transmisiones si las primeras $k-1$ transmisiones sufren algún error y la k -ésima llega sin errores detectables. Dada la suposición de independencia, este evento sucede con probabilidad $\mathbf{P}[X=k] = p^{k-1}(1-p)$. El número promedio de transmisiones será

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k p_k = (1-p) \sum_{k=0}^{\infty} k p^{k-1} = (1-p) \sum_{k=0}^{\infty} \frac{d}{dp} p^k \\ &= (1-p) \frac{d}{dp} \sum_{k=0}^{\infty} p^k = (1-p) \frac{d}{dp} \frac{1}{1-p} = \frac{1}{1-p} \end{aligned}$$

El segundo momento se puede calcular igualmente fácil

$$\begin{aligned} E[X^2] &= \sum_{k=1}^{\infty} k^2 p_k = (1-p) p \sum_{k=0}^{\infty} k^2 p^{k-2} = (1-p) p \sum_{k=0}^{\infty} \left(\frac{d^2}{dp^2} p^k + k p^{k-2} \right) = (1-p) p \frac{d^2}{dp^2} \sum_{k=0}^{\infty} p^k + E[X] \\ &= (1-p) p \frac{d^2}{dp^2} \frac{1}{1-p} + \frac{1}{1-p} = \frac{2p}{(1-p)^2} + \frac{1}{1-p} = \frac{1+p}{(1-p)^2} \end{aligned}$$

de donde $V[X] = E[X^2] - E[X]^2 = p/(1-p)^2$.

En efecto, si la probabilidad de error es cero, el número promedio de transmisiones es uno y la varianza es cero, pues con probabilidad uno sólo se necesita una transmisión. A medida que aumenta p , tanto el promedio como la varianza aumentan, aunque la varianza aumenta más

rápidamente. Si p es uno, se requerirá un número infinito de transmisiones o, lo que es lo mismo, el paquete nunca llegará bien si el enlace es un lazo de cabuya.

Un segundo modelo de tráfico basado en la *va* geométrica puede construirse a partir del modelo anterior (basado en la *va* de Bernoulli) si contamos el número de unidades de tiempo (o *slots*) que debemos esperar hasta ver la llegada de la siguiente celda. Si en el primer slot que observamos llegó una celda, lo cual ocurre con probabilidad p , debimos esperar cero unidades. Para esperar una unidad de tiempo será necesario que en el primer slot no venga ninguna celda y en el segundo venga una celda, lo cual ocurre con probabilidad $p(1-p)$ si cada unidad de tiempo es independiente de las demás. En general, si en los primeros k slots no llegaron celdas y la primera celda llegó en el slot $k+1$, debimos esperar k unidades, lo cual, bajo la suposición de independencia, ocurre con probabilidad $\mathbf{P}[X=k] = p(1-p)^k, k=0,1,2,\dots$. Nótese que ésta es otra de las cuatro maneras de definir una distribución geométrica. Haciendo $Y = X+1$ y $q=1-p$, notamos que $\mathbf{P}[Y=k] = q^{k-1}(1-q), k=1,2,3,\dots$, como en el ejemplo anterior, de manera que $E[Y] = 1/(1-q)$ y, por consiguiente, usando la definición 41(a), $E[X] = 1/p - 1 = (1-p)/p$. Similarmente, por la definición 41(c), como $V[Y] = q/(1-q)^2$, entonces $V[X] = (1-p)/p^2$. La Figura 50 muestra estas cantidades.

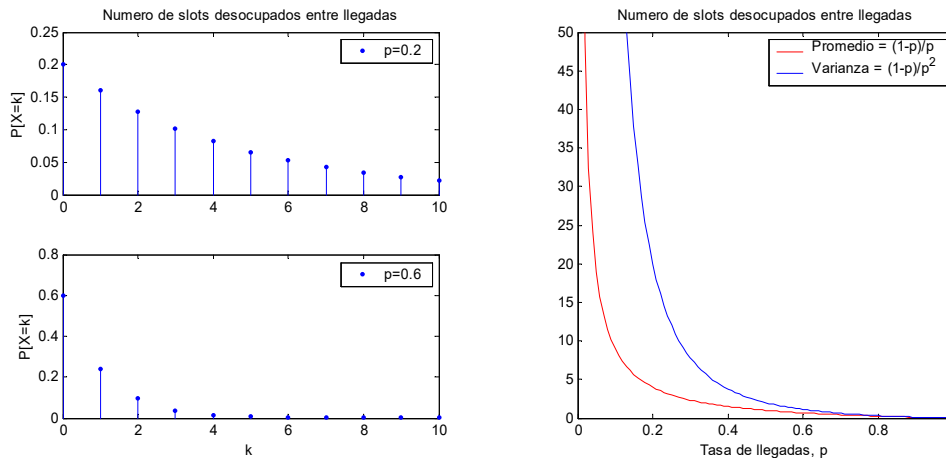


Figura 50. Distribución, promedio y varianza del número de slots desocupados entre llegadas bajo un modelo de tráfico geométrico

Nótese que en estos modelos geométricos es absolutamente necesario que los distintos experimentos de Bernoulli se realicen de manera independiente entre ellos. En el modelo de errores de transmisión, ¿será posible que la presencia de errores en la transmisión de una trama sea independiente de las tramas anteriores o siguientes? Si la transmisión se hace por un par trenzado no blindado de baja categoría y los errores se deben a la ignición de un motor eléctrico, los errores NO son independientes. Pero si la transmisión se hace a través de un satélite en horas nocturnas y los errores se deben al ruido galáctico, los errores SI pueden ser independientes: El modelo exige independencia y el analista deberá determinar si el modelo es aplicable o no. Igualmente, en el modelo de tráfico, si las celdas vienen de un gran número de fuentes independientes en las que cada una participa con una fracción muy pequeña del tráfico de manera que ninguna de ellas pueda generar celdas en unidades de tiempo cercanas entre sí, la suposición de independencia puede tener sentido. Pero si es un número pequeño de fuentes, cada una de las

cuales puede generar ráfagas de celdas en breves instantes de tiempo, será necesario revisar cuidadosamente la validez de la suposición de independencia.

- (c) La variable aleatoria binomial surge de hacer n repeticiones independientes de un experimento de Bernoulli y contar cuántas veces sucedió el resultado 1. Supongamos, por ejemplo, que se transmite una trama de n bits sobre un enlace no confiable y se cuenta el número de bits que llegan con error cuando los errores se dan en cada bit independientemente con probabilidad p . La probabilidad de que no haya ningún error es, claramente, $\mathbf{P}[X=0] = (1-p)^n$. La probabilidad de que solamente se dañe el i -ésimo bit es $p(1-p)^{n-1}$, de manera que la probabilidad de que se dañe un solo bit es

$$\mathbf{P}[X = 1] = P \left[\bigcup_{i=1}^n \{ \text{error sólo en el bit } i \} \right] = \sum_{i=1}^n p(1-p)^{n-1} = np(1-p)^{n-1}$$

donde la segunda igualdad obedece al tercer axioma de las probabilidades. De la misma manera, la probabilidad de que sólo se dañen los bits i y j es $p^2(1-p)^{n-2}$, por lo que la probabilidad de que se dañen exactamente dos bits es

$$\mathbf{P}[X = 2] = P \left[\bigcup_{i=1}^{n-1} \bigcup_{j=i+1}^n \{ \text{error sólo en los bits } (i, j) \} \right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^n p^2(1-p)^{n-2} = \frac{n(n-1)}{2} p^2(1-p)^{n-2}$$

En general, una combinación particular de sólo k errores se da con probabilidad $p^k(1-p)^{n-k}$, que es la misma probabilidad de que se dañen los k primeros bits o los k últimos, o los k de la mitad, o k de ellos tomados de dos en dos, etc. Como hay $\binom{n}{k} = n!/(k!(n-k)!)$ formas posibles de combinar k

bits con errores entre n bits transmitidos, $\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$.

El número promedio de bits recibidos con error es

$$E[X] = \sum_{k=0}^n kp_k = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np$$

De manera semejante podemos calcular el segundo momento,

$$E[X^2] = \sum_{k=1}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^j (1-p)^{n-1-j} \Big|_{m=n-1}^{j=k-1} = np((n-1)p+1)$$

de donde la varianza del número de bits equivocados es

$$V[X] = [(np)^2 + np(1-p)] - (np)^2 = np(1-p)$$

Nótese que la variable aleatoria binomial es la suma de n variables aleatorias de Bernoulli independientes. Como veremos en el próximo capítulo, eso justifica el hecho de que la media y la varianza de la distribución binomial sean n veces la media y la varianza de la distribución de Bernoulli, respectivamente.

Siguiendo con la serie de modelos de tráfico, podemos considerar una trama TDM (Time Division Multiplexing) de n slots, donde cada slot se comporta según los modelos de tráfico descritos en los modelos Bernoulli y geométrico. Entonces la variable aleatoria X =Número de celdas en una trama, está binomialmente distribuida con parámetros (n,p) . Igualmente, si consideramos un multiplexor que concentra n enlaces ATM como los descritos antes, donde el tráfico en cada

enlace es independiente de los demás enlaces, el número de celdas que llegan por unidad de tiempo tiene una distribución binomial con los mismos parámetros. La Figura 51 muestra la distribución, el promedio y la varianza del número de celdas que llegan en una trama de 32 slots.

Como en el caso de la variable aleatoria geométrica, cada vez que se quiera aplicar el modelo de la variable aleatoria binomial debemos justificar la suposición de independencia de los experimentos de Bernoulli subyacentes.

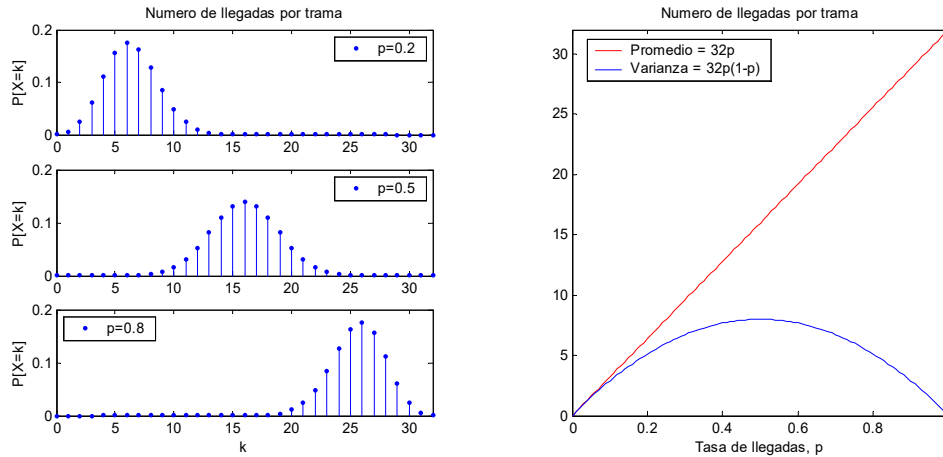


Figura 51. Distribución, promedio y varianza del número de celdas que llegan en una trama de 32 slots bajo un modelo de tráfico binomial

(d) Considérese un multiplexor que concentra un gran número de usuarios, de manera que los paquetes de datos pueden llegar en cualquier instante (modelo de tiempo continuo). Definamos la $va X$ como el número de llegadas que hay en un período de t segundos. Para caracterizar la $va X$, dividimos el intervalo de t segundos en n subintervalos contiguos y no superpuestos de longitud Δt , donde $t = n\Delta t$, y hacemos dos suposiciones básicas:

- a medida que la longitud del subintervalo Δt se hace más y más pequeña, la probabilidad de más de una llegada en un subintervalo tiende a cero y la probabilidad de una sola llegada en un subintervalo se hace proporcional a la longitud del intervalo, con factor de proporcionalidad λ :

$$P[k \text{ llegadas en } \Delta t] = \begin{cases} \lambda\Delta t + o(\Delta t) & k = 1 \\ 1 - \lambda\Delta t + o(\Delta t) & k = 0 \\ o(\Delta t) & k > 1 \end{cases}$$

donde $o(\Delta t)$ -ómicron de Δt - es cualquier función que tienda a cero más rápidamente que Δt :

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$$

de manera que $o(\Delta t) \pm o(\Delta t) = o(\Delta t)$, $o(\Delta t) \cdot o(\Delta t) = o(\Delta t)$, $\Delta t \cdot o(\Delta t) = o(\Delta t)$, etc. La distribución anterior indica que las llegadas simultáneas son improbables y que en cada subintervalo infinitesimal tenemos un experimento de Bernoulli en el que puede haber una llegada con probabilidad $\lambda\Delta t$ o ninguna llegada con probabilidad $1 - \lambda\Delta t$.

- El número de llegadas en un intervalo de tiempo dado es independiente del número de llegadas en cualquier otro intervalo de tiempo no superpuesto con el primero. En particular, el número de llegadas en cualquiera de los subintervalos de longitud Δt en que dividimos el tiempo es independiente del número de llegadas en cualquier otro intervalo anterior o posterior.

Para que X tome el valor k puede ocurrir que en k de los n subintervalos haya habido una sola llegada y en los restantes $n-k$ subintervalos no hayan habido llegadas, o que las k llegadas hayan sucedido en menos de k subintervalos. En este último caso, hubo más de una llegada en por lo menos un subintervalo, lo cual sucede con alguna probabilidad que tiende a cero más rápidamente que Δt , $o(\Delta t)$:

$$P[X = k] = o(\Delta t) + P \left[\bigcup_{I \subset \{1, 2, \dots, n\}, |I|=k} \left\{ \begin{array}{l} \text{una llegada en c/u de los subintervalos de } I, \\ \text{cero llegadas en los restantes } n-k \text{ subintervalos} \end{array} \right\} \right]$$

Dada la suposición de independencia en intervalos no superpuestos, así se trate de intervalos infinitesimales, la anterior expresión toma la siguiente forma:

$$P[X = k] = o(\Delta t) + \binom{n}{k} (\lambda \Delta t + o(\Delta t))^k (1 - \lambda \Delta t + o(\Delta t))^{n-k}$$

Expandiendo las potencias en el segundo término de la derecha y agrupando todas las funciones $o(\Delta t)$ en una sola,

$$\begin{aligned} P[X = k] &= o(\Delta t) + \binom{n}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} = o\left(\frac{t}{n}\right) + \frac{n!}{k!(n-k)!} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &= o\left(\frac{t}{n}\right) + \frac{n!}{n^k (n-k)!} \frac{(\lambda t)^k}{k!} \frac{1}{\left(1 - \lambda t/n\right)^k} \left(1 - \frac{\lambda t}{n}\right)^n = o\left(\frac{t}{n}\right) + \left(\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-k+1}{n}\right) \frac{(\lambda t)^k}{k!} \frac{1}{\left(1 - \lambda t/n\right)^k} \left(1 - \frac{\lambda t}{n}\right)^n \end{aligned}$$

Tomando el límite cuando n tiende a infinito (y Δt tiende a cero de manera que $t = n\Delta t$ siga constante), obtenemos $o(t/n) \rightarrow 0$, $(n-i)/n \rightarrow 1$, $(1 - \lambda t/n)^k \rightarrow 1$ y $(1 - \lambda t/n)^n \rightarrow e^{-\lambda t}$, de manera que

$$P[X = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Esto es, bajo las suposiciones anteriores, el número de llegadas en t segundos tiene una distribución Poisson con parámetro $\rho = \lambda t$. Durante cerca de un siglo éste ha sido el modelo de tráfico por excelencia en el diseño y análisis de redes de comunicaciones, aunque en las últimas dos décadas se ha acumulado una gran cantidad de evidencia que muestra que, en redes modernas de comunicaciones conmutadas por paquetes, la suposición de independencia en intervalos no superpuestos no es válida cuando se habla de la llegada de paquetes (aunque aún puede serlo cuando se habla del establecimiento de flujos o sesiones). Más aún, en muchos casos hay evidencia empírica que muestra cierta dependencia aún entre intervalos muy separados en el tiempo, fenómeno conocido como “dependencia de largo rango”, LRD –long range dependence–. Sin embargo, como veremos más adelante, la simplicidad del modelo Poisson (que supone independencia aún a nivel infinitesimal) permite obtener expresiones cerradas para muchas medidas de desempeño, gracias a lo cual sigue siendo utilizado como una primera aproximación en el dimensionamiento de la capacidad de las redes de comunicaciones y en el diseño de algoritmos de control para las mismas.

El valor esperado del número de paquetes que llegan en t segundos es

$$E[X] = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k \frac{\rho^k}{k!} e^{-\rho} = \rho \sum_{k=1}^{\infty} \frac{\rho^{k-1}}{(k-1)!} e^{-\rho} = \rho$$

que, con ρ definido como λt , indica que λ es el número promedio de llegadas por segundo o la tasa promedio de llegadas, que es uno de los parámetros más importantes en la caracterización de tráfico (el único parámetro en el caso de modelos de Poisson). El segundo momento es

$$\begin{aligned} E[X^2] &= \sum_{k=0}^{\infty} k^2 \frac{\rho^k}{k!} e^{-\rho} = \sum_{k=1}^{\infty} k((k-1)+1) \frac{\rho^k}{k!} e^{-\rho} = \rho \sum_{k=1}^{\infty} ((k-1)+1) \frac{\rho^{k-1}}{(k-1)!} e^{-\rho} \\ &= \rho \sum_{k=1}^{\infty} (k-1) \frac{\rho^{k-1}}{(k-1)!} e^{-\rho} + \rho \sum_{k=1}^{\infty} \frac{\rho^{k-1}}{(k-1)!} e^{-\rho} = \rho E[X] + \rho = \rho^2 + \rho \end{aligned}$$

de donde $V[X] = E[X^2] - E[X]^2 = \rho$. La varianza de una variable aleatorio de Poisson es igual a su valor medio.

La Figura 52 muestra la distribución del número de llegadas en un segundo cuando el tráfico obedece a un modelo de Poisson. La gráfica de la media y la varianza respecto a ρ son sólo dos líneas a 45°.

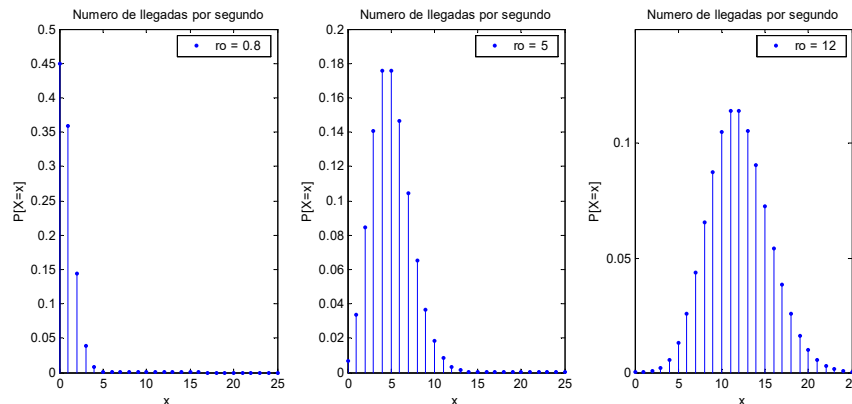


Figura 52. Distribución del número de llegadas en un segundo bajo un modelo de tráfico Poisson

- (e) Una trama TDM tiene n slots, numerados de 0 a $n-1$. Los paquetes llegan al multiplexor TDM en instantes completamente aleatorios, independientemente de la sincronización de la trama. Sea la va X el slot que se está transmitiendo de la trama que se está transmitiendo en el instante en que llega un paquete. Como no hay ninguna razón que permita imaginar que un paquete tenga alguna preferencia por un slot o un grupo de slots particular²⁴, parece razonable suponer que $\mathbf{P}[X=k] = 1/n, k=0,1,\dots,n-1$.

²⁴ Consideramos cada paquete independientemente de los demás. Dado un proceso de llegadas particular, podría haber alguna preferencia si condicionamos en el slot que le correspondió al paquete anterior, por ejemplo.

Su valor promedio es $E[X] = \sum_{k=0}^{n-1} k \frac{1}{n} = \frac{1}{n} \frac{n(n-1)}{2} = \frac{n-1}{2}$ y su segundo momento es $E[X^2] = \frac{1}{n} \sum_{k=0}^{n-1} k^2 = \frac{(2n-1)(n-1)}{6}$, de manera que su varianza es $V[X] = E[X^2] - E[X]^2 = \frac{(n^2-1)}{12}$.

Nótese la naturaleza de los ejemplos anteriores en los que cada *va* se convierte en un modelo probabilístico adecuado. Los experimentos de Bernoulli consisten en observar uno de dos posibles resultados, a cada uno de los cuales se les asocia el valor 0 ó 1 (error o no-error en un bit, ocupación o desocupación de un enlace, falla u operatividad de un dispositivo, presencia o ausencia de un paquete, etc.). Las variables binomial, geométrica y de Poisson modelan repeticiones independientes de un experimento de Bernoulli. En el modelo geométrico, se repite el experimento independientemente hasta obtener alguno de los dos resultados. En el modelo binomial el experimento se repite independientemente n veces y se cuenta el número de ocasiones en que ocurrió el resultado favorable. El modelo de Poisson es el límite consistente en un número infinito de repeticiones independientes durante un período finito de tiempo. El modelo uniforme obedece al principio de la máxima incertidumbre: Si tenemos un conjunto de proposiciones excluyentes a las cuales queremos asignar una distribución de probabilidad, debemos tener en cuenta qué sabemos de ellas. Si conocemos cuál es la cierta, le debemos asignar un valor de probabilidad igual a uno y las demás proposiciones tendrán probabilidad cero, pues no tenemos ninguna incertidumbre. Si algún conocimiento previo nos permite favorecer algunas proposiciones más que otras, podremos asignarles mayor probabilidad. Pero si no tenemos ninguna información que nos permita favorecer a ninguna de las proposiciones sobre las demás, nuestra incertidumbre será máxima y lo más conveniente será asignar las probabilidades uniformemente. Si lo hiciésemos de otra manera, estaríamos suponiendo una información que no poseemos.

44. Algunas Variables Aleatorias Continuas

- (a) Una variable aleatoria X **uniforme continua** con parámetros reales (a, b) toma valores en el intervalo $[a, b]$ y su pdf es $f_X(x) = 1/(b-a)$, $x \in [a, b]$. Su valor esperado es $(a+b)/2$ y su varianza es $(b-a)^2/12$.
- (b) Una variable aleatoria X **exponencial** con parámetro real positivo, $\lambda > 0$, toma valores entre los reales no negativos, y su pdf es $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Su valor esperado es $1/\lambda$ y su varianza es $1/\lambda^2$.
- (c) Una variable aleatoria X **Normal** (o **Gaussiana**) tiene parámetros (μ, σ^2) , donde μ es un número real y σ^2 es un número real no negativo, toma valores en los reales, y su pdf es

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), x \in \mathbb{R}$$

Su valor esperado es μ y su varianza es σ^2 .

- (d) Una variable aleatoria X con distribución de **Pareto** tiene parámetros positivos (a,b) , toma valores en los reales mayores o iguales a b , y su pdf es

$$f_X(x) = \frac{a}{b} \left(\frac{b}{x}\right)^{a+1}, x \geq b$$

Si $a > 1$, su valor esperado es $ab/(a-1)$; si no, su valor esperado es infinito. Si $a > 2$, su varianza es $ab^2/((a-2)(a-1)^2)$; si no, su varianza es infinita.

- (e) Una variable aleatoria X con distribución de **Cauchy** tiene parámetros reales (a,b) , $b > 0$, toma valores reales, y su pdf es

$$f_X(x) = \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}, x \in \mathbb{R}$$

Ni la media ni la varianza de la distribución de Cauchy están definidas.

- (f) Una variable aleatoria X con distribución de **Laplace** tiene un parámetro real positivo a , toma valores reales, y su pdf es

$$f_X(x) = \frac{a}{2} e^{-a|x|}, x \in \mathbb{R}$$

Su valor esperado es cero y su varianza es $2a^{-2}$.

- (g) Una variable aleatoria X con distribución de **Erlang** tiene parámetros (n,λ) , donde n es un entero positivo y λ es un real positivo. Toma valores reales no negativos y su pdf es

$$f_X(x) = \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}, x \geq 0$$

Su valor esperado es n/λ y su varianza es n/λ^2 .

- (h) Una variable aleatoria X con distribución **Gamma** tiene parámetros reales positivos (a,λ) , toma valores reales no negativos y su pdf es

$$f_X(x) = \frac{\lambda(\lambda x)^{a-1} e^{-\lambda x}}{\Gamma(a)}, x \geq 0 \quad \text{donde} \quad \Gamma(a) = \int_0^\infty s^{a-1} e^{-s} ds \quad (\text{función Gamma})$$

Su valor esperado es a/λ y su varianza es a/λ^2 .

- (i) Una variable aleatoria X con distribución de **Weibull** tiene dos parámetros (a,λ) , ambos reales positivos, toma valores reales no negativos y su pdf es

$$f_X(x) = a\lambda^a x^{a-1} \exp(-(\lambda x)^a), x \geq 0$$

Su valor esperado es $\Gamma((a+1)/a)/\lambda$ y su varianza es $((\Gamma((a+2)/a) - \Gamma((a+1)/a))^2)/\lambda^2$.

- (j) Una variable aleatoria X con distribución **Chi-cuadrado** (χ^2) tiene un parámetro real positivo, a , toma valores reales no negativos y su pdf es

$$f_X(x) = \frac{x^{a/2-1} \exp(-x/2)}{2^{a/2} \Gamma(a/2)}, x \geq 0$$

Su valor esperado es a y su varianza es $2a$.

- (k) Una variable aleatoria X con distribución de **t-de-Student** tiene un parámetro real positivo a , toma valores reales y su pdf es

$$f_X(x) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\sqrt{a\pi}\Gamma\left(\frac{a}{2}\right)} \left(1 + \frac{x^2}{a}\right)^{-\frac{a+1}{2}}, x \in \mathbb{R}$$

Su valor esperado es 0 y su varianza es $a/(a-2)$ para $a > 2$.

- (l) Una variable aleatoria X con distribución de **Rayleigh** tiene un parámetro real positivo, a , toma valores no negativos y su pdf es

$$f_X(x) = \frac{x}{a^2} \exp\left(-\frac{1}{2}\left(\frac{x}{a}\right)^2\right), x \geq 0$$

Su valor esperado es $a(\pi/2)^{1/2}$ y su varianza es $(2 - \pi/2)a^2$.

- (m) Una variable aleatoria X con distribución **Rice** tiene dos parámetros reales positivos, (ν, σ) , toma valores no negativos y su pdf es

$$f_X(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right), x \geq 0$$

donde $I_0(z)$ es la función modificada de Bessel del primer tipo y de orden cero. Su valor esperado es $\sigma\sqrt{\pi/2}L_{1/2}(-\nu/2\sigma^2)$ y su varianza es

$2\sigma^2 + \nu^2 - (\pi\sigma^2/2)L_{1/2}^2(-\nu/2\sigma^2)$, donde $L_{1/2}(x)$ es el polinomio de Laguerre,

$$L_{1/2}(x) = \exp(x/2)[(1-x)I_0(-x/2) - xI_1(-x/2)].$$

Las anteriores distribuciones son la base de algunos de los modelos más ampliamente usados y, por tal motivo, es importante que el lector aprenda a usar estos modelos en los contextos adecuados en los que se pueden utilizar. A continuación damos algunos ejemplos de los ocho primeros modelos probabilísticos en redes de comunicaciones y demostramos los resultados obtenidos respecto a la media y la varianza.

- (a) A un multiplexor estadístico llegan paquetes de longitud fija en instantes aleatorios e independientes de tiempo. En el instante de su llegada, el paquete b encuentra el enlace de salida ocupado transmitiendo el paquete a , y una larga cola de paquetes delante de él esperando ser transmitidos. Se mide el tiempo que transcurre desde la llegada de b hasta que a termina de ser transmitido, X , o “tiempo residual de servicio de a ”. Como a y b no son paquetes consecutivos (hubo un gran número de llegadas entre ellos) y como lo único que conocemos respecto al proceso de llegadas es que los tiempos entre llegadas son aleatorios e independientes, parece razonable suponer que b no tiene ninguna preferencia por llegar hacia el comienzo, el final o la mitad del tiempo de servicio de a . Y, como el rango de posibles valores de X es el intervalo $[0, T]$, donde T es el tiempo de transmisión de un paquete, el principio de máxima incertidumbre sugiere escoger la distribución uniforme para X , $f_X(x) = 1/T$ para $0 \leq x \leq T$. El valor medio de esta

distribución es $\frac{1}{T} \int_0^T t dt = \frac{t^2}{2T} \Big|_0^T = \frac{T}{2}$ y el segundo momento es $\frac{1}{T} \int_0^T t^2 dt = \frac{t^3}{3T} \Big|_0^T = \frac{T^2}{3}$, de

manera que su varianza es $(T^2/3) - (T^2/4) = T^2/12$. La Figura 53 muestra la *pdf*, la media y la varianza del tiempo residual de servicio de un paquete.

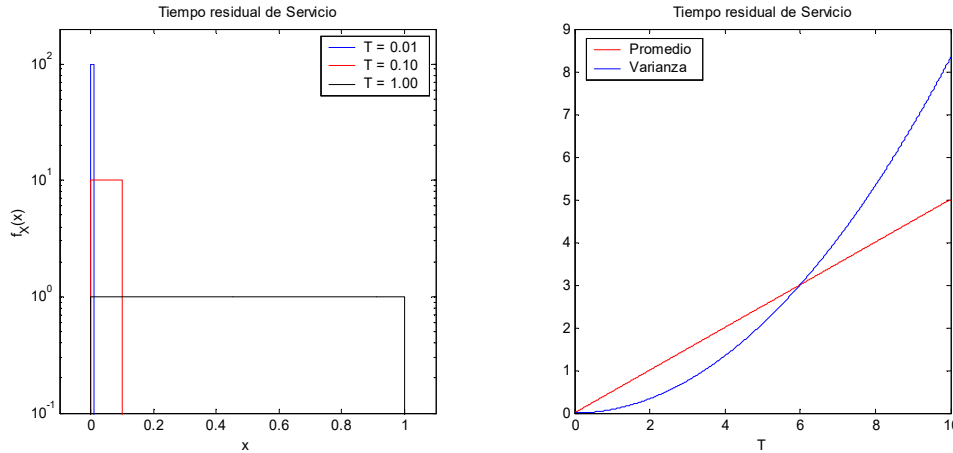


Figura 53. Función de densidad de probabilidad, promedio y varianza del tiempo residual de servicio según un modelo uniforme

(b) A un multiplexor llegan paquetes según un proceso de Poisson como el descrito en la definición

43(d), es decir, la probabilidad de que hayan k llegadas en t segundos es $\mathbf{P}[X = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

. Sea T la variable aleatoria “tiempo que toca esperar hasta ver la próxima llegada”. Consideremos el evento $T > t$, que corresponde al caso en el que, desde que empezamos a ver, han transcurrido t segundos sin que haya llegado aún ningún paquete. La probabilidad de dicho evento es la misma probabilidad de que en t segundos haya habido cero llegadas que, de acuerdo con la suposición de llegadas tipo Poisson, corresponde a $\mathbf{P}[T > t] = e^{-\lambda t}$. La probabilidad del evento complementario es $F_T(t) = \mathbf{P}[T \leq t] = 1 - e^{-\lambda t}$. La derivada de esta *CDF* es $f_T(t) = \lambda e^{-\lambda t}$, que es la *pdf* de una variable aleatoria exponencial. Su valor esperado es

$$E[T] = \lambda \int_0^{\infty} t e^{-\lambda t} dt = \lambda \left[-\frac{1 + \lambda t}{\lambda^2} e^{-\lambda t} \right]_0^{\infty} = \frac{1}{\lambda} \quad \text{y su } 2^{\text{do}} \text{ momento es}$$

$$E[T^2] = \lambda \int_0^{\infty} t^2 e^{-\lambda t} dt = \left[-\left(t^2 + (2t/\lambda) + (2/\lambda^2) \right) e^{-\lambda t} \right]_0^{\infty} = \frac{2}{\lambda^2}, \text{ de manera que su varianza es}$$

$1/\lambda^2$. Estos resultados refuerzan la idea de que el parámetro λ es la tasa promedio de llegada de paquetes.

Obsérvese que esta variable aleatoria es el modelo probabilístico de los tiempos entre llegadas cuando el tráfico obedece a un proceso de Poisson que, como dijimos, es el modelo de tráfico preferencialmente utilizado en redes de comunicaciones. Por consiguiente, la variable aleatoria exponencial es uno de los modelos probabilísticos más usados en redes de comunicaciones. La razón de su amplio uso es fácil de ver en la misma derivación que acabamos de hacer: nótese que medimos el tiempo que tardamos en ver la llegada del próximo paquete ¡sin tener en cuanto hace cuánto tiempo llegó el paquete anterior! La variable aleatoria exponencial es la única variable

continua que no tiene memoria: la distribución del tiempo que falta para ver la llegada del siguiente paquete es la misma independientemente del tiempo que ha transcurrido desde la llegada del paquete anterior. Esto es, la distribución de T sigue siendo $f_T(t) = \lambda e^{-\lambda t}$ así hayamos empezado a medir desde que llegó el paquete anterior o desde media hora después de que llegó el paquete anterior (dado que en esa media hora no ha llegado ningún paquete, por supuesto). Esta falta de memoria, que demostraremos formalmente en la definición 51, facilita enormemente el análisis de redes de comunicación, como veremos en el capítulo de teoría de colas.

La Figura 54 muestra la *pdf*, la media y la varianza del tiempo entre llegada de paquetes cuando el número de llegadas en cierto período de tiempo se modela mediante una distribución de Poisson.

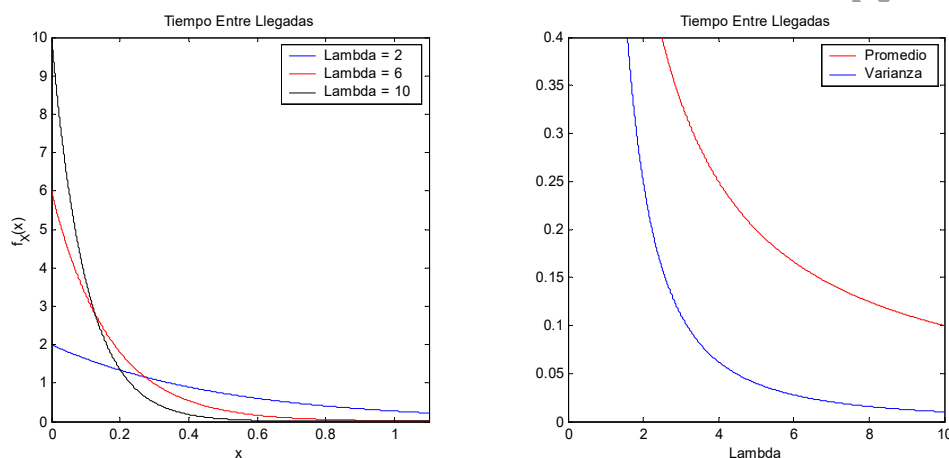


Figura 54. Función de densidad de probabilidad, promedio y varianza del tiempo entre llegada de paquetes según un modelo exponencial

- (c) Cuando una resistencia metálica de R ohmios se encuentra a una temperatura de T kelvins, sus electrones se mueven de manera aleatoria generando un voltaje de *ruido térmico* con nivel DC (media) cero y potencia (varianza) $2R(\pi kT)^2/3h$ W, donde k es la constante de Boltzmann y h es la constante de Planck (sugiriendo la presencia de fenómenos termodinámicos cuánticos). Supongamos, de una manera muy simplificada, que el movimiento de cada electrón en una resistencia R de 6.37 megohmios a 290 kelvins produce una caída de $+\Delta$ voltios con probabilidad 0.5 y $-\Delta$ voltios con probabilidad 0.5 y que cada electrón se mueve independientemente de los demás. Si existen n electrones libres en la resistencia, el voltaje producido será $V = (2X - n)\Delta$, donde X es una variable aleatoria binomial con parámetros $(n, 1/2)$, correspondiente al número de electrones que producen $+\Delta$ voltios. Aplicando los resultados de la definición 41 y de la definición 43(c), el valor medio del ruido térmico es cero y la varianza es $n\Delta^2$. Si hacemos que n crezca y Δ disminuya de manera que $n\Delta^2 = 2R(\pi kT)^2/3h = 1$ V², como predice la física, la probabilidad de obtener un voltaje de $(2k-n)\Delta$ voltios, con $0 \leq k \leq n$, es $\binom{n}{k} 2^{-n}$. Dividiendo esta probabilidad por $2\Delta = 2/\sqrt{n}$, que es el mínimo cambio en el voltaje, obtenemos la siguiente densidad de probabilidad:

$$\frac{P\left[(2k-n)/\sqrt{n} \leq V < (2(k+1)-n)/\sqrt{n}\right]}{2/\sqrt{n}} = \sqrt{n} \binom{n}{k} 2^{-(n+1)}$$

la cual se muestra en la Figura 55 (barras) y que se compara con la expresión

$f_V(v) = \frac{1}{\sqrt{2\pi}} \exp(-v^2/2)$ (línea continua), conocida como “Campana de Gauss”, que es

la *pdf* de una variable Gaussiana con media 0 y varianza 1.

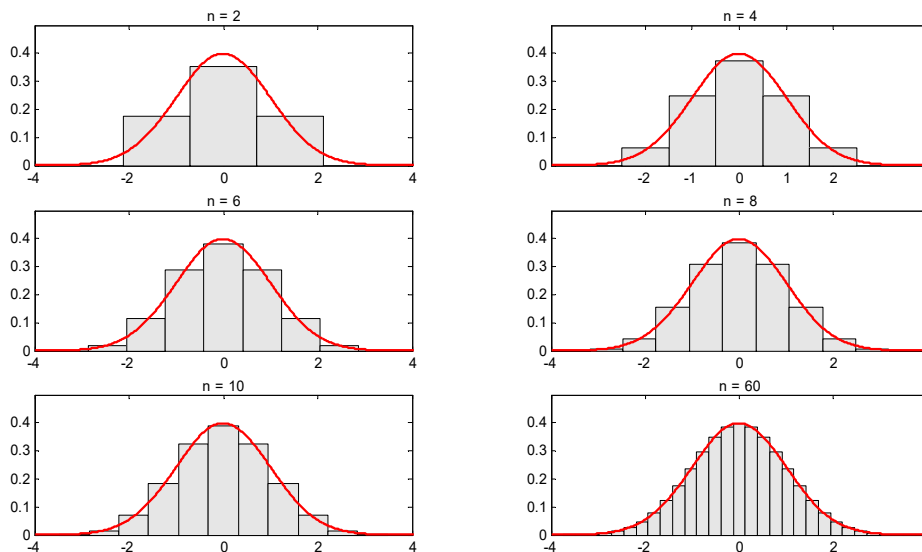


Figura 55. Función de densidad de probabilidad del ruido térmico producido por n partículas, donde cada partícula genera $+1/\sqrt{n}$ voltios con probabilidad $1/2$ o $-1/\sqrt{n}$ voltios con probabilidad $1/2$. Se compara con la función de densidad de probabilidad Gaussiana

Claramente, a medida que consideramos más y más electrones, la *pdf* del ruido térmico se hace más cercana a la distribución Gaussiana. Por supuesto, lo más razonable es considerar un número infinito de electrones, cada uno participando con un infinitésimo del voltaje de ruido, de manera que el modelo Gaussiano resulta apenas natural para modelar el ruido térmico en una resistencia metálica, tal como la impedancia de entrada del amplificador de radiofrecuencia en un sistema de comunicaciones.

Como en el ejemplo anterior, si X representa la suma de N componentes aleatorios independientes en la que cada componente contribuye con una pequeña fracción de la suma, la *pdf* de X se aproxima a la distribución Gaussiana a medida que N aumenta, ¡independientemente de la distribución de los componentes individuales! Este es el teorema del límite central propuesto por Laplace en 1810, que estudiaremos con cuidado en la definición 81. De hecho, dado el determinismo que imperaba en esa época, la aleatoriedad sólo se usaba para modelar los errores experimentales de medición que, en términos de observaciones astronómicas, Gauss asoció con su famosa “campana” pues, evidentemente, se trataba de la suma de muchos errores debidos a la dispersión y la difracción de la luz con cada partícula de

la atmósfera. Debido a la naturaleza de estos modelos de error, la distribución Gaussiana resultó la más normal de las distribuciones y, por esa razón, también se le conoce como distribución normal, $\mathcal{N}(\mu, \sigma^2)$. Hoy se sabe que las distribuciones más "normales" en la naturaleza son las que tienen colas pesadas (definición 48), como la distribución de Pareto que se describirá a continuación de ésta distribución Gaussiana.

El valor medio de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$ (que se lee "normalmente distribuida con parámetros μ y σ^2 ") es

$$\begin{aligned} E[X] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) \exp(-y^2/2) dy \Bigg|_{y=\frac{x-\mu}{\sigma}} \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy + \mu \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) = \mu \end{aligned}$$

Pues, en la última expresión, la primera integral es cero por tratarse de una función con simetría impar y la expresión entre paréntesis del segundo término es la probabilidad total de una variable $\mathcal{N}(0,1)$. Para hallar la varianza de X partamos de la probabilidad total:

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = 1$$

Multipliquemos ambos lados por $\sigma\sqrt{2\pi}$ y derivemos respecto a σ :

$$\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma^3} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \sqrt{2\pi}$$

Y, finalmente, multipliquemos a ambos lados por $\sigma^2/\sqrt{2\pi}$ para obtener

$$V[X] = E[(X - \mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right) dx = \sigma^2$$

De donde los parámetros de una variable normal son su media y su varianza. La figura 2.20 muestra algunas *pdfs* Gaussianas, donde se nota el efecto del valor esperado μ como un parámetro de posición y el efecto de la varianza σ^2 como un parámetro de forma que describe la concentración de la distribución alrededor de su valor esperado.

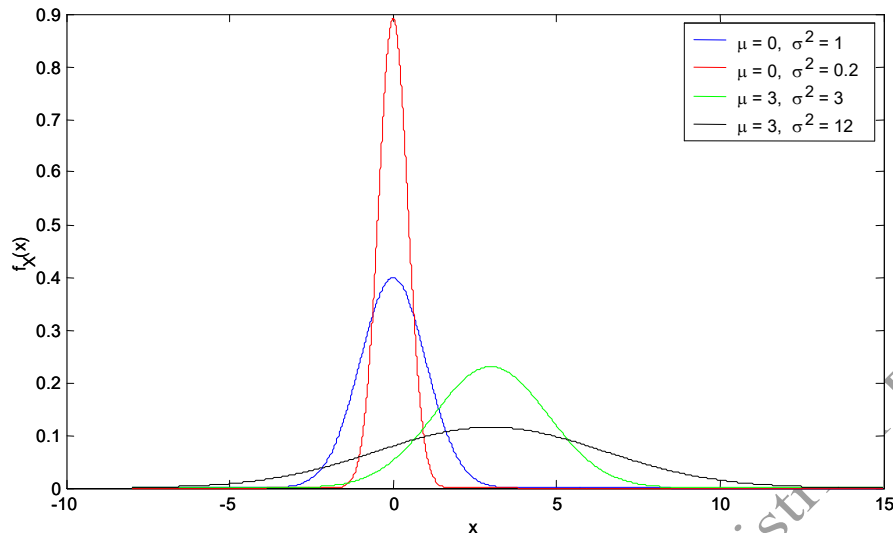


Figura 56. Funciones de densidad de probabilidad Gaussianas

- (d) Las distribuciones que hemos visto hasta ahora se caracterizan porque la probabilidad de que las variables aleatorias tomen valores muy grandes es muy pequeña, de manera que el efecto total de dichos valores es despreciable. Sin embargo, en redes modernas de comunicaciones (como en todos los sistemas que recientemente se han caracterizado como “complejos”), se hacen cada vez más comunes algunas variables aleatorias que pueden tomar valores muy grandes con probabilidad no despreciable, de manera que, cuando finalmente se presentan estos valores, su efecto puede ser determinante. Este es el caso del tamaño de los archivos que se intercambian por la red (la gran mayoría son muy pequeños pero la pequeña fracción de archivos grandes son los que consumen la mayoría de recursos en la red), la duración de una conexión http (la gran mayoría de conexiones son breves, pero las pocas conexiones duraderas son las que más ocupan a los servidores web), etc. De estas cantidades se dice que tienen “cola pesada” (ver definiciones 0, 48 y 51), y una de las distribuciones más utilizadas para modelarlas probabilísticamente es la distribución de Pareto, la cual se usó originalmente para describir la concentración de riquezas (la gran mayoría de personas son pobres, pero las pocas personas ricas que existen poseen la gran mayoría de la riqueza del mundo)²⁵. En efecto, la Figura 57 compara una distribución exponencial con parámetro $\lambda = 1/3$ y una distribución de Pareto con parámetros $a=1.5$ y $b=1$, de manera que ambas tienen el mismo valor promedio $\mu=3$, aunque la segunda tiene varianza infinita. Un cálculo muy simple muestra que la probabilidad de que la variable de Pareto sea superior a n veces su valor esperado es $[(a-1)/(na)]^a = (3n)^{-3/2}$, mientras que la probabilidad de que la variable exponencial sea superior a n veces su valor esperado es $\exp(-n)$. Esto es, aunque la probabilidad de que la variable exponencial supere la media es casi el doble de que la variable Pareto también lo haga, la probabilidad de que la variable exponencial supere 9 veces la media es menos de una millonésima de la probabilidad de que la variable Pareto también lo haga!

²⁵ Este fenómeno de cola pesada (o “ley-de-potencia” en la cola de la distribución) ha resultado tan ubicuo, que muchos científicos empiezan a considerar una explicación general basada en la auto-organización en puntos críticos al borde del caos o basada en la tolerancia altamente optimizada,

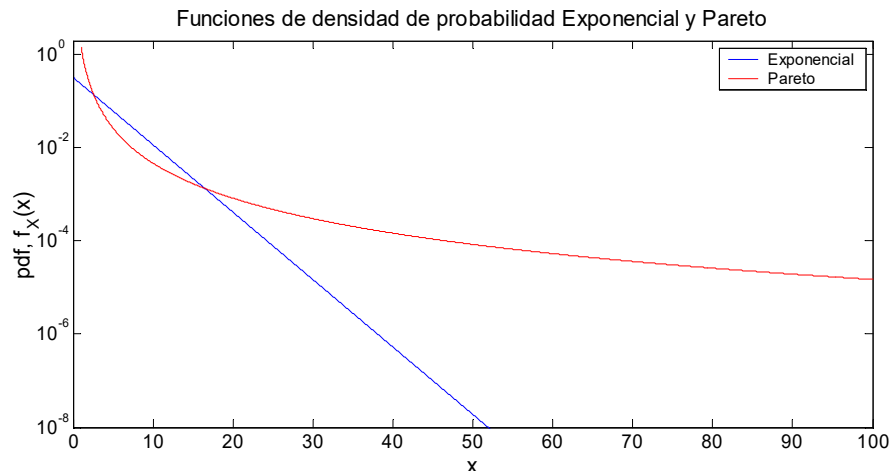


Figura 57. Funciones de densidad de probabilidad Exponencial y de Pareto

El valor medio de una variable $X \sim \text{Pareto}(a,b)$ es

$$E[X] = \int_b^{\infty} x \frac{a}{b} \left(\frac{b}{x}\right)^{a+1} dx = ab^a \int_b^{\infty} x^{-a} dx = -\frac{ab^a}{a-1} (x^{-a+1}) \Big|_b^{\infty} = \begin{cases} \frac{ab}{a-1} & a > 1 \\ \infty & a \leq 1 \end{cases}$$

Y su segundo momento es

$$E[X^2] = \int_b^{\infty} x^2 \frac{a}{b} \left(\frac{b}{x}\right)^{a+1} dx = -\frac{ab^a}{a-2} (x^{-a+2}) \Big|_b^{\infty} = \begin{cases} \frac{ab^2}{a-2} & a > 2 \\ \infty & a \leq 2 \end{cases}$$

De manera que la varianza es

$$V[X] = E[X^2] - E^2[X] = \begin{cases} \frac{ab^2}{(a-2)(a-1)^2} & a > 2 \\ \infty & a \leq 2 \end{cases}$$

Obsérvese que, en el rango $1 < a \leq 2$, una v.a. de Pareto tiene media finita y varianza infinita, de donde surgen las características de ley de potencia que hacen tan interesante esta distribución para representar los fenómenos de complejidad observados recientemente en redes de comunicaciones.

- (e) La distribución de Cauchy (o de Lorentz, como se le conoce en física) resuelve la ecuación diferencial que describe algunos sistemas de resonancia forzada, tales como el ensanchamiento de las líneas espectroscópicas debido a fenómenos de resonancia. En redes de telecomunicaciones el interés en la distribución de Cauchy es de tipo estadístico porque, al ser semejante a la distribución normal cerca al máximo de la distribución (el *modo*), como muestra la Figura 58, la robustez de las pruebas de hipótesis que asumen normalidad se puede probar con datos tomados de una distribución Cauchy. Además, la razón X/Y de dos v.a. gaussianas independientes X y Y tiene una distribución Cauchy.

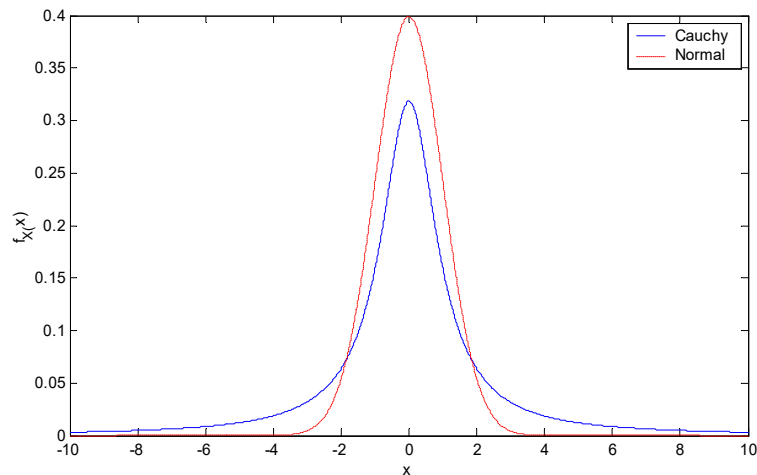


Figura 58. La distribución de Cauchy es la versión con cola pesada de la distribución Gaussiana

- (f) Considere dos paquetes de longitud exponencialmente distribuida con promedio L bits, independientes entre ellos, que empiezan a transmitirse simultáneamente en dos canales de la misma capacidad, C . Sea T_1 el tiempo de transmisión del primer paquete y T_2 el tiempo de transmisión del segundo paquete. La diferencia entre los tiempos de transmisión, $T = T_1 - T_2$, tiene una distribución Laplaciana:

$$f_T(t) = \int_{\max(0,-t)}^{\infty} f_{T_1}(t+s)f_{T_2}(s)ds = \lambda^2 e^{-\lambda t} \int_{\max(0,-t)}^{\infty} e^{-2\lambda s} ds = \frac{\lambda}{2} e^{-\lambda t} e^{-2\lambda \max(0,-t)} = \frac{\lambda}{2} e^{-\lambda|t|}$$

donde $1/\lambda = L/C$ es el tiempo promedio de transmisión de un paquete. En efecto, más adelante veremos cómo la independencia de los eventos asociados con cada variable hace que la *pdf* conjunta $f_{T_1, T_2}(t_1, t_2)$ sea el producto de las *pdf* marginales $f_{T_1}(t_1)f_{T_2}(t_2)$, de manera que la expresión anterior es, sencillamente, la evaluación de la probabilidad total (definición 19). La distribución de Laplace se muestra en la Figura 59.

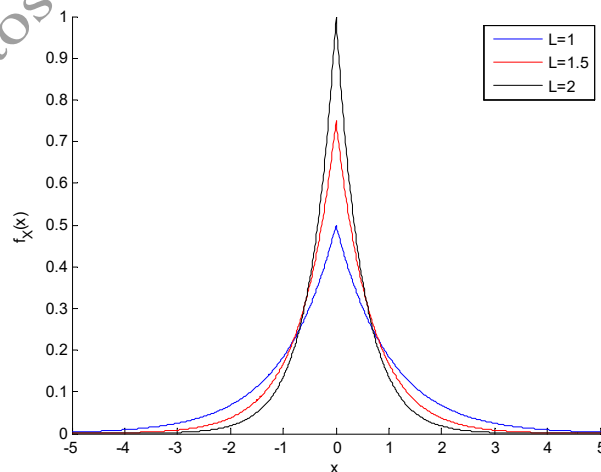


Figura 59. La distribución de Laplace es una versión simétrica de la distribución exponencial

El valor medio de una variable Laplaciana es

$$E[T] = \frac{\lambda}{2} \int_{-\infty}^{\infty} t e^{-\lambda|t|} dt = \frac{\lambda}{2} \left(\int_0^{\infty} t e^{-\lambda t} dt + \int_{-\infty}^0 t e^{-\lambda|t|} dt \right) = 0$$

y su segundo momento es

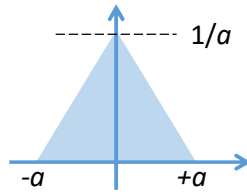
$$V[T] = E[T^2] = \frac{\lambda}{2} \int_{-\infty}^{\infty} t^2 e^{-\lambda|t|} dt = \lambda \int_0^{\infty} t^2 e^{-\lambda t} dt = \frac{2}{\lambda^2}$$

(g & h) Cómo vimos, el tiempo entre llegadas consecutivas de paquetes que obedecen a un proceso de Poisson es una variable exponencial. Cabe preguntarse por el tiempo que tomará la llegada de n paquetes, que corresponderá a la suma de n variables aleatorias exponenciales independientes e idénticamente distribuidas. La suma obedecerá a la distribución de una variable aleatoria de Erlang(n, λ), donde λ es el parámetro del proceso Poisson subyacente. Si el número de términos que se suman se puede interpolar a valores no enteros, se obtiene la distribución Gamma.

45. La distribución Gaussiana

En 1904, Henri Poincaré intentó explicar con humor el uso generalizado de la distribución Gaussiana: “Los físicos creen que la distribución Gaussiana fue demostrada por las matemáticas y los matemáticos creen que la distribución Gaussiana fue descubierta experimentalmente por los físicos”. Lo cierto es que diferentes formas de esta distribución han sido propuestas desde el siglo XVIII, por lo que ha habido cerca de 300 años para estudiarla, aprovecharla y hasta para abusar de ella. De hecho, en 1733, DeMoivre la descubrió como una fórmula rápida para calcular la probabilidad del número de caras y sellos en una gran cantidad de monedas (Figura 55), que era un empleo típico de los matemáticos en las cortes europeas. Sin embargo, pasaron más de 70 años antes de que la distribución Gaussiana cobrara toda su importancia. Entonces se suponía que el uso de la aleatoriedad era exclusivamente para considerar errores de medición, como la dispersión de los rayos de luz provenientes de una estrella cuando atraviesan las microturbulencias de la atmósfera terrestre antes de llegar al lente de un telescopio. En 1755 Thomas Simpson propuso una función de densidad de probabilidad triangular para el error de medición, que tiene su base en el intervalo $[-a, a]$ y su altura máxima, $1/a$, en cero. Con esta distribución, el error se podía reducir tomando muchas muestras y promediándolas. En 1770 Laplace generalizó esta propuesta postulando que la distribución del error debe ser simétrica alrededor de cero y debe tender a cero a medida que el valor absoluto del error tiende a infinito, algo que ya Galileo había intuido 150 años antes (que los errores de medición eran simétricos y que los errores más pequeños eran más frecuentes). En 1780 Laplace mismo introdujo la distribución de Laplace como función de densidad del error para capturar sus postulados de manera concreta. La Figura 60 muestra las propuestas de Simpson y Laplace para las distribuciones de los errores de medición.

$$f_E(x) = \frac{1}{a} \max[0, \min(1+x/a, 1-x/a)]$$



$$f_E(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

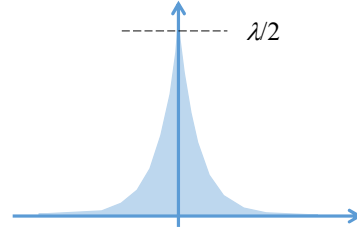


Figura 60. Funciones de densidad de probabilidad para el error de Simpson y Laplace

En 1805 Legendre publicó el método de los mínimos cuadrados para regresión lineal. Dado un conjunto de medidas $\{(x_i, y_i), i=1, 2, \dots, n\}$ que se quieren ajustar a una línea recta $\hat{y}_i = ax_i + b$,

se considera el error cuadrado promedio de las medidas, $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - (ax_i + b))^2$, y se

encuentran los valores de a y b que minimizan el MSE , $(a^*, b^*) = \arg \min_{(a,b)} \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2$. Al

derivar el MSE con respecto a a y a b e igualar a cero se obtiene el mínimo (pues el MSE es una

parábola como función de a y de b), $\frac{\partial}{\partial a} MSE = \frac{\partial}{\partial b} MSE = 0$, lo que conduce a la solución

$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i & 1 \\ \frac{1}{n} \sum_{i=1}^n x_i^2 & \frac{1}{n} \sum_{i=1}^n x_i \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n y_i \\ \frac{1}{n} \sum_{i=1}^n x_i y_i \end{bmatrix}$$

Con los resultados de Legendre y Laplace, y la hermosa función de DeMoivre, Gauss realizó el siguiente análisis:

Supongamos que $\{X_i, i=1 \dots n\}$ son n medidas erróneas de una cantidad Y , de manera que a cada medida corresponde un error $E_i = Y - X_i$. Supongamos también que todos los errores son independientes e idénticamente distribuidos con $pdf \phi(x)$, que es una función que satisface los postulados de Laplace.

Por independencia, la pdf de la secuencia de errores, $E = \{E_i, i=1 \dots n\}$ es $f(E) = \prod_{i=1}^n \phi(E_i)$ (ver

definiciones 65 y 75). Se quiere que el valor promedio propuesto por Simpson, $\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$, sea el

punto de las mediciones donde se produzca el error más probable, esto es, $\frac{\partial}{\partial \bar{x}} f(E) = 0$ o, lo que es

lo mismo, $\frac{\partial}{\partial \bar{x}} \log[f(E)] = 0$. Expandiendo $f(E)$, la condición necesaria para maximizar la probabilidad del error con el promedio de las mediciones es

$$\frac{1}{\phi(E_1)} \frac{\partial}{\partial \bar{x}} \phi(E_1) + \frac{1}{\phi(E_2)} \frac{\partial}{\partial \bar{x}} \phi(E_2) + \dots + \frac{1}{\phi(E_n)} \frac{\partial}{\partial \bar{x}} \phi(E_n) = 0$$

Mediante la regla de la cadena,

$$\left(\frac{\partial}{\partial \bar{x}} \phi(E_i) = \frac{\partial}{\partial E_i} \phi(E_i) \frac{\partial}{\partial \bar{x}} E_i = \phi'(E_i) \frac{\partial}{\partial \bar{x}} E_i = -\frac{1}{n} \phi'(E_i) \right)$$

Obtenemos

$$\frac{\phi'(E_1)}{\phi(E_1)} + \frac{\phi'(E_2)}{\phi(E_2)} + \dots + \frac{\phi'(E_n)}{\phi(E_n)} = 0$$

Como las medidas X_2, X_3, X_4, \dots aparecen únicamente en la suma $\sum_{i=2}^n X_i$, Gauss consideró la suma como una constante, de manera que los errores son $E_1 = (n-1)d$, $E_i = -d$, $i = 2, \dots, n$ y, por consiguiente,

$$\frac{\phi'((n-1)d)}{\phi((n-1)d)} = (1-n) \frac{\phi'(-d)}{\phi(-d)}$$

de donde Gauss dedujo que para todo $x \in \mathbb{R}$ y algunas constantes a y c ,

$$\frac{\phi'(x)}{\phi(x)} = cx \Rightarrow \phi(x) = ae^{cx^2/2}$$

Para satisfacer las condiciones de Laplace, es necesario que $c < 0$, que en notación de hoy escribimos $c = -1/\sigma^2$. La constante a se obtiene por normalización:

$$\int_{-\infty}^{\infty} \phi(x) dx = 1 \Rightarrow \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)$$

que es la distribución Gaussiana o normal. Resumiendo, la distribución Gaussiana fue diseñada para que el promedio aritmético de las medidas sea el que minimice el error cuadrado promedio, suponiendo errores simétricos. Este modelo resulta muy apropiado para considerar errores de medición con pequeñas varianzas (la ganancia de un jugador de dados, la altura física de los seres humanos, el ruido térmico en una resistencia eléctrica, etc.). Sin embargo, son muchos los ejemplos que se pueden encontrar en los que este modelo diseñado no aplica: el número de personas afectadas por un apagón, el tamaño de los incendios forestales, las cascadas de los eventos de congestión en tráfico aéreo, los impactos de meteoritos, las muertes y pérdidas económicas por desastres naturales o artificiales, las variaciones en el mercado de valores, el uso de las palabras del español, la población de las ciudades, los ingresos y riqueza de las compañías y los individuos, la citación de artículos, las publicaciones por autor, los tamaños de los archivos en un disco duro, la utilización de la CPU en un computador, etc. Todos estos son ejemplos de variables aleatorias cuyas distribuciones exhiben colas pesadas y, por lo tanto, la alta variabilidad no corresponde con las condiciones para las que se diseñó la distribución gaussiana. Nunca podremos subestimar la importancia de esta distribución, como veremos al estudiar, por ejemplo, el teorema del límite central (definición **xx**). Sin embargo, es importante tener en cuenta que es bastante fácil (y bastante común) abusar de ella.

46. Algunos Ejemplos muy Simplificados de Modelos Probabilísticos de Eficiencia en Redes de Comunicaciones Basados en Variables Aleatorias

- (a) Con tráfico tipo Poisson, longitud fija de paquetes y un gran número de usuarios, la eficiencia del protocolo Aloha en la utilización efectiva del enlace es $\rho e^{-2\rho}$, donde ρ es la intensidad de tráfico. Esta eficiencia tiene un valor máximo de 0.184 cuando $\rho=0.5$.
- (b) Bajo las mismas condiciones, la eficiencia del protocolo Aloha ranurado es $\rho e^{-\rho}$, que tiene un valor máximo de 0.368 cuando $\rho=1$.
- (c) Si en el protocolo Aloha ranurado se considera un número finito de usuarios, n , la máxima eficiencia que se puede conseguir es $[(n-1)/n]^{n-1}$, cuando la intensidad de tráfico es 1. Esta eficiencia tiende a 0.368 a medida que n tiende a infinito.
- (d) La máxima eficiencia del protocolo de retransmisión Stop&Wait es $(L(1-BER)^{L+2h})/(L + h + (h+2Ct_p))$, donde L es la longitud (constante) de los paquetes en bits, h es el número de bits en el encabezado que se les añade, y el canal se caracteriza por la tasa de errores, BER, el retardo de propagación, t_p , y la velocidad de transmisión, C .
- (e) Bajo las mismas condiciones, la máxima eficiencia del protocolo de retransmisión GoBack-N es $(L(1-BER)^{L+2h})/(L + h + p(h+2Ct_p))$, donde $p=1-(1-BER)^{L+2h}$ es la probabilidad de que se dañe al menos un bit de una trama o de su reconocimiento. Y la máxima eficiencia del protocolo de retransmisión Selective-Repeat es $(L(1-BER)^{L+2h})/(L + h)$.

Como hemos mencionado, el modelado probabilístico no es un formalismo matemático que se pueda aprender como aprendimos, por ejemplo, los axiomas que definen la probabilidad. Al contrario, es un tipo de arte que sólo se puede llegar a conocer después de haber estudiado muchos ejemplos y haber acumulado mucha práctica, a veces con éxito y a veces no. En los ejemplos sencillos que vienen a continuación notaremos algunos de los aspectos fundamentales de este modelado. Por ejemplo, es necesario conocer muy bien el sistema que se quiere modelar y establecer con claridad y sin ambigüedades el problema que se quiere resolver respecto a dicho sistema. Luego se deben identificar los elementos que participan en el sistema, incluyendo las variables que describen tanto a los elementos como a las interacciones entre ellos. En esta etapa se incluyen todas las restricciones y las suposiciones que se deban hacer para que el modelo sea analíticamente tratable sin alejarlo demasiado de la realidad. Entonces es cuando se aplican los aspectos particulares de las teorías matemáticas apropiadas. Los siguientes pretenden ser ejemplos muy sencillos de este proceso, en los que se desea estimar la eficiencia en el uso de un canal de comunicaciones, ya sea compartido (ejemplos (a), (b), y (c)) o dedicado (ejemplos (d) y (e)).

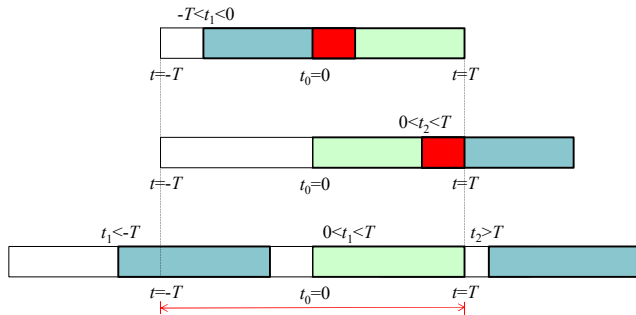
- (a) Muchos medios de comunicación no conectan dos puntos de manera exclusiva, sino que deben compartirse entre diferentes usuarios, por lo que se conocen como medios de acceso múltiple (redes satelitales, algunas redes de área local, redes de radio, etc.). Para controlar el acceso a este tipo de medios existe la capa MAC (Medium Access Control), para la cual se han desarrollado diferentes tipos de protocolos que van desde algunos perfectamente organizados como *round-robin TDM* (en el que cada nodo tiene una ranura de tiempo exclusiva para sus transmisiones) hasta otros completamente aleatorios como *Aloha* (en el que cada nodo envía sus paquetes en el momento en que se generan, con la esperanza de que no “colisionen” con las transmisiones de otros paquetes). Este último, el más simple de todos los protocolos de acceso múltiple, se adoptó en la Universidad de Hawaii en la década de los 70’s y es la base de la mayoría de protocolos de acceso aleatorio más utilizados en diferentes tipos de redes. El protocolo se basa en que los usuarios, que transmiten un paquete cada vez que desean, puedan detectar si la interferencia generada por otros nodos afectó la recepción de su paquete, en cuyo caso esperan un tiempo aleatorio antes de reintentar la retransmisión del mismo paquete.

Supongamos las siguientes condiciones: El enlace tiene una capacidad de C bps y lo comparte un gran número de usuarios, cada uno participando con una fracción muy pequeña del tráfico total, de manera que la suma de las transmisiones nuevas y las retransmisiones forma un proceso de Poisson con un promedio de λ paquetes por segundo. Todos los paquetes tienen la misma longitud, L , y el mismo tiempo de transmisión, $T = L/C$. Nos preguntamos por la eficiencia con que se puede explotar el enlace, donde definimos la eficiencia como la tasa efectiva de paquetes por segundo que puede transmitir el canal, normalizada por la máxima tasa que se podría lograr si no hubieran colisiones y el enlace permaneciera ocupado transmitiendo paquetes,

$$\text{Eficiencia} = \frac{\text{Tasa efectiva}}{\text{Tasa ideal}} = \frac{\lambda P[\text{éxito}]}{C/L}$$

En efecto, si el medio puede permanecer el 100% del tiempo transmitiendo paquetes, lograría transmitir C/L paquetes por segundo. Pero, en realidad, de los λ paquetes que en total producen los usuarios, sólo transmite efectivamente aquellos que no sufren colisiones, esto es, aquellos que tienen éxito, $\lambda P[\text{éxito}]$. Como el tiempo de transmisión es $T=L/C$, la forma que toma la expresión anterior para la eficiencia es $E = \rho P[\text{éxito}]$, donde $\rho = \lambda T$ es la intensidad de tráfico. Así pues, sólo falta determinar la probabilidad de que un paquete no sufra colisiones, $P[\text{éxito}]$.

Para que un paquete tenga éxito es necesario que nadie haya empezado a transmitir durante los T segundos anteriores al inicio de su transmisión (con eso él no colisiona con nadie que ya haya estado transmitiendo) ni durante los T segundos que dura su transmisión (con eso nadie colisiona con él durante su transmisión). Esto es, para que un paquete tenga éxito, hay un intervalo de $2T$ segundos durante los cuales no debe iniciarse la transmisión de ningún otro paquete: $P[\text{éxito}] = P[0 \text{ paquetes inicien transmisión en } 2T \text{ segundos}]$ (ver Figura 61).



- a) Para que un paquete tenga éxito es necesario que nadie haya empezado a transmitir durante los T segundos anteriores al inicio de su transmisión (con eso él no colisiona con nadie que ya haya estado transmitiendo)
- b) Para que un paquete tenga éxito es necesario que nadie haya empezado a transmitir durante los T segundos que dura su transmisión (con eso nadie colisiona con él durante su transmisión).
- c) Esto es, para que un paquete tenga éxito, hay un intervalo de $2T$ segundos durante los cuales no debe iniciarse la transmisión de ningún otro paquete

Figura 61. Para que un paquete que inicia transmisión en $t_0=0$ tenga éxito es necesario que nadie más intente empezar a transmitir en el intervalo $[-T, T]$

De acuerdo con la suposición de tráfico Poisson,

$$P[\text{éxito}] = [(\lambda 2T)^0 / 0!] e^{-\lambda 2T} = e^{-2\rho}$$

En consecuencia, la eficiencia del protocolo Aloha bajo las condiciones mencionadas depende de la intensidad de tráfico, ρ , así:

$$\text{Eficiencia} = \rho e^{-2\rho}.$$

Esta eficiencia, que se grafica en la Figura 62, tiene un máximo en el punto en el que $\frac{d}{d\rho} \text{Eficiencia} = (1 - 2\rho)e^{-2\rho} = 0$, que corresponde a $\rho = 1/2$, donde la eficiencia toma el valor $e^{-1/2} = 0.184$.

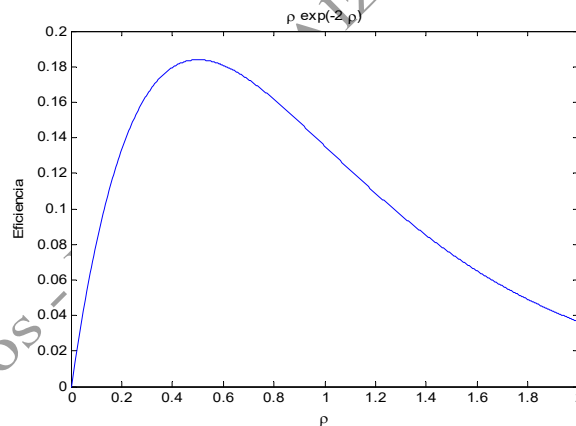


Figura 62. Eficiencia del protocolo Aloha como función de la intensidad de tráfico

- (b) Supongamos que, en el ejemplo anterior, hacemos que una estación emita una señal cada $T=L/C$ segundos indicando que quienes quieran transmitir pueden empezar a hacerlo. Cuando un usuario quiera transmitir un paquete, en vez de hacerlo inmediatamente, espera a la señal de temporización. Esto es, el canal se ha dividido en franjas de tiempo (slots) que supondremos exactamente iguales al tiempo de transmisión de un paquete, T . De esta manera, para que un paquete tenga éxito, es necesario que nadie haya decidido empezar a transmitir durante el período de T segundos anteriores al inicio de su transmisión: $P[\text{éxito}] = [(\lambda T)^0 / 0!] e^{-\lambda T} = e^{-\rho}$. La eficiencia es, entonces, $E = \rho P[\text{éxito}] = \rho e^{-\rho}$, que se maximiza cuando $\frac{d}{d\rho} \text{Eficiencia} = (1 - \rho)e^{-\rho} = 0$, es decir, cuando $\rho=1$, en cuyo caso la eficiencia es $e^{-1}=0.368$, como muestra la Figura 63. Este es el protocolo Aloha ranurado.

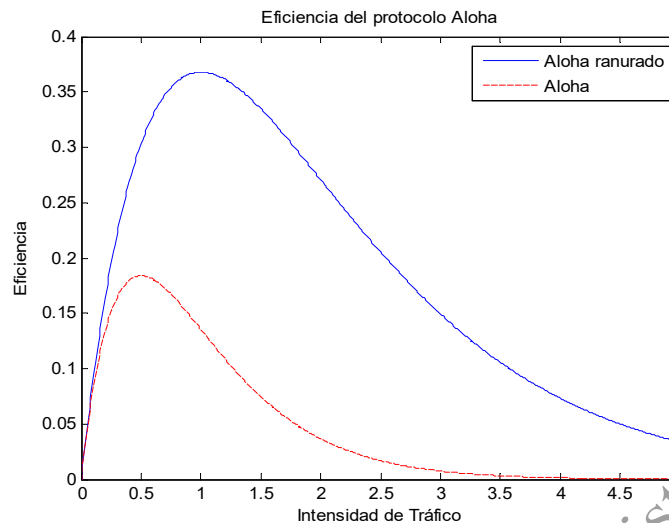


Figura 63. Eficiencia del protocolo Aloha ranurado

- (c) En los dos ejemplos anteriores supusimos un número infinito de usuarios, cada uno participando con un infinitésimo del tráfico. ¿Qué pasa si consideramos un número finito, n , de usuarios? Por supuesto, debemos considerar otro modelo de tráfico. Por ejemplo, si el canal se divide en ranuras de tiempo, podemos suponer que cada usuario genera un paquete en cada slot con probabilidad p , independientemente entre usuarios e independientemente entre slots. En este caso, en cada slot pueden ocurrir tres casos diferentes: Que el slot se pierda porque nadie lo use, que el slot se aproveche efectivamente porque sólo un usuario decide usarlo, o que el slot se pierda porque más de un usuario lo quiere usar y se produzca una colisión. Sea N el número de usuarios que deciden transmitir en un slot. Como vimos, N tiene una distribución binomial con parámetros (n, p) . La eficiencia es la fracción de tiempo que aprovechamos el canal para la transmisión efectiva de paquetes, esto es, la fracción de slots que se aprovechan exitosamente. En nuestra interpretación frecuentista, esta fracción de slots tiende exactamente a la probabilidad de que un slot se aproveche correctamente: $E = P[N=1] = \binom{n}{1}p(1-p)^{n-1} = np(1-p)^{n-1}$. Esta es una función cóncava en p , así que para encontrar el valor de p que maximiza la eficiencia basta con igualar la derivada de E respecto a p a cero y despejar p :

$$\frac{d}{dp} E = n(1-p)^{n-1} - n(n-1)p(1-p)^{n-2} = n(1-p)^{n-2}(1-np) = 0$$

Que tiene dos soluciones, $p=1/n$ y $p=1$, de las cuales la de interés para nosotros es $p=1/n$.

Remplazando este valor de p en la expresión de la eficiencia, obtenemos la eficiencia máxima,

$$E = [(n-1)/n]^{n-1}.$$

Nótese que esta expresión se basa en suponer un modelo binomial de tráfico, mientras que la expresión anterior para Aloha ranurado surgió de suponer un modelo Poisson. Sin embargo, si en el modelo binomial hacemos tender n a infinito y p a cero de manera que np permanezca igual a la intensidad de tráfico deseada, obtenemos un modelo de Poisson. En efecto, si hacemos $n \rightarrow \infty$ con $np=1$, obtenemos $[(n-1)/n]^{n-1} \rightarrow e^{-1} = 0.368$, como corresponde al modelo anterior de aloha ranurado.

- (d) Consideremos un enlace dedicado punto a punto caracterizado por una capacidad de transmisión, C bits/segundo, una tasa de errores de bit, BER , y un tiempo de propagación, t_p segundos. Sobre ese enlace enviamos paquetes de L bits desde el extremo transmisor hasta el extremo receptor. Supongamos que usamos un código detector de errores (CRC, por ejemplo) capaz de detectar todos los posibles errores (!). Dicho código, junto con el campo de numeración de secuencia de paquetes y otros campos de control, requieren un encabezado de h bits para cada paquete, con lo que se construye una trama de $L+h$ bits.

Una forma simple y efectiva de corregir los errores de transmisión es a través de los protocolos de solicitud automática de retransmisión (*ARQ*): el módulo receptor detecta las tramas con errores y solicita automáticamente la retransmisión del paquete correspondiente. El protocolo *ARQ* más simple de todos se denomina *Stop&Wait* y se basa en la idea de que no se debe transmitir ningún nuevo paquete hasta no estar completamente seguro que el paquete anterior llegó correctamente. Para esto, por cada trama recibida correctamente, el nodo receptor devuelve un reconocimiento positivo indicando que recibió bien el último paquete y espera el siguiente paquete. Este reconocimiento se convierte en un permiso para que el nodo transmisor transmita el siguiente paquete (Figura 64(a)). Si algún paquete llega con errores, el nodo receptor devuelve un reconocimiento negativo solicitando la retransmisión del último paquete (Figura 64(b)). Y, si después de un tiempo prudente, t_{out} , el transmisor no recibe ningún reconocimiento positivo ni negativo, supone que hubo un error y retransmite el último paquete enviado (Figura 64(c)). Los reconocimientos son tramas de control que no contienen ningún paquete y, por lo tanto, sólo tienen h bits de longitud.

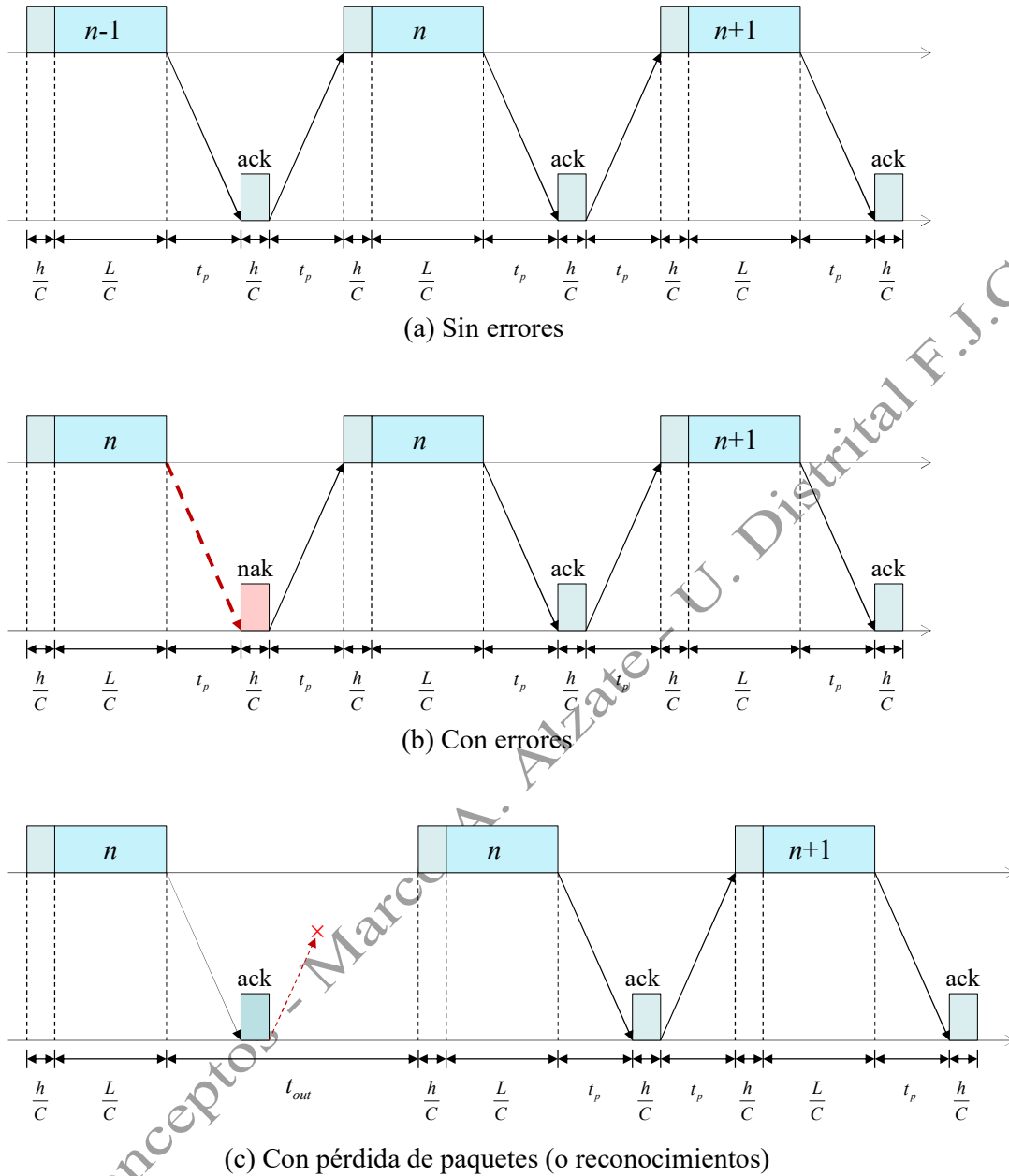


Figura 64. Secuencia de eventos en el protocolo *Stop&Wait*

Claramente, es necesario que t_{out} sea significativamente mayor que $2t_p + h/C$ para evitar retransmisiones innecesarias. Sin embargo, para encontrar la máxima eficiencia de este protocolo, supondremos que t_{out} es exactamente igual a $2t_p + h/C$. De esta manera, una transmisión, exitosa o no, toma un tiempo $t_1 = (L+2h)/C + 2t_p$. La probabilidad de que no haya errores es $(1-BER)^{L+2h}$, suponiendo que la presencia de errores en un bit es independiente de la presencia de errores en los bits vecinos (lo cual no puede ser cierto en la mayoría de casos, pero es una libertad que nos tomamos para mantener el modelo tratable). Entonces, la probabilidad de que una trama se dañe y sea necesario retransmitirla es $p=1 - (1-BER)^{L+2h}$ y, como se vio en la

definición 43(b), se necesitan en promedio $1/(1-p)$ transmisiones para que una trama llegue correctamente. En consecuencia, el tiempo promedio que toma la transmisión correcta de una trama es $t_T = t_1/(1-p)$. Si pudiéramos transmitir bits de usuario a la máxima velocidad del canal, resulta que hubiéramos enviado Ct_T bits en el tiempo en el que, en realidad, solamente enviamos L bits. Por lo tanto, la eficiencia de este protocolo es $E = L/Ct_T$. Multiplicando arriba y abajo por $L+h$ y expandiendo los términos de t_T y p , encontramos la siguiente expresión para la eficiencia del protocolo *Stop&Wait*:

$$E_{S\&W} = \frac{L}{L+h} \cdot (1-BER)^{L+2h} \cdot \frac{L+h}{L+h+(h+2t_p C)}$$

Esta expresión está compuesta por tres términos: El primero, $L/(L+h)$, se refiere a la reducción en eficiencia debida al encabezado, para la cual no se puede hacer nada más que reducir al máximo la longitud del encabezado que se añade en el nivel de enlace a cada paquete del nivel de red. El segundo término, $(1-BER)^{L+2h}$, corresponde a la reducción en eficiencia debida a los errores de transmisión, para la cual no se puede hacer nada más que reducir el BER mediante la selección apropiada de las técnicas de modulación y codificación a nivel físico. El término que queda es la reducción en eficiencia debida propiamente al protocolo *Stop&Wait* y se caracteriza por un parámetro muy importante en el análisis de desempeño de los sistemas de comunicación: El “*bandwidth-delay product*”, $2t_p C$, que indica cuántos bits podríamos transmitir durante el tiempo de propagación. Con el propósito de ver el efecto de este parámetro, consideremos 5 tipos de enlaces, así:

Tipo de enlace	Tiempo de propagación (Distancia, velocidad)	Capacidad	Tasa de errores
Línea telefónica	5 μ s (1 km a 2×10^8 m/s)	56 Kbps	10^{-5}
ADSL	5 μ s (1 km a 2×10^8 m/s)	512 Kbps	10^{-5}
LAN inalámbrica	1 μ s (300 m a 3×10^8 m/s)	10 Mbps	10^{-4}
Fibra óptica	5 ms (100 km a 2×10^8 m/s)	1 Gbps	10^{-6}
Satélite	240 ms (72000 km a 3×10^8 m/s)	100 Mbps	10^{-5}

Suponiendo un encabezado de 64 bits, podemos graficar la eficiencia de *Stop&Wait* para cada uno de esos cinco enlaces, en función de la longitud del paquete, como muestra la Figura 65

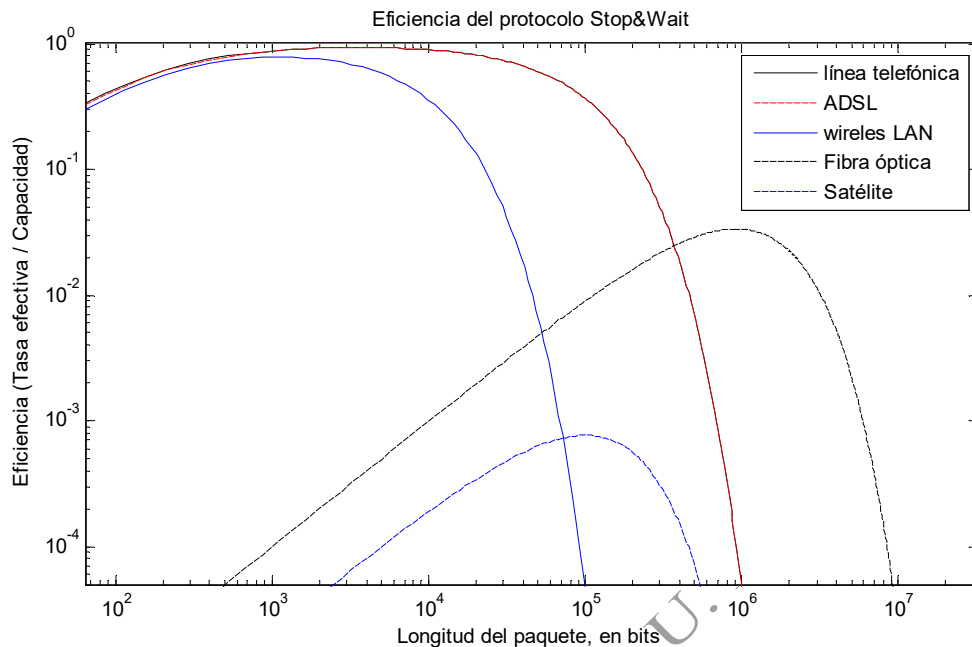
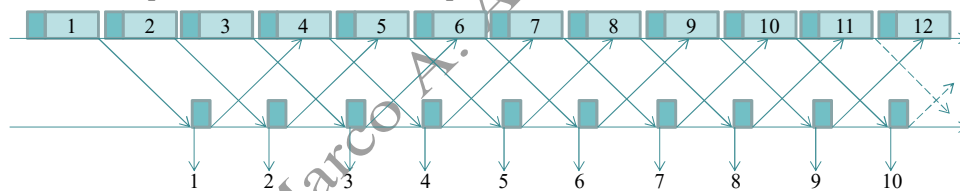


Figura 65. Eficiencia del protocolo *Stop&Wait* para los cinco enlaces de ejemplo

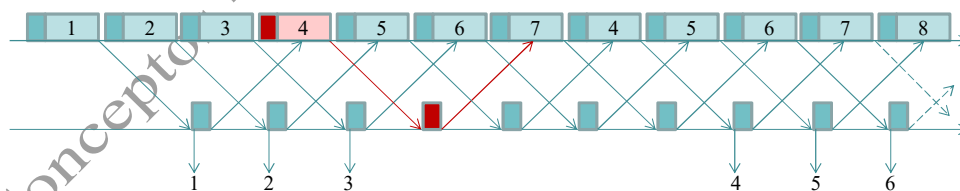
En cada caso hay una longitud óptima de paquete, por debajo de la cual dominan la ineficiencia debida al encabezado (si $2t_pC$ es pequeña comparada con el encabezado) o la ineficiencia debida al protocolo (si $2t_pC$ es grande comparada con el encabezado), y por encima de la cual domina la ineficiencia debida a los errores. Para paquetes entre 2500 y 5000 bits de longitud, la eficiencia del protocolo *Stop&Wait* puede ser hasta del 93% en la línea telefónica y la línea ADSL, en las que $2t_pC$ es de sólo algunos pocos bits (línea ADSL) o una fracción de bit (línea telefónica). En el enlace *wireless* LAN, la máxima eficiencia que se alcanza es cercana al 80% con paquetes cercanos a 1000 bits, donde $2t_pC$ es de algunas decenas de bits. Pero en el enlace de fibra óptica se alcanzaría a transmitir 10 millones de bits durante el tiempo de propagación, mientras que a través del satélite se alcanzaría a transmitir 48 millones de bits. Con paquetes de datos de semejante longitud, la probabilidad de error es casi uno. Por eso en el canal de fibra óptica apenas se alcanza una máxima eficiencia del 3% con paquetes de un millón de bits, y en el canal satelital la máxima eficiencia alcanzable es de menos del 0.1% con paquetes de cerca de cien mil bits. Claramente, para enlaces con un gran *bandwidth-delay product* se hace necesario diseñar protocolos más apropiados.

- (e) Puesto que la ineficiencia del protocolo *stop&wait* (cuando el *bandwidth-delay product*, $2t_pC$, no es un número despreciable de bits) se debe a que durante los tiempos de propagación el canal se mantiene ocioso, una alternativa sería permitir que el transmisor siguiera transmitiendo nuevos paquetes mientras espera un reconocimiento. Sin embargo, ante la presencia de un error, el transmisor debe saber cuáles tramas debe reenviar, por lo cual debe mantener un buffer con los paquetes que ya ha transmitido pero que no han sido reconocidos por el receptor. Cada reconocimiento positivo libera un espacio en el buffer (porque ya existe seguridad de que el paquete correspondiente llegó correctamente) y cada paquete transmitido ocupa un espacio en el buffer (porque debemos tener disponible ese paquete por si llegara a hacer falta retransmitirlo).

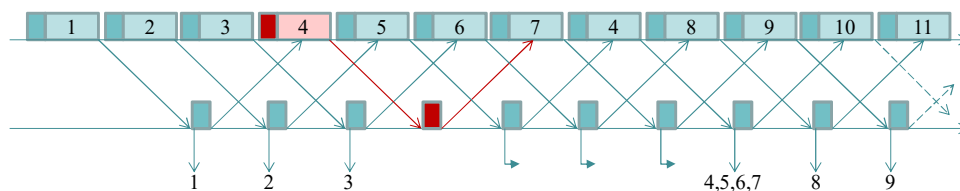
Así pues, el buffer se maneja como una “ventana deslizante” (*sliding window*). El manejo de este buffer aumenta la complejidad del protocolo en el transmisor pero permite aprovechar mejor los tiempos de propagación. En el receptor podemos hacer dos cosas: (1) Aceptar sólo los paquetes que llegan en orden de manera que, si llega un paquete fuera de secuencia, se descarta y se envía un reconocimiento negativo. (2) Aceptar todos los paquetes que lleguen sin errores de manera que, si llega un paquete fuera de secuencia, lo almacena para entregarlo posteriormente al nivel superior, cuando haya completado todos los paquetes intermedios. En el primer caso, cada paquete que el receptor acepta puede entregarlo inmediatamente al nivel superior, por lo que no necesita ninguna ventana deslizante que almacene los paquetes recibidos fuera de orden, aunque, con cada error, el transmisor debe volver a enviar la ventana entera (*GoBack-N*). En el segundo caso, como al nivel superior se deben entregar los paquetes en orden, aquellos que lleguen fuera de secuencia deben almacenarse en un buffer local, con el fin de entregarlos al nivel superior cuando se complete una secuencia ordenada de paquetes. Por eso el receptor también necesita un buffer para almacenar los paquetes recibidos que no ha podido entregar al nivel superior, el cual se maneja como una ventana deslizante que se abre con cada paquete que entrega al nivel superior y se cierra con cada paquete que acepta en desorden. La ventaja es que, en caso de error, en vez de retransmitir toda la ventana, el transmisor sólo debe retransmitir el paquete en problemas y continuar donde iba (*SelectiveRepeat*). El manejo del buffer en el receptor aumenta la complejidad del protocolo, pero permite aprovechar las transmisiones exitosas que se hayan hecho entre el paquete que se dañó y la notificación de dicho error en el transmisor. La Figura 66 muestra el comportamiento de estos dos protocolos.



(a) *GoBack-N* y *SelectiveRepeat* sin errores de transmisión



(b) *GoBack-N* con un error de transmisión



(c) *SelectiveRepeat* con un error de transmisión

Figura 66. Secuencias de eventos en los protocolos *Go-Back-N* y *SelectiveRepeat*

En *GoBack-N*, una transmisión con error toma $t_a = (L+2h)/C + 2t_p$ mientras que una transmisión exitosa sólo toma $t_b = (L+h)/C$. Como la probabilidad de error es $p=1 - (1-BER)^{L+2h}$, de las $1/(1-p)$ transmisiones que toca hacer en promedio hasta que un paquete llegue bien, una corresponde a la transmisión correcta y las otras $p/(1-p)$ corresponden a transmisiones con errores, de manera que el tiempo promedio que toma la transmisión de un paquete hasta que llegue correctamente a su destino es $t_T = t_b + pt_a/(1-p) = [L+h+p(h+2t_pC)]/[C(1-p)]$. Entonces, en el tiempo en que se podrían transmitir $t_T C$ bits de usuario, sólo se transmiten L bits, por lo que la eficiencia es $L/t_T C$, ó

$$E_{GB-N} = \frac{L}{L+h} \cdot (1-BER)^{L+2h} \cdot \frac{L+h}{L+h+p(h+2t_pC)}$$

Comparando con la eficiencia de *Stop&Wait*, los dos primeros términos no cambian pues los errores y el encabezado son inevitables, pero lo que si logramos hacer es que la ineficiencia debida al *bandwidth-delay product* sólo se haga presente cuando haya errores de transmisión, esto es, con probabilidad $p=1 - (1-BER)^{L+2h}$. Con respecto a *SelectiveRepeat*, tanto las transmisiones correctas como las transmisiones con errores toman un tiempo $t = (L+h)/C$, por lo que la el término en la eficiencia debido al protocolo es uno:

$$E_{SR} = \frac{L}{L+h} \cdot (1-BER)^{L+2h}$$

La figura 2.29 muestra la eficiencia de *GoBack-N* y de *SelectiveRepeat*, en función de la longitud del paquete, para los mismos enlaces de la figura 2.27.

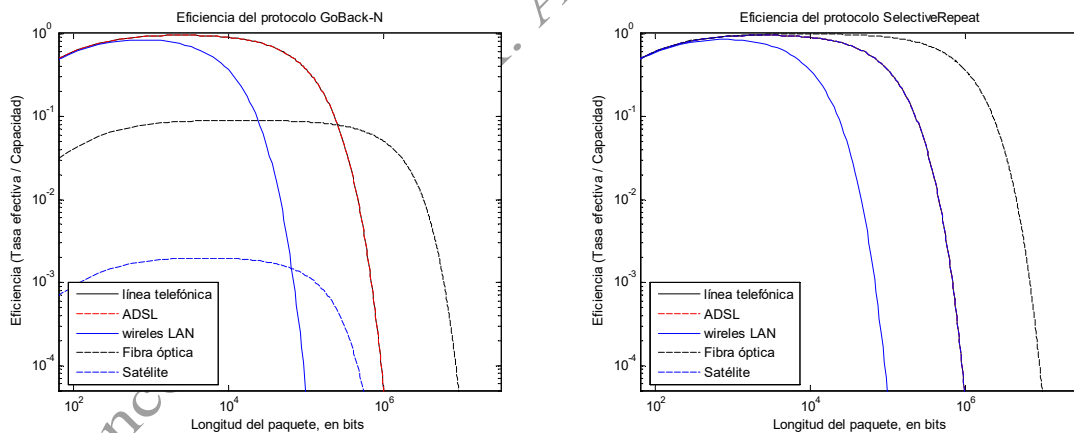


Figura 67. Eficiencia de los protocolos *GoBack-N* y *SelectiveRepeat*

Nótese cómo la diferencia en las expresiones de eficiencia para los tres protocolos se encuentra en la manera como afecta el término asociado con el *bandwidth-delay-product*, $2t_pC+h$. En *Stop&Wait* aparece multiplicado por uno, ya que su efecto se sufre en todas las transmisiones. En *GoBack-N* aparece multiplicado por p , ya que su efecto se sufre sólo en presencia de errores, que ocurren con probabilidad p . Y en *SelectiveRepeat* aparece multiplicado por cero, ya que su efecto nunca se sufre.

47. Desigualdades en la Cola de una Distribución

Sea X una v.a. cualquiera. A las probabilidades de la forma $P[X \geq a]$ ó $P[|X| \geq a]$ se les conoce como “Cola de la Distribución de X ”. Estas colas se pueden acotar sin necesidad de calcularlas exactamente:

(a) *Desigualdad de Markov:* Sea X una variable aleatoria no negativa con valor esperado $E[X] < \infty$. Para cualquier $\alpha > 0$, se cumple que

$$P[X > \alpha] \leq \frac{E[X]}{\alpha}$$

(b) *Desigualdad de Chebyshev:* Sea X una variable aleatoria con valor esperado $E[X] < \infty$ y varianza $V[X] < \infty$. Para cualquier $\alpha > 0$, se cumple que

$$P[|X - E[X]| > \alpha] \leq \frac{V[X]}{\alpha^2}$$

(c) *Cota de Chernov:* Sea X una variable aleatoria. Para cualquier $\alpha > 0$, se cumple que

$$P[X > \alpha] \leq \min_{s>0} e^{-\alpha s} E[e^{sX}]$$

Entre sus muchas aplicaciones, las cotas de la cola de una distribución resultan muy útiles para determinar garantías de calidad de servicio en redes con servicios diferenciados. Por ejemplo, una medida de calidad de servicio indicaría que menos del 1% de los paquetes sufrirá un retardo superior a 100 ms; esto es, si D es la variable aleatoria que representa el retardo sufrido por un paquete, la medida de calidad de servicio está dada como una cota en la cola de su distribución: $P[D \geq 0.1] \leq 0.01$. Por esta razón, las cotas en la cola de la distribución de una v.a. resultan de gran importancia, al punto que un desarrollo teórico tan fundamental en ingeniería de redes con QoS como la “Capacidad equivalente” de una fuente de tráfico, por ejemplo, se basa en las cotas de Chernov.

Resulta fácil verificar la validez de estas cotas. En efecto, para una variable aleatoria no negativa, podemos partir de la definición misma del valor esperado, $E[X]$, así:

$$\begin{aligned} E[X] &= \int_0^{\infty} x dF_X(x) = \int_0^{\alpha} x dF_X(x) + \int_{\alpha}^{\infty} x dF_X(x) \\ &\geq \int_{\alpha}^{\infty} x dF_X(x) \geq \alpha \int_{\alpha}^{\infty} dF_X(x) = \alpha P[X > \alpha] \end{aligned}$$

que es la desigualdad de Markov. Esta desigualdad es muy interesante pues, por ejemplo, permite ver que para cualquier variable aleatoria no negativa con valor esperado finito, la probabilidad de que la variable exceda n veces su valor esperado siempre será menor o igual a $1/n$.

Ahora, para cualquier variable X con valor medio y varianza finitos, podemos construir la variable no negativa $(X - E[X])^2$, cuyo valor esperado es la varianza de X , y aplicarle la desigualdad de Markov, con lo que obtenemos la desigualdad Chebyshev:

$$P[(X - E[X])^2 \geq \alpha^2] \leq \frac{V[X]}{\alpha^2} \Rightarrow P[|X - E[X]| \geq \alpha] \leq \frac{V[X]}{\alpha^2}$$

La cota de Chernov tiene una gran aplicabilidad en la teoría de las grandes desviaciones, de donde deriva gran parte de la formalidad de la ingeniería de redes con calidad de servicio. Para verificarla podemos partir de una relación muy general,

$$e^{s(x-a)} \geq u(x-a), \quad s > 0$$

donde $u(x)$ es el escalón unitario, igual a cero para valores negativos de x e igual a 1 para otros valores de x , como se aprecia en la Figura 68.

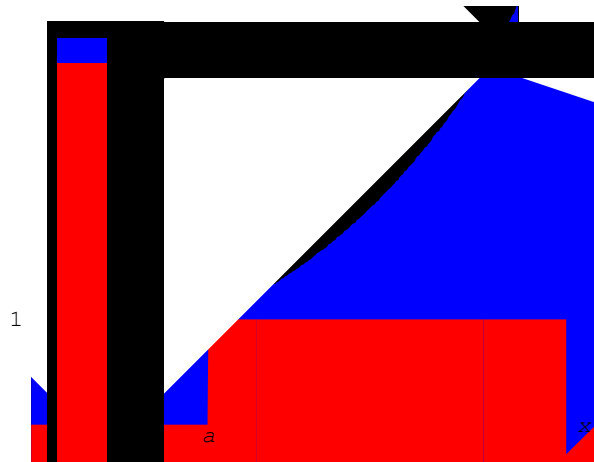


Figura 68. $e^{s(x-a)} \geq u(x)$ para cualquier $s > 0$

De donde podemos verificar las siguientes expresiones:

$$P[X > a] = \int_a^\infty dF_X(x) = \int_{-\infty}^\infty u(x-a) dF_X(x) \leq \int_{-\infty}^\infty e^{s(x-a)} dF_X(x) = e^{-as} E[e^{sX}]$$

Como dicha desigualdad es válida para cualquier valor no-negativo de s , podemos minimizar con respecto a $s > 0$ para obtener la cota de Chernov.

Aunque tendremos oportunidad de usar estas desigualdades recurrentemente, es interesante mostrar un ejemplo sencillo para notar su poder y sus limitaciones: Se quiere asegurar que, en promedio, de cada mil paquetes no más de uno experimente un retardo superior o igual a 100 ms. ¿Cuánto debe ser el promedio del retardo de los paquetes para garantizar esta condición? Como el retardo D es una v.a. no negativa, aplica la desigualdad de Markov según la cual $P[D \geq 0.1] \leq 10E[D] = 0.001$, por lo que podemos ofrecer la garantía de retardo máximo si diseñamos para un retardo promedio de 0.1 ms, independientemente de la distribución de D . Pero si el promedio resulta ser, digamos, de 0.2 ms, ¿Cuánto debe valer la desviación estándar para garantizar la misma condición? Según la desigualdad de Chebyshev, $P[D \geq 0.1] = P[D - 0.0002 \geq 0.0998] = P[|D - 0.0002| \geq 0.0998] \leq V[D]/0.0998^2 = 0.001$, por lo que necesitamos que la desviación estándar no supere los 3.2 ms, independientemente de la distribución de D (nótese que la positividad del retardo nos permitió la segunda igualdad, pues $|D - 0.0002| \geq 0.0998$ significa que $D \geq 0.1$ o que $D \leq -0.0996$, pero la segunda desigualdad es imposible por la positividad del retardo).

Claramente, estas cotas pueden ser muy poco estrictas, en la medida en que el servicio que realmente estemos ofreciendo puede ser mucho mejor a la garantía que ofrecemos. Si conocemos la distribución del retardo, podríamos hacer algún diseño más eficiente usando cotas de Chernov. Por ejemplo, si

sabemos que el retardo obedece a una distribución normal²⁶ con media μ y desviación estándar $\mu/3$, basta con tener un retardo promedio inferior a 44.7 ms para asegurar que la probabilidad de superar los 100 ms no sea mayor que 1/1000. En efecto, la cota de Chernov resulta ser $\exp(s(\mu(90+5\mu s)-9)/90)$, que se minimiza con $s=9(1-10\mu)/10\mu^2$, lo que conduce a una cota mínima igual a $\exp(-9(10\mu-1)^2/(200\mu^2))$, que es igual a 1/1000 cuando $\mu=0.0447$.

Para este último caso podemos calcular exactamente la probabilidad de superar los 100 ms de retardo, pues para variables $N(\mu, \sigma^2)$ se puede calcular fácilmente la cola de la distribución mediante la función $Q[(x-\mu)/\sigma] = 1 - F_X(x)$, ampliamente tabulada o fácilmente calculable numéricamente: Para un retardo promedio de 44.7 ms y una desviación estándar de 14.9 ms, la probabilidad de superar los 100 ms es $0.000103 < 0.001$. En este caso, aún la cota de Chernov resulta ser poco estricta.

Otro caso interesante es el de la variable $X \sim \text{binomial}(n, p)$, para la cual $E[X] = np$, $V[X] = np(1-p)$ y

$$E[e^{sX}] = \sum_{k=0}^n e^{sk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (pe^s)^k (1-p)^{n-k} = (pe^s + (1-p))^n = (1-p(1-e^s))^n$$

de manera que

$$\frac{d}{ds} e^{-as} E[e^{sX}] = e^{-as} [1-p(1-e^s)]^{n-1} [(n-a)pe^s - a(1-p)]$$

Para que la anterior derivada sea cero se necesita $as \rightarrow \infty$, $e^s = -(1-p)/p$ ó $e^s = a(1-p)/(p(n-a))$. El único resultado aceptable es el tercero, en cuyo caso obtenemos que

$$\min_{s>0} e^{-as} E[e^{sX}] = \left[\frac{a(1-p)}{p(n-a)} \right]^{-a} \left[\frac{n(1-p)}{n-a} \right]^n$$

¿Cuál es la probabilidad de que una variable binomial supere dos veces su valor medio? De acuerdo con la cota de Markov,

$$P[X > 2np] \leq np/(2np) = 0.5$$

La cota de Chebyshev dice que

$$P[|X - np| > np] = P[X - np > np] + P[X - np < -np] = P[X > 2np] \leq np(1-p)/(np)^2 = (1-p)/np$$

Y la cota de Chernov dice que

$$P[X > 2np] \leq 2^{-2np} [(1-p)/(1-2p)]^{n(1-2p)}$$

Por ejemplo, si transmitimos un paquete de 10^5 bits en un enlace de radio con $\text{BER} = 10^{-4}$, en promedio se dañarán 10 bits con una desviación estándar de 3.162 bits. ¿Cuál es la probabilidad de que se dañen más de 20 bits?

Markov : $P[\text{Se dañen más de 20 bits}] \leq 0.5$

Chebyshev : $P[\text{Se dañen más de 20 bits}] \leq 0.09999$

Chernov : $P[\text{Se dañen más de 20 bits}] \leq 0.021$

La cota de Markov, claro, es correcta. Pero es casi 5 veces mayor que la cota de Chebyshev, la cual es, a su vez, casi 5 veces mayor que la cota de Chernov. Pero aún la cota de Chernov es casi 13 veces mayor a la probabilidad de que se dañen más de 20 bits, que es 0.0016.

²⁶ Para una variable $X \sim N(\mu, \sigma^2)$ no es difícil encontrar que $E[e^{sX}] = \exp(\mu s + \sigma^2 s^2/2)$. Ver definición 53.

Es oportuno hacer en este momento una primera justificación de la interpretación de la probabilidad que se prefiere en la ingeniería: el límite de la frecuencia relativa. Sea un espacio de probabilidad (Ω, \mathcal{F}, P) donde A es un evento con probabilidad p . Como cada ejecución del experimento aleatorio para ver si ocurre o no ocurre el evento A es un experimento independiente de Bernoulli, la distribución de probabilidad de la frecuencia relativa $f_N(A)$ es semejante a la binomial:

$$P\left[f_N(A) = \frac{k}{N}\right] = \binom{N}{k} p^k (1-p)^{N-k}$$

De manera que $E[f_N(A)] = p$ y $V[f_N(A)] = p(1-p)/N$. Entonces la cota de Chebyzhev nos dice que

$$P\left[|f_N(A) - p| > \varepsilon\right] \leq \frac{p(1-p)}{N\varepsilon^2}$$

Como $p(1-p)$ se maximiza con $p=0.5$, en cuyo caso $p(1-p)$ es $1/4$, tenemos un resultado muy interesante:

$$P\left[|f_N(A) - p| > \varepsilon\right] \leq \frac{1}{4N\varepsilon^2} = \delta$$

Así pues, si queremos hacer suficientes experimentos para asegurar que

$$P\left[|f_N(A) - p| \leq \varepsilon\right] > 1 - \delta$$

para dos valores pequeños ε y δ , será suficiente con escoger $N > 1/4\delta\varepsilon^2$. Es de esperar que este resultado nos exija realizar bastantes más experimentos de los que son realmente necesarios, pero podemos estar absolutamente seguros, gracias a Chebyshev, que con este valor de N ya estaremos suficientemente cerca de la probabilidad real (con un error menor a ε) con una alta probabilidad (mayor a $1-\delta$)²⁷.

48. Distribuciones con Cola Pesada

Sea X una v.a. con CDF F . Se dice que X tiene una distribución con cola pesada si $\lim_{x \rightarrow \infty} e^{\mu x} (1 - F(x)) = \infty \quad \forall \mu > 0$, esto es, si el decrecimiento de la cola de la distribución (la probabilidad de que la variable tome valores mayores a x para valores grandes de x , $P[X > x] = 1 - F(x)$) es más lento que exponencial. Como un decrecimiento hiperbólico es más lento que exponencial, a veces el concepto de cola pesada se particulariza al caso en que $1 - F(x)$ toma la forma cx^{-a} cuando $x \rightarrow \infty$, para $0 < a < 2$ y $c > 0$.

Nótese cómo, por ejemplo, la distribución exponencial tiene una cola que decae exponencialmente rápido:

$$\Pr[X > x] = 1 - F(x) = \int_x^\infty \lambda e^{-\lambda u} du = e^{-\lambda x}, \quad x \geq 0$$

mientras que la distribución Gaussiana tiene una cola que decae aún más rápidamente:

²⁷ Expresiones como ésta ($P[|X - \mu| < \varepsilon] > 1 - \delta$) implican que, si X es una estimación de μ , el resultado es aproximadamente correcto con alta probabilidad: PAC (*Probably Approximately Correct*). En 1984 Leslie Valiant hizo de este principio PAC el contexto teórico para *machine learning*.

$$Q(x) \triangleq \Pr[X > x] = 1 - F(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/2} du \leq \frac{1}{2} e^{-x^2/2}, \quad x > 0$$

Sin embargo, considere la distribución de Pareto:

$$\Pr[X > x] = 1 - F(x) = \int_x^{\infty} \frac{a}{b} \left(\frac{b}{u}\right)^{a+1} du = \left(\frac{b}{x}\right)^a, \quad x \geq b$$

que tiene exactamente la forma hiperbólica mencionada en la definición, lo cual implica una forma mucho más lenta de decaer. Con esta distribución podemos ver algunas de las características principales de las distribuciones con cola pesada. Por ejemplo, como se vio en la definición 44(d), nótese que si $a < 2$, la varianza de la distribución es infinita y, si $a < 1$, el valor esperado de la distribución también es infinito. Esta es una característica fundamental de las distribuciones con cola pesada: Una altísima variabilidad. En otras palabras, la cola pesada conduce a valores extremadamente grandes con una probabilidad no despreciable, de manera que al tomar muestras de una distribución así, la mayoría de ellas serán “pequeñas” pero algunas pocas de ella tendrán valores muy grandes. A medida que el parámetro a tiende a uno por la derecha, se reduce la velocidad con que la media muestral tiende al valor esperado.

Podemos verificar la ubicuidad de este fenómeno analizando la longitud de los archivos en nuestro disco duro. Invitamos al lector a que lea las longitudes de todos los archivos en su disco duro y grafique las frecuencias relativas de los eventos $A(k) = \{\text{Un archivo tiene una longitud mayor que o igual a } 1024 \cdot (2^k - 1) \text{ bytes y menor que } 1024 \cdot (2^{k+1} - 1) \text{ bytes}\}$, $k \in \{0, 1, 2, \dots\}$. La Figura 69 muestra el resultado obtenido en el computador portátil del autor. Del espacio ocupado en el disco, la mitad la ocupan los 12118 archivos más grandes (el 5.3% de los archivos) y la otra mitad la ocupan los 218607 archivos más pequeños (el 94.7% de los archivos). Si esta distribución caracteriza los archivos que intercambian por internet, se empezaría a explicar por qué las características del tráfico moderno son tan variables. De aquí la importancia que tiene el estudio de este tipo de variables aleatorias en el modelado de redes de comunicaciones. De hecho, muchas otras medidas en estas redes, tales como la duración de una sesión TCP, la longitud de un período de silencio en una conversación VoIP, o los tiempos de actividad e inactividad de una sesión http, tienen distribuciones con cola pesada. Las conclusiones que se hagan sobre el desempeño de la red con base en muestras de este tipo de variables aleatorias pueden ser equivocadas si no se hace un muy juicioso estudio estadístico para determinar la significancia de los resultados. Otras variables con cola pesada incluyen el número de personas afectadas por un apagón, el tamaño de los incendios forestales, los eventos de congestión en cascada del tráfico aéreo, el impacto de meteoritos en la superficie terrestre, las muertes en desastres naturales, las variaciones del mercado de valores, el uso de las palabras del español, la población de las ciudades, la riqueza de los individuos, la citación de artículos científicos, etc.

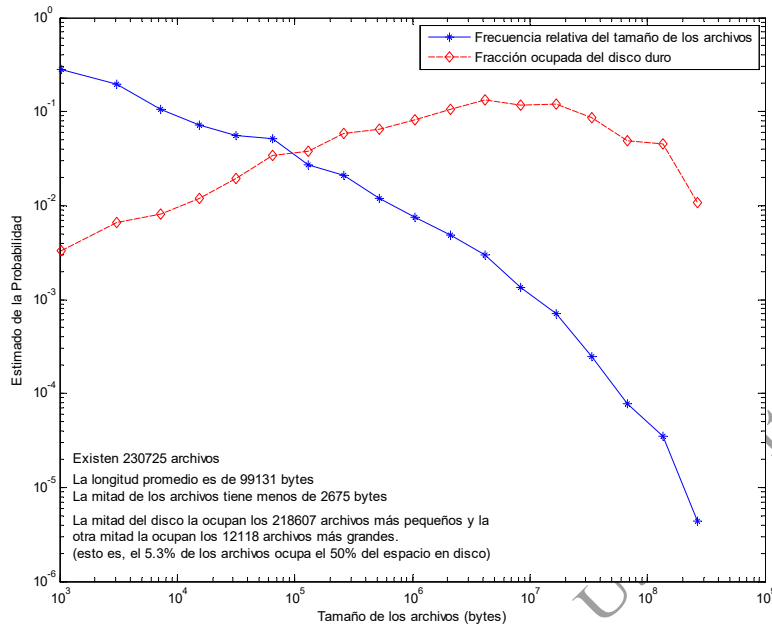


Figura 69. Frecuencia relativa de la longitud de los archivos en un disco duro

49. Distribución Condicional

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad donde se encuentra definida la variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ y el evento $B \in \mathcal{F}$. La función de distribución acumulativa condicional de X dado B es $F_{X|B}(x|B) = \mathbf{P}(\{X \leq x\} \cap B) / \mathbf{P}(B)$. Su derivada es la pdf condicional de X dado B y, si X es discreta, las áreas debajo de los impulsos de la pdf condicional es la pmf condicional de X dado B :

$$f_{X|B}(x|B) = \frac{d}{dt} F_{X|B}(x|B) \qquad p_{k|B} = \frac{P[\{X = x_k\} \cap B]}{P[B]}$$

El evento condicional B bien puede estar definido en términos de la misma variable aleatoria X aunque, en general, puede ser cualquier evento del espacio de probabilidad original, $(\Omega, \mathcal{F}, \mathbf{P})$.

Como se demostró en la definición 18, $F_{X|B}(x|B)$ es una función de distribución acumulativa con las mismas propiedades que cualquier otra CDF:

- (a) La CDF condicional es no-negativa: $F_{X|B}(x|B) \geq 0 \quad \forall x \in \mathbb{R}$
- (b) La CDF condicional es no-decreciente: si $x_1 < x_2$ entonces $F_{X|B}(x_1|B) \leq F_{X|B}(x_2|B)$
- (c) La CDF condicional es acotada: $F_{X|B}(-\infty|B) = 0, F_{X|B}(\infty|B) = 1$.
- (d) La CDF condicional es continua por la derecha: $F_{X|B}(x^+|B) = F_{X|B}(x|B)$.

Lo mismo podemos decir de las funciones condicionales de densidad y distribución de probabilidad:

$$\begin{aligned}
 (a) \quad & f_{X|B}(x|B) \geq 0 \quad \forall x \in \mathbb{R} && p_{k|B} \geq 0 \\
 (b) \quad & F_{X|B}(x|B) = \int_{-\infty}^x f_{X|B}(a|B) da && F_{X|B}(x|B) = \sum_{k: x_k \leq x} p_{k|B} \\
 (c) \quad & \int_{-\infty}^{\infty} f_{X|B}(a|B) da = 1 && \sum_k p_{k|B} = 1
 \end{aligned}$$

En efecto, al condicionar en el evento B sólo se ha restringido el espacio muestral de (Ω, \mathcal{F}, P) al espacio $(B, \mathcal{H} = \mathcal{F} \cap B, Q(\cdot) = P(\cdot|B))$, de manera que los eventos $A(x) = \{\omega \in \Omega : X(\omega) \leq x\}, x \in \mathbb{R}$ se restringen a $A_B(x) = \{\omega \in B : X(\omega) \leq x\}$ con probabilidad $Q(A_B(x)) = P(A(x) \cap B) / P(B) = P(A_B(x)) / P(B) = F_{X|B}(x|B)$.

A manera de ejemplo, supongamos que el evento B es $\{\omega \in \Omega : a < X(\omega) \leq b\}$ o, en nuestra notación simplificada, $B = \{a < X \leq b\}$. ¿Cuánto es $F_{X|B}(x|B)$?

$$F_{X|B}(x|B) = P[X \leq x | a < X \leq b] = \frac{P[X \leq x, a < X \leq b]}{P[a < X \leq b]} = \begin{cases} \frac{P[\Phi]}{F_X(b) - F_X(a)} = 0 & x \leq a \\ \frac{P[a < X \leq x]}{P[a < X \leq b]} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a < x \leq b \\ \frac{P[a < X \leq b]}{P[a < X \leq b]} = \frac{F_X(b) - F_X(a)}{F_X(b) - F_X(a)} = 1 & x > b \end{cases}$$

Derivando respecto a x ,

$$f_{X|a < X \leq b}(x|a < X \leq b) = \begin{cases} 0 & x \leq a \\ \frac{f_X(x)}{\int_a^b f_X(s) ds} & a < x \leq b \\ 0 & x > b \end{cases}$$

o, en el caso de una variable discreta,

$$p_{k|a < X \leq b} = \begin{cases} 0 & x_k \leq a \\ \frac{p_k}{\sum_{a < x_j \leq b} p_j} & a < x_k \leq b \\ 0 & x_k > b \end{cases}$$

Por ejemplo, si lanzamos un dado y sabemos que cayó un número par, ¿Cuál es la distribución de probabilidad del número obtenido? $P[X=k|X \text{ es par}] = P[X=k] / (1/2) = 0.3$ si k es par o cero si k es impar.

Como un último ejemplo, si X es gaussiana con media μ y varianza σ^2 , ¿cuál será su distribución dado que $|X - \mu| \leq k\sigma$? Sabemos que $P[|X - \mu| \leq k\sigma] = P[\mu - k\sigma \leq X \leq \mu + k\sigma] = F_X(\mu + k\sigma) - F_X(\mu - k\sigma)$. Esta es la misma probabilidad de que una variable gaussiana con media cero y varianza uno esté en el intervalo $[-k, k]$. Si llamamos $G(x)$ a la CDF de la variable $\mathcal{N}(0, 1)$, la pdf de X dado $|X - \mu| \leq k\sigma$ es

$$f_x(x|X-\mu \leq k\sigma) = \begin{cases} \frac{\exp(-(x-\mu)^2 / 2\sigma^2)}{\sigma(2G(k)-1)\sqrt{2\pi}} & \text{si } x \in [\mu - k\sigma, \mu + k\sigma] \\ 0 & \text{otro caso} \end{cases}$$

que se conoce como la distribución gaussiana truncada.

50. Momentos condicionales

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad donde se encuentra definida la variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ y el evento $B \in \mathcal{F}$. El n -ésimo momento condicional de la variable aleatoria dado el evento B es $E[X^n | B] = \int_{\mathbb{R}} x^n dF_{X|B}(x|B)$. El n -ésimo momento central condicional de la variable aleatoria dado el evento B es $E[(X - E[X|B])^n | B] = \int_{\mathbb{R}} (x - E[X|B])^n dF_{X|B}(x|B)$, donde $E[X|B]$ es el primero momento condicional de X dado B .

Sea, por ejemplo, una variable aleatoria Gaussiana con media cero y varianza σ^2 , X . ¿Cuáles serán el valor esperado y la varianza condicionales de X dado que $X > 0$? Por la definición 49 sabemos que

$$f_{X|X>0}(x|X > 0) = \begin{cases} 0 & x \leq 0 \\ \frac{f_X(x)}{\int_0^{\infty} f_X(s) ds} & x > 0 \end{cases}$$

donde $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ y $\int_0^{\infty} f_X(s) ds = 1/2$. Esto es,

$$f_{X|X>0}(x|X > 0) = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}, x > 0$$

En consecuencia

$$E[X|X > 0] = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^{\infty} x e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx$$

Esta integral se puede calcular fácilmente haciendo el cambio de variable $y = (x/\sigma)^2/2$ de manera que $dy = x dx / \sigma^2$:

$$E[X|X > 0] = \sigma \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y} dy = \sigma \sqrt{\frac{2}{\pi}}$$

ya que $\int_0^{\infty} e^{-y} dy = -e^{-y} \Big|_0^{\infty} = 1$. De acuerdo con este resultado, podemos calcular

$$V[X|X > 0] = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^{\infty} (x - \sigma \sqrt{2/\pi})^2 e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx$$

$$V[X|X > 0] = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^2 e^{-\frac{1}{2} \left(\frac{x}{\sigma}\right)^2} dx - \sigma^2 \frac{2}{\pi}$$

Como el integrando tiene simetría par con respecto a $x=0$, podemos integrar de $-\infty$ a $+\infty$ y dividir por 2:

$$V[X|X > 0] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2} \left(\frac{x}{\sigma}\right)^2} dx - \sigma^2 \frac{2}{\pi}$$

donde podemos reconocer la varianza de X en el primer término, σ^2 , de donde

$$V[X|X > 0] = \sigma^2 \left(1 - \frac{2}{\pi}\right)$$

51. Memoria de una Distribución

Dada una variable aleatoria X con CDF $F_X(x)$, podemos considerar la distribución condicional de la cola de la distribución, $P[X > x+s | X > s]$, esto es, cuál es la probabilidad de que la variable sea mayor a $x+s$ dado que ya sabemos que es mayor a s . Si esta probabilidad depende de s , se dice que la distribución tiene memoria. En otro caso, la distribución carece de memoria. La única distribución discreta sin memoria es la geométrica. La única distribución continua sin memoria es la exponencial.

Para ver la importancia de esta definición, consideremos por ejemplo tres modelos diferentes para el tiempo de transmisión de un paquete:

Modelo 1: T_1 es una v.a. uniformemente distribuida entre 0 y S : $f_1(t) = \begin{cases} \frac{1}{S} & 0 \leq t \leq S \\ 0 & \text{otro } t \end{cases}$

Modelo 2: T_2 es una v.a. exponencialmente distribuida con parámetro λ : $f_2(t) = \lambda e^{-\lambda t}$, $t \geq 0$

Modelo 3: T_3 es una v.a. con distribución de Pareto, con parámetros a, b : $f_3(t) = \frac{a}{b} \left(\frac{b}{x}\right)^{a+1}$ $t \geq b$

La pregunta que nos haremos es simple: Si el paquete que está en servicio empezó a ser transmitido hace τ segundos, ¿cuánto tiempo le falta para terminar su transmisión? Esta es otra variable aleatoria, el tiempo remanente, cuya distribución es la cola condicional de la distribución original. Esto es, si T es el tiempo de transmisión y el paquete lleva τ segundos siendo transmitido, lo que queremos saber es con qué probabilidad faltan más de t segundos para terminar la transmisión:

$$P[T > t + \tau | T > \tau] = \frac{P[T > t + \tau]}{P[T > \tau]}$$

En el modelo 1 tenemos $P[T > s] = \max(0, \min(1, 1 - s/S))$, de manera que $P[T > t + \tau | T > \tau] = \max(0, \min(1, 1 - t/(S - \tau)))$. Esto es, la distribución del tiempo remanente está uniformemente distribuida entre 0 y $S - \tau$, de manera que entre mayor tiempo hayamos esperado, menor tiempo nos

falta por esperar. De hecho el tiempo promedio de espera se reduce de $S/2$ a $(S-\tau)/2$, de manera que el tiempo que falta se acerca a cero a medida que τ se acerca a S . Esta dependencia entre el tiempo que hemos esperado y el que nos falta por esperar se conoce como la **memoria de la distribución uniforme**, que es negativa porque entre más hemos esperado menos falta por esperar.

En el modelo 2 tenemos que $P[T>s] = e^{-\lambda s}$, de manera que $P[T>t+\tau|T>\tau] = e^{-\lambda(t+\tau)}/e^{-\lambda\tau} = e^{-\lambda t}$: El tiempo que falta por esperar sigue siendo una variable aleatoria exponencial con parámetro λ , independientemente de que ya hayamos esperado 1 milisegundo, 1 minuto ó 1 año. De hecho, originalmente el tiempo promedio de espera era $1/\lambda$ y, después de esperar τ segundos, el tiempo promedio del tiempo que falta por esperar sigue siendo $1/\lambda$, independientemente de τ . Esta propiedad tan particular se conoce como la **falta de memoria de la distribución exponencial**.

En el modelo 3 tenemos que $P[T>s] = (b/s)^a$, de manera que $P[T>t+\tau|T>\tau] = [b/(t+\tau)]^a$. Esto es, la probabilidad de que debamos esperar t segundos más dado que ya hemos esperado τ segundos ¡aumenta con t !, de manera que tiende a uno a medida que τ tiende a infinito: entre mayor tiempo hayamos esperado, mayor tiempo nos falta por esperar. De hecho, suponiendo que $a>1$, el tiempo promedio que falta por esperar después de haber esperado τ segundos pasa de $ab/(a-1)$ a $\tau/(a-1)$, que tiende a infinito si τ tiende a infinito. Esta dependencia entre el tiempo que hemos esperado y el que nos falta por esperar se conoce como la **memoria de la distribución Pareto**. El hecho de que el tiempo remanente aumente con el tiempo que llevamos es un efecto de la **cola pesada** de la distribución, una de las **leyes de potencia** que tienen grandes implicaciones en la ingeniería de tráfico en redes de comunicaciones.

Las distribuciones exponencial y geométrica son las únicas distribuciones que no tienen memoria. En efecto, como necesitamos que $P[T>t+\tau|T>\tau] = P[T>t+\tau]/P[T>\tau] = P[T>t]$, será necesario que $P[T>t+\tau]$ se pueda expresar como el producto $P[T>t]P[T>\tau]$. En el caso continuo, esto exige que la cola de la distribución sea de la forma $P[T>t] = e^{-\lambda t}$. En el caso discreto, esto exige que la cola de la distribución sea $P[X>n] = p^n$.

52. Función Característica

Sea X una variable aleatoria. La función característica de X es una función de los reales en los complejos, definida de la siguiente manera

$$\phi_X(\omega) = E[e^{j\omega X}] = \int_{\mathbb{R}} e^{j\omega x} dF_X(x)$$

donde $j = \sqrt{-1}$.

Nótese que, con excepción del signo del exponente, la función característica de una variable aleatoria es la transformada de Fourier de su función de densidad de probabilidad o, tratándose de variables discretas, la transformada de Fourier en tiempo discreto de su función de distribución de probabilidad:

$$\phi_X(\omega) = \begin{cases} \int_{-\infty}^{\infty} f_X(x)e^{j\omega x} dx & \text{continua} \\ \sum_{k=-\infty}^{\infty} p_k e^{jk\omega} & \text{discreta} \end{cases}$$

Posiblemente recordemos cómo ésta era una herramienta muy útil en la soluciones de ecuaciones diferenciales (o de diferencia), así como en el cálculo de integrales (o sumas) de convolución en sistemas lineales. Pues bien, más adelante encontraremos este tipo de ecuaciones en el estudio de múltiples variables aleatorias y, en ese sentido, la función característica cumple un papel igualmente importante. En esas operaciones, si recuerda bien, se operaba en el dominio de la frecuencia y, posteriormente, se retornaba al dominio del tiempo. Ese retorno se consigue mediante la fórmula inversa de la función característica:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(\omega)e^{-j\omega x} d\omega$$

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_X(\omega)e^{-jk\omega} d\omega$$

Los cambios de signo en el exponente complejo con respecto a la transformada de Fourier y la transformada inversa obedecen a una convención muy simple: en teoría de señales los vectores suelen ser columnas y en teoría de probabilidades los vectores suelen ser filas, lo cual exige cambiar el vector que se transpone y se conjuga al calcular el producto interno representado por la transformada.

Como un ejemplo de la utilidad de la función característica, nótese que si $\phi_X(\omega)$ es n veces diferenciable en el origen ($\omega=0$), los n -ésimos momentos de X resultan fáciles de calcular:

$$E[X^n] = j^{-n} \frac{d^n}{d\omega^n} \phi_X(0)$$

Como se puede demostrar fácilmente al derivar la definición 52.

Por ejemplo, consideremos una variable aleatoria X exponencialmente distribuida con parámetro λ , cuya función característica es

$$\Phi_X(\omega) = \lambda \int_0^{\infty} e^{(j\omega - \lambda)x} dx = \frac{\lambda}{j\omega - \lambda} e^{(j\omega - \lambda)x} \Big|_0^{\infty} = \frac{\lambda}{\lambda - j\omega}$$

El cálculo de las estadísticas de primer y segundo orden resultan en $E[X] = \frac{1}{j} \frac{j\lambda}{(\lambda - j\omega)^2} \Big|_{\omega=0} = \frac{1}{\lambda}$

y $E[X^2] = \frac{1}{j^2} \frac{-2\lambda}{(\lambda - j\omega)^3} \Big|_{\omega=0} = \frac{2}{\lambda^2}$, con lo cual $V[X] = \frac{1}{\lambda^2}$.

Igualmente fácil es considerar una variable Y normalmente distribuida con parámetros μ y σ^2 :

$$\Phi_Y(\omega) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 + j\omega y} dy = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-\mu-j\omega\sigma^2)^2 + (\omega^2\sigma^4 - 2j\omega\mu\sigma^2)}{2\sigma^2}} dy = \dots$$

$$e^{-\frac{\omega^2 \sigma^4 - 2j\omega\mu\sigma^2}{2\sigma^2}} \left[\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-\mu-j\omega\sigma^2)^2}{2\sigma^2}} dy \right] = e^{j\omega\mu - \frac{\omega^2 \sigma^2}{2}}$$

(este truco de completar el cuadrado para obtener una integral conocida es muy típico al trabajar con la distribución normal). De estos resultados se encuentran fácilmente los momentos de la distribución Gaussiana:

$$E[Y] = \left. \frac{1}{j} \frac{d}{d\omega} e^{j\omega\mu - \frac{\omega^2 \sigma^2}{2}} \right|_{\omega=0} = \mu \quad E[Y^2] = \left. -\frac{d^2}{d\omega^2} e^{j\omega\mu - \frac{\omega^2 \sigma^2}{2}} \right|_{\omega=0} = \mu^2 + \sigma^2$$

de donde $V[Y] = \sigma^2$.

Por último, veamos el caso de una variable de Poisson, U , con intensidad ρ :

$$\Phi_U(j\omega) = \sum_{k=0}^{\infty} e^{j\omega k} \frac{\rho^k}{k!} e^{-\rho} = e^{-\rho} \sum_{k=0}^{\infty} \frac{(\rho e^{j\omega})^k}{k!} = e^{-\rho} \exp(\rho e^{j\omega}) = \exp(\rho(e^{j\omega} - 1))$$

$$E[U] = \left. \frac{1}{j} \frac{d}{d\omega} \exp(\rho(e^{j\omega} - 1)) \right|_{\omega=0} = \rho e^{j\omega} \exp(\rho(e^{j\omega} - 1)) \Big|_{\omega=0} = \rho$$

$$E[U^2] = \left. -\frac{d^2}{d\omega^2} \exp(\rho(e^{j\omega} - 1)) \right|_{\omega=0} = \rho e^{j\omega} \exp(\rho(e^{j\omega} - 1)) (1 + \rho e^{j\omega}) \Big|_{\omega=0} = \rho(1 + \rho)$$

$$V[U] = E[U^2] - E^2[U] = \rho$$

Cuando se trata de variables continuas o de variables discretas, la función característica (transformada de Fourier) se puede generalizar a la función generadora de momentos (transformada de Laplace) o a la función generadora de probabilidad (transformada Z). En este capítulo las mencionaremos brevemente aunque su utilidad principal se verá en los próximos capítulos al considerar distribuciones conjuntas de múltiples variables aleatorias.

53. Función Generadora de Momentos

Sea X una variable aleatoria continua. La función generadora de momentos de X es una función de los complejos en los complejos definida de la siguiente manera

$$M_X(s) = E[e^{sX}] = \int_{\mathbb{R}} e^{sx} dF_X(x) = \int_{\mathbb{R}} e^{sx} f_X(x) dx$$

Cuando $s=j\omega$, obtenemos la función característica de la distribución.

El nombre de Función generadora de momentos es fácil de comprender:

$$\frac{d^n}{ds^n} M_X(0) = \frac{d^n}{ds^n} E[e^{sX}] \Big|_{s=0} = E \left[\frac{d^n}{ds^n} e^{sX} \right] \Big|_{s=0} = E[X^n e^{sX}] \Big|_{s=0} = E[X^n]$$

de manera que el n -ésimo momento de X es la n -ésima derivada de la función generadora de momentos, evaluada en $s=0$.

Consideremos, por ejemplo, la función generadora de momentos de algunas variables conocidas:

1. Exponencial(λ): $M_X(s) = E[e^{sX}] = \lambda \int_0^{\infty} e^{-(\lambda-s)x} dx = \frac{\lambda}{\lambda-s}, \quad s < \lambda$
2. Gaussiana(μ, σ): $M_X(s) = \exp(s(\mu + \sigma^2 s/2))$ -de la manera que se calculó $\Phi_X(\omega)$ en 52-
3. Pareto(a, b): $M_X(s) = E[e^{sX}] = ab^a \int_b^{\infty} e^{sx} x^{-a-1} dx = a(-bs)^a \Gamma(-a, -bs), \quad s < 0$

donde $\Gamma(u, v)$ es la función gamma incompleta inferior, $\Gamma(u, v) = \int_v^{\infty} e^{-x} x^{u-1} dx$

4. Uniforme(a, b): $M_X(s) = E[e^{sX}] = \frac{1}{b-a} \int_a^b e^{sx} dx = \frac{e^{sb} - e^{sa}}{s(b-a)}$

Nótese que la función generadora de momentos fue la que apareció naturalmente al demostrar la desigualdad de Chernoff: $P[X > a] \leq \min_{s > 0} e^{-as} M_X(s)$ (definición 47).

54. Función Generadora de Probabilidad

Sea X una variable aleatoria discreta que toma valores en los enteros no negativos, con $p_k = \text{Prob}[X=k], k=0,1,2,\dots$. La función generadora de probabilidad de X es una función de los complejos en los complejos definida de la siguiente manera

$$G_X(z) = E[z^X] = \sum_{k=0}^{\infty} p_k z^k$$

Cuando $z=e^{j\omega}$, obtenemos la función característica de la distribución.

El nombre de Función generadora de probabilidad es fácil de comprender. Considérese una variable aleatoria discreta que toma valores en los enteros, de manera que $p_k = P[X=k]$ para $k=0,1,2,\dots$

$$\begin{aligned} \left. \frac{d^n}{dz^n} G_X(0) = \frac{d^n}{dz^n} \sum_{k=0}^{\infty} p_k z^k \right|_{z=0} &= \sum_{k=0}^{\infty} p_k \left. \frac{d^n}{dz^n} z^k \right|_{z=0} = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} p_k z^{k-n} \Big|_{z=0} = n! p_n \\ \Rightarrow p_n &= \frac{1}{n!} \frac{d^n}{dz^n} G_X(0) \end{aligned}$$

de manera que la distribución de probabilidad se puede obtener inmediatamente de la función generadora de probabilidad mediante diferenciación y normalización.

También ocurren cosas interesantes cuando evaluamos en $z=1$:

$$G_X(1) = \sum_{k=0}^{\infty} p_k = 1 \quad \frac{d}{dz} G_X(1) = \sum_{k=1}^{\infty} k p_k = E[X] \quad \frac{d^2}{dz^2} G_X(1) = \sum_{k=2}^{\infty} k(k-1) p_k = E[X(X-1)]$$

$$\frac{d^n}{dz^n} G_X(1) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1) p_k = E[X(X-1)\cdots(X-n+1)]$$

lo que indica que la función generadora de probabilidad también es, de algún modo, una función generadora de momentos.

Consideremos, por ejemplo, la función generadora de probabilidad de algunas variables conocidas:

1. Bernoulli(p): $G_X(z) = (1-p) + pz = 1 + p(z-1)$
2. Geométrica(p): $G_X(z) = \sum_{k=0}^{\infty} z^k p^k (1-p) = (1-p) \sum_{k=0}^{\infty} (zp)^k = (1-p)/(1-zp)$, $|zp| < 1$
3. Binomial(n, p): $G_X(z) = \sum_{k=0}^n z^k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (zp)^k (1-p)^{n-k} = (1 + p(z-1))^n$
4. Poisson(λ): $G_X(z) = \sum_{k=0}^{\infty} z^k \frac{\rho^k}{k!} e^{-\rho} = e^{-\rho} \sum_{k=0}^{\infty} \frac{(\rho z)^k}{k!} = e^{-\rho} e^{\rho z} = \exp(\rho(z-1))$

La utilidad de las funciones característica, generadora de momentos y generadora de probabilidades se podrá apreciar mejor al considerar variables aleatorias conjuntamente distribuidas, que es el tema del siguiente capítulo.

IV. Vectores Aleatorios Bidimensionales

55. Campo- σ de Borel en el Plano, $\mathcal{B}(\mathbb{R}^2)$

El campo- σ de Borel del plano real, $\mathcal{B}(\mathbb{R}^2)$, es el mínimo campo- σ que contiene a todos los subconjuntos de la forma

$$A(x,y) = \{(a,b) \in \mathbb{R}^2 : a \leq x, b \leq y\}, \forall (x,y) \in \mathbb{R}^2.$$

Los subconjuntos de \mathbb{R}^2 que pertenecen a $\mathcal{B}(\mathbb{R}^2)$ se denominan "conjuntos de Borel".

Ya vimos cómo los conjuntos de la forma $(-\infty, x]$ resultaron fundamentales para pasar de espacios de probabilidad arbitrarios $(\Omega, \mathcal{F}, \mathbf{P})$ a espacios más fáciles de manejar definidos por una variable aleatoria, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(x))$. De la misma manera, al trabajar con dos resultados numéricos diferentes, X_1 y X_2 , asociados con un mismo experimento aleatorio $(\Omega, \mathcal{F}, \mathbf{P})$, podemos evitar el tener que describir explícitamente el espacio de probabilidad original si construimos un nuevo espacio implícito para el resultado (X_1, X_2) donde el espacio muestral está dado por el plano cartesiano \mathbb{R}^2 , el campo-sigma de eventos es el campo-sigma de Borel de \mathbb{R}^2 , y donde la medida de probabilidad está dada por una función de \mathbb{R}^2 en \mathbb{R} , $F_{X,Y}(x,y) = \mathbf{P}[\{\omega \in \Omega : X_1(\omega) \leq x\} \cap \{\omega \in \Omega : X_2(\omega) \leq y\}]$, $(x,y) \in \mathbb{R}^2$. En este caso hablaremos del vector aleatorio (X_1, X_2) , definido en 56, y de su función de distribución acumulativa conjunta $F_{X,Y}(x,y)$, definida en 57. Efectivamente, como es imposible pensar en asignar probabilidades a cada elemento del conjunto potencia de \mathbb{R}^2 , nos limitamos a los conjuntos que pertenezcan a $\mathcal{A}(\mathbb{R}^2)$. Nuevamente, el campo sigma $\mathcal{A}(\mathbb{R}^2)$ incluye puntos, líneas, curvas, polígonos, círculos y otros subconjuntos razonables de \mathbb{R}^2 . Limitándonos a uniones numerables de este tipo de eventos en \mathbb{R}^2 , podemos construir el espacio de probabilidad $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2), F_{X,Y}(\cdot, \cdot))$, sobre el cual podremos aplicar toda la lógica booleana sin llegar a inconsistencias.

La siguiente figura muestra un conjunto elemental de Borel, $A(8,8)$, que es el producto cartesiano de los intervalos $\{x \leq 8\} \times \{y \leq 8\}$. Con complementos, uniones contables e intersecciones contables de estos conjuntos básicos podemos generar casi cualquier subconjunto imaginable de \mathbb{R}^2 .

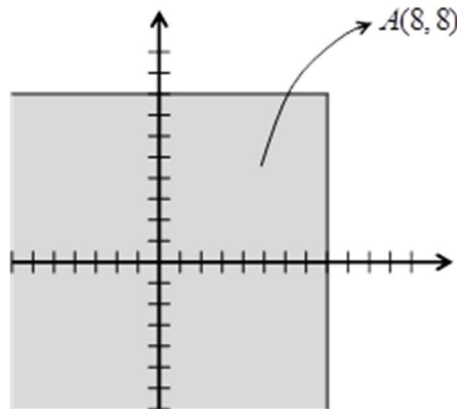


Figura 70. $A(8,8) = \{(a,b) \in \mathbb{R}^2 : a \leq 8, b \leq 8\}$

Por ejemplo, un rectángulo con esquinas en $\{(x_0, y_0), (x_0, y_1), (x_1, y_1), (x_1, y_0)\}$ se podría expresar así: $R(x_0, y_0, x_1, y_1) = A(x_1, y_1) \cap A(x_1, y_0)^c \cap A(x_0, y_1)^c$. Nótese que en este conjunto de Borel hemos incluido las fronteras superior y derecha del rectángulo, pero no sus fronteras inferior e izquierda. Para incluirlas, simplemente las unimos a R , pues un segmento horizontal de recta entre (x_0, a) y (x_1, a) es otro conjunto

de Borel: $L_h(a, x_0, x_1) = \bigcap_{n=1}^{\infty} R(x_0, a - \frac{1}{n}, x_1, a)$, al igual que un segmento vertical de recta entre (b, y_0) y

(b, y_1) : $L_v(b, y_0, y_1) = \bigcap_{n=1}^{\infty} R(b - \frac{1}{n}, y_0, b, y_1)$. En consecuencia, el rectángulo cerrado (con sus

fronteras) es $R_f(x_0, y_0, x_1, y_1) = R(x_0, y_0, x_1, y_1) \cup L_h(y_0, x_0, x_1) \cup L_v(x_0, y_0, y_1)$.

Consideremos ahora la unión de una secuencia de rectángulos como el anterior,

$$R_n(x, y) = \bigcup_{k=1}^n R_f \left(x + \frac{k-1}{n}, y - \sqrt{1 - \left(\frac{k}{n}\right)^2}, x + \frac{k}{n}, y + \sqrt{1 - \left(\frac{k}{n}\right)^2} \right),$$

como muestra la siguiente figura.

Claramente, basta con tomar el límite cuando $n \rightarrow \infty$ para encontrar que un semicírculo es un conjunto de Borel en \mathbb{R}^2 , pues es la unión contable de rectángulos (infinitos rectángulos de área infinitesimal, pero contables).

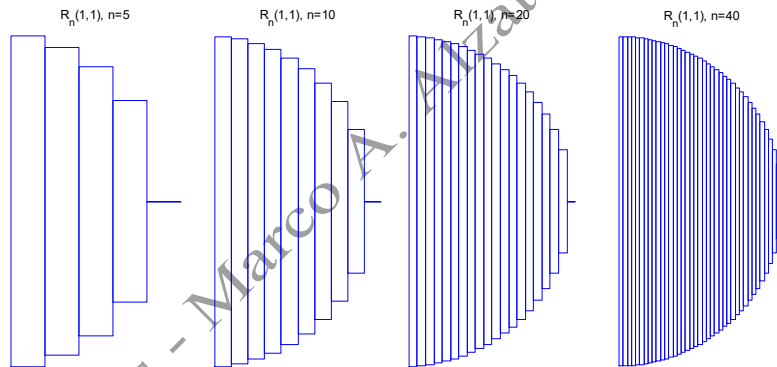


Figura 71. Un semicírculo es un conjunto de Borel en \mathbb{R}^2

Como último ejemplo, si $g: [0, 1] \rightarrow \mathbb{R}$ es una función monótonamente creciente, su gráfica en el plano es un conjunto de Borel: $F = \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{n-1} R(k/n, (k+1)/n, f(k/n), f((k+1)/n))$.

De igual manera podemos ver que todas las curvas suaves en el plano, todos los puntos individuales (como conjuntos unitarios), todos los triángulos, y todos los objetos regulares e irregulares de $n < \infty$ lados son conjuntos de Borel, así como sus complementos, uniones contables e intersecciones contables. Ni siquiera subconjuntos “patológicos” como el conjunto bidimensional de Cantor (que

nada tiene de patológico²⁸), mostrado en la siguiente figura –ver definición 13–, dejan de ser subconjuntos de Borel en \mathbb{R}^2 aunque se trate de conjunto incontables de puntos aislados en el plano unitario.

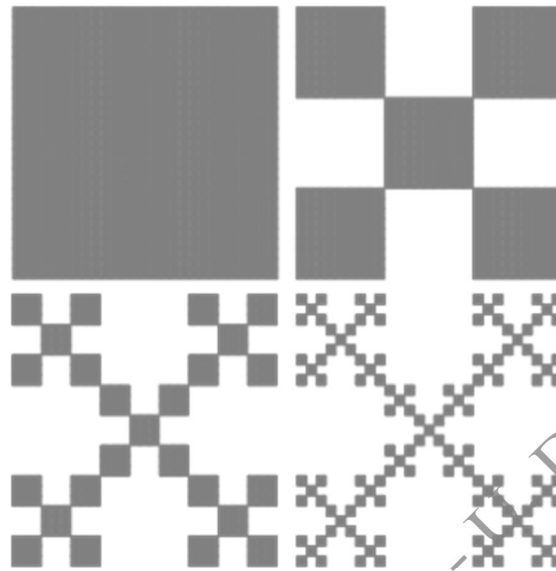


Figura 72. Primeras cuatro etapas en la construcción del conjunto de Cantor en el plano

56. Vector Aleatorio Bidimensional

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se definen dos funciones $X_1: \Omega \rightarrow \mathbb{R}$ y $X_2: \Omega \rightarrow \mathbb{R}$. Entonces $\mathbf{X}(\omega) = [X_1(\omega) \ X_2(\omega)]^T$ es un vector aleatorio bidimensional si $\forall \mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$, el evento $A(\mathbf{x}) = \{\omega \in \Omega : X_1(\omega) \leq x_1\} \cap \{\omega \in \Omega : X_2(\omega) \leq x_2\}$ es un evento medible, $A(\mathbf{x}) \in \mathcal{F}$.

El vector aleatorio bidimensional le asigna a cada elemento del espacio muestral un elemento del producto cartesiano $\mathbb{R} \times \mathbb{R}$, de manera tal que las imágenes de los eventos en \mathcal{F} resultan ser conjuntos de Borel en \mathbb{R}^2 . Esta es, entonces, una extensión natural de la definición 29. De hecho, nótese que cada uno de los componentes del vector aleatorio son variables aleatorias unidimensionales definidas en el mismo espacio de probabilidad pues, de acuerdo con la definición, $A(x_1, \infty) = \{\omega \in \Omega : X_1(\omega) \leq x_1\} \in \mathcal{F}$, y $A(\infty, x_2) = \{\omega \in \Omega : X_2(\omega) \leq x_2\} \in \mathcal{F}$, lo cual hace de X_1 y X_2 dos variables aleatorias unidimensionales de acuerdo con la definición 29. Pero lo que hace de $\mathbf{X} = [X_1 \ X_2]^T$ un vector aleatorio es que para cada conjunto de Borel B en \mathbb{R}^2 se puede definir la probabilidad $\mathbf{P}[\mathbf{X} \in B]$, pues siempre se tratará de uniones contables, intersecciones contables y complementos de eventos de la forma $A(\mathbf{x}) = \{\omega \in \Omega : X_1(\omega) \leq x_1\} \cap \{\omega \in \Omega : X_2(\omega) \leq x_2\}$. La Figura 73 describe el vector aleatorio como función vectorial del espacio muestral de un experimento.

²⁸ La condición de “patológico” es de comienzos del siglo XX. Hoy sabemos que muchas estructuras naturales se pueden modelar mediante algoritmos de construcción semejantes a las de Cantor (incluyendo algunos modelos de tráfico en redes de comunicaciones).

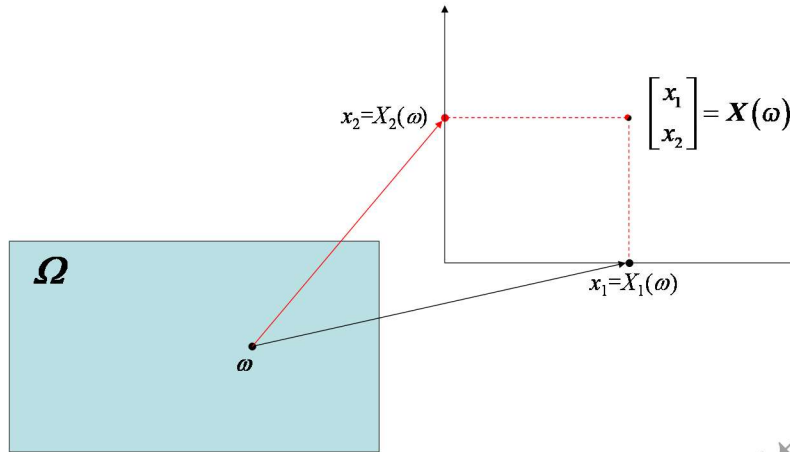


Figura 73. Vector aleatorio bidimensional

Nótese que tanto \mathbf{X} como X_1 y X_2 se definen en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$. Si cada función X_1 y X_2 estuvieran definidos en espacios de probabilidad diferentes, $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ y $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$, sería necesario definir \mathbf{X} en el espacio muestral $\Omega_1 \times \Omega_2$, en cuyo caso el campo de eventos debería redefinirse como el mínimo campo sigma que contenga a $\mathcal{F}_1 \times \mathcal{F}_2$, y la medida de probabilidad debería redefinirse cuidadosamente de acuerdo con la naturaleza del experimento conjunto (por ejemplo, $\mathbf{P}(A, B) = \mathbf{P}_1(A)\mathbf{P}_2(B) \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2$ si los experimentos fuesen completamente independientes). Este aspecto tiene muchas "tecnicidades" matemáticas que, desde la pragmaticidad de la ingeniería, solemos evitar definiendo un espacio de probabilidad subyacente común. Después de todo ya hemos dicho que la definición de una variable aleatoria $X \sim F_X(x)$ define un espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X(\cdot))$ en sí mismo y la existencia de otro espacio subyacente $(\Omega, \mathcal{F}, \mathbf{P})$ puede dejarse implícita. De la misma manera, la definición de una variable aleatoria bi-dimensional (X_1, X_2) define un espacio de probabilidad $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathbf{P}(\{\omega \in \Omega : X_1(\omega) \leq x_1\} \cap \{\omega \in \Omega : X_2(\omega) \leq x_2\}))$ en sí mismo, y la existencia de otro espacio subyacente $(\Omega, \mathcal{F}, \mathbf{P})$ puede dejarse implícita.

¿Por qué considerar dos variables de manera conjunta en vez de aplicar “dos veces” la teoría desarrollada en el capítulo anterior? Considere un experimento como el de transferir un archivo de un servidor a un cliente a través de una conexión de red. Podemos considerar dos variables aleatorias: La longitud del archivo, L , y el tiempo de transferencia, T . Podríamos definir eventos relacionados con cada uno de estas variables aleatorias tales como $A = \{\text{el archivo no supera 100 KB}\}$ o $B = \{\text{La transferencia toma 1 s o menos}\}$. Ambos eventos pertenecen al campo de Borel de los reales y sus probabilidades se podrían encontrar mediante la teoría desarrollada en el capítulo anterior en términos de las CDF de cada variable, $P[A] = F_L(100 \text{ KB})$ y $P[B] = F_T(1 \text{ s})$. Pero es apenas razonable querer encontrar, por ejemplo, la probabilidad de que el archivo no supere 100 KB dado que el tiempo de transferencia no supera un segundo, $P[A|B] = P[A \cap B]/P[B]$. Para esto necesitamos evaluar $P[A \cap B]$, que es exactamente la probabilidad del evento $\{\omega \in \Omega : L(\omega) \leq 100 \text{ KB}\} \cap \{\omega \in \Omega : T(\omega) \leq 1 \text{ s}\}$ que define al vector bi-valuado $[L \ T]^T$. En general, las distribuciones $F_L(l)$ y $F_T(t)$ no contienen suficiente información para evaluar esta probabilidad conjunta (a menos que L y T fueran variables independientes que, obviamente, no es el caso), por lo cual necesitamos una nueva función de

distribución que relacione las dos variables aleatorias de manera conjunta, $F_{L,T}(l,t)$, correspondiente al espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ con $\Omega = \{(l,t) \in \mathbb{R}^2: l \geq 0, t \geq 0\}$, $\mathcal{F} = \mathcal{B}(\mathbb{R}_+^2)$ donde \mathbb{R}_+ son los reales no negativos, y con $\mathbf{P}: \mathcal{B}(\mathbb{R}_+^2) \rightarrow [0, 1]$ dada por $F_{L,T}: \mathbb{R}_+^2 \rightarrow [0, 1]$. A continuación definiremos precisamente esta función de distribución de probabilidad acumulativa conjunta y elaboraremos un poco más el ejemplo propuesto.

57. Función de Distribución de Probabilidad Acumulativa Conjunta, JCDF

Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad sobre el cual se define un vector aleatorio $(X,Y): \Omega \rightarrow \mathbb{R}^2$. La Función acumulativa de distribución de probabilidad de (X,Y) es la función $F_{X,Y}: \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como $F_{X,Y}(x,y) = \mathbf{P}(\{\omega \in \Omega: X(\omega) \leq x, Y(\omega) \leq y\})$, $\forall x,y \in \mathbb{R}^2$. Le diremos JCDF por la sigla en inglés de Joint Cumulative Distribution Function.

Esta es la extensión directa del concepto de CDF cuando se involucran dos variables aleatorias. La JCDF es la función que nos permite pasar del espacio de probabilidad subyacente $(\Omega, \mathcal{F}, \mathbf{P})$ al espacio $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), F_{X,Y}(\cdot, \cdot))$ que es exactamente lo que queremos ganar al considerar las variaciones conjuntas de dos variables aleatorias. Al final, el espacio de probabilidad subyacente se hace irrelevante cuando es la JCDF la que constituye el modelo probabilístico del sistema bajo estudio.

En el caso en que medíamos la longitud de un archivo, L , y su tiempo de transferencia, T , además de las distribuciones marginales $F_L(l) = P[L \leq l]$ y $F_T(t) = P[T \leq t]$ podemos hablar de la distribución conjunta $F_{L,T}(l,t) = P[L \leq l, T \leq t]$. Esta distribución conjunta captura todas las relaciones que pueda haber entre las dos variables, mientras que las dos distribuciones marginales por aparte dejan mucha información por conocer. Por ejemplo, supongamos que la transferencia se hace a través de un canal de 1 Mbps (1000000 bits/segundo), que L está uniformemente distribuida entre 500 bytes y 40000 bytes, y que T está uniformemente distribuida entre 5 ms y 2000 ms. La probabilidad de que L supere 28000 bytes es aproximadamente 0.3 y la probabilidad de que T sea inferior a 200 ms es aproximadamente 0.1. Sin embargo, la probabilidad de que ocurran ambos eventos es cero, porque si L supera 28000 bytes, sólo el tiempo de transmisión será 224 ms, de manera que T debe superar 224 ms. Ese hecho no está contemplado en el par de distribuciones marginales $F_L(l)$ y $F_T(t)$, por lo que se hace necesario considerar la distribución conjunta $F_{L,T}(l,t)$, la cual debe revelar el hecho de que $P[T < 200 \text{ ms}, L > 28000 \text{ bytes}] = 0$. Por ejemplo, aunque L y T sean uniformes en $[500, 40000]$ bytes y en $[5, 2000]$ ms, respectivamente, su región de variación conjunta no es el rectángulo $[500, 40000] \times [5, 2000]$ sino una región diferente en el plano \mathbb{R}^2 , por ejemplo como la mostrada en la Figura 74.

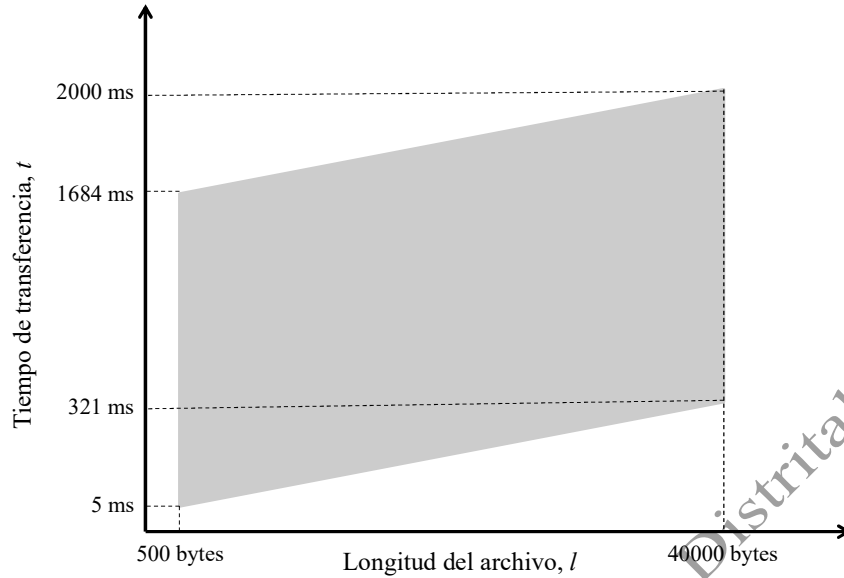


Figura 74. Aunque L está distribuida en $[500, 40000]$ bytes y T está distribuida en $[5, 2000]$ ms, L y T están conjuntamente distribuidas en la región mostrada, diferente al rectángulo $[500, 40000]$ bytes \times $[5, 2000]$ ms

Es fácil imaginar muchos ejemplos semejantes. Por ejemplo, ¿Cómo caracterizar la altura y el peso de una persona sino como dos variables aleatorias conjuntas, esto es, un vector bidimensional? ¿Qué tal los valores de los activos y los pasivos de una empresa? ¿O la dosis de insulina y el nivel de azúcar en la sangre? Es necesario tratar dos variables aleatorias como un vector aleatorio bidimensional conjuntamente distribuido.

58. Propiedades de la JCDF

Sean X y Y dos variables aleatorias con JCDF $F_{X,Y}(\cdot, \cdot)$. Entonces,

- La JCDF es no-negativa: $F_{X,Y}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
- La JCDF es no-decreciente en ambos argumentos: si $x_0 \leq x_1$ y $y_0 \leq y_1$, entonces

$$F_{X,Y}(x_0, y_0) \leq F_{X,Y}(x_0, y_1) \leq F_{X,Y}(x_1, y_1)$$

$$F_{X,Y}(x_0, y_0) \leq F_{X,Y}(x_1, y_0) \leq F_{X,Y}(x_1, y_1)$$

$$F_{X,Y}(x_0, y_0) + F_{X,Y}(x_1, y_1) \geq F_{X,Y}(x_1, y_0) + F_{X,Y}(x_0, y_1)$$
- La JCDF es acotada: $F_{X,Y}(-\infty, y) = 0 \quad \forall y \in \mathbb{R}$, $F_{X,Y}(x, -\infty) = 0 \quad \forall x \in \mathbb{R}$,
 $F_{X,Y}(x, \infty) = F_X(x) \quad \forall x \in \mathbb{R}$, $F_{X,Y}(\infty, y) = F_Y(y) \quad \forall y \in \mathbb{R}$, $F_{X,Y}(\infty, \infty) = 1$.
- La JCDF es continua por la derecha en ambos argumentos: $F_{X,Y}(x^+, y) = F_{X,Y}(x, y)$
 $\forall x, y \in \mathbb{R}$, $F_{X,Y}(x, y^+) = F_{X,Y}(x, y) \quad \forall x, y \in \mathbb{R}$, $F_{X,Y}(x^+, y^+) = F_{X,Y}(x, y) \quad \forall x, y \in \mathbb{R}$.

Como en el caso de una variable aleatoria, estas propiedades se pueden demostrar muy fácilmente, con base en regiones como las descritas en la definición 55. Por ejemplo, la propiedad (a) surge del

segundo axioma de las probabilidades. Para probar la propiedad (b), basta considerar los siguientes eventos mutuamente excluyentes en \mathbb{R}^2 , donde $x_0 \leq x_1$ y $y_0 \leq y_1$, como muestra la Figura 75.

$$A = \{(x,y) \in \mathbb{R}^2 : x \leq x_0, y \leq y_0\}, \quad B = \{(x,y) \in \mathbb{R}^2 : x \leq x_0, y_0 < y \leq y_1\}, \quad C = \{(x,y) \in \mathbb{R}^2 : x_0 < x \leq x_1, y \leq y_0\}, \quad D = \{(x,y) \in \mathbb{R}^2 : x_0 < x \leq x_1, y_0 < y \leq y_1\}$$

Si definimos $P[A] = P[\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}]$ y, de manera similar, definimos $P[B]$, $P[C]$ y $P[D]$, tenemos que $F_{X,Y}(x_0, y_0) = P[A]$, $F_{X,Y}(x_0, y_1) = P[A] + P[B]$, $F_{X,Y}(x_1, y_0) = P[A] + P[C]$ y $F_{X,Y}(x_1, y_1) = P[A] + P[B] + P[C] + P[D]$. Como el segundo axioma indica que todas estas probabilidades son mayores o iguales a cero,

$$P[A] \leq P[A] + P[B] \leq P[A] + P[B] + P[C] + P[D], \text{ esto es, } F_{X,Y}(x_0, y_0) \leq F_{X,Y}(x_0, y_1) \leq F_{X,Y}(x_1, y_1)$$

$$P[A] \leq P[A] + P[C] \leq P[A] + P[B] + P[C] + P[D], \text{ esto es, } F_{X,Y}(x_0, y_0) \leq F_{X,Y}(x_1, y_0) \leq F_{X,Y}(x_1, y_1)$$

$$2P[A] + P[B] + P[C] + P[D] \geq 2P[A] + P[B] + P[C], \text{ esto es,}$$

$$F_{X,Y}(x_0, y_0) + F_{X,Y}(x_1, y_1) \geq F_{X,Y}(x_1, y_0) + F_{X,Y}(x_0, y_1)$$

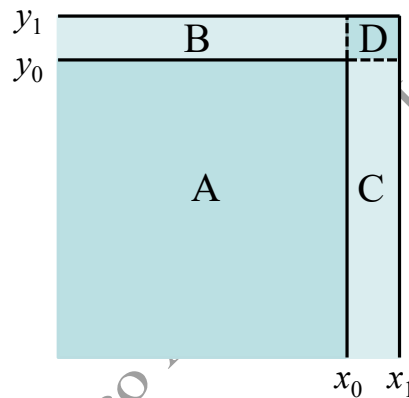


Figura 75. $A = \{(x,y) \in \mathbb{R}^2 : x \leq x_0, y \leq y_0\}$, $B = \{(x,y) \in \mathbb{R}^2 : x \leq x_0, y_0 < y \leq y_1\}$,
 $C = \{(x,y) \in \mathbb{R}^2 : x_0 < x \leq x_1, y \leq y_0\}$, $D = \{(x,y) \in \mathbb{R}^2 : x_0 < x \leq x_1, y_0 < y \leq y_1\}$

Las demás demostraciones son semejantes, inspiradas en las demostraciones presentadas en la definición 31.

Lo más interesante es que cualquier función de \mathbb{R}^2 en \mathbb{R} que satisfaga las cuatro propiedades anteriores es una JCDF válida para algún vector aleatorio bidimensional en algún espacio de probabilidad subyacente, el cual, en principio, podría omitirse, pues todo el experimento aleatorio se puede asociar con el espacio $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), F_{X,Y}(\cdot, \cdot))$, definido completamente por la JCDF²⁹. El modelamiento probabilístico se reduce, entonces, a encontrar una JCDF adecuada a la naturaleza del experimento. La Figura 76 muestra dos funciones que satisfacen las anteriores cuatro propiedades, correspondientes a un par de variables conjuntamente continuas (izquierda) y conjuntamente discretas (derecha).

²⁹ Es importante saber que el espacio de probabilidad subyacente TIENE que existir de acuerdo con la naturaleza del experimento aleatorio y, por si fuese necesario, siempre deberíamos estar en condiciones de especificarlo. Pero, mientras no sea estrictamente necesario (suele no serlo), el vector aleatorio $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), F_{X,Y}(\cdot, \cdot))$ ya constituye un modelo probabilístico perfectamente válido sin que seamos explícitos en la especificación de (Ω, \mathcal{F}, P) .

También puede haber variables conjuntas en las que una sea discreta y otra sea continua o en las que una o ambas sean mixtas (combinaciones convexas de una distribución continua y una distribución discreta). La Figura 77 muestra una JCDF de una variable continua que toma valores en el intervalo $[0,1]$ y una variable aleatoria discreta que toma valores en el conjunto $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Nótese cómo la función mostrada también satisface las propiedades de ser no-negativa, no-decreciente, acotada y continua por la derecha en ambos argumentos.

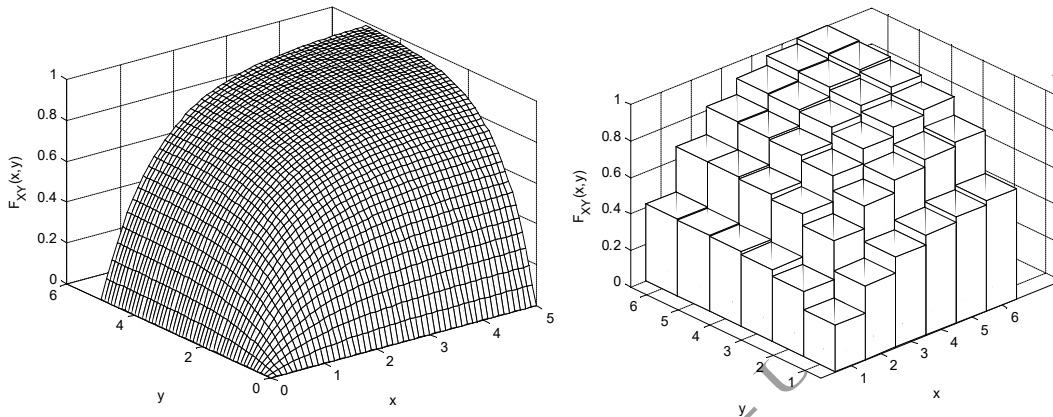


Figura 76. JCDF de dos variables aleatorias continuas (a) y discretas (b)

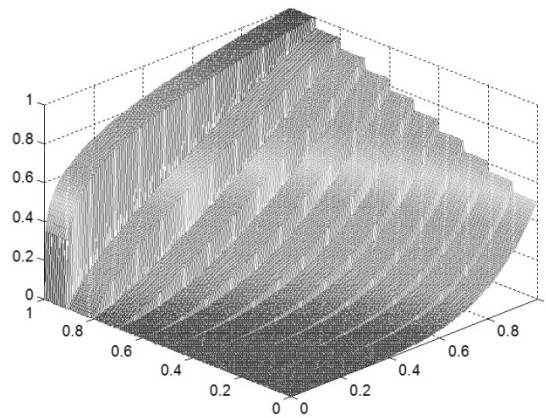


Figura 77. JCDF de una variable aleatoria continua y otra discreta

Como un ejemplo final, considérese la función $g(x,y) = (1 - e^{-(x+y)})u(x)u(y)$, donde $u(\cdot)$ es el escalón unitario que vale uno si su argumento es mayor o igual a cero y vale cero en otro caso (Figura 78). ¿Es g una JCDF válida? Es evidente que $g(x,y) \geq 0 \forall (x,y) \in \mathbb{R}^2$ porque tanto $(1 - e^{-(x+y)})$ como $u(x)$ y $u(y)$ son no negativas para x, y no negativas, mientras que $u(x)$ o $u(y)$ son cero cuando $(1 - e^{-(x+y)})$ sea negativa (esto es, cuando $x+y < 0$). Por otro lado, para un valor fijo de $y \geq 0$ con $0 < \alpha = e^{-y} \leq 1$, es claro que $h(x; \alpha) = 1 - \alpha e^{-x}$ es no-decreciente (y, para un valor fijo de x con $\beta = e^{-x}$, es claro que $h(y; \beta) = 1 - \beta e^{-y}$ es no-decreciente). Además, cuando x ó y sean negativos, $g(x,y)$ toma su mínimo valor, que es cero, y cuando $x+y$ tienda a infinito, $g(x,y)$ tiende a su valor supremo, que es uno. Por último, $g(x,y)$ es

continua para $x, y > 0$, de manera que es continua por la derecha para $x, y > 0$. Cuando $x=0$ y $y > 0$, $g(0, y) = 1 - e^{-y}$ es el límite de la función cuando su segundo argumento tiende a y por la derecha; cuando $y=0$ y $x > 0$, $g(x, 0) = 1 - e^{-x}$ es el límite de la función cuando su primer argumento tiende a x por la derecha. Así pues, podemos imaginar dos variables aleatorias X y Y con JCDF dada por $F_{XY}(x, y) = (1 - e^{-(x+y)})u(x)u(y)$. $[X, Y]^T$ será un vector aleatorio bidimensional perfectamente válido, aunque no especifiquemos el experimento subyacente en el que se definen las funciones determinísticas X y Y .

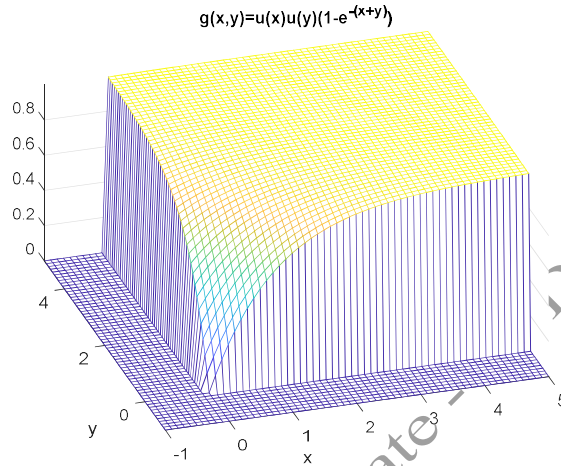


Figura 78. Una función de \mathbb{R}^2 en \mathbb{R} que satisface las propiedades de una JCDF

59. Probabilidades de algunos conjuntos de Borel en \mathbb{R}^2

Sean X y Y dos variables aleatorias con JCDF $F_{X,Y}(\cdot, \cdot)$. La probabilidad de que las dos variables caigan dentro de un conjunto de Borel B en \mathbb{R}^2 es

$$P[(X, Y) \in B] = \iint_B d^2 F_{X,Y}(x, y)$$

Los siguientes son algunos ejemplos significativos:

$$P[X \in (-\infty, x], Y \in (-\infty, y]] = F_{XY}(x, y),$$

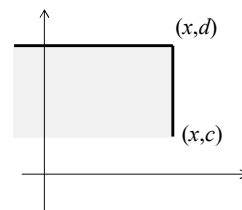
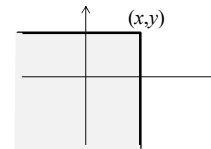
por la definición de JCDF

$$P[X \in (-\infty, x], Y \in (c, d]] = F_{XY}(x, d) - F_{XY}(x, c),$$

porque $(X \in (-\infty, x]) \cap (Y \in (-\infty, d]) =$

$$((X \in (-\infty, x]) \cap (Y \in (-\infty, c])) \cup ((X \in (-\infty, x]) \cap (Y \in (c, d]))$$

Al aplicar probabilidad a ambos conjuntos, y por el tercer axioma, $F_{XY}(x, d) = F_{XY}(x, c) +$



$P[X \in (-\infty, x], Y \in (c, d]]$, de donde despejamos la probabilidad de interés

$$P[X \in (a, b], Y \in (-\infty, y]] = F_{XY}(b, y) - F_{XY}(a, y),$$

porque $(X \in (-\infty, b]) \cap (Y \in (-\infty, y]) =$

$$((X \in (-\infty, a]) \cap (Y \in (-\infty, y])) \cup ((X \in (a, b]) \cap (Y \in (-\infty, y]))$$

Al aplicar probabilidad a ambos conjuntos, y por el tercer axioma, $F_{XY}(b, y) = F_{XY}(a, y) +$

$P[X \in (a, b], Y \in (-\infty, y]]$, de donde despejamos la probabilidad de interés

$$P[X \in (a, b], Y \in (c, d]] = F_{XY}(b, d) - F_{XY}(a, d) - \dots$$

$$\dots F_{XY}(b, c) + F_{XY}(a, c),$$

Porque $(X \in (-\infty, b]) \cap (Y \in (c, d]) =$

$$((X \in (-\infty, a]) \cap (Y \in (c, d])) \cup ((X \in (a, b]) \cap (Y \in (c, d]))$$

Al aplicar probabilidad a ambos conjuntos, y por el tercer axioma, $F_{XY}(b, d) - F_{XY}(b, c) = F_{XY}(a, d) - F_{XY}(a, c) + F_{XY}(a, c) + P[X \in (a, b], Y \in (c, d]]$, de donde despejamos la probabilidad de interés

$$P[X=x, Y=y] = F_{XY}(x^+, y^+) - F_{XY}(x^+, y^-) - \dots$$

$$\dots F_{XY}(x^+, y^+) + F_{XY}(x^-, y^-),$$

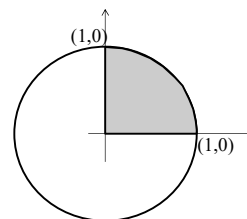
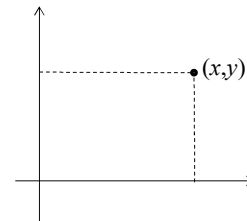
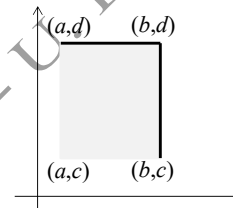
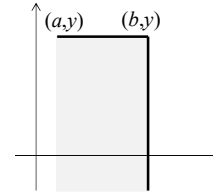
porque, si usamos el resultado anterior con $b=x, d=y, a=x-1/m$ y $c=y-1/n$, bastaría con intersectar los rectángulos para m desde 1 hasta ∞ y para n desde 1 hasta ∞ para obtener el resultado.

$$P[(X, Y) \in B] = \iint_B d^2 F_{X,Y}(x, y) = \int_0^1 \int_0^{\sqrt{1-x^2}} d^2 F_{X,Y}(x, y)$$

Porque

$$B = \lim_{n \rightarrow \infty} \bigcup_{k=0}^{n-1} \left\{ \left(X \in \left(\frac{k}{n}, \frac{k+1}{n} \right] \right) \cap \left(Y \in \left(0, \sqrt{1 - \left(\frac{k}{n} \right)^2} \right] \right) \right\}$$

Que es la unión contable de eventos mutuamente excluyentes.



60. Función de Distribución de Probabilidad Acumulativa Marginal

Sean X y Y dos variables aleatorias con JCDF $F_{X,Y}(\cdot, \cdot)$. Entonces la CDF de cada una de las respectivas variables aleatorias unidimensionales, o CDF marginal, se define como

$$F_X(x) = F_{XY}(x, \infty) \quad \forall x \in \mathbb{R}, \quad F_Y(y) = F_{XY}(\infty, y) \quad \forall y \in \mathbb{R}$$

Esta definición cobra todo el sentido cuando consideramos la definición de los eventos asociados en el espacio subyacente original:

$$\begin{aligned} F_X(x) &= P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(\{\omega \in \Omega : X(\omega) \leq x\} \cap \Omega) \\ &= P(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : Y(\omega) \leq \infty\}) \\ &= P(\{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq \infty\}) \\ &= F_{X,Y}(x, \infty) \end{aligned}$$

Así, como mencionábamos en las definiciones anteriores, la JCDF $F_{XY}(x, y)$ incluye toda la información proporcionada por las distribuciones marginales $F_X(x)$ y $F_Y(y)$, hasta el punto que las marginales se pueden extraer de la conjunta, como acabamos de ver. Pero puede incluir mucha más información adicional sobre las variaciones conjuntas de ambas variables.

Finalmente, nótese que, en la definición de las propiedades de la JCDF (definición 58), la propiedad (c) es exactamente esta definición 60 sobre distribución acumulativa marginal.

61. Función de Densidad de Probabilidad Conjunta (jpdf) y Función de Distribución de Probabilidad Conjunta (jpmf)

Sean X y Y dos variables aleatorias con JCDF $F_{X,Y}(\cdot, \cdot)$. La función de densidad de probabilidad conjunta (jpdf) se define como

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$

Si X y Y son variables discretas con valores $\{x_i, i \in \mathbb{N}\}$ y $\{y_j, j \in \mathbb{N}\}$ respectivamente, la jpdf toma la siguiente forma

$$f_{XY}(x, y) = \sum_i \sum_j p_{ij} \delta(x - x_i) \delta(y - y_j)$$

Donde $p_{ij} = \text{Prob}[X=x_i, Y=y_j]$ es la función de distribución de probabilidad conjunta (jpmf por las siglas en inglés de Joint Probability Mass Function)

Las anteriores definiciones implican las siguientes propiedades:

$$(1) f_{XY}(x, y) \geq 0 \quad \forall (x, y) \in \mathbb{R}^2$$

$$(2) F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du \quad \forall (x, y) \in \mathbb{R}^2$$

$$(3) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$$

Las cuales toman las siguientes formas cuando se trata de variables conjuntamente discretas con jpmf p_{ij} ,

$$(1') p_{i,j} \geq 0$$

$$(2') F_{X,Y}(x,y) = \sum_{i:x_i \leq x} \sum_{j:y_j \leq y} p_{ij}$$

$$(3') \sum_i \sum_j p_{ij} = 1$$

Cualquier función $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ que satisfaga las anteriores propiedades (1) y (3) es la jpdf de algún par de variables aleatorias conjuntamente continuas. Cualquier mapa $q: \mathbb{N}^2 \rightarrow \mathbb{R}$ que satisfaga las propiedades (1') y (3') es la jpmf de algún par de variables aleatorias conjuntamente discretas.

Como de costumbre, un ingeniero no puede tener una función de \mathbb{R}^2 en \mathbb{R} sin pensar inmediatamente en sus derivadas parciales; aquí hemos satisfecho esa imperiosa necesidad.

La propiedad (1) surge porque la JCDF es no-decreciente en ambos argumentos. La propiedad (2) es el teorema fundamental del cálculo y la propiedad (3) es el primer axioma, $P[\Omega]=1$. Las propiedades (1'), (2') y (3') surgen de las anteriores cuando la jpdf toma la forma $f_{XY}(x,y) = \sum_i \sum_j p_{ij} \delta(x-x_i) \delta(y-y_j)$.

La última frase de la definición es muy interesante, porque uno puede inventar un par de variables aleatorias conjuntas si inventa una función que satisfaga las dos propiedades mencionadas. Por ejemplo, las tres funciones de la Figura 79 satisfacen las propiedades, por lo que cada una de ellas es la jpdfs de un par de variables aleatorias conjuntas válidas.

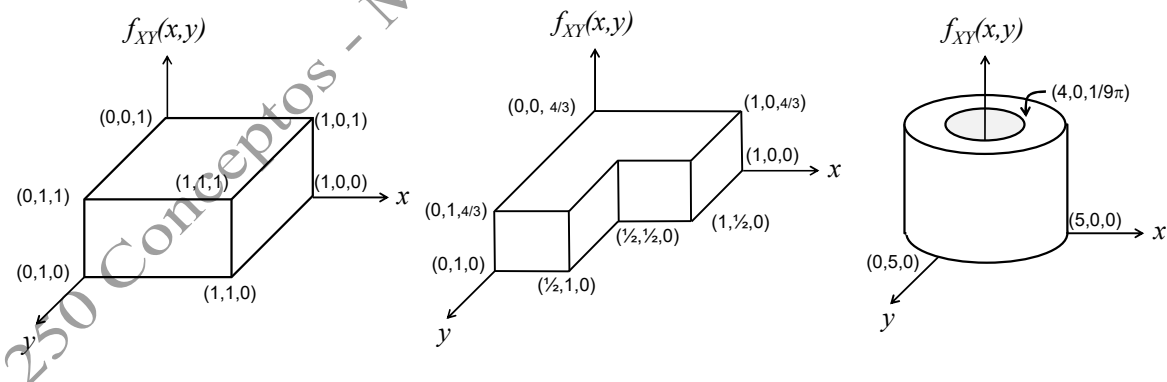


Figura 79. Tres funciones de \mathbb{R}^2 en \mathbb{R} que satisfacen las propiedades necesarias para se jpdfs válidas

La función
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \rho\frac{2(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right),$$

que corresponde a la forma más general de la jpdf de dos variables aleatorias conjuntamente

gaussianas, también satisface las dos propiedades $\iint f_{XY}(x,y)dxdy = 1$ y $f_{XY}(x,y) \geq 0$. Esta jpdf se muestra en la **Figura 80** cuando $\mu_x=1, \mu_y=-1, \sigma_x^2=0.25, \sigma_y^2=1$ y $\rho=0.6$.

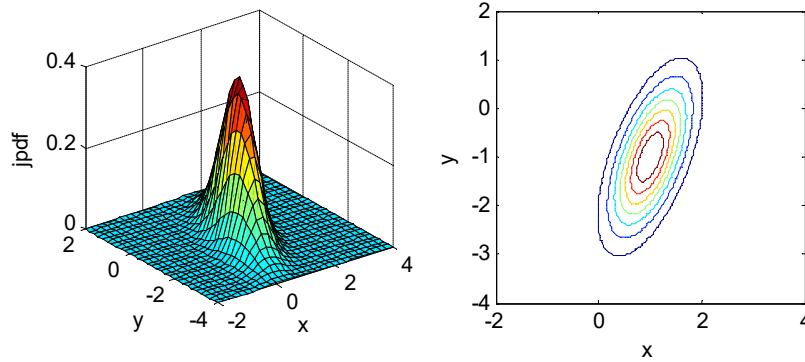


Figura 80. Distribución bivariada conjuntamente gaussiana

Por ejemplo, ¿Cuál es la JCDF en el caso de la segunda jpdf de la Figura 79?

$$F_{XY}(x,y) = \int_0^x \int_0^y f_{XY}(a,b) db da = \begin{cases} \frac{4}{3} \int_0^x \int_0^y db da = \frac{4xy}{3} & ((0 \leq x \leq 1) \cap (0 \leq y \leq 1/2)) \cup ((0 \leq x \leq 1/2) \cap (1/2 < y \leq 1)) \\ \frac{4xy}{3} - \frac{4}{3} \int_{1/2}^x \int_{1/2}^y db da = \frac{2x+2y-1}{3} & (1/2 < x \leq 1) \cap (1/2 < y \leq 1) \end{cases}$$

Por supuesto, este cálculo obedece a que la probabilidad de que las dos variables aleatorias caigan dentro de cualquier conjunto de Borel de \mathbb{R}^2 , B , es $P[(X,Y) \in B] = \iint_B d^2 F_{X,Y}(x,y)$.

62. Funciones marginales de densidad de probabilidad y de distribución de probabilidad

Sean X y Y dos variables aleatorias con jpdf $f_{X,Y}(\cdot, \cdot)$. Entonces,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$

Sean X y Y dos variables aleatorias con jpmf $p_{i,j}$. Entonces,

$$P[x = x_i] = \sum_j p_{i,j}, \quad P[y = y_j] = \sum_i p_{i,j}$$

En efecto, como en la definición 61 vimos que $F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) dv du$ y en la definición 60 vimos que $F_X(x) = F_{X,Y}(x, \infty)$, basta con derivar esta última expresión con respecto a x y reemplazar la primera expresión para obtener el resultado:

$$f_X(x) = \frac{d}{dx} F_{X,Y}(x, \infty) = \frac{d}{dx} \left[\int_{-\infty}^x \int_{-\infty}^{\infty} f_{X,Y}(u,y) dy du \right] = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

Si consideramos variables mutuamente discretas, con $f_{X,Y}(x,y) = \sum_i \sum_j p_{ij} \delta(x-x_i) \delta(y-y_j)$, obtenemos

$$\begin{aligned} \sum_i P[x = x_i] \delta(x - x_i) &= \frac{d}{dx} \left[\int_{-\infty}^x \int_{-\infty}^{\infty} \sum_i \sum_j p_{ij} \delta(u - x_i) \delta(y - y_j) dy du \right] \\ &= \int_{-\infty}^{\infty} \sum_i \sum_j p_{ij} \delta(x - x_i) \delta(y - y_j) dy = \sum_i \left(\sum_j p_{ij} \right) \delta(x - x_i) \end{aligned}$$

Igualando los coeficientes de $\delta(x-x_i)$ obtenemos el resultado deseado, $P[x = x_i] = \sum_j p_{ij}$.

Por ejemplo, consideremos la siguiente función de densidad de probabilidad conjunta:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)$$

que corresponde a dos variables conjuntamente normales. Para encontrar las marginales, observemos que el argumento del exponente en $f_{XY}(x, y)$ se puede completar como la suma de dos cuadrados:

$$\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} = \left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2 + (1-\rho^2)\frac{(y-\mu_2)^2}{\sigma_2^2}$$

De manera que

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2 + (1-\rho^2)\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right] dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right) \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2\right] dx \\ &= \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right) \left(\frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-(\mu_1 + \rho\sigma_1(y-\mu_2)/\sigma_2)}{\sigma_1\sqrt{1-\rho^2}}\right)^2\right] dx\right) \\ &= \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right) \end{aligned}$$

Nótese que el último factor de la derecha en la tercera fila es uno, pues se trata de la probabilidad total de una variable aleatoria gaussiana con media $\mu_1 + \rho\sigma_1(y-\mu_2)/\sigma_2$ y varianza $\sigma_1^2(1-\rho^2)$. Así pues, Y es una variable aleatoria gaussiana con media μ_2 y varianza σ_2^2 . De igual manera se encuentra que X es una variable aleatoria gaussiana con media μ_1 y varianza σ_1^2 .

63. Probabilidades condicionales

Sean X y Y dos variables aleatorias con función de distribución acumulativa conjunta $F_{XY}(x, y)$ y con función de densidad de probabilidad conjunta $f_{XY}(x, y) = \partial^2 F_{XY}(x, y) / \partial x \partial y$. Entonces la **función de distribución acumulativa condicional** de X dado que $Y=y$ es

$$F_{X|Y}(x|y) = \frac{P[X \leq x, Y = y]}{P[Y = y]} = \begin{cases} \frac{\sum_{x_i \leq x} P_{X,Y}(x_i, y)}{P_Y(y)} & \text{Si } X \text{ y } Y \text{ son conjuntamente discretas} \\ \frac{\int_{-\infty}^x f_{XY}(s, y) ds}{f_Y(y)} & \text{Si } X \text{ y } Y \text{ son conjuntamente continuas} \end{cases}$$

La **función de densidad de probabilidad condicional** de X dado que $Y=y$ es

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Y la **función de distribución de probabilidad condicional** de X dado que $Y=y$ es

$$P_{X|Y}(x|y) = P[X = x | Y = y] = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

La notación $P[A|B]$ ya es bastante clara para nosotros a partir de la definición 18. Cuando se tiene una distribución conjunta, es posible definir el evento A en términos de sólo una de las variables y el evento B en términos de la otra variable. Por ejemplo, de acuerdo con la definición 18,

$$P[X \leq x | Y \leq y] = \frac{F_{X,Y}(x, y)}{F_Y(y)}$$

Sin embargo, es más común condicionar en una medición ya realizada, por ejemplo, en el evento $Y = y$. En este caso se habla de la **función de distribución acumulativa condicional de X dado que $Y = y$** ,

$$F_{X|Y}(x|y) = \frac{P[X \leq x, Y = y]}{P[Y = y]}$$

Las formas particulares del numerador y el denominador en la expresión anterior varían de acuerdo con el tipo de variables involucradas en la distribución conjunta. Por ejemplo, tratándose de variables aleatorias conjuntamente discretas, podemos hablar de la **función de distribución de probabilidad condicional** de X dado Y ,

$$P[X = x | Y = y] = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

donde $P_{X,Y}(x, y) = P[X=x, Y=y]$ es la función de distribución de probabilidad conjunta de X y Y (jpmf), y $P_Y(y) = P[Y=y]$ es la función de distribución de probabilidad marginal de Y (pmf). En este caso, la función de distribución acumulativa condicional de X dado que $Y = y$ es

$$F_{X|Y}(x|y) = \frac{\sum_{x_i \leq x} P_{X,Y}(x_i, y)}{P_Y(y)}$$

Tratándose de variables aleatorias conjuntamente continuas, definimos la **función de densidad de probabilidad condicional** de X dado Y ,

$$f_{X|Y}(x|y) = \frac{d}{dx} F_{X|Y}(x|y)$$

Para expresar de mejor manera estas funciones de densidad y de distribución acumulativa, consideremos la condición $y < Y < y + \Delta y$ en vez de la condición $Y = y$:

$$F_{x|y}(x|y < Y \leq y + \Delta y) = \frac{P[X \leq x, y < Y \leq y + \Delta y]}{P[y < Y \leq y + \Delta y]} = \frac{\int_{-\infty}^x \int_y^{y+\Delta y} f_{XY}(s, t) dt ds}{\int_y^{y+\Delta y} f_Y(t) dt} \approx \frac{\int_{-\infty}^x (\Delta y f_{XY}(s, y) ds)}{\Delta y f_Y(y)}$$

donde la aproximación surge de la interpretación de la densidad de probabilidad, $f_X(x)\Delta x \approx P(x < X \leq x + \Delta x)$, como se discutió en la definición 33. Esta aproximación es más precisa entre más pequeño sea el incremento Δy de manera que, tomando el límite cuando $\Delta y \rightarrow 0$, obtenemos

$$F_{x|y}(x|Y = y) = \frac{\int_{-\infty}^x f_{XY}(s, y) ds}{f_Y(y)}$$

Al derivar esta última expresión con respecto a x obtenemos la función de densidad de probabilidad condicional,

$$f_{x|y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Consideremos un ejemplo muy simple: Un usuario de una red de comunicaciones puede disponer de tres niveles de ancho de banda, 1 (bajo), 2 (medio) o 3 (alto), y puede experimentar tres niveles de retardo, 1 (bajo), 2 (medio) o 3 (alto). Al observar las frecuencias relativas de cada una de estas condiciones para muchos usuarios de la red, se hicieron las siguientes estimaciones de la función de distribución de probabilidad conjunta:

P(nivel BW, nivel Retardo)	Nivel de retardo 1	Nivel de retardo 2	Nivel de retardo 3
Nivel BW 1	0.01	0.07	0.24
Nivel BW 2	0.07	0.22	0.10
Nivel BW 3	0.20	0.08	0.01

De acuerdo con la definición 62, las distribuciones marginales se obtienen simplemente sumando las filas o las columnas:

$$\begin{aligned} P[\text{BW 1}] &= 0.32, & P[\text{BW 2}] &= 0.39, & P[\text{BW 3}] &= 0.29 \\ P[\text{Retardo 1}] &= 0.28, & P[\text{Retardo 2}] &= 0.37, & P[\text{Retardo 3}] &= 0.35 \end{aligned}$$

Y, si usamos estas marginales como condición, obtenemos las probabilidades condicionales. Por ejemplo, dividiendo cada columna por las marginales de cada retardo obtenemos la probabilidad de cada ancho de banda condicionado a cada retardo:

P(BW Retardo)	Retardo 1	Retardo 2	Retardo 3
BW 1	0.01/0.28=0.0357	0.07/0.37=0.1892	0.24/0.35=0.6857
BW 2	0.07/0.28=0.2500	0.22/0.37=0.5946	0.10/0.35=0.2857
BW 3	0.20/0.28=0.7143	0.08/0.37=0.2162	0.01/0.35=0.0286

Y dividiendo cada fila por las marginales de cada ancho de banda obtenemos la probabilidad de cada retardo condicionado a cada ancho de banda:

P(Retardo BW)	Retardo 1	Retardo 2	Retardo 3
BW 1	0.01/0.32=0.03125	0.07/0.32=0.21875	0.24/0.32=0.75
BW 2	0.07/0.39=0.1795	0.22/0.39=0.5641	0.10/0.39=0.2564
BW 3	0.20/0.29=0.6896	0.08/0.29=0.2759	0.01/0.29=0.0345

64. Probabilidad total y Regla de Bayes

Al aplicar el teorema de la definición 19 a dos variables aleatorias conjuntamente distribuidas, obtenemos

$$dF_X(x) = \int_{y \in \mathbb{R}} dF_{X|Y}(x|y) dF_Y(y)$$

Y la regla de la definición 20 resulta

$$dF_{Y|X}(y|x) = \frac{dF_{X|Y}(x|y) dF_Y(y)}{\int_{s \in \mathbb{R}} dF_{X|Y}(x|s) dF_Y(s)}$$

En particular, dependiendo de las formas discreta o continua de cada variable, el teorema de la probabilidad total toma las siguientes expresiones:

	Y discreta	Y continua
X discreta	$P[X = x_i] = \sum_j P[X = x_i Y = y_j] P[Y = y_j]$	$P[X = x_i] = \int_{-\infty}^{\infty} P[X = x_i Y = y] f_Y(y) dy$
X continua	$f_X(x) = \sum_j f_{X Y}(x y_j) P[Y = y_j]$	$f_X(x) = \int_{-\infty}^{\infty} f_{X Y}(x y) f_Y(y) dy$

O, en términos de la CDF,

	Y discreta	Y continua
X discreta	$F_X(x) = \sum_{i: x_i \leq x} \sum_j P[X = x_i Y = y_j] P[Y = y_j]$	$F_X(x) = \sum_{i: x_i \leq x} \int_{-\infty}^{\infty} P[X = x_i Y = y] f_Y(y) dy$
X continua	$F_X(x) = \sum_j F_{X Y}(x y_j) P[Y = y_j]$	$F_X(x) = \int_{\mathbb{R}} F_{X Y}(x y) f_Y(y) dy$

La regla de Bayes también toma distintas formas dependiendo del tipo de variables involucradas:

	Y discreta	Y continua
X discreta	$P[Y = y_j X = x_i] = \frac{P[X = x_i Y = y_j]P[Y = y_j]}{\sum_k P[X = x_i Y = y_k]P[Y = y_k]}$	$f_{Y X}(y x_i) = \frac{P[X = x_i Y = y]f_Y(y)}{\int_{\mathbb{R}} P[X = x_i Y = s]f_Y(s)ds}$
X continua	$P[Y = y_j X = x] = \frac{f_{X Y}(x y_j)P[Y = y_j]}{\sum_k f_{X Y}(x y_k)P[Y = y_k]}$	$f_{Y X}(y x) = \frac{f_{X Y}(x y)f_Y(y)}{\int_{\mathbb{R}} f_{X Y}(x s)f_Y(s)ds}$

Veamos un ejemplo sencillo.

En el extremo transmisor de una fibra óptica existe un diodo emisor de luz (LED) que se prende para transmitir un uno y se apaga para transmitir un cero. En el extremo receptor existe un fotodetector que cuenta cuántos fotones salen de la fibra durante el período de transmisión de un bit (ver Figura 81).



Figura 81. Sistema de comunicación óptica

Una de las variables aleatorias que participan en este problema es $X \in \{0,1\}$, el bit generado en el transmisor. Se trata de una variable aleatoria de Bernoulli que corresponde al estado del LED: $X=1$ si el LED está prendido, lo cual ocurre con probabilidad p , y $X=0$ si el LED está apagado, lo cual ocurre con probabilidad $1-p$. Esta variable no la podemos observar desde el extremo receptor de la fibra, pero es exactamente la variable que deseamos conocer. El parámetro p de la distribución de Bernoulli a la que obedece X se conoce como la probabilidad "a priori" de que el LED esté encendido, pues habla de nuestra incertidumbre con respecto a X antes de haber visto una señal correlacionada con X que nos reduzca la incertidumbre. La otra variable aleatoria en este problema es $N \in \{0,1,2,\dots\}$, el número de fotones que inciden en el fotodetector durante el período de un bit. Para ella tenemos un modelo que determina el tipo de distribución según el estado del LED: Cuando el LED está encendido, el número de fotones que llegan al detector es una variable aleatoria geoméricamente distribuida con parámetro u_1 . Cuando el LED está apagado, el número de fotones que llegan al detector es una variable aleatoria geoméricamente distribuida con parámetro u_0 . Estas distribuciones condicionales se llaman "likelihood" –o verosimilitud-. Por supuesto, lo que queremos es saber cuál es la probabilidad de que el LED esté prendido o apagado una vez hemos contado el número de fotones que llegan al fotodetector, que es lo que se denomina probabilidad "a posteriori", pues corresponde a la incertidumbre sobre el estado del LED una vez hemos contado el número de fotones

recibidos en el fotodetector. La pregunta que nos queremos hacer es simple: Si llegan $N=k$ fotones al detector, ¿qué decisión deberíamos tomar con respecto al estado del LED, esto es, con respecto al bit transmitido? Este es un tipo de problema muy general en muchas disciplinas que, por lo mismo, recibe diferentes nombres: En teoría de señales se llama el problema de la detección, en Machine Learning se llama el problema de la clasificación, en estadística se llama el problema de la prueba de hipótesis, en administración se llama el problema de la decisión, etc.

La solución del problema depende del criterio de optimalidad que escogamos. Un criterio elemental es tratar de minimizar la probabilidad de equivocarnos, esto es, debemos escoger el estado más probable del LED dada la medición del número de fotones. En otras palabras, queremos escoger el estado del LED que maximiza la probabilidad a posteriori, o criterio MAP (Maximum A Posteriori). La idea es simple: Conocemos los likelihoods

$$P[N = k | X = 0] = (1 - u_0) u_0^k, \quad k = 0, 1, 2, \dots$$

$$P[N = k | X = 1] = (1 - u_1) u_1^k, \quad k = 0, 1, 2, \dots$$

con $0 < u_0 < u_1$. Con estas distribuciones condicionales y las probabilidades a priori podemos calcular las probabilidades a posteriori usando la regla de Bayes:

$$\begin{aligned} P[X = 0 | N = k] &= \frac{P[N = k | X = 0] P[X = 0]}{P[N = k | X = 0] P[X = 0] + P[N = k | X = 1] P[X = 1]} \\ &= \frac{(1 - p) u_0^k (1 - u_0)}{(1 - p) u_0^k (1 - u_0) + p u_1^k (1 - u_1)} \end{aligned}$$

$$\begin{aligned} P[X = 1 | N = k] &= \frac{P[N = k | X = 1] P[X = 1]}{P[N = k | X = 0] P[X = 0] + P[N = k | X = 1] P[X = 1]} \\ &= \frac{p u_1^k (1 - u_1)}{(1 - p) u_0^k (1 - u_0) + p u_1^k (1 - u_1)} \end{aligned}$$

El criterio de optimalidad consiste en escoger el estado del LED más probable dado el número de fotones que inciden en el detector, lo cual se puede indicar mediante la siguiente notación:

$$\begin{array}{c} \text{"0"} \\ P[X=1|N=k] \lesseqgtr P[X=0|N=k] \\ \text{"1"} \end{array}$$

que se lee así: "Si, dado que $N=k$, la probabilidad de $X=1$ es mayor a la probabilidad de $X=0$, escoja el estimador $\hat{X} = 1$. En otro caso, escoja el estimador $\hat{X} = 0$ ". De esta manera maximizamos la probabilidad a posteriori o, dicho de otra manera, minimizamos la probabilidad de equivocarnos. Como los denominadores de las probabilidades a posteriori son iguales para $X=1$ y para $X=0$, basta con hacer la comparación en los numeradores:

$$\begin{array}{c} \text{"0"} \\ p u_1^k (1 - u_1) \lesseqgtr (1 - p) u_0^k (1 - u_0) \\ \text{"1"} \end{array}$$

$$\left(\frac{u_1}{u_0}\right)^k \underset{"1"}{\overset{"0"}{\leq}} \frac{1-p}{p} \frac{1-u_0}{1-u_1}$$

Tomando logaritmo natural a ambos lados

$$k(\ln(u_1) - \ln(u_0)) \underset{"1"}{\overset{"0"}{\leq}} \ln(1-p) - \ln(p) + \ln(1-u_0) - \ln(1-u_1)$$

Y, por último, el criterio de decisión para minimizar la probabilidad de equivocarnos al detectar el bit transmitido es

$$k \underset{"1"}{\overset{"0"}{\leq}} \gamma$$

donde

$$\gamma = \frac{\ln(1-p) - \ln(p) + \ln(1-u_0) - \ln(1-u_1)}{\ln(u_1) - \ln(u_0)}$$

Si llegan más de γ fotones al detector, es más probable que el LED haya estado prendido, pero si llegan menos de γ fotones, lo más probable es que el LED haya estado apagado. Por supuesto, hay una probabilidad de equivocarnos que corresponde a los casos en que el LED esté prendido y lleguen menos de γ fotones, o que el LED esté apagado y lleguen más de γ fotones. En ambos casos tomaremos la decisión equivocada de acuerdo con nuestro criterio:

$$P[Error] = (1-p)P[N > \gamma | X=0] + pP[N < \gamma | X=1] = (1-p)u_0^{\text{ceil}(\gamma)} + p(1-u_1)^{\text{ceil}(\gamma)}$$

El valor de comparación que escogimos es el que minimiza dicho error. La siguiente figura muestra la probabilidad de error en función del valor de comparación, γ , cuando $p=0.7$, $u_0=0.95$ y $u_1=0.99$. Efectivamente, el valor óptimo γ^* corresponde a $(\ln(1-p) + \ln(1-u_0) - \ln(p) - \ln(1-u_1)) / (\ln(u_1) - \ln(u_0)) \approx 18.5$, de manera que si llegan 18 fotones o menos se debe suponer que $X=0$, pero si llegan 19 fotones o más se debe suponer que $X=1$.

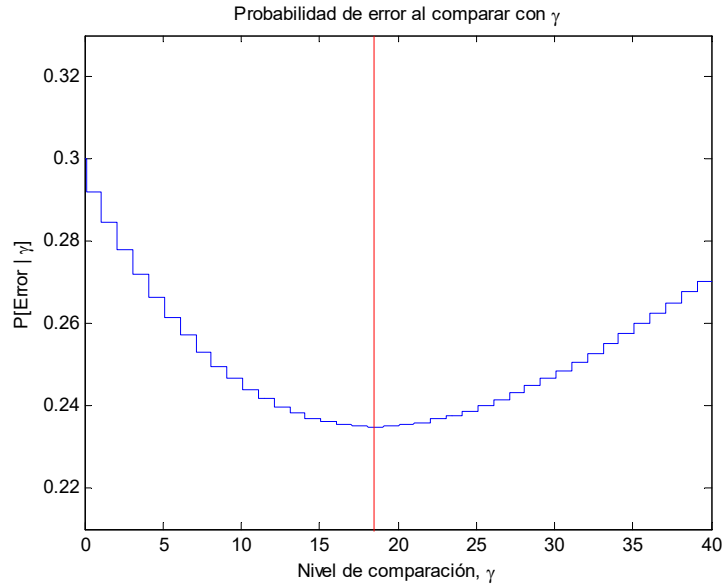


Figura 82. Criterio MAP para minimizar la probabilidad de error

Nótese que, una vez escogido el nivel de comparación óptimo, el sistema de transmisión de este ejemplo se convierte en un canal binario asimétrico, parecido al de la Figura 25, pero con diferentes probabilidades para los falsos negativos y para los falsos positivos:

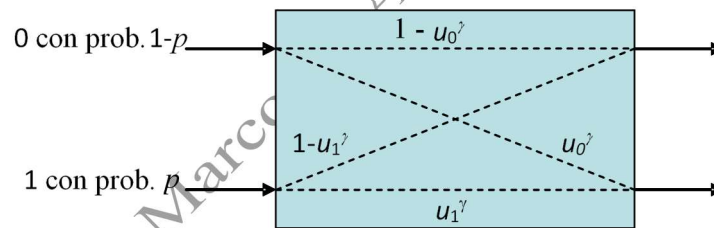


Figura 83. Canal binario asimétrico obtenido del ejemplo anterior

65. Independencia de variables aleatorias

Sean X y Y dos variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$. X y Y son independientes si, para cualquier par de conjuntos de Borel $B_X \in \mathcal{B}(\mathbb{R})$ y $B_Y \in \mathcal{B}(\mathbb{R})$,

$$P[\{\omega \in \Omega: X(\omega) \in B_X\} \cap \{\omega \in \Omega: Y(\omega) \in B_Y\}] = P[\{\omega \in \Omega: X(\omega) \in B_X\}]P[\{\omega \in \Omega: Y(\omega) \in B_Y\}] \quad (4)$$

La noción de independencia de la definición 21 se extiende así, de manera inmediata, al espacio de las variables aleatorias. Sin embargo, al tener que considerar todos los conjuntos de Borel B_X y B_Y , resulta muy difícil –si no imposible– determinar la independencia de dos variables aleatorias. Afortunadamente, como la medida de probabilidad del espacio original se reduce a la JCDF de las dos variables, se puede demostrar que la definición anterior equivale a la siguiente afirmación: Dos

variables aleatorias X y Y son independientes si y sólo si la JCDF es el producto de las CDF marginales:

$$F_{XY}(x, y) = F_X(x)F_Y(y) \quad (5)$$

De hecho, la Ecuación (5) surge de la definición misma de independencia (Ecuación (4)) si consideramos los eventos $A_X = \{\omega \in \Omega: X(\omega) \leq x\}$ y $A_Y = \{\omega \in \Omega: Y(\omega) \leq y\}$, para cualesquier x y y reales. Pero, como cualquier conjunto de Borel se puede expresar como la unión e intersección contable de conjuntos de esta forma (o de sus complementos), también podemos deducir la Ecuación (4) de la Ecuación (5). Por ejemplo, considere los siguientes dos eventos:

$$A_X = \{\omega \in \Omega: x_0 < X(\omega) \leq x_1\}, A_Y = \{\omega \in \Omega: y_0 < Y(\omega) \leq y_1\}$$

con $P[A_X] = F_X(x_1) - F_X(x_0)$ y $P[A_Y] = F_Y(y_1) - F_Y(y_0)$. Por la definición de la JCDF,

$$P[A_X \cap A_Y] = F_{XY}(x_1, y_1) - F_{XY}(x_1, y_0) + F_{XY}(x_0, y_0) - F_{XY}(x_0, y_1)$$

Así que, aplicando (9),

$$P[A_X \cap A_Y] = F_X(x_1) F_Y(y_1) - F_X(x_1) F_Y(y_0) + F_X(x_0) F_Y(y_0) - F_X(x_0) F_Y(y_1)$$

$$P[A_X \cap A_Y] = [F_X(x_1) - F_X(x_0)][F_Y(y_1) - F_Y(y_0)]$$

$$P[A_X \cap A_Y] = P[A_X] P[A_Y]$$

En principio, al considerar cualquier conjunto de Borel como union contable de rectángulos semejantes, un procedimiento semejante al anterior se podría usar para demostrar que la Ecuación (4) se puede deducir de la Ecuación (5).

Igualmente, la independencia de dos variables aleatorias conjuntamente continuas se puede definir mediante la relación

$$f_{XY}(x, y) = f_X(x)f_Y(y) \quad (6)$$

o, tratándose de variables conjuntamente discretas,

$$P[X = x, Y = y] = P[X = x]P[Y = y] \quad (7)$$

pues basta con derivar la JCDF una vez con respecto a x y otra vez con respecto a y .

De esta manera, las ecuaciones (5), (6) y (7) son definiciones equivalentes de independencia entre variables aleatorias, aunque aún existen más definiciones de independencia que, en ocasiones, pueden constituir pruebas más fáciles de aplicar:

- X y Y son independientes si y sólo si $F_{X|Y}(x|y) = F_X(x) \forall x, y \in \mathbb{R}$
- X y Y son independientes si y sólo si $f_{X|Y}(x|y) = f_X(x) \forall x, y \in \mathbb{R}$
- X y Y son independientes si y sólo si $F_{X|Y}(x|y_1) = F_{X|Y}(x|y_2) \forall x, y_1, y_2 \in \mathbb{R}$
- X y Y son independientes si y sólo si $f_{X|Y}(x|y_1) = f_{X|Y}(x|y_2) \forall x, y_1, y_2 \in \mathbb{R}$

En caso de variables mutuamente discretas podemos aplicar las siguientes definiciones/pruebas de independencia:

- X y Y son independientes si y sólo si $P[X=x_i|Y=y_j] = P[X=x_i] \forall i, j \in \mathbb{Z}$
- X y Y son independientes si y sólo si $P[X=x_i|Y=y_j] = P[X=x_i|Y=y_k] \forall i, j, k \in \mathbb{Z}$

Por último, recordemos las funciones de variables aleatorias tratadas en las definiciones 37, 38 y 39. Considerando dos variables aleatorias independientes, X y Y , podemos preguntarnos por la independencia de otras dos variables aleatorias $V = g(X)$ y $W = h(Y)$, donde g y h son funciones de los reales en los reales que transforman conjuntos de Borel en conjuntos de Borel. Sean dos conjuntos de Borel A_V y A_W cuyas imágenes inversas son $A_X = g^{-1}(A_V)$ y $A_Y = h^{-1}(A_W)$, respectivamente. Entonces,

$$P[V \in A_V, W \in A_W] = P[X \in A_X, Y \in A_Y] = P[X \in A_X] P[Y \in A_Y] = P[V \in A_V] P[W \in A_W]$$

de manera que V y W también son independientes.

A veces las variables aleatorias en cuestión están definidas en espacios de probabilidad diferentes que ni siquiera se encuentran relacionados entre ellos, esto es, X está definida en $(\Omega_X, \mathcal{F}_X, \mathbf{P}_X)$ y Y está definida en $(\Omega_Y, \mathcal{F}_Y, \mathbf{P}_Y)$. Por ejemplo, X puede ser el número de paquetes que circulan por una red durante una hora específica y Y puede ser la edad del sobrino mayor del ingeniero que opera la red. Intuitivamente resulta fácil imaginar que las dos variables son independientes, aunque no podemos aplicar la definición 65 por carecer de un espacio de probabilidad subyacente en común. Sin embargo basta con definir un nuevo experimento con espacio muestral $\Omega = \Omega_X \times \Omega_Y$ y campo sigma de eventos $\mathcal{F} = \sigma(\{A \times B, A \in \mathcal{F}_X, B \in \mathcal{F}_Y\})$. Para asignar la medida de probabilidad, como sólo tenemos $F_X(x)$ y $F_Y(y)$, podríamos definir $P[A \times B] = P_X[A] P_Y[B] \forall A \in \mathcal{F}_X, B \in \mathcal{F}_Y$, lo cual conduce a $F_{XY}(x,y) = F_X(x)F_Y(y)$: dos variables aleatorias independientes.

Consideremos los siguientes dos ejemplos: Sea $P[X=m, Y=n] = c/(m!n!)$, $0 \leq m, n \leq 10$. Sumando los 121 posible términos del rango de (X, Y) podemos encontrar el valor de la constante c :

$$\sum_{m=0..10} \sum_{n=0..10} P[X=m, Y=n] = 1$$

$$c \sum_{m=0..10} \sum_{n=0..10} 1/(m!n!) = c \sum_{m=0..10} (1/m!) \sum_{n=0..10} 1/(n!) = c [\sum_{n=0..10} (1/n!)]^2 = 1$$

de manera que $c = 1/(\sum_{n=0..10} 1/n!)^2$.

Al remplazar este valor en la *jpmf*,

$$P[X=m, Y=n] = (\sqrt[2]{c} / m!) (\sqrt[2]{c} / n!) = P[X=m] P[Y=n]$$

de donde X y Y son independientes. Otro ejemplo, muy parecido en principio, es en el que X y Y tienen la misma distribución conjunta $P[X=m, Y=n] = c_1/(m!n!)$, para los mismos valores de m y n pero sujetos a $0 \leq m+n \leq 10$. En este caso $P[X=10] = P[X=10, Y=0] = c_1/10!$ y $P[Y=10] = P[X=0, Y=10] = c_1/10!$, pero $P[X=10, Y=10] = 0 \neq (c_1/10!)^2$: En este caso X y Y ya no son independientes. Por supuesto, lo debimos sospechar al saber que $c \neq c_1$. ¿Cuál es el valor de c_1 ? Lo obtendremos de la normalización $\sum_{m=0..10} \sum_{n=0..10-m} P[X=m, Y=n] = 1$.

Otro ejemplo interesante es el que ya vimos en la definición 62. Allí demostramos que la distribución conjunta

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)$$

Corresponde a las distribuciones marginales

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) \quad \text{y} \quad f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)$$

Claramente, no es cierto que $f_{XY}(x, y) = f_X(x)f_Y(y)$, a menos que el parámetro adicional ρ sea cero. Este parámetro, llamado coeficiente de correlación (definición 69), será muy importante de aquí en adelante. En general, que ρ sea cero es una condición necesaria pero no suficiente para que las dos variables sean independientes. Sin embargo, para dos variables conjuntamente Gaussianas, que el coeficiente de correlación sea cero y que las variables sean independientes son dos condiciones equivalentes, pues cualquiera de ellas implica a la otra.

¿Cuál es la distribución condicional de Y dado X ? Veamos:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{XY}(x, y)}{f_X(x)} = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)}{\frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)} \\ &= \frac{1}{(\sigma_2\sqrt{1-\rho^2})\sqrt{2\pi}} \exp\left(\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right) \\ &= \frac{1}{(\sigma_2\sqrt{1-\rho^2})\sqrt{2\pi}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\rho\frac{x-\mu_1}{\sigma_1} - \frac{y-\mu_2}{\sigma_2}\right)^2\right) \\ &= \frac{1}{(\sigma_2\sqrt{1-\rho^2})\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y - \left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]\right)^2\right) \end{aligned}$$

Esto es, dado que la variable aleatoria X toma el valor x , la variable aleatoria Y es condicionalmente gaussiana con media $\mu_2 + \rho\sigma_2(x-\mu_1)/\sigma_1$ y varianza $\sigma_2^2(1-\rho^2)$. Similarmente, por simetría, dado que la variable aleatoria Y toma el valor y , la variable aleatoria X es condicionalmente gaussiana con media $\mu_1 + \rho\sigma_1(y-\mu_2)/\sigma_2$ y varianza $\sigma_1^2(1-\rho^2)$.

La independencia entre variables es deseable cuando debemos hacer diferentes mediciones de una cantidad y quisiéramos que esas mediciones sean valores independientes entre ellos. Entonces las sumas o los promedios de las medidas observadas tendrán propiedades interesantes que facilitarán mucho el análisis estadístico de las mediciones. Sin embargo, la falta de independencia también es deseable cuando deseamos recopilar información sobre una variable aleatoria (que no podemos medir

directamente) a partir de las medidas que hemos hecho de otra variable aleatoria, tal como en el ejemplo anterior de la transmisión por fibra óptica.

66. Una función de dos variables aleatorias

Sean X y Y dos variables aleatorias con JCDF $F_{XY}(x,y)$ y definamos una nueva variable aleatoria mediante la función $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ así: $Z = g(X,Y)$. Por supuesto, g debe ser una función que asocie un conjunto de Borel en \mathbb{R} , $\mathcal{B}(\mathbb{R})$, a cada conjunto de Borel en \mathbb{R}^2 , $\mathcal{B}(\mathbb{R}^2)$. Para cada $z \in \mathbb{R}$ definamos el conjunto de Borel en \mathbb{R}^2 , $B(z) = \{(x,y) \in \mathbb{R}^2 : g(x,y) \leq z\}$. Entonces

$$F_Z(z) = \iint_{B(z)} d^2 F_{XY}(x,y) = \begin{cases} \iint_{B(z)} f_{XY}(x,y) dy dx & \text{continuas} \\ \sum_{i,j: (x_i, y_j) \in B(z)} P_{XY}(x_i, y_j) & \text{discretas} \end{cases}$$

donde $f_{XY}(x,y) = \partial^2 F_{XY}(x,y) / \partial x \partial y$ es la jpdf de X y Y y $P_{XY}(x,y)$ es la jpmf de X y Y .

La Figura 84 ilustra el concepto propuesto. La variable aleatoria Z es una función $g(X,Y)$ de las variables aleatorias X y Y . En el espacio de probabilidad subyacente, un mismo evento medible se puede expresar de dos maneras diferentes: $C = \{\omega \in \Omega : Z(\omega) \leq z\} = \{\omega \in \Omega : g(X(\omega), Y(\omega)) \leq z\}$. La primera expresión se refiere al evento que se transforma en el segmento semi-infinito del eje Z mostrado en la gráfica de la izquierda. La segunda expresión se refiere al evento que se transforma en el anillo circular del plano (X,Y) mostrado en la gráfica de la derecha. De cualquier forma que se exprese el evento C , se trata del mismo evento C , por lo tanto, de la misma probabilidad.

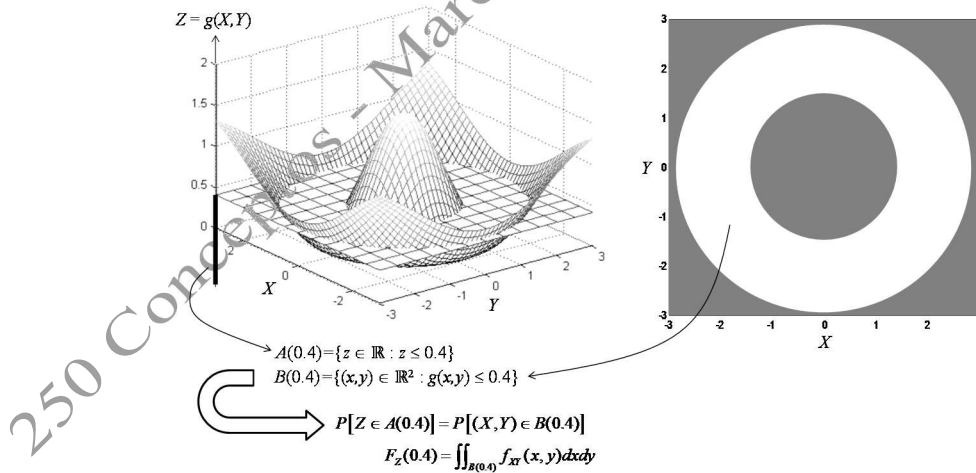


Figura 84. CDF de una función de dos variables aleatorias

Es interesante notar que no podemos dar una expresión genérica semejante para la pdf de Z , ya que la derivada de $F_Z(z) = \iint_{B(z)} f_{XY}(x,y) dy dx$ con respecto a z depende del conjunto $B(z)$. Sin embargo,

si la jpdf de X y Y es de la forma $\sum_i \sum_j p_{i,j} \delta(x-x_i) \delta(y-y_j)$, esto es, si se trata de dos variables conjuntamente discretas, la anterior relación se expresa como $F_Z(z) = \sum_{(i,j):g(x_i,y_j) \leq z} p_{i,j}$, lo cual conduce

$$a) P[Z = z_k] = \sum_{(i,j):g(x_i,y_j)=z_k} p_{i,j}.$$

Como hicimos en las definiciones 39 y 40, en muchas ocasiones no nos interesa tanto la distribución de $Z = g(X,Y)$ como algunas estadísticas. Por ejemplo, es fácil ver que

$$E[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) d^2 F_{XY}(x,y)$$

$$V[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g(x,y) - E[Z])^2 d^2 F_{XY}(x,y)$$

Un ejemplo muy interesante es cuando $g(X,Y) = X+Y$, como veremos en la siguiente definición.

67. Suma de dos variables aleatorias: Convolución.

Sean X y Y dos variables aleatorias con JCDF $F_{XY}(x,y)$ y definamos una nueva variable aleatoria $Z = X + Y$. Podemos conocer la distribución de Z :

$$F_Z(z) = \iint_{x+y \leq z} d^2 F_{XY}(x,y)$$

$$dF_Z(z) = \int_{x=-\infty}^{\infty} d^2 F_{XY}(x, z-x) = \int_{y=-\infty}^{\infty} d^2 F_{XY}(z-y, y)$$

Si X y Y son independientes, $d^2 F_{XY}(x,y) = dF_X(x) dF_Y(y)$ y, en consecuencia, la distribución de la suma es la convolución de las distribuciones de X y Y ,

$$dF_Z(z) = \int_{x=-\infty}^{\infty} dF_X(x) dF_Y(z-x) = \int_{y=-\infty}^{\infty} dF_Y(y) dF_X(z-y)$$

Un caso particular de la definición 66 es cuando tenemos la relación $Z = g(X,Y) = X+Y$, en cuyo caso, de acuerdo con dicha definición, $F_Z(z) = \iint_{x+y \leq z} d^2 F_{XY}(x,y)$. En este caso sí es fácil derivar con respecto a z para obtener

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(z-x|x) dx$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-y, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(z-y|y) f_Y(y) dy$$

que no es más que una expresión de la probabilidad total. Si $f_{XY}(x,y)$ es un conjunto de impulsos de Dirac en \mathbb{Z}_+^2 , esto es, si las variables aleatorias X y Y son conjuntamente discretas con valores en los enteros no negativos, las anteriores expresiones toman la siguiente forma:

$$P[Z = k] = \sum_{i=0}^{\infty} P[X = i, Y = k - i] = \sum_{i=0}^{\infty} P[X = i]P[Y = k - i | X = i]$$

$$P[Z = k] = \sum_{j=0}^{\infty} P[X = k - j, Y = j] = \sum_{j=0}^{\infty} P[X = k - j | Y = j]P[Y = j]$$

(Otra vez, una expresión del teorema de la probabilidad total). Un caso de sumo interés es la suma de variables aleatorias independientes, en cuyo caso las probabilidades condicionales de las expresiones anteriores son iguales a las respectivas probabilidades marginales:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy$$

$$P[Z = k] = \sum_{i=0}^{\infty} P[X = i]P[Y = k - i] = \sum_{j=0}^{\infty} P[X = k - j]P[Y = j]$$

Tal vez el lector recuerde estas ecuaciones como la integral y la suma de convolución de sus cursos de pregrado en señales y sistemas: Si consideramos la distribución de una variable como una señal de entrada que pasa a través de un sistema LTI (lineal e invariante en el tiempo) cuya respuesta al impulso es la distribución de la otra variable, a la salida tendremos la distribución de la suma (Figura 85).

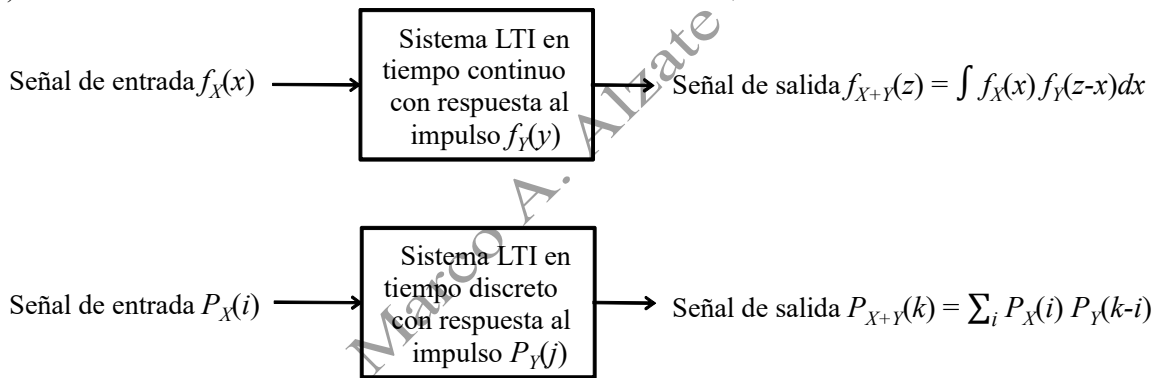


Figura 85. Integral y suma de convolución, distribución de la suma de dos variables aleatorias independientes.

Veamos algunos ejemplos interesantes sobre suma de variables aleatorias independientes:

1. Consideremos un multiplexor que concentra el tráfico proveniente de dos flujos de Poisson como el descrito en la definición 43(d) y los saca a través de una única línea de salida (Figura 86). Sea X el número de paquetes que llegan del primer flujo durante un período de t segundos y sea Y el número de paquetes que llegan del segundo flujo durante el mismo período. Si llamamos Z al número total de paquetes que llegan durante esos t para ser transmitidos por la línea de salida encontramos que $Z = X + Y$. X y Y son variables aleatorias independientes con las siguientes distribuciones

$$P[X = k] = \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t} \quad P[Y = m] = \frac{(\lambda_2 t)^m}{m!} e^{-\lambda_2 t}$$

¿Cuál es la distribución de Z ?

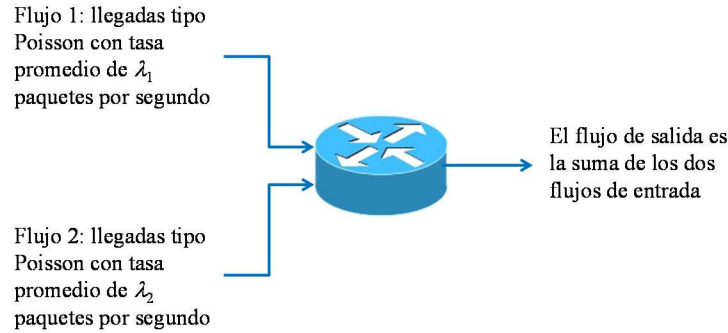


Figura 86. Multiplexor de paquetes

Queremos averiguar la distribución $P[Z=n]$, para lo cual usamos el teorema de la probabilidad total (definición 19), particionando el espacio muestral de acuerdo con el número de llegadas por la primera línea:

$$P[Z = n] = \sum_{k=0}^{\infty} P[X = k]P[Y = n - k | X = k]$$

Teorema de la probabilidad total

$$P[Z = n] = \sum_{k=0}^{\infty} P[X = k]P[Y = n - k]$$

X y Y son independientes

$$P[Z = n] = \sum_{k=0}^n \frac{\rho_1^k}{k!} e^{-\rho_1} \frac{\rho_2^{n-k}}{(n-k)!} e^{-\rho_2} \Big|_{\rho_1 = \lambda_1 t, \rho_2 = \lambda_2 t}$$

La suma sólo va hasta $k=n$ porque Y es una variable no negativa.

$$P[Z = n] = \frac{e^{-\rho_1 - \rho_2}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \rho_1^k \rho_2^{n-k} = \frac{e^{-(\rho_1 + \rho_2)}}{n!} \sum_{k=0}^n \binom{n}{k} \rho_1^k \rho_2^{n-k}$$

Multiplicando y dividiendo por $n!$ y reordenando, reconocemos la suma de la expansión binomial Z es una variable de Poisson con promedio igual a la suma de los promedios de X y Y

$$P[Z = n] = \frac{e^{-(\rho_1 + \rho_2)}}{n!} (\rho_1 + \rho_2)^n = \frac{\rho^n}{n!} e^{-\rho} \Big|_{\rho = \rho_1 + \rho_2}$$

El multiplexaje de flujos independientes de Poisson produce otro flujo de Poisson

$$P[Z = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \Big|_{\lambda = \lambda_1 + \lambda_2}$$

Como conclusión, repitamos la última frase: **El multiplexaje de flujos independientes de Poisson produce otro flujo de Poisson.** ¿Qué ocurrirá cuando demultiplexamos un flujo de Poisson? Veamos.

- Una aplicación importante de ingeniería de tráfico es el balanceo de carga mediante dispersión de tráfico. Supongamos, por ejemplo, un multiplexor al que llega un flujo de Poisson con intensidad de λ paquetes por segundo que debe ser repartido entre dos líneas de salida, como muestra la Figura 87.

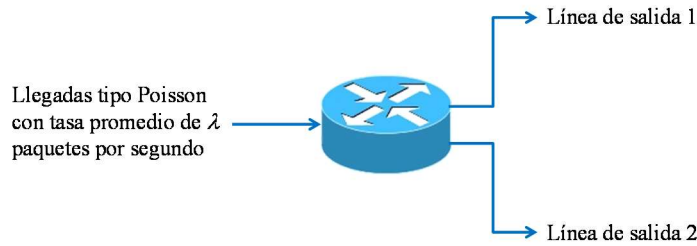


Figura 87. Demultiplexor de paquetes

Un mecanismo posible de dispersión equitativa de tráfico es enviar el primer paquete por la línea de salida 1, el segundo paquete por la línea 2, el tercer paquete por la línea 1 y así sucesivamente, alternando las líneas de salida con cada nuevo paquete que llegue. En este caso, el tiempo entre dos salidas consecutivas por cada una de las líneas de salida corresponde a la suma de dos tiempos entre llegadas consecutivas por la línea de entrada, como muestra la Figura 88.

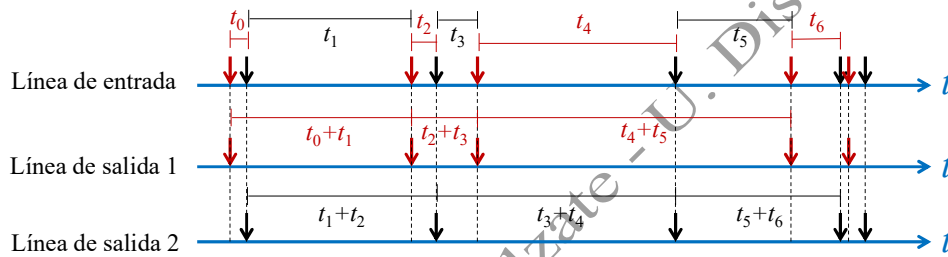


Figura 88. Esquema de dispersión de tráfico por turnos estrictos

Sea $\{t_0, t_1, t_2, t_3, \dots\}$ la secuencia de tiempos entre llegadas consecutivas de paquetes a la línea de entrada. De acuerdo con el concepto de llegadas tipo Poisson de la definición 43(d), esta es una secuencia de variables aleatorias independientes exponencialmente distribuidas con parámetro λ . La secuencia de tiempos entre salidas consecutivas de paquetes por la línea de salida 1 es $\{t_0+t_1, t_2+t_3, \dots\}$, esto es cada tiempo entre salidas es la suma dos variables iid: Si T es el tiempo entre dos salidas consecutivas por una de las líneas de salida y $f_E(\cdot)$ es la pdf exponencial,

$$f_T(t) = \int_{-\infty}^{\infty} f_E(\tau) f_E(t-\tau) d\tau = \lambda^2 \int_0^t e^{-\lambda\tau} e^{-\lambda(t-\tau)} d\tau = \lambda^2 \int_0^t e^{-\lambda t} d\tau = \lambda^2 t e^{-\lambda t}, \quad t \geq 0$$

No se trata de una variable aleatoria exponencial, de manera que el proceso de salida por cada línea no es un proceso de Poisson. Efectivamente, la distribución que obtuvimos es la distribución Erlang con parámetros 2 y λ . En general, la distribución de Erlang con parámetros k y λ corresponde a

$$f_k(t) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}$$

que se puede interpretar como la suma de k variables aleatorias exponenciales iid con promedio $1/\lambda$ o, lo que es lo mismo, el tiempo que toca esperar hasta ver el k -ésimo evento en un proceso de Poisson con tasa promedio de λ eventos por segundo.

¿Porqué el multiplexaje de procesos de Poisson es otro proceso de Poisson mientras el demultiplexaje de un proceso de Poisson no resultó en dos procesos de Poisson? Nótese que en el ejemplo anterior establecimos que el multiplexaje de procesos de Poisson **independientes** es otro proceso de Poisson. Lo que acabamos de construir fue dos procesos dependientes, pues los tiempos entre salidas por cada línea de salida tienen un término en común (t_0+t_1 no puede ser independiente de t_1+t_2 por el término en común t_1). Veamos qué pasa si restablecemos la independencia entre los procesos de salida.

3. Considere el mismo demultiplexor de la Figura 87 pero con el siguiente mecanismo de demultiplexaje: Con cada nuevo paquete de entrada el demultiplexor lanza una moneda para decidir el correspondiente enlace de salida. Si la moneda cae en cara (con probabilidad p), el paquete saldrá por el enlace de salida 1. Si cae en sello (con probabilidad $1-p$), el paquete saldrá por el enlace de salida 2.

Para encontrar la distribución del tiempo entre salidas de paquetes consecutivos por el enlace de salida 1 cuando usamos este nuevo mecanismo, consideremos la cola de la distribución, $P[T > t]$. Para que transcurran t segundos sin que salga ningún paquete por la línea 1 es necesario que no haya llegado ningún paquete a la línea de entrada o que hayan llegado uno o más pero todos hayan sido dirigidos a la línea 2:

$$P[T > t] = \sum_{k=0}^{\infty} P[k \text{ llegadas a la línea de entrada}] P[k \text{ sellos}] = \sum_{k=0}^{\infty} \left(\frac{(\lambda t)^k}{k!} e^{-\lambda t} \right) ((1-p)^k)$$

$$= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{((1-p)\lambda t)^k}{k!} = e^{-\lambda t} e^{(1-p)\lambda t} = e^{-(p\lambda)t}$$

Esta es la cola de la distribución de una variable aleatoria exponencial con tasa $(p\lambda)$,

$$F_T(t) = 1 - e^{-(p\lambda)t}, \quad f_T(t) = p\lambda e^{-(p\lambda)t}$$

Lo cual indica que el proceso de salidas por la línea 1 es Poisson con tasa $p\lambda$ (y el proceso de salidas por la línea 2 es Poisson con tasa $(1-p)\lambda$). Si la multiplexión se hace al azar, restableciendo la independencia entre los flujos de salida, los procesos de salida por cada línea son de Poisson con la tasa de entrada multiplicada por la probabilidad de seleccionar la respectiva línea de salida. Como conclusión, repitamos la última frase: **El demultiplexaje independiente de un flujo de Poisson produce otros flujos de Poisson.**

68. Suma de dos variables aleatorias: Función generadora de momentos, función generadora de probabilidad y función característica

Sean X y Y dos variables aleatorias independientes, con funciones generadoras de momentos $M_X(s)$ y $M_Y(s)$, y definamos una nueva variable aleatoria $W = X + Y$. Entonces la función generadora de momentos de W es $M_W(s) = M_X(s)M_Y(s)$. Similarmente, la función generadora de probabilidad de W es $G_W(z) = G_X(z)G_Y(z)$ y su función característica es $\Phi_W(\omega) = \Phi_X(\omega)\Phi_Y(\omega)$.

En efecto, como

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x) dx$$

entonces

$$M_W(s) = E[e^{sW}] = \int_{-\infty}^{\infty} e^{sw} f_W(w) dw$$

$$M_W(s) = \int_{-\infty}^{\infty} e^{sw} \int_{-\infty}^{\infty} f_X(x) f_Y(w-x) dx dw$$

$$M_W(s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{sx} f_X(x) e^{s(w-x)} f_Y(w-x) dx dw$$

$$M_W(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx \int_{-\infty}^{\infty} e^{s(w-x)} f_Y(w-x) dw$$

$$M_W(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx \int_{-\infty}^{\infty} e^{sy} f_Y(y) dy$$

$$M_W(s) = E[e^{sX}] E[e^{sY}]$$

$$M_W(s) = M_X(s) M_Y(s)$$

Si se trata de dos variables conjuntamente discretas, como

$$P_W(w) = \sum_{x=-\infty}^{\infty} P_X(x) P_Y(w-x)$$

entonces

$$G_W(z) = E[z^W] = \sum_{w=-\infty}^{\infty} z^w P_W(w)$$

$$G_W(z) = \sum_{w=-\infty}^{\infty} z^w \sum_{x=-\infty}^{\infty} P_X(x) P_Y(w-x)$$

$$G_W(z) = \sum_{w=-\infty}^{\infty} \sum_{x=-\infty}^{\infty} z^x P_X(x) z^{w-x} P_Y(w-x)$$

$$G_W(z) = \sum_{x=-\infty}^{\infty} z^x P_X(x) \sum_{w=-\infty}^{\infty} z^{w-x} P_Y(w-x)$$

$$G_W(z) = \sum_{x=-\infty}^{\infty} z^x P_X(x) \sum_{y=-\infty}^{\infty} z^y P_Y(y)$$

$$G_W(z) = E[z^X] E[z^Y]$$

$$G_W(z) = G_X(z) G_Y(z)$$

Al evaluar la expresión $M_W(s) = M_X(s)M_Y(s)$ en $s=j\omega$ o al evaluar la expresión $G_W(z) = G_X(z)G_Y(z)$ en $z=e^{j\omega}$, obtenemos los correspondientes resultados para la función característica, $\Phi_W(\omega) = \Phi_X(\omega)\Phi_Y(\omega)$.

Consideremos, por ejemplo, dos variables independientes de Poisson con parámetros ρ_1 y ρ_2 . Es fácil deducir la función generadora de probabilidad de la distribución de Poisson:

$$E[z^X] = \sum_{n=0}^{\infty} z^n P[X=n] = \sum_{n=0}^{\infty} z^n \frac{\rho^n}{n!} e^{-\rho} = e^{-\rho} \sum_{n=0}^{\infty} \frac{(z\rho)^n}{n!}$$

La última sumatoria es la definición formal de la función exponencial:

$$E[z^X] = e^{-\rho} \sum_{n=0}^{\infty} \frac{(z\rho)^n}{n!} = e^{-\rho} e^{z\rho} = e^{-\rho(1-z)}$$

En consecuencia, el producto de las funciones generadoras de probabilidad de dos variables aleatorias de Poisson X_1 y X_2 , con parámetros ρ_1 y ρ_2 , respectivamente, es

$$E[z^{X_1}]E[z^{X_2}] = e^{-\rho_1(1-z)}e^{-\rho_2(1-z)} = e^{-(\rho_1+\rho_2)(1-z)}$$

que es la función generadora de probabilidad de una variable aleatoria de Poisson con parámetro $\rho_1+\rho_2$. Con esto hemos vuelto a demostrar que la suma de variables aleatorias de Poisson independientes es una variable aleatoria de Poisson, notando que resulta más fácil verificarlo desde la función generadora de probabilidad que desde la función de distribución de probabilidad.

Igualmente, la función generadora de momentos de una variable aleatoria exponencial con promedio $1/\lambda$ es

$$E[e^{sx}] = \lambda \int_0^{\infty} e^{sx} e^{-\lambda x} dx = \frac{\lambda}{\lambda-s} e^{-(\lambda-s)x} \Big|_0^{\infty} = \frac{\lambda}{\lambda-s} \quad \text{si } \lambda > s$$

De manera que la suma de dos variables aleatorias exponenciales independientes e idénticamente distribuidas tiene como función generadora de momentos $(\lambda/(\lambda-s))^2$, si $\lambda > s$. En general, la función generadora de momentos de la distribución Erlang(k, λ) es $(\lambda/(\lambda-s))^k$.

Finalmente, consideremos la suma de dos variables aleatorias gaussianas independientes. La función generadora de momentos de una variable aleatoria gaussiana con media μ y varianza σ^2 es

$$E[e^{sx}] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2 - 2\sigma^2 sx}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-s\sigma)^2 - s\sigma^2(s\sigma^2+2\mu)}{2\sigma^2}} dx = \dots$$

$$E[e^{sx}] = \frac{e^{s(s\sigma^2+2\mu)/2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-s\sigma)^2}{2\sigma^2}} dx = e^{s(s\sigma^2+2\mu)/2}$$

Si sumamos dos variables aleatorias gaussianas independientes X_1 y X_2 con medias μ_1 y μ_2 respectivamente, y con varianzas σ_1^2 y σ_2^2 respectivamente, obtenemos que la función generadora de momentos de la suma es

$$E[e^{s(X_1+X_2)}] = e^{s(s\sigma_1^2+2\mu_1)/2} e^{s(s\sigma_2^2+2\mu_2)/2} = e^{s(s(\sigma_1^2+\sigma_2^2)+2(\mu_1+\mu_2))/2}$$

que es la función generadora de momentos de un variable aleatoria Gaussiana con media $(\mu_1 + \mu_2)$ y varianza $(\sigma_1^2 + \sigma_2^2)$: La suma de variables aleatorias Gaussianas independientes también es una variable aleatoria Gaussiana.

Es interesante considerar también las transformaciones asociadas directamente con la distribución conjunta de dos variables aleatorias, como veremos en breve (definición 74).

69. Suma de dos variables aleatorias: Covarianza, correlación y coeficiente de correlación

Sean X y Y dos variables aleatorias con JCDF $F_{XY}(x,y)$ y definamos una nueva variable aleatoria $Z = X + Y$. Entonces

$$E[Z] = E[X] + E[Y]$$

$$V[Z] = E[X] + E[Y] + 2Cov[X,Y]$$

donde $Cov[X,Y] = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$ es la covarianza entre X y Y . En la definición de la covarianza aparece el término $E[XY]$, que se conoce como la correlación entre X y Y , $Corr[X,Y]$. El coeficiente de correlación entre X y Y se define mediante la expresión $\rho_{XY} = Cov[X,Y]/\sigma_X\sigma_Y$. Si $Cov[X,Y] = 0$ se dice que X y Y son no-correlacionados. Si $E[XY] = 0$ se dice que X y Y son ortogonales. Si X y Y son independientes, $E[XY] = E[X]E[Y]$ y, en consecuencia, X y Y son no-correlacionados. Si X y Y son independientes y $E[X]E[Y] = 0$, X y Y son ortogonales. Puede haber variables no-correlacionadas que no sean independientes.

A veces la distribución de la suma de dos variables aleatorias puede ser una información más amplia que la que realmente necesitamos pues, en ocasiones, basta con conocer sólo las principales estadísticas de la suma. Por ejemplo, consideremos el valor esperado de la suma:

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) d^2 F_{XY}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x d^2 F_{XY}(x,y) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y d^2 F_{XY}(x,y) \\ &= \int_{x=-\infty}^{\infty} x \int_{y=-\infty}^{\infty} d^2 F_{XY}(x,y) + \int_{y=-\infty}^{\infty} y \int_{x=-\infty}^{\infty} d^2 F_{XY}(x,y) = \int_{x=-\infty}^{\infty} x dF_X(x) + \int_{y=-\infty}^{\infty} y dF_Y(y) \\ &= E[X] + E[Y] \end{aligned}$$

El valor esperado de la suma de dos variables aleatorias es la suma de los valores esperados, sin importar si las variables son independientes o no. Consideremos ahora la varianza:

$$\begin{aligned} V[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x+y) - (E[X] + E[Y]))^2 d^2 F_{XY}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x - E[X]) + (y - E[Y]))^2 d^2 F_{XY}(x,y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x - E[X])^2 + (y - E[Y])^2 + 2(x - E[X])(y - E[Y])) d^2 F_{XY}(x,y) \\ &= V[X] + V[Y] + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) d^2 F_{XY}(x,y) \\ &= V[X] + V[Y] + 2E[(X - E[X])(Y - E[Y])] \\ &= V[X] + V[Y] + 2Cov[X,Y] \end{aligned}$$

La varianza de la suma de dos variables aleatorias es la suma de las varianzas más un término adicional, dos veces la **covarianza** entre las variables aleatorias X y Y :

$$Cov[X,Y] = E[(X-E[X])(Y-E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]] = E[XY] - E[X]E[Y]$$

La covarianza extiende el concepto de varianza a una forma de variación conjunta, pues $V[X] = \text{Cov}(X, X)$. Así como $V[X]$ mide la dispersión de X respecto a su valor medio, la covarianza mide cuánto varían conjuntamente dos variables aleatorias con respecto a sus correspondientes valores medios. Por ejemplo, si valores altos de X tienden a presentarse simultáneamente con valores altos de Y y los valores bajos de X tienden a presentarse simultáneamente con valores bajos de Y , la covarianza tendrá un valor positivo (la varianza de la suma será mayor que la suma de las varianzas). De otra manera, si los valores altos de una variable se dan simultáneamente con los valores bajos de la otra, la covarianza tendrá un valor negativo (la varianza de la suma será menor que la suma de las varianzas). Por eso decimos que la covarianza mide el grado de relación lineal entre dos variables.

Al definir $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$ también definimos la **correlación** entre X y Y , $\text{Corr}[X, Y] = E[XY]$. Si alguno de los sumandos X o Y (o ambos) tienen valor esperado cero, la covarianza y la correlación resulta idénticas.

Si la covarianza es cero (en cuyo caso se dice que X y Y son no-correlacionadas), la varianza de la suma será la suma de las varianzas. Este caso ocurre, por ejemplo, si X y Y son independientes, aunque puede ocurrir también sin necesidad de independencia. En efecto, si X y Y son independientes,

$$E[XY] = \iint xy d^2 F_{X,Y}(x, y) = \iint xy dF_X(x) dF_Y(y) = \int x dF_X(x) \int y dF_Y(y) = E[X]E[Y]$$

y, en consecuencia, $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 0$. Como dijimos, no es necesariamente cierto que si X y Y son no-correlacionadas entonces sean independientes, como ilustra el siguiente caso:

Sea X una variable aleatoria uniformemente distribuida en el intervalo $[-1, 1]$ y sea $Y = X^2$. Siendo Y una función de X , se trata de dos variables dependientes (dado el valor de X , Y queda completamente determinada y, dado el valor de Y , el rango de X se reduce a sólo dos posibles valores, $\pm\sqrt{Y}$). Sin embargo, X y Y son dos variables no correlacionadas, pues

$$E[X] = 0$$

$$E[Y] = E[X^2] = \frac{1}{2} \int_{-1}^1 x^2 dx = \frac{1}{3}$$

$$E[XY] = E[X^3] = \frac{1}{2} \int_{-1}^1 x^3 dx = 0.$$

En consecuencia, $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 0$: X y Y son no-correlacionadas, pero no son independientes.

Un concepto de gran interés es el del coeficiente de correlación, definido como

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

donde $\sigma_X = \sqrt{V[X]}$ es la desviación estándar de X y $\sigma_Y = \sqrt{V[Y]}$ es la desviación estándar de Y , y ambas se suponen mayores a cero (si alguna de ellas es cero, $\text{Cov}[X, Y] = 0$ y, entonces, ρ_{XY} es igual a cero por definición). Es fácil notar que, al normalizar así la covarianza, se obtiene un número entre -1 y 1 . En efecto, como el valor esperado de una variable aleatoria no negativa es un número no negativo, $E[(a(X-E[X])+(Y-E[Y]))^2] \geq 0 \quad \forall a \in \mathbb{R}$; Expandiendo el cuadrado y el valor esperado, $a^2 V[X] + 2a \text{Cov}[X, Y] + V[Y] \geq 0 \quad \forall a \in \mathbb{R}$. Como ésta es una forma cuadrática en a , tendrá un mínimo en

donde la derivada respecto a a se hace cero: $2a^*V[X]+2Cov[X,Y]=0 \Rightarrow a^* = -Cov[X,Y]/V[X]$. Reemplazando este valor de a en la desigualdad original, $Cov^2[X,Y]/V[X] - 2Cov^2[X,Y]/V[X] + V[Y] \geq 0$, esto es, $V[Y] \geq Cov^2[X,Y]/V[X]$. Dividiendo ambos lados de la desigualdad por $V[Y]$ obtenemos el resultado básico, $\rho_{XY}^2 \leq 1$ o, lo que es lo mismo,

$$-1 \leq \rho_{XY} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} \leq 1$$

¿Cuándo se logran estos valores extremos del coeficiente de correlación? Si $Y=aX+b$, $Cov[X,Y] = aV[X]$ y $V[Y] = a^2V[X]$, de manera que $\rho_{XY} = a/\sqrt{a^2} = a/|a| = \pm 1$ según el signo de a . Entonces $\rho_{XY} = +1$ si y sólo si Y se incrementa linealmente con X y $\rho_{XY} = -1$ si y sólo si Y se decrementa linealmente con X (sólo demostramos la implicación en un sentido, pero la implicación en el otro sentido se demuestra de manera semejante, aunque es un poco más elaborado). Nuevamente, esto indica que si los valores altos de X tienden a coincidir con los valores altos de Y y los valores bajos de X tienden a coincidir con los valores bajos de Y , el coeficiente de correlación será positivo. Si los valores altos de X tienden a coincidir con los valores bajos de Y y los valores bajos de X tienden a coincidir con los valores altos de Y , el coeficiente de correlación será negativo. Cuando el coeficiente de correlación es cero, se dice que X y Y son ortogonales³⁰. Como vimos en la definición 65, dos variables aleatorias conjuntamente gaussianas son independientes si son ortogonales, aunque en general la ortogonalidad no implica independencia mientras que la independencia sí implica ortogonalidad.

Al hablar de ortogonalidad se invoca un concepto geométrico. Y es que, desde un punto de vista geométrico, la covarianza, la correlación y el coeficiente de correlación toman significados mucho más claros e intuitivos. En el capítulo VI describiremos una interpretación geométrica de estos conceptos a la luz del álgebra lineal: Las variables aleatorias como vectores en un espacio vectorial abstracto (espacio de Hilbert). Veremos, por ejemplo, que si $-1 \leq \rho_{XY} \leq 1$ es porque ρ_{XY} se puede interpretar como el coseno del ángulo que forman los vectores X y Y y esa relación es un principio “geométrico” fundamental, conocido como desigualdad de Cauchy-Schwarz.

70. Dos funciones de dos variables aleatorias

Sean X y Y dos variables aleatorias definidas en un mismo espacio de probabilidad (Ω, \mathcal{A}, P) y sean $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ y $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ dos funciones en el plano. Al evaluar g y h en las variables aleatorias X y Y obtenemos dos nuevas variables aleatorias $V = g(X, Y)$ y $W = h(X, Y)$. Si $D(v, w) = \{(x, y) \in \mathbb{R}^2 \mid g(x, y) \leq v, h(x, y) \leq w\}$ entonces

$$F_{VW}(v, w) = \iint_{D(v, w)} dF_{XY}(x, y)$$

³⁰ Aquí aparece otro truco mnemónico: Dos variables aleatorias no-correlacionadas no necesariamente deben tener una correlación cero, pero deben tener una covarianza cero; y dos variables cuya correlación es cero no necesariamente son no-correlacionadas, pero son ortogonales. Si $E[X]E[Y]=0$, covarianza y correlación son iguales, de manera que, en ese caso, X y Y son ortogonales si y sólo si son no correlacionadas. En muchas ocasiones, al usar la correlación, se supone que se trabaja con variables de media cero o, explícitamente, se le resta el valor medio a cada variable aleatoria. Después de todo, las variables aleatorias X y Y tienen la misma covarianza y el mismo coeficiente de correlación que las variables aleatorias $(X-E[X])$ y $(Y-E[Y])$.

Nuevamente se trata, en principio, de considerar el evento infinitesimal $(x, x+dx) \times (y, y+dy)$ en el plano xy y la región correspondiente en el plano vw bajo las transformaciones g y h . La definición anterior parte de que la probabilidad de que (V, W) esté en la segunda región es la misma probabilidad de que (X, Y) esté en la primera región, como indica la Figura 89. Después basta con integrar en la región del plano (x, y) que hace que $g(x, y)$ sea menor que v y $h(x, y)$ sea menor que w para obtener la JCDF de V y W .

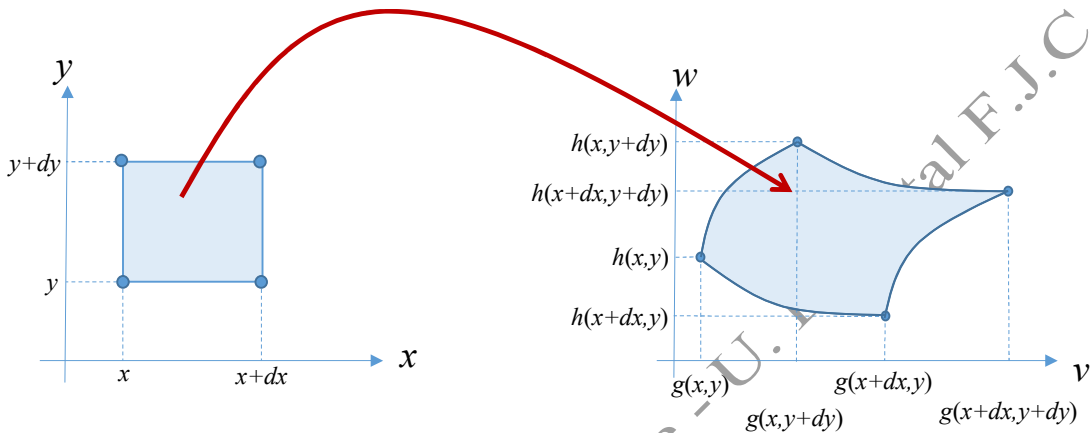


Figura 89. Dos funciones de dos variables

Si se trata de variables conjuntamente discretas, la anterior expresión se reduce a la relación obvia que describe la jpmf de V y W en términos de la jpmf de X y Y ,

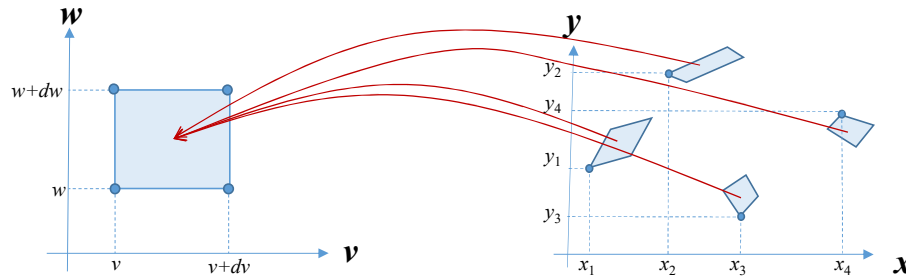
$$P(V = v, W = w) = \sum_{\substack{j: g(x_j, y_j) = v \\ h(x_j, y_j) = w}} P(X = x_j, Y = y_j)$$

Si se trata de variables conjuntamente continuas, la expresión anterior correspondería con las jpdf así:

$$f_{VW}(v, w) = \sum_{\substack{j: g(x_j, y_j) = v \\ h(x_j, y_j) = w}} \frac{f_{XY}(x_j, y_j)}{\left| \begin{array}{cc} \frac{\partial g(x_j, y_j)}{\partial x} & \frac{\partial g(x_j, y_j)}{\partial y} \\ \frac{\partial h(x_j, y_j)}{\partial x} & \frac{\partial h(x_j, y_j)}{\partial y} \end{array} \right|}$$

donde el determinante en el denominador es el Jacobiano de la transformación, $J(x_j, y_j)$, y la doble barra vertical se refiere al valor absoluto del determinante.

En efecto, si consideramos un rectángulo infinitesimal $(v, v+dv) \times (w, w+dw)$ en el plano (v, w) , alrededor de los puntos (x_j, y_j) tales que $g(x_j, y_j) = v$ y $h(x_j, y_j) = w$, habrá regiones irregulares alrededor de cada uno de esos puntos que se transforman en el rectángulo $(v, v+dv) \times (w, w+dw)$ y que tienden a ser paralelogramos infinitesimales a medida que dv y dw se hacen más y más pequeños, como muestra la Figura 90.



$$v = g(x_1, y_1) = g(x_2, y_2) = g(x_3, y_3) = g(x_4, y_4)$$

$$w = h(x_1, y_1) = h(x_2, y_2) = h(x_3, y_3) = h(x_4, y_4)$$

Figura 90. Regiones en (x, y) que se transforman en $(v, v+dv) \times (w, w+dw)$

Por cada (x_j, y_j) que satisface $v=g(x_j, y_j)$ y $w=h(x_j, y_j)$, existe un paralelogramo que se transforma en el rectángulo $(v, v+dv) \times (w, w+dw)$ y cuya área es $dv dw / |J(x_j, y_j)|$. De esta manera, la masa de probabilidad que hay en $(v, v+dv) \times (w, w+dw)$ es la suma de las probabilidades que hay en cada uno de los paralelogramos:

$$f_{vw}(v, w) dv dw = \sum_{j=1}^n f_{xy}(x_j, y_j) \frac{dv dw}{|J(x_j, y_j)|}$$

Y sólo falta dividir por $dv dw$ para obtener el resultado. De hecho, recordemos cómo en la definición 37 la longitud del intervalo dx se transformaba en el eje $y=g(x)$ a la longitud $dy/|g'(x)|$. Igualmente, el área de la región $dx dy$ se transforma en el plano (v, w) en el área $dv dw / |J(x, y)|$. Simplemente consideremos la aproximación de Taylor,

$$g(x + dx, y) = g(x, y) + \frac{\partial}{\partial x} g(x, y) dx + o(dx^2) \quad h(x + dx, y) = h(x, y) + \frac{\partial}{\partial x} h(x, y) dx + o(dx^2)$$

$$g(x, y + dy) = g(x, y) + \frac{\partial}{\partial y} g(x, y) dy + o(dy^2) \quad h(x, y + dy) = h(x, y) + \frac{\partial}{\partial y} h(x, y) dy + o(dy^2)$$

El ejemplo más simple es el de una transformación lineal, $V = aX + bY$, $W = cX + dY$, donde a, b, c y d son constantes reales. El jacobiano es $ad - bc$. Si este Jacobiano no es cero, la transformación inversa es $X = (dV - bW) / (ad - bc)$ y $Y = (aW - cV) / (ad - bc)$, de manera que la jpdf de V y W es

$$f_{vw}(v, w) = \frac{1}{|ad - bc|} f_{xy} \left(\frac{dV - bW}{ad - bc}, \frac{aW - cV}{ad - bc} \right)$$

Otro ejemplo típico en señales de telecomunicaciones es cuando disponemos de dos amplitudes aleatorias (los datos de entrada) y las modulamos sobre una única portadora mediante dos fases en cuadratura. Si los datos son dos variables gaussianas X y Y de media cero e igual varianza σ^2 , la señal modulada toma la forma $V(t) = X \cos(\omega t) + Y \sin(\omega t)$, donde $f = \omega / 2\pi$ es la frecuencia de la portadora. Es interesante averiguar por la amplitud (A) y la fase (Θ) de la señal modulada, $V(t) = A \cos(\omega t - \Theta)$. ¿Cómo están distribuidas? ¿Son independientes? Primero, es fácil encontrar un par de transformaciones que permitan el cambio de variables de X, Y a A, Θ .

$$\begin{aligned}
 V(t) &= X \cos(\omega t) + Y \sin(\omega t) = \sqrt{X^2 + Y^2} \left(\frac{X}{\sqrt{X^2 + Y^2}} \cos(\omega t) + \frac{Y}{\sqrt{X^2 + Y^2}} \sin(\omega t) \right) \\
 &= A (\cos(\Theta) \cos(\omega t) + \sin(\Theta) \sin(\omega t)) \left| \begin{array}{l} A = \sqrt{X^2 + Y^2} \\ \Theta = \cos^{-1}(X/A) = \tan^{-1}(Y/X) \end{array} \right. \\
 &= A \cos(\omega t - \Theta)
 \end{aligned}$$

de manera que la transformación es $A=g(X,Y)=(X^2+Y^2)^{1/2}$ y $\Theta=\tan^{-1}(Y/X)$. Dado un par (a, θ) , los valores correspondientes de x y y son $x = a \cos(\theta)$ y $y = a \sin(\theta)$. El jacobiano de esta transformación en un punto (x,y) transformado en (a, θ) es

$$\left[\begin{array}{cc} \frac{\partial g(x,y)}{\partial x} = \frac{x}{a} = \cos(\theta) & \frac{\partial g(x,y)}{\partial y} = \frac{y}{a} = \sin(\theta) \\ \frac{\partial h(x,y)}{\partial x} = -\frac{y}{a^2} = -\frac{1}{a} \sin(\theta) & \frac{\partial h(x,y)}{\partial y} = \frac{x}{a^2} = \frac{1}{a} \cos(\theta) \end{array} \right] = \frac{1}{a}$$

por lo que la distribución conjunta de A y Θ es

$$f_{A,\Theta}(a, \theta) = af_{XY}(a \cos(\theta), a \sin(\theta)) = \frac{a}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(x^2 + y^2)\right) \Bigg|_{\substack{x=a \cos(\theta) \\ y=a \sin(\theta)}} = \frac{a}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right)$$

La distribución conjunta no depende explícitamente de θ . Para determinar si son independientes, debemos encontrar las distribuciones marginales de A y de Θ :

$$f_A(a) = \int_0^{2\pi} f_{A\Theta}(a, \theta) d\theta = \frac{a}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right) \int_0^{2\pi} d\theta = \frac{a}{\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right)$$

$$f_\Theta(\theta) = \int_0^\infty f_{A\Theta}(a, \theta) da = \frac{1}{2\pi\sigma^2} \int_0^\infty a \exp\left(-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right) da = \frac{1}{2\pi}$$

La amplitud A tiene una distribución de Rayleigh con parámetro σ y la fase Θ tiene una distribución uniforme entre 0 y 2π . Lo más interesante es que la distribución conjunta es el producto de las marginales, de manera que A y Θ son independientes.

71. Algunas distribuciones conjuntas

La función de distribución conjunta más elemental es la de dos variables conjuntamente Bernoulli:

$$p_{00} = P[X=0, Y=0], \quad p_{01} = P[X=0, Y=1], \quad p_{10} = P[X=1, Y=0], \quad p_{11} = P[X=1, Y=1]$$

donde cada una es marginalmente Bernoulli con $P[X=1] = p_{10} + p_{11} = 1 - P[X=0]$ y $P[Y=1] = p_{01} + p_{11} = 1 - P[Y=0]$. Las probabilidades condicionales son Bernoulli con

$$\begin{aligned}
 P[X=1|Y=0] &= p_{10}/(p_{10} + p_{00}) = 1 - P[X=0|Y=0], & P[X=1|Y=1] &= p_{11}/(p_{01} + p_{11}) = 1 - P[X=0|Y=1], \\
 P[Y=1|X=0] &= p_{01}/(p_{01} + p_{00}) = 1 - P[Y=0|X=0], & P[Y=1|X=1] &= p_{11}/(p_{10} + p_{11}) = 1 - P[Y=0|X=1]
 \end{aligned}$$

La distribución conjunta de dos variables de Poisson tiene tres parámetros, λ_1 , λ_2 y λ_{12} , de acuerdo con la siguiente expresión

$$P[X = k, Y = m] = \sum_{j=0}^{\min(k,m)} \frac{\lambda_{12}^j \lambda_1^{k-j} \lambda_2^{m-j}}{j!(k-j)!(m-j)!} \exp(-(\lambda_1 + \lambda_2 + \lambda_{12}))$$

La distribución conjunta de dos variables normales tiene cinco parámetros, μ_1 , μ_2 , σ_1 , σ_2 y ρ , de acuerdo con la siguiente expresión:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right)$$

La distribución multinomial se refiere a n repeticiones independientes de un experimento con k posibles resultados, donde se cuenta el número de veces que ocurrió cada uno de los posibles resultados. Con tres posibles resultados, la probabilidad de que haya ocurrido i veces el resultado x_0 , j veces el resultado x_1 y $n-i-j$ veces el otro resultado es

$$P[X = i, Y = j] = \frac{n!}{i!j!(n-i-j)!} p_0^i p_1^j (1-p_0-p_1)^{n-i-j}$$

que es una forma de distribución binomial conjunta.

72. Valor esperado y varianzas condicionales

Sean X y Y dos variables aleatorias con JCDF $F_{XY}(x,y)$. Podemos construir dos funciones de Y , $h(Y) = E[X|Y]$, llamada la **esperanza condicional** de X dado el valor de Y , y $g(Y) = V[X|Y]$, llamada la **varianza condicional** de X dado el valor de Y .

Una propiedad interesante de la esperanza condicional es que $E[E(X|Y)] = E[X]$. Una propiedad interesante de la varianza condicional es que $E[V[Y|X]] + V[E[Y|X]] = V[Y]$.

Veamos brevemente la primera función de Y :

$$E[X|Y = y_0] = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y_0) = h(y_0)$$

Si no se especifica el valor de Y , obtenemos una función de la variable aleatoria Y ,

$$E[X|Y] = \int_{-\infty}^{\infty} x dF_{X|Y}(x|Y) = h(Y)$$

y, al promediar sobre Y , $E[E[X|Y]] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x dF_{X|Y}(x|y)\right) dF_Y(y) = \int_{-\infty}^{\infty} x dF_X(x) = E[X]$

Entonces $h(Y) = E[X|Y]$ es otra variable aleatoria cuyo valor esperado es $E[X]$.

Por otro lado, la definición natural de la varianza condicional de X dado que Y toma el valor y es $V[X|Y=y] = E[(X-E[X|Y=y])^2 | Y=y]$. Como estamos integrando sobre todos los valores de X , esta es una función de y y, por lo mismo, $V[X|Y]$ es una variable aleatoria dada como función de Y . En particular nótese que

$$\begin{aligned}
 V[Y] &= E[(Y - E[Y])^2] && \text{por definición} \\
 &= E[(Y - E[Y|X] + E[Y|X] - E[Y])^2] && \text{sumando y restando } E[Y|X] \\
 &= E[E[(Y - E[Y|X]) + (E[Y|X] - E[Y])^2|X]] && \text{porque } E[E[Z|X]] = E[Z] \\
 &= E[E[(Y - E[Y|X])^2 + (E[Y|X] - E[Y])^2 + 2(Y - E[Y|X])(E[Y|X] - E[Y])|X]] && \text{expandiendo} \\
 &= E[V[Y|X]] + V[E[Y|X]] && \text{porque } 2E[E[(Y - E[Y|X])(E[Y|X] - E[Y])|X]] = 0
 \end{aligned}$$

Si el último paso requiere más explicación, nótese que $E[Yg(X)|X] = g(X)E[Y|X]$ para cualquier $g: \mathbb{R} \rightarrow \mathbb{R}$. En efecto,

$$E[Yg(X)|X = x] = \int_{-\infty}^{\infty} yg(x)dF_{Y|X}(y|x) = g(x) \int_{-\infty}^{\infty} ydF_{Y|X}(y|x) = g(x)E[Y|X = x]$$

Aunque las definiciones de la esperanza y la varianza condicionales son simples y obvias, las propiedades que hemos deducido tendrán importantes aplicaciones en la teoría de estimación y de detección, como veremos en 165.

El último ejemplo de la definición 65 mostraba que, si X y Y son conjuntamente Gaussianas con $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ y coeficiente de correlación ρ , la media y varianza condicionales de Y dado X son

$$E[Y | X=x] = \mu_y + \rho\sigma_y(x-\mu_x)/\sigma_x \quad \text{y} \quad V[Y | X=x] = \sigma_y^2(1-\rho^2)$$

73. Momentos conjuntos de dos variables aleatorias

Sean X y Y dos variables aleatorias con JCDF $F_{XY}(x,y)$. El momento conjunto (m,n) de X y Y se define como

$$E[X^m Y^n] = \iint_{\mathbb{R}^2} x^m y^n d^2 F_{XY}(x,y)$$

El momento conjunto central de orden (m,n) de X y Y se define como

$$E[(X - E[X])^m (Y - E[Y])^n] = \iint_{\mathbb{R}^2} (x - E[X])^m (y - E[Y])^n d^2 F_{XY}(x,y)$$

De acuerdo con la definición 66, se trata de encontrar el valor medio de las funciones $Z_1 = X^m Y^n$ y $Z_2 = (X - E[X])^m (Y - E[Y])^n$. El momento conjunto $(1,0)$ corresponde a $E[X]$ y el momento conjunto $(0,1)$ es $E[Y]$. En la definición 69 estudiamos el momento conjunto de orden $(1,1)$, $E[XY]$ o **correlación**, y el momento conjunto central de orden $(1,1)$ o **covarianza**, $\text{Cov}[X,Y] = E[(X - E[X])(Y - E[Y])]$, que se puede expresar como $\text{Cov}[X,Y] = E[XY] - E[X]E[Y]$. Estas son dos de las principales estadísticas en el análisis de dos variables aleatorias pues miden la dependencia lineal entre las dos variables. Como vimos en la definición 69, si las dos variables son independientes la correlación es el producto entre los valores esperados de cada una de las variables:

$$E[XY] = \iint_{\mathbb{R}^2} xy d^2 F_{X,Y}(x,y) = \iint_{\mathbb{R}^2} xy dF_X(x) dF_Y(y) = \int_{\mathbb{R}} x dF_X(x) \int_{\mathbb{R}} y dF_Y(y) = E[X]E[Y]$$

de manera que la covarianza se hace cero. Sin embargo, que la correlación sea el producto de los valores esperados no necesariamente indica que las variables sean independientes. Dos variables aleatorias son “**no-correlacionadas**” si su covarianza es cero, y son “**ortogonales**” si su correlación es cero. Un parámetro de gran interés es el **coeficiente de correlación**, ρ , definido como

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

donde $\sigma_X^2 = \text{Var}[X]$ y $\sigma_Y^2 = \text{Var}[Y]$. Como demostramos en la definición 69, $|\rho_{XY}| \leq 1$ con igualdad si $Y = \alpha X + b$.

La independencia, por supuesto, también simplifica el cálculo de los momentos conjuntos:

$$E[X^m Y^n] = \iint_{\mathbb{R}^2} x^m y^n d^2 F_{X,Y}(x, y) = \iint_{\mathbb{R}^2} x^m y^n dF_X(x) dF_Y(y) = \int_{\mathbb{R}} x^m dF_X(x) \int_{\mathbb{R}} y^n dF_Y(y) = E[X^m] E[Y^n]$$

Los momentos de orden 1 y 2 (medias, varianzas, covarianzas, correlaciones) surgen de suponer relaciones lineales entre variables. Por ejemplo, si $Z = X + Y$, obtenemos que $V[Z] = V[X] + V[Y] + 2\text{Cov}[X, Y]$. Cuando consideramos relaciones no lineales obtenemos momentos de orden superior. Por ejemplo, si $Z = (X + Y)X$, obtenemos que la varianza de Z , $V[Z] = E[X^4] - E^2[X^2] + 2E[X^3 Y] - 2E[X^2]E[XY] + E[X^2 Y^2] - E^2[XY]$, involucra los momentos conjuntos (4,0), (2,0), (3,1), (1,1) y (2,2) de X y Y . Sin embargo, como en la definición 42, los momentos de orden superior se suelen agrupar en estadísticas más significativas como skewness, kurtosis o cumulantes conjuntos. En este curso estaremos casi completamente limitados a momentos de orden 2, aunque usaremos momentos de orden superior al tratar procesos estocásticos no gaussianos, como la deconvolución ciega con entradas no gaussianas (definición xx) o los procesos fractales (definición xx).

74. Función conjunta generadora de momentos, función conjunta generadora de probabilidad, función característica conjunta

Sean X y Y dos variables aleatorias conjuntas. La función conjunta generadora de momentos se define como

$$M_{X,Y}(s_x, s_y) = E[\exp(s_x X + s_y Y)]$$

La función conjunta generadora de probabilidad se define como

$$G_{X,Y}(z_x, z_y) = E[z_x^X z_y^Y]$$

La función conjunta característica se define como

$$\Phi_{X,Y}(\omega_x, \omega_y) = E[\exp(j(\omega_x X + \omega_y Y))]$$

Consideremos sólo la función característica conjunta (que es la función generadora de momentos evaluada en $s_x = j\omega_x$ y $s_y = j\omega_y$, o la función generadora de probabilidad evaluada en $z_x = e^{j\omega_x}$ y $z_y = e^{j\omega_y}$). Evidentemente, la función característica se trata de una versión conjugada de la transformada de Fourier, por lo que se puede invertir fácilmente en los casos de variables conjuntamente discretas o conjuntamente continuas:

Discretas:

$$\Phi_{X,Y}(\omega_x, \omega_y) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \exp(j(x\omega_x + y\omega_y)) P_{XY}(x, y)$$

$$P_{XY}(x, y) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Phi_{X,Y}(\omega_x, \omega_y) \exp(-j(x\omega_x + y\omega_y)) d\omega_x d\omega_y$$

Continuas:

$$\Phi_{X,Y}(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(j(x\omega_x + y\omega_y)) f_{X,Y}(x, y) dx dy$$

$$f_{X,Y}(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-j(x\omega_x + y\omega_y)) \Phi_{X,Y}(\omega_x, \omega_y) d\omega_x d\omega_y$$

Claro, esta invertibilidad implica que toda la información sobre X y Y , que está contenida en $F_{XY}(x,y)$, también está contenida en $\Phi_{XY}(\omega_x, \omega_y)$. Por ejemplo, las funciones características marginales se encuentran fácilmente a partir de la conjunta:

$$\Phi_{X,Y}(\omega_x, 0) = E[\exp(j(\omega_x X + 0))] = E[\exp(j(\omega_x X))] = \Phi_X(\omega_x)$$

Más específicamente,

$$\Phi_{X,Y}(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(j(x\omega_x + y\omega_y)) d^2 F_{X,Y}(x, y)$$

$$\Phi_{X,Y}(\omega_x, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(jx\omega_x) d^2 F_{X,Y}(x, y)$$

$$\Phi_{X,Y}(\omega_x, 0) = \int_{-\infty}^{\infty} \exp(jx\omega_x) dF_X(x) \int_{-\infty}^{\infty} dF_{Y|X}(y|x)$$

$$\Phi_{X,Y}(\omega_x, 0) = \int_{-\infty}^{\infty} \exp(jx\omega_x) dF_X(x) = \Phi_X(\omega_x)$$

Similarmente, $\Phi_{X,Y}(0, \omega_y) = \Phi_Y(\omega_y)$.

Por otro lado, recordemos que cuando teníamos una combinación lineal de dos variables aleatorias, $Z = aX + bY$, el teorema de la probabilidad total daba la distribución de Z en términos de una complicada integral:

$$dF_Z(z) = \int_{x=-\infty}^{\infty} dF_{Y|X}\left(\frac{z-ax}{b} | x\right) dF_X(x)$$

En cambio, la función característica de Z es mucho más fácil de encontrar:

$$\Phi_Z(\omega) = E[e^{j\omega Z}] = E[e^{j\omega(aX+bY)}] = E[e^{j((a\omega)X+(b\omega)Y)}] = \Phi_{XY}(a\omega, b\omega)$$

Si X y Y son independientes, la función característica conjunta es el producto de las

$$\begin{aligned}\Phi_{X,Y}(\omega_x, \omega_y) &= E \left[\exp(j(\omega_x X + \omega_y Y)) \right] = E \left[\exp(j\omega_x X) \exp(j\omega_y Y) \right] = \dots \\ &= E \left[\exp(j\omega_x X) \right] E \left[\exp(j\omega_y Y) \right] = \Phi_X(\omega_x) \Phi_Y(\omega_y)\end{aligned}$$

En particular, si X y Y son independientes y $Z=X+Y$, la función característica de Z es simplemente el producto de las funciones características marginales. En efecto, ya sabíamos que la transformada de Fourier convierte la convolución en el tiempo en el producto en la frecuencia.

Veamos el ejemplo típico en el que X y Y son gaussianas e independientes. Como vimos en la definición 52,

$$\Phi_X(\omega) = e^{j\omega\mu_x - \frac{\omega^2\sigma_x^2}{2}} \quad \Phi_Y(\omega) = e^{j\omega\mu_y - \frac{\omega^2\sigma_y^2}{2}}$$

De manera que

$$\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega) = e^{j\omega\mu_x - \frac{\omega^2\sigma_x^2}{2}} e^{j\omega\mu_y - \frac{\omega^2\sigma_y^2}{2}} = e^{j\omega(\mu_x + \mu_y) - \frac{\omega^2(\sigma_x^2 + \sigma_y^2)}{2}}$$

Esto es, Z es otra variable aleatoria Gaussiana cuya media es la suma de las medias de X y Y , y cuya varianza es la suma de las varianzas de X y Y .

Si no son independientes, ¿cuál será la función característica conjunta? Ahora existe un parámetro adicional, el coeficiente de correlación ρ . Llevando a cabo la doble integración –el proceso es largo y debemos recordar el truco de completar cuadrados– obtenemos:

$$\Phi_{X,Y}(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(j(x\omega_x + y\omega_y)) f_{XY}(x, y) dx dy$$

$$\text{donde } f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right)$$

$$\Rightarrow \Phi_{X,Y}(\omega_x, \omega_y) = \exp\left[j(\mu_x\omega_x + \mu_y\omega_y) - \frac{1}{2}(\sigma_x^2\omega_x^2 + \sigma_y^2\omega_y^2 + 2\rho\sigma_x\omega_x\sigma_y\omega_y)\right]$$

Nuevamente podemos ver que, en el caso de variables gaussianas, el no-correlación ($\rho=0$) es una condición equivalente a la independencia. Pero, más interesante aún, podemos ver cómo se comporta la suma de dos variables aleatorias no independientes, $Z = aX+bY$:

$$\Phi_Z(\omega) = \Phi_{XY}(a\omega, b\omega) = \exp\left[j\omega(a\mu_x + b\mu_y) - \frac{\omega^2}{2}(\sigma_x^2 a^2 + \sigma_y^2 b^2 + 2\rho ab\sigma_x\sigma_y)\right]$$

Z es, entonces, una variable aleatoria gaussiana con valor medio $E[Z]=a\mu_x+b\mu_y$ y con varianza $V[Z]=a^2\sigma_x^2+b^2\sigma_y^2+2\rho ab\sigma_x\sigma_y$.

Una última propiedad interesante es la forma tan fácil como se pueden generar momentos conjuntos a partir de la función característica conjunta, tomando derivadas parciales y evaluando en $\omega_x=\omega_y=0$:

$$\begin{aligned}\Phi_{XY}(\omega_x, \omega_y) &= E\left[\exp(j(X\omega_x + Y\omega_y))\right] = E\left[e^{jX\omega_x} e^{jY\omega_y}\right] \\ \frac{\partial^m}{\partial \omega_x^m} \Phi_{XY}(\omega_x, \omega_y) &= E\left[(jX)^m e^{jX\omega_x} e^{jY\omega_y}\right] \\ \frac{\partial^n}{\partial \omega_y^n} \frac{\partial^m}{\partial \omega_x^m} \Phi_{XY}(\omega_x, \omega_y) &= E\left[(jX)^m e^{jX\omega_x} (jY)^n e^{jY\omega_y}\right] = j^{m+n} E\left[X^m Y^n e^{jX\omega_x} e^{jY\omega_y}\right] \\ \Rightarrow E\left[X^m Y^n\right] &= \frac{1}{j^{m+n}} \frac{\partial^n}{\partial \omega_y^n} \frac{\partial^m}{\partial \omega_x^m} \Phi_{XY}(\omega_x, \omega_y) \Bigg|_{\substack{\omega_x=0 \\ \omega_y=0}}\end{aligned}$$

Por ejemplo, sean X y Y dos variables aleatorias conjuntamente Gaussianas con medias μ_x y μ_y , varianzas σ_x^2 y σ_y^2 , y coeficiente de correlación ρ . La función característica conjunta es

$$\Phi_{X,Y}(\omega_x, \omega_y) = \exp\left[j(\mu_x \omega_x + \mu_y \omega_y) - \frac{1}{2}(\sigma_x^2 \omega_x^2 + \sigma_y^2 \omega_y^2 + 2\rho \sigma_x \omega_x \sigma_y \omega_y)\right]$$

Derivando,

$$\frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) = (j\mu_x - \sigma_x^2 \omega_x - \rho \sigma_x \sigma_y \omega_y) \exp\left[j(\mu_x \omega_x + \mu_y \omega_y) - \frac{1}{2}(\sigma_x^2 \omega_x^2 + \sigma_y^2 \omega_y^2 + 2\rho \sigma_x \omega_x \sigma_y \omega_y)\right]$$

Al evaluar en $\omega_x = \omega_y = 0$ obtenemos

$$\frac{1}{j} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) \Bigg|_{\substack{\omega_x=0 \\ \omega_y=0}} = \mu_x$$

que es el momento (1,0). De igual manera,

$$\frac{1}{j} \frac{\partial}{\partial \omega_y} \Phi_{X,Y}(\omega_x, \omega_y) \Bigg|_{\substack{\omega_x=0 \\ \omega_y=0}} = \mu_y$$

que es el momento (0,1). De otro lado,

$$\begin{aligned}\frac{\partial}{\partial \omega_y} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) &= \left[(j\mu_x - \sigma_x^2 \omega_x - \rho \sigma_x \sigma_y \omega_y)(j\mu_y - \sigma_y^2 \omega_y - \rho \sigma_x \sigma_y \omega_x) - \rho \sigma_x \sigma_y\right] \dots \\ &\dots \exp\left[j(\mu_x \omega_x + \mu_y \omega_y) - \frac{1}{2}(\sigma_x^2 \omega_x^2 + \sigma_y^2 \omega_y^2 + 2\rho \sigma_x \omega_x \sigma_y \omega_y)\right]\end{aligned}$$

Al evaluar en $\omega_x = \omega_y = 0$ obtenemos

$$\left. \frac{\partial}{\partial \omega_y} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) \right|_{\substack{\omega_x=0 \\ \omega_y=0}} = [(j\mu_x)(j\mu_y) - \rho\sigma_x\sigma_y]$$

Esto es, el momento (1,1) de las variables $[X, Y]^T$ es

$$E[XY] = \frac{1}{j^2} \frac{\partial}{\partial \omega_y} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) \Big|_{\substack{\omega_x=0 \\ \omega_y=0}} = \rho\sigma_x\sigma_y + \mu_x\mu_y$$

de manera que $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = \rho\sigma_x\sigma_y$, como se esperaba. Podemos calcular así momentos de orden superior. Por ejemplo, volviendo a dividir respecto a ω :

$$\begin{aligned} \frac{\partial^2}{\partial \omega_y^2} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(\omega_x, \omega_y) &= \exp \left[j(\mu_x\omega_x + \mu_y\omega_y) - \frac{1}{2}(\sigma_x^2\omega_x^2 + \sigma_y^2\omega_y^2 + 2\rho\sigma_x\sigma_y\omega_x\omega_y) \right] \dots \\ &\dots \left\{ \left[(j\mu_x - \sigma_x^2\omega_x - \rho\sigma_x\sigma_y\omega_y)(j\mu_y - \sigma_y^2\omega_y - \rho\sigma_x\sigma_y\omega_x) - \rho\sigma_x\sigma_y \right] (j\mu_y - \sigma_y^2\omega_y - \rho\sigma_x\sigma_y\omega_x) + \dots \right. \\ &\left. \dots \left[-\sigma_y^2(j\mu_x - \sigma_x^2\omega_x - \rho\sigma_x\sigma_y\omega_y) - \rho\sigma_x\sigma_y(j\mu_y - \sigma_y^2\omega_y - \rho\sigma_x\sigma_y\omega_x) \right] \right\} \\ \frac{\partial^2}{\partial \omega_y^2} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(0, 0) &= -j\mu_y [\mu_x\mu_y + \rho\sigma_x\sigma_y] - j[\sigma_y^2\mu_x + \rho\sigma_x\sigma_y\mu_y] \end{aligned}$$

de manera que el momento (1,2) de $[X, Y]^T$ es

$$E[XY^2] = \frac{1}{j^3} \frac{\partial^2}{\partial \omega_y^2} \frac{\partial}{\partial \omega_x} \Phi_{X,Y}(0, 0) = \mu_x [\mu_y^2 + \sigma_y^2] + 2\rho\sigma_x\sigma_y\mu_y$$

75. Extensión a más de dos variables aleatorias

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, podemos definir un vector n -dimensional de variables aleatorias $(X_1, X_2, X_3, \dots, X_n)$ si cada una de ellas asocia un número real con cada elemento del espacio muestral $\omega \in \Omega$ satisfaciendo la medibilidad de los eventos de interés $A(x_1, x_2, \dots, x_n) = \{\omega \in \mathbb{R}^n : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, X_3(\omega) \leq x_3, \dots, X_n(\omega) \leq x_n\}$. Siendo así, en el espacio $(\Omega, \mathcal{F}, \mathbf{P})$ podemos medir la probabilidad de los eventos $A(x_1, x_2, \dots, x_n)$ mediante la **función de distribución acumulativa conjunta n -dimensional**, $F_{\underline{X}}: \mathbb{R}^n \rightarrow \mathbb{R}$ definida mediante

$$F_{\underline{X}}(\underline{x}) = P[\{\omega \in \Omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\}]$$

Mediante el vector aleatorio logramos asociar el espacio de probabilidad original $(\Omega, \mathcal{F}, \mathbf{P})$ con el espacio de probabilidad $(\mathbb{R}^n, \mathcal{A}(\mathbb{R}^n), F_{\underline{X}}(\cdot))$ donde los resultados del experimento son vectores n -dimensionales. La JCDF tiene propiedades ya obvias: Es no negativa, es continua por la derecha en cada dimensión, es no decreciente en cada dirección y, por supuesto,

$$F_{\underline{X}}(\infty, \infty, \dots, \infty) = 1$$

$$F_{\underline{X}}(\infty, \infty, \dots, x_i, \dots, \infty) = F_{X_i}(x_i)$$

$$F_{\underline{X}}(\infty, \infty, \dots, \infty, x_i, \infty, \dots, \infty, x_j, \infty, \dots, \infty) = F_{X_i, X_j}(x_i, x_j)$$

$$F_{\underline{X}}(x_1, x_2, \dots, -\infty, \dots, x_n) = 0$$

Si derivamos la JCDF respecto a cada uno de sus argumentos obtenemos la jpdf, **función de densidad de probabilidad conjunta n -dimensional**,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{\partial^n F_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

Con sus propiedades ya obvias,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n = 1$$

$$P[\underline{X} \in A] = \iiint_A \cdots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

$$f_{X_2, \dots, X_n}(x_2, \dots, x_n) = \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1$$

Si se trata de variables discretas con posibles valores enteros, la jpdf es un conjunto de impulsos de Dirac cuyas áreas debajo de las curvas conforman la jpmf, **función conjunta de distribución de probabilidad**,

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P[X_1 = x_1, \dots, X_n = x_n]$$

Con sus propiedades ya obvias,

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$$

$$\sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} \cdots \sum_{x_n=-\infty}^{\infty} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = 1$$

$$P[\underline{X} \in A] = \sum_A \sum \cdots \sum p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

$$p_{X_2, \dots, X_n}(x_2, \dots, x_n) = \sum_{x_1=-\infty}^{\infty} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

Las **distribuciones condicionales** son

$$p_{X_n | X_1, \dots, X_{n-1}}(x_n | x_1, \dots, x_{n-1}) = \frac{p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{p_{X_1, X_2, \dots, X_{n-1}}(x_1, x_2, \dots, x_{n-1})}$$

$$f_{X_n | X_1, \dots, X_{n-1}}(x_n | x_1, \dots, x_{n-1}) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{f_{X_1, X_2, \dots, X_{n-1}}(x_1, x_2, \dots, x_{n-1})}$$

Mas generalmente, si $I = \{1, 2, \dots, n\}$, $J = \{j_1, j_2, \dots, j_m\} \subset I$ y $K = I \setminus J$

El **vector de valor medio** es

$$E[\underline{X}] = [E[X_1], E[X_2], \dots, E[X_n]]^T$$

Donde $E[X_i] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i d^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$

La **matriz de covarianzas** es

$$Cov[\underline{X}] = \begin{bmatrix} V(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & V(X_n) \end{bmatrix}$$

Donde

$$Cov(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - E[X_i])(x_j - E[X_j]) d^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

Y $V(X_i) = E[(X_i - E[X_i])^2] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - E[X_i])^2 d^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$

Dado un conjunto de n variables aleatorias $\{X_1, X_2, \dots, X_n\}$ y una función $g: \mathbb{R}^n \rightarrow \mathbb{R}$, podemos definir una nueva variable aleatoria mediante $Y = g(X_1, X_2, \dots, X_n)$, una **función de n variables aleatorias**. Si $A(y)$ es la región en \mathbb{R}^n de los argumentos que hacen $g(x_1, x_2, \dots, x_n) \leq y$,

$$F_Y(y) = P[(X_1, X_2, \dots, X_n) \in A(y)]$$

$$F_Y(y) = \int \dots \int_{A(y)} dF^n(x_1, x_2, \dots, x_n)$$

De manera que

$$E[Y] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) dF^n(x_1, \dots, x_n)$$

Si tuviésemos n variables $Y_1 = g_1(X_1, X_2, \dots, X_n)$, $Y_2 = g_2(X_1, X_2, \dots, X_n)$, ..., $Y_n = g_n(X_1, X_2, \dots, X_n)$, podemos considerar la jpdf de las variables aleatorias $\{Y_1, Y_2, \dots, Y_n\}$ mediante la transformación,

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \dots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}}$$

Donde el determinante en el denominador es el Jacobiano de la transformación y los valores de \underline{x} son los que hacen $\underline{y} = g(\underline{x})$. Tal vez el ejemplo más sencillo es la transformación lineal, $\underline{Y} = \mathbf{A}\underline{X}$, donde \underline{X} y \underline{Y} son vectores aleatorios de dimensión n y \mathbf{A} es una matriz invertible constante de tamaño $n \times n$. En este caso, $f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(\underline{x})/|\mathbf{A}|$, donde $\underline{x} = \mathbf{A}^{-1}\underline{y}$.

Una secuencia de n variables aleatorias X_1, X_2, \dots, X_n es independiente si su función de distribución conjunta es el producto de las distribuciones marginales:

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n)$$

$$d^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = dF_{X_1}(x_1)dF_{X_2}(x_2) \dots dF_{X_n}(x_n)$$

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \dots f_{X_n}(x_n)$$

$$P[X_1=x_1, X_2=x_2, \dots, X_n=x_n] = P[X_1=x_1]P[X_2=x_2] \dots P[X_n=x_n]$$

Siendo este el caso, las probabilidades condicionales no dependen de la condición:

$$F_{X_n|X_1, X_2, \dots, X_{n-1}}(x_n|x_1, x_2, \dots, x_{n-1}) = F_{X_n}(x_n)$$

$$dF_{X_n|X_1, X_2, \dots, X_{n-1}}(x_n|x_1, x_2, \dots, x_{n-1}) = dF_{X_n}(x_n)$$

$$f_{X_n|X_1, X_2, \dots, X_{n-1}}(x_n|x_1, x_2, \dots, x_{n-1}) = f_{X_n}(x_n)$$

$$P[X_n=x_n|X_1=x_1, X_2=x_2, \dots, X_{n-1}=x_{n-1}] = P[X_n=x_n]$$

Una distribución común es la multinomial: Un experimento tiene m posibles resultados, digamos los resultado 1, 2, ..., m , de manera que la probabilidad de obtener el k -ésimo resultado es p_k , para $k=1, 2, \dots, m$. Por supuesto, $\sum_{k=1, 2, \dots, m} p_k = 1$. Ahora repetimos el experimento independientemente n veces y contamos cuántas veces se obtuvo el resultado k . Esto es, X_k es una variable aleatoria que representa el número de veces que ocurrió el resultado k en n repeticiones independientes del experimento. El conjunto $\{X_1, X_2, \dots, X_m\}$ tiene una distribución conjunta multinomial:

$$P[X_1 = x_1, X_2 = x_2, \dots, X_m = x_m] = \frac{n!}{x_1! x_2! \dots x_m!} p_1^{x_1} p_2^{x_2} \dots p_m^{x_m}$$

Donde $\sum_{k=1}^m x_k = n$ y cada x_k está entre 0 y n . Si $m=2$, tendremos la conocida distribución binomial.

Otra distribución supremamente útil es la Gaussiana multivariada. Las m variables aleatorias las agrupamos en un vector m -dimensional, $\underline{X}=[X_1, X_2, \dots, X_m]^T$, de manera que una observación retorna un vector m -dimensional, $\underline{x}=[x_1, x_2, \dots, x_m]^T$. Si la función de densidad de probabilidad conjunta es

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^m} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) \right]$$

donde $\underline{\mu}$ es el vector de valor medio con entradas $\mu_k = E[X_k]$, y Σ es la matriz de covarianzas con entradas $\Sigma_{ij} = \text{Cov}(X_i, X_j)$, decimos que el vector \underline{X} está compuesto por m variables conjuntamente gaussianas. Obsérvese que, en el caso $m=2$, obtenemos el caso de la definición 71,

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left(-\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right) \right)$$

Correspondiente a

$$f_{XY}(x, y) = \frac{1}{2\pi \sqrt{\begin{vmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{vmatrix}}} \exp \left(-\frac{1}{2} \begin{bmatrix} x-\mu_x & y-\mu_y \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} \right)$$

Donde $E[X] = \mu_x$, $E[Y] = \mu_y$, $V[X] = \sigma_x^2$, $V[Y] = \sigma_y^2$ y $\text{Cov}[XY] = \rho\sigma_x\sigma_y$.

Como un ejemplo simple, consideremos la transformación lineal $\underline{Y} = \underline{A}\underline{X}$ para un vector conjuntamente Gaussiano \underline{X} y una matriz invertible \underline{A} . Entonces $f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(\underline{A}^{-1}\underline{y})/|\underline{A}|$, como vimos antes. Esto es,

$$\begin{aligned}
 f_{\underline{x}}(\underline{x}) &= \frac{1}{\sqrt{|\Sigma_x|(2\pi)^n}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_x)^T \Sigma_x^{-1}(\underline{x}-\underline{\mu}_x)\right] \Bigg|_{\substack{\underline{x}=A^{-1}\underline{y} \\ \underline{\mu}_x=A^{-1}\underline{\mu}_y}} \\
 \Rightarrow f_{\underline{y}}(\underline{y}) &= \frac{1}{\sqrt{|A|^2|\Sigma_x|(2\pi)^n}} \exp\left[-\frac{1}{2}\left(A^{-1}(\underline{y}-\underline{\mu}_y)\right)^T \Sigma_x^{-1}\left(A^{-1}(\underline{y}-\underline{\mu}_y)\right)\right] \\
 &= \frac{1}{\sqrt{|A|^2|\Sigma_x|(2\pi)^n}} \exp\left[-\frac{1}{2}(\underline{y}-\underline{\mu}_y)^T \left(A^{-T}\Sigma_x^{-1}A^{-1}\right)(\underline{y}-\underline{\mu}_y)\right] \\
 &= \frac{1}{\sqrt{|\Sigma_y|(2\pi)^n}} \exp\left[-\frac{1}{2}(\underline{y}-\underline{\mu}_y)^T \Sigma_y^{-1}(\underline{y}-\underline{\mu}_y)\right]
 \end{aligned}$$

Esto es, \underline{Y} es un vector aleatorio gaussiano con vector de valor medio $\underline{\mu}_y = A\underline{\mu}_x$ y matriz de covarianza $\Sigma_y = A\Sigma_x A^T$.

250 Conceptos - Marco A. Alzate - U. Distrital F.J.C.

V. Secuencias de Variables Aleatorias

76. Secuencias de variables aleatorias independientes e idénticamente distribuidas (iid)

Una secuencia de variables aleatorias $\{X_i, i=1,2,\dots\}$ es una secuencia iid si todas las variables de la secuencia tienen la misma CDF marginal ($F_{X_i}(x) = F(x)$) y si para cualquier subconjunto finito o infinito de ellas la distribución conjunta es el producto de las distribuciones marginales, esto es, dado cualquier subconjunto de índices I , la distribución conjunta de la subsecuencia $\{X_i, i \in I\}$ es

$$F_{X_{I(1)}, X_{I(2)}, \dots}(x_{I(1)}, x_{I(2)}, \dots) = \prod_i F(x_{I(i)})$$

Ya hemos visto muchas secuencias semejantes, como una secuencia de lanzada de dados o una secuencia de números ganadores en la lotería. Estas secuencias pueden ser muy comunes en ingeniería. El modelo más común es la observación repetida de una cantidad constante, donde las medidas están contaminadas por ruido. Por ejemplo, cuando hacemos varias observaciones de la posición de una estrella a través de un telescopio, pero la luz se dispersa en la atmósfera de manera que cada medida es una cantidad iid. O cuando hacemos varias mediciones en el receptor durante el tiempo de transmisión de un bit, pero cada medición viene contaminada con ruido blanco gaussiano (ver definición 83). Lo importante en estos casos es que basta una sola distribución para describir el comportamiento de cualquier subconjunto de muestras, pues la distribución de la secuencia completa es el producto de las distribuciones marginales, como se mostró en la definición.

77. Suma de variables aleatorias iid y suma normalizada

Si $\{X_i, i=1,2,\dots\}$ es una secuencia de variables aleatorias iid, la secuencia de sumas

$$S_n = \sum_{i=1}^n X_i$$

es otra secuencia de variables aleatorias a la que llamaremos la suma de las $\{X_i\}$.

A la secuencia $M_n = S_n/n$ la llamaremos la suma normalizada. Algunas propiedades interesantes de estas secuencias son

$$\Phi_{S_n}(\omega) = (\Phi_X(\omega))^n$$

$$\Phi_{M_n}(\omega) = (\Phi_X(\omega/n))^n$$

$$E[S_n] = nE[X]$$

$$E[M_n] = E[X]$$

$$V[S_n] = nV[X]$$

$$V[M_n] = V[X]/n$$

En efecto,

$$\Phi_{S_n}(\omega) = E[e^{j\omega S_n}] = E\left[e^{j\omega \sum_{i=1}^n X_i}\right] = E\left[\prod_{i=1}^n e^{j\omega X_i}\right] = \prod_{i=1}^n E[e^{j\omega X_i}] = (\Phi_X(\omega))^n$$

$$\Phi_{M_n}(\omega) = E[e^{j\omega M_n}] = E\left[e^{j\frac{\omega}{n} S_n}\right] = \Phi_{S_n}\left(\frac{\omega}{n}\right) = \left(\Phi_X\left(\frac{\omega}{n}\right)\right)^n$$

$$E[S_n] = \sum_{i=1}^n E[X_i] = nE[X]$$

$$\begin{aligned} V[S_n] &= E[(S_n - nE[X])^2] = E\left[\left(\sum_{i=1}^n (E[X_i] - E[X])\right)^2\right] = E\left[\sum_{i=1}^n (E[X_i] - E[X]) \sum_{j=1}^n (E[X_j] - E[X])\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n V[X_i] = nV[X] \end{aligned}$$

$$E[M_n] = E\left[\frac{1}{n} S_n\right] = \frac{1}{n} E[S_n] = E[X]$$

$$V[M_n] = V\left[\frac{1}{n} S_n\right] = \frac{1}{n^2} V[S_n] = \frac{1}{n} V[X]$$

78. Suma de un número aleatorio de variables aleatorias iid

Si $\{X_i, i=1,2,\dots\}$ es una secuencia de variables aleatorias iid y N es una variable aleatoria positiva que toma valores enteros, la cantidad

$$S_N = \sum_{i=1}^N X_i$$

se conoce como una suma aleatoria y tiene las siguientes propiedades:

- $\Phi_{S_N}(\omega) = G_N(\Phi_X(\omega))$ donde $G_N(z)$ es la función generadora de probabilidad de N
- $E[S_N] = E[N]E[X]$
- $V[S_N] = E[N]V[X] + E^2[X]V[N]$

Estos resultados son fáciles de obtener recurriendo a la esperanza condicional, $E[X] = E_Y[E_X[X|Y]]$, como vimos en la definición 72.

$$\Phi_{S_N}(\omega) = E[e^{j\omega S_N}] = E_N\left[E_{S_N}\left[e^{j\omega S_N} | N\right]\right]$$

Dado $N=n$, obtenemos

$$E_{S_N} \left[e^{j\omega S_N} \mid N = n \right] = E \left[\exp \left(j\omega \sum_{i=1}^n X_i \right) \right] = E \left[\prod_{i=1}^n \exp(j\omega X_i) \right] = \prod_{i=1}^n E \left[e^{j\omega X_i} \right] = \left(E \left[e^{j\omega X} \right] \right)^n$$

donde el producto salió del valor esperado debido a la independencia de las X_i . El valor esperado que se está elevando a la potencia n es la función característica de X , una variable genérica con la misma distribución de cada una de las X_i .

$$E_{S_N} \left[e^{j\omega S_N} \mid N = n \right] = \left(\Phi_X(\omega) \right)^n$$

De manera que

$$\Phi_{S_N}(\omega) = E \left[e^{j\omega S_N} \right] = E_N \left[\Phi_X(\omega)^N \right] = G_N(z) \Big|_{z=\Phi_X(\omega)}$$

El término de la derecha de la segunda igualdad es precisamente la definición de la función generadora de probabilidad de la variable aleatoria N , por lo que obtenemos el resultado deseado.

Con el mismo poderoso truco podemos obtener las estadísticas de S_N :

$$E[S_N] = E_N \left[E_{S_N} [S_N \mid N] \right]$$

El valor esperado condicional de S_N dado $N=n$ es

$$E_{S_N} [S_N \mid N = n] = E \left[\sum_{i=1}^n X_i \right] = nE[X]$$

De manera que

$$E_N \left[E_{S_N} [S_N \mid N] \right] = E_N [NE[X]] = E[N]E[X]$$

Donde la última igualdad se debe a la independencia entre N y los $\{X_i\}$. De acuerdo con la definición 72,

$$E[S_N] = E[N]E[X]$$

Por otro lado,

$$V[S_N] = E \left[(S_N - E[S_N])^2 \right] = E \left[(S_N - E[N]E[X])^2 \right] = E_N \left[E_{S_N} \left[(S_N - E[N]E[X])^2 \mid N \right] \right]$$

Cuando $N=n$, el valor esperado condicional es

$$E_{S_N} \left[(S_N - E[N]E[X])^2 \mid N = n \right] = E \left[(S_n - E[N]E[X])^2 \right]$$

Sumando y restando $nE[X]$,

$$\begin{aligned} E \left[(S_n - E[N]E[X])^2 \right] &= E \left[\left((S_n - nE[X]) - (nE[X] - E[N]E[X]) \right)^2 \right] \\ &= E \left[(S_n - nE[X])^2 - 2E[X](S_n - nE[X])(n - E[N]) + E[X]^2 (n - E[N])^2 \right] \\ &= E \left[(S_n - nE[X])^2 + E[X]^2 (n - E[N])^2 \right] - 2E[X](n - E[N])E[S_n - nE[X]] \\ &= E \left[(S_n - nE[X])^2 + E[X]^2 (n - E[N])^2 \right] \end{aligned}$$

De manera que

$$E_{S_N} \left[(S_N - E[N]E[X])^2 | N \right] = E \left[(S_N - NE[X])^2 + E[X]^2 (N - E[N])^2 \right]$$

De donde

$$V[S_N] = E[N]V[X] + E[X]^2 V[N]$$

79. Convergencia de secuencias de variables aleatorias

Una secuencia de variables aleatorias $\{X_n, n=1,2,\dots\}$ en un espacio de probabilidad (Ω, \mathcal{F}, P) puede converger a una variable aleatoria X de diferentes maneras:

- De forma segura si $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \forall \omega \in \Omega$. Decimos que $X_n \xrightarrow{s} X$ (s por "sure").
- De forma casi segura si $P \left[\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right] = 1$. Decimos que $X_n \xrightarrow{a.s.} X$ (a.s. por "almost sure") o que $X_n \xrightarrow{w.p.1} X$ (w.p.1 por "with probability 1").
- En probabilidad si $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P[|X_n - X| > \varepsilon] = 0$
- En r-ésimo momento si $\lim_{n \rightarrow \infty} E[(X_n - X)^r] = 0$
- En distribución si $\forall x \in \mathbb{R}, \lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$

Lo primero que debemos notar es que, cuando decimos que una secuencia de variables aleatorias $\{X_1, X_2, \dots\}$ tiende a otra variable aleatoria X (el límite de la secuencia), podemos querer decir diferentes cosas y debemos ser muy explícitos en aclarar a qué nos referimos, aunque siempre lo expresemos como $X_n \rightarrow X$ o como $\lim_{n \rightarrow \infty} X_n = X$. Para empezar a discutir estos posibles significados, recordemos a qué nos referimos cuando decimos que una secuencia de números reales $\{x_1, x_2, x_3, \dots\}$ tiende a otro número real x , lo cual se expresa mediante $x_n \rightarrow x$ o mediante $\lim_{n \rightarrow \infty} x_n = x$. En este caso, el significado es sólo uno y es muy fácil de explicar:

$$\forall \varepsilon > 0 \exists N(\varepsilon) \in \mathbb{N} \mid \forall n > N(\varepsilon), |x_n - x| < \varepsilon$$

La idea es simple: siempre encontraremos un término de la secuencia a partir del cual la diferencia entre los elementos de la secuencia y el límite sean menores que cualquier nivel de error que escojamos. Entre más pequeño sea ε , más grande será $N(\varepsilon)$, pero siempre existirá ese $N(\varepsilon)$ para cualquier $\varepsilon > 0$, por pequeño que sea. Esta es la única interpretación posible para la expresión $x_n \rightarrow x$. Por ejemplo, si $x_n = 1/n, n=1,2,3,\dots$ es fácil ver que x_n tiende a cero porque

$$\forall \varepsilon > 0 \exists N(\varepsilon) = \lceil 1/\varepsilon \rceil \mid \forall n > \lceil 1/\varepsilon \rceil, |1/n - 0| < \varepsilon$$

Eso es todo respecto a la convergencia de secuencias de números reales. ¿Qué tal una secuencia de funciones? Después de todo, una variable aleatoria es una función del espacio muestral en los reales. Sea, pues, una secuencia de funciones $\{\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots\}$ donde $\underline{x}_n = \{x_n(t), t \in \mathbb{R}\}$ y una función límite $\underline{x} = \{x(t), t \in \mathbb{R}\}$. ¿Qué puede querer decir $\underline{x}_n \rightarrow \underline{x}$ ó $\lim_{n \rightarrow \infty} \underline{x}_n = \underline{x}$? Bueno, ahora sí puede haber varias interpretaciones. La primera es la convergencia punto-a-punto: Dado un instante t_0 , podemos construir

una secuencia de números reales $y_n = x_n(t_0)$ y verificamos que $y_n \rightarrow y$, donde $y = x(t_0)$. Esto es, una secuencia de funciones $\{x_1, x_2, x_3, \dots\}$ converge punto-a-punto a la función x si

$$\forall \varepsilon > 0, \forall t_0 \in \mathbb{R} \exists N(\varepsilon, t_0) \in \mathbb{N} \mid \forall n > N(\varepsilon, t_0), |x_n(t_0) - x(t_0)| < \varepsilon$$

que parece ser el significado más natural de la convergencia de una secuencia de funciones. Por ejemplo, si $x_n = \{x_n(t) = t/n, t \in \mathbb{R}^+\}$ y $x = \{x(t) = 0, t \in \mathbb{R}^+\}$, resulta claro que

$$\forall \varepsilon > 0, \forall t_0 \in \mathbb{R}^+ \exists N(\varepsilon, t_0) = \lceil t_0/\varepsilon \rceil \mid \forall n > \lceil t_0/\varepsilon \rceil, |t_0/n - 0| = t_0/n < \varepsilon$$

Entonces decimos que x_n tiende a x punto-a-punto.

Pero existe al menos una forma más estricta y otra menos estricta de convergencia de funciones. La convergencia uniforme de una secuencia de funciones $\{x_1, x_2, x_3, \dots\}$ a una función límite x se refiere a que

$$\forall \varepsilon > 0 \exists N(\varepsilon) \in \mathbb{N} \mid \forall t_0 \in \mathbb{R}, \forall n > N(\varepsilon), |x_n(t_0) - x(t_0)| < \varepsilon$$

Tanto en la convergencia punto-a-punto como en la convergencia uniforme, escogemos cualquier instante t_0 y construimos la secuencia de números reales $\{x_1(t_0), x_2(t_0), x_3(t_0), \dots\}$ para verificar la convergencia de dicha secuencia de números, esto es, esperando que a partir de algún elemento de la secuencia N ($\forall n > N$), la diferencia entre los elementos de la secuencia $x_n(t_0)$ y el límite $x(t_0)$ sea suficientemente pequeña ($|x_n(t_0) - x(t_0)| < \varepsilon$). La diferencia es que en la convergencia punto-a-punto dicho número N no sólo depende del error ε sino que también depende del instante t_0 . En la convergencia uniforme podemos escoger un único N par todos los instantes t . Por ejemplo, si $x_n = \{x_n(t) = t/n, t \in [a, b] \subset \mathbb{R}^+\}$ y $x = \{x(t) = 0, t \in [a, b] \subset \mathbb{R}^+\}$, resulta claro que

$$\forall \varepsilon > 0, \exists N(\varepsilon) = \lceil b/\varepsilon \rceil \mid \forall t_0 \in [a, b] \forall n > \lceil b/\varepsilon \rceil, |t_0/n - 0| < \varepsilon$$

Entonces decimos que x_n tiende a x uniformemente en $[a, b]$. Por supuesto, la convergencia uniforme es más estricta que la convergencia punto-a-punto pues en la convergencia punto-a-punto me debes dar ε y t_0 para que yo pueda encontrar $N(\varepsilon, t_0)$ mientras que en la convergencia uniforme basta con que me des ε para que yo pueda encontrar un $N(\varepsilon)$ que sirva para cualquier t_0 . Pero, en ambos casos, cada secuencia de números reales $\{x_n(t)\}$ converge a $x(t)$, cualquiera que sea el valor de t . Veamos ahora otra forma de convergencia de secuencias de funciones menos estricta que la convergencia punto-a-punto. Partamos de que las funciones son vectores en un espacio vectorial de Hilbert y, como tales, se les puede dotar de una norma, $\|x\|$, como vimos en la definición **¡Error! No se encuentra el origen de la referencia..** Podemos ahora considerar la norma del error, $d_n = \|x_n - x\|$, que es una secuencia de números, y ver que converja a cero. Esta es la convergencia en la norma:

$$\forall \varepsilon > 0 \exists N(\varepsilon) \in \mathbb{N} \mid \forall n > N(\varepsilon), \|x_n - x\| < \varepsilon$$

Nótese que ya no hace falta que la secuencia de funciones converja en cada instante a la función límite. Puede haber algunos valores de t para los cuales $x_n(t)$ no converge a $x(t)$, siempre que estas diferencias puntuales no afecten la norma del error, que es una medida global. Algunas normas típicas son

$$L_1: \|x\|_1 = \int_{-\infty}^{\infty} |x(t)| dt$$

$$L_2: \|x\|_2^2 = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

$$L_\infty: \|x\|_\infty = \max_{t \in \mathbb{R}} |x(t)|$$

Cuando usamos la norma L_2 , decimos que la secuencia de funciones converge en el sentido del error cuadrado promedio. Por ejemplo, consideremos la siguiente secuencia de funciones:

$$\underline{x}_n = \left\{ x_n(t) = \frac{1}{2} + \sum_{k=1}^n \text{sinc}(k-1/2) \cos((2k-1)t), \quad t \in [-\pi, \pi] \right\}$$

Y el siguiente posible límite:

$$\underline{x} = \{x(t) = \mathbf{1}(t \in [-\pi/2, \pi/2])\}$$

Es fácil ver que no puede haber convergencia punto-a-punto, mucho menos uniforme, porque $x_n(\pi/2) = 1/2$ para todo $n \in \mathbb{N}$ cuando $x(\pi/2) = 0$ y $x_n(-\pi/2) = 1/2$ para todo $n \in \mathbb{N}$ cuando $x(-\pi/2) = 1$. Esto es, las secuencias de números reales $\{x_1(\pi/2), x_2(\pi/2), x_3(\pi/2), \dots\}$ y $\{x_1(-\pi/2), x_2(-\pi/2), x_3(-\pi/2), \dots\}$ son iguales a $\{1/2, 1/2, 1/2, \dots\}$ que no convergen a 0 ni a 1, como debería ser para la convergencia uniforme o punto-a-punto. Sin embargo, si vemos la secuencia de funciones, es claro que hay algún otro tipo de convergencia:

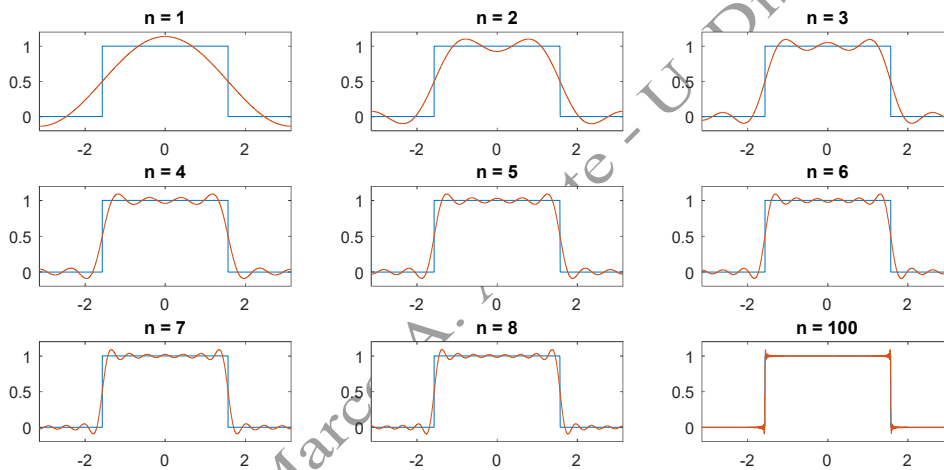


Figura 91. Convergencia en la norma 2, o convergencia en el sentido del error cuadrado promedio

Se trata exactamente de que $\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} |x(t) - x_n(t)|^2 dt = 0$. Esto es, aunque no haya convergencia punto-a-punto, la energía del error tiende a cero a medida que incluimos más y más términos en la secuencia de funciones. En este ejemplo la secuencia de funciones se refiere a versiones truncadas de la serie de Fourier y la convergencia en la norma pero no punto-a-punto se conoce como fenómeno de Gibbs.

Una relación interesante es que, si una secuencia de funciones converge uniformemente, también converge punto-a-punto. Y si converge punto-a-punto, también converge en la norma L_1 . Y, si converge en la norma L_n , también converge en la norma L_{n+1} . Estas relaciones no se dan en el sentido contrario.

Pues bien, siendo que las variables aleatorias son funciones del espacio muestral Ω en los reales, hay muchos tipos de convergencia para ellas. Veamos cada uno de ellos:

Convergencia segura: Se trata de la convergencia uniforme o punto-a-punto de la secuencia de variables aleatorias X_n a la variable límite X . Esto es, para cualquier $\omega \in \Omega$, la secuencia de números reales $\{X_1(\omega), X_2(\omega), X_3(\omega), \dots\}$ converge al número real $X(\omega)$:

$$\forall \omega \in \Omega, \forall \varepsilon > 0, \exists N(\varepsilon, \omega) \in \mathbb{N} \mid \forall n > N(\varepsilon, \omega), |X_n(\omega) - X(\omega)| < \varepsilon$$

Por ejemplo, si X es una variable aleatoria distribuida en el intervalo $[-1, 1]$ según la CDF $F_X(x)$, podemos construir la secuencia de variables aleatorias $\{X_n = X^2(1+1/n), n=1, 2, \dots\}$. Entonces, $\forall \omega \in \Omega, \forall \varepsilon > 0 \exists N = \lceil X^2(\omega)/\varepsilon \rceil$ tal que $\forall n > N |X_n(\omega) - X^2(\omega)| = |X^2(\omega)(1+1/n) - X^2(\omega)| = X^2(\omega)/n < X^2(\omega)/N$, por lo que la convergencia es punto a punto. Como la convergencia se da para todo $\omega \in \Omega$, es una convergencia segura.

La verdad es que, como este tipo de convergencia no incluye probabilidades, la convergencia segura no es de nuestro interés, pues se da sobre Ω sin importar el espacio de probabilidad (Ω, \mathcal{F}, P) (obsérvese que en la definición no usamos $F_X(x)$ para nada). Las siguientes formas de convergencia si incluyen la medida de probabilidad y son todas muy importantes en el modelado probabilístico.

Convergencia casi segura: Se trata de la convergencia uniforme o punto-a-punto de la secuencia de variables aleatorias X_n a la variable límite X pero no en cualquier $\omega \in \Omega$ sino sólo para los ω que pertenecen a un evento de probabilidad uno:

$$\exists A \subset \Omega, A \in \mathcal{F}, A \neq \Omega, \text{ con } P(A)=1, \mid \forall \omega \in A, \forall \varepsilon > 0, \exists N(\varepsilon, \omega) \in \mathbb{N} \mid \forall n > N(\varepsilon, \omega), |X_n(\omega) - X(\omega)| < \varepsilon$$

Por eso $X_n \xrightarrow{a.s.} X$ significa que $P\left[\omega \in \Omega: \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right] = 1$: La secuencia converge con probabilidad 1.

Por ejemplo, si X es una variable aleatoria distribuida exponencialmente con parámetro λ , podemos construir la secuencia $\{X_n = (1+1/n)/X, n=1, 2, 3, \dots\}$. Para cualquier $\omega \in \Omega$ tal que $X(\omega) > 0$, es evidente que $X_n(\omega)$ converge a $1/X(\omega)$. Por supuesto, $P[\omega \in \Omega \mid X(\omega) > 0] = 1$ pero X puede valer 0.

La convergencia casi segura es la más estricta de las convergencias de interés para secuencias de variables aleatorias. Pero, por supuesto, existen formas menos estrictas y muy útiles.

Convergencia en probabilidad: En las anteriores formas de convergencia buscábamos que los valores precisos de las variables aleatorias en la secuencia convergieran a los valores precisos de la variable límite: $X_n(\omega) \rightarrow X(\omega)$ para todo $\omega \in \Omega$ (convergencia segura) o para los $\omega \in A$ para un conjunto A de probabilidad 1 (convergencia casi segura). Ahora la secuencia de números reales que queremos que converjan no son los valores de las variables aleatorias sino las probabilidades de la diferencia:

$$\forall \varepsilon > 0, \forall \delta > 0, \exists N(\varepsilon, \delta) \in \mathbb{N} \mid \forall n > N(\varepsilon, \delta), P[|X_n - X| < \varepsilon] > 1 - \delta$$

Esto es, si construimos la secuencia de números reales $r_n = P[|X_n - X| > \varepsilon]$, debemos verificar que $r_n \rightarrow 0$ para cualquier $\varepsilon > 0$. Como la convergencia ya no es en los valores de las variables sino en sus probabilidades, decimos que la secuencia de variables aleatorias converge en probabilidad.

Por ejemplo, si X está uniformemente distribuida en el intervalo $[0,1]$ y construimos la secuencia $\{X_n = X \cdot (1+1/n), n=1,2,3,\dots\}$, notaremos que $P[|X_n - X| > \varepsilon] = P[X/n > \varepsilon] = P[X > n\varepsilon]$. Esta probabilidad se acerca a cero a medida que $n\varepsilon$ se acerca a 1. De hecho, $P[X > n\varepsilon] = 0 \quad \forall n > \lceil 1/\varepsilon \rceil$. En consecuencia, $\{X_1, X_2, X_3, \dots\}$ converge a X en probabilidad.

Convergencia en el m-ésimo momento: Ya no nos interesan los valores precisos de las variables ni de la probabilidad de los errores. Ahora nos interesan las estadísticas del error:

$$\forall \varepsilon > 0, \exists N(\varepsilon) \in \mathbb{N} \mid \forall n > N(\varepsilon), E[|X_n - X|^m] < \varepsilon$$

Esto es, si construimos la secuencia de números reales $r_n = E[|X_n - X|^m]$, debemos verificar que $r_n \rightarrow 0$.

Por ejemplo, si X esta uniformemente distribuida en el intervalo unitario $[0, 1]$ y construimos la secuencia de variables aleatorias $\{X_n = X \cdot (1+1/n), n=1,2,3,\dots\}$, notaremos que el error $|X_n - X|$ es igual a X/n , que está uniformemente distribuido entre 0 y $1/n$ con promedio $E[|X_n - X|] = E[X/n] = E[X]/n = 1/(2n)$. El segundo momento es

$$E[(X_n - X)^2] = \int_0^{1/n} (x - 1/2n)^2 dx = \frac{1}{3} \left(x - \frac{1}{2n} \right)^3 \Big|_0^{1/n} = \frac{1}{12n^3} \xrightarrow{n \rightarrow \infty} 0$$

por lo que podemos decir que la secuencia $\{X_n, n=1,2,\dots\}$ tiende a X en el 2^{do} momento. En este caso $m=2$, se trata del promedio del error cuadrado, por lo que decimos que la convergencia es en el sentido del error cuadrado promedio.

Convergencia en distribución: Este es el menos estricto de los modos de convergencia, tanto que muchos autores opinan que no se debería considerar un modo de convergencia de secuencias de variables aleatorias porque no convergen sus valores, ni las probabilidades del error, ni las estadísticas del error. Lo que converge es simplemente su función de distribución... pero vimos tres maneras distintas como una función puede converger. En este caso hablamos de una convergencia punto-a-punto de la función de distribución:

$$\forall x_0 \in \mathbb{R}, \forall \varepsilon > 0, \exists N(x_0, \varepsilon) \in \mathbb{N} \mid \forall n > N(x_0, \varepsilon), |F_{X_n}(x_0) - F_X(x_0)| < \varepsilon$$

Claro, sabemos que dos variables aleatorias igualmente distribuidas pueden tener valores muy distintos.

Por ejemplo, otra vez, si X esta uniformemente distribuida en el intervalo unitario $[0, 1]$ y construimos la secuencia de variables aleatorias $\{X_n = X \cdot (1+1/n), n=1,2,3,\dots\}$, notamos que

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{1+1/n} & \text{si } 0 \leq x \leq 1 + \frac{1}{n} \\ 1 & \text{si } x > 1 + \frac{1}{n} \end{cases} \xrightarrow{n \rightarrow \infty} F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Es interesante notar que usamos el mismo ejemplo en casi todas las formas de convergencia. Esto se debe a que unas formas de convergencia implican otras:

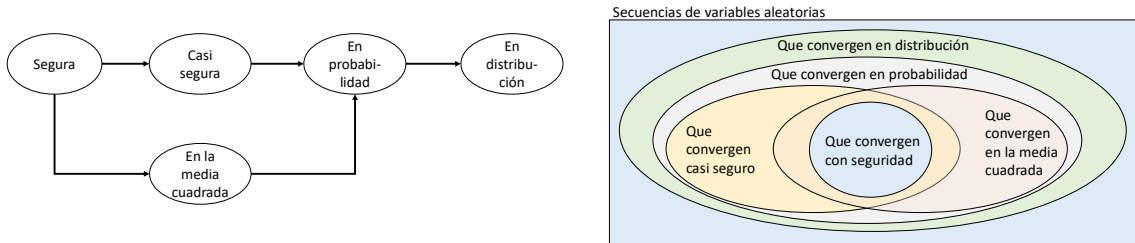


Figura 92. Implicaciones entre los modos de convergencia

Las siguientes definiciones muestran algunos ejemplos muy importantes de varios modos de convergencia.

80. Leyes de los grandes números (WLLN y SLLN)

Considere una secuencia de variables aleatorias iid $\{X_n, n=1,2,\dots\}$ con valor medio finito $E[X]=\mu_X$ y varianza finita $V[X]=\sigma_X^2$. Una ley débil de los grandes números dice que $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left[\left|\frac{1}{n} \sum_{i=1}^n X_n - \mu_X\right| < \varepsilon\right] = 1$. Esto es, la suma normalizada tiende al valor medio en probabilidad. Más aún, una segunda ley débil de los grandes números dice que $\lim_{n \rightarrow \infty} E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - \mu_X\right)^2\right] = 0$. Esto es, la suma normalizada tiende al valor medio en el sentido del error cuadrado promedio. Podemos ser aún más estrictos y notar que $P\left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_n = E[X]\right] = 1$. Esto es, la suma normalizada tiende al valor medio con probabilidad 1. La convergencia en probabilidad se conoce como Ley Débil de los Grandes Números (WLLN), la convergencia en la media cuadrada es otra Ley Débil de los Grandes Números, aunque a veces se le llama teorema de Markov. La última igualdad es la ley fuerte de los grandes números (SLLN).

Efectivamente, si $S_n = \sum_{i=1}^n X_i$ y $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ es fácil ver que

$$E[S_n] = \sum_{i=1}^n E[X_i] = nE[X] = n\mu_X \text{ (donde } \mu_X \text{ es el valor esperado de cada una de las } X_i)$$

$$\begin{aligned} V[S_n] &= E[(S_n - n\mu_X)^2] = E\left[\left(\sum_{i=1}^n (X_i - \mu_X)\right)^2\right] = E\left[\sum_{i=1}^n (X_i - \mu_X) \sum_{j=1}^n (X_j - \mu_X)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n V[X_i] = n\sigma_X^2 \end{aligned}$$

donde σ_X^2 es la varianza de cada una de las X_i y la penúltima igualdad es por la independencia de la secuencia $\{X_n, n=1,2,\dots\}$. Al considerar la suma normalizada $M_n = S_n/n$, y de acuerdo con las propiedades (b) y (d) de la definición 41,

$$E[M_n] = \mu_X \text{ y } V[M_n] = \sigma_X^2/n$$

Si aplicamos a M_n la desigualdad de Chebyzhev (definición 47(b)) obtenemos

$$P\left[|M_n - \mu_X| \geq \varepsilon\right] \leq \frac{\sigma_X^2/n}{\varepsilon^2}. \text{ En términos de nuestras expresiones,}$$

$$P\left[|M_n - \mu_X| < \varepsilon\right] \geq 1 - \delta = 1 - \frac{\sigma_X^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 1$$

Esto es, $\forall \varepsilon > 0, \forall \delta > 0, \exists N(\varepsilon, \delta) = \lceil \sigma_X^2/(\delta\varepsilon^2) \rceil \mid \forall n > N(\varepsilon, \delta), P[|X_n - \mu_X| < \varepsilon] > 1 - \delta$, que es la definición de convergencia de M_n a μ_X en probabilidad.

Por otro lado, volviendo a las expresiones $E[M_n] = \mu_X$ y $V[M_n] = \sigma_X^2/n$, queda claro que, al hacer que n tienda a infinito, el valor esperado sigue siendo μ_X y la varianza tiende a cero. De nuevo,

$$\lim_{n \rightarrow \infty} E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - \mu_X\right)^2\right] = 0$$

por lo que la suma normalizada tiende al valor medio en el sentido del error cuadrado promedio a medida que consideramos más y más muestras independientes.

También es posible demostrar la forma más estricta de convergencia para la suma normalizada dada por la ley fuerte de los grandes números:

$$P\left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]\right] = 1. \text{ Esto es, la suma normalizada tiende al valor medio con probabilidad}$$

1. Sin embargo, la demostración la dejaremos para después, pues necesitamos más herramientas conceptuales para ella (definición xx).

Es interesante notar que la suposición de que la varianza es finita no es estrictamente necesaria para que se cumpla la WLLN, aunque la demostración es mucho más elaborada. También existen versiones en que las variables no son iid pero tienen todas el mismo valor esperado.

Las leyes de los grandes números garantizan que un gran número de experimentos conducirán a la estabilidad de los promedios a la larga, de manera que un gran número de caras seguidas al lanzar muchas veces un dado se verán compensadas con un número suficiente de sellos para que el promedio siga siendo de una cara en la mitad de las lanzadas. Pero debemos ser cuidadosos: Un gran número de caras seguidas no aumentan la probabilidad de que la próxima lanzada sea un sello, pues cada lanzada es un experimento independiente de los anteriores. Sin embargo, un gran número de caras seguidas seguirá siendo un número despreciable de caras a medida que lancemos más y más veces la moneda. Es importante no confundirse con esta interpretación de las leyes de los grandes números: Es aplicable sólo a grandes números de repeticiones del experimento o a grandes números de variables aleatorias.

81. Teorema del límite central

Considere secuencia de variables aleatorias iid $\{X_n, n=1,2,\dots\}$ con valor medio finito

$E[X]=\mu_X$ y varianza finita $V[X]=\sigma_X^2$. Sea $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ la suma normalizada, con

media μ_X y varianza σ_X^2/n . Considere la variable con media cero y varianza 1

$Z_n = \frac{M_n - \mu_X}{\sigma_X / \sqrt{n}}$ La función de densidad de probabilidad de Z_n satisface

$$\lim_{n \rightarrow \infty} f_{Z_n}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \forall z \in \mathbb{R}$$

Es decir, la suma normalizada tiende a una variable Gaussiana en distribución.

En efecto, lo primero por notar es que

$$E[Z_n] = \frac{E[M_n] - \mu_X}{\sigma_X / \sqrt{n}} = 0 \quad \text{y} \quad V[Z_n] = \frac{V[M_n]}{\sigma_X^2/n} = 1$$

Ahora vamos a hacer dos procesos: En el primero encontraremos la función generadora de momentos de una variable aleatoria $\mathcal{N}(0,1)$ –Gaussiana con media cero y varianza 1– y en el segundo encontraremos la función generadora de momentos de Z_n . Al notar que la segunda tiende a la primera punto-a-punto, sabremos que la distribución de Z_n tiende a ser gaussiana $\mathcal{N}(0,1)$: Convergencia en distribución.

Sea $Y \sim \mathcal{N}(0,1)$, esto es, $f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)$. Entonces

$$M_Y(s) = E[e^{sY}] = \int_{-\infty}^{\infty} e^{sy} f_Y(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(sy - y^2/2) dy$$

Como es costumbre al trabajar con estas integrales, podemos completar el cuadrado añadiendo y restando términos apropiados:

$$-\frac{1}{2}(y-s)^2 = -\frac{1}{2}(y^2 - 2ys + s^2) = (-\frac{1}{2}y^2 + sy) - \frac{1}{2}s^2 \quad \Rightarrow \quad -\frac{1}{2}y^2 + sy = -\frac{1}{2}(y-s)^2 + \frac{1}{2}s^2$$

De donde

$$\begin{aligned} M_Y(s) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}(y-s)^2 + \frac{1}{2}s^2) dy \\ &= \exp(\frac{1}{2}s^2) \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}(y-s)^2) dy \right] \end{aligned}$$

pero la integral de la derecha es la probabilidad de que una variable $\mathcal{N}(s,1)$ tome un valor en el intervalo $(-\infty, \infty)$, que es 1, de manera que

$$M_Y(s) = \exp\left(\frac{1}{2}s^2\right)$$

Ahora vamos al segundo proceso y consideremos la versión centrada y normalizada de la media muestral de n variables iid con media μ y varianza σ^2 ,

$$Z_n = \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i\right) - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

Su función generadora de momentos es

$$M_{Z_n}(s) = E\left[e^{sZ_n}\right] = E\left[\exp\left(\frac{s}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}\right)\right] = E\left[\prod_{i=1}^n \exp\left(\frac{s}{\sqrt{n}} \frac{X_i - \mu}{\sigma}\right)\right]$$

Por tratarse de variables independientes, la esperanza del producto es el producto de las esperanzas:

$$M_{Z_n}(s) = \prod_{i=1}^n E\left[\exp\left(\frac{s}{\sqrt{n}} \frac{X_i - \mu}{\sigma}\right)\right]$$

Pero, como las X_i son variables idénticamente distribuidas,

$$M_{Z_n}(s) = E^n \left[\exp\left(\frac{s}{\sqrt{n}} \frac{X_1 - \mu}{\sigma}\right) \right]$$

Ahora usamos la expansión en serie de Taylor $e^x = 1 + x + \frac{x^2}{2} + \dots$

$$M_{Z_n}(s) = E^n \left[1 + \frac{s}{\sqrt{n}} \left(\frac{X_1 - \mu}{\sigma}\right) + \frac{s^2}{2n} \left(\frac{X_1 - \mu}{\sigma}\right)^2 + \dots \right]$$

$$M_{Z_n}(s) = \left(1 + \frac{s^2}{2n} + \dots\right)^n \approx \left(1 + \frac{s^2}{2n}\right)^n$$

Donde el valor aproximado obedece a que, para valores grandes de n , los términos con potencias mayores de n se hacen despreciables en comparación con los primeros dos términos. Entonces

$$\lim_{n \rightarrow \infty} M_{Z_n}(s) = \lim_{n \rightarrow \infty} \left(1 + \frac{s^2}{2n}\right)^n = \exp\left(\frac{1}{2}s^2\right)$$

que es la función generadora de momentos de la distribución gaussiana $\mathcal{N}(0,1)$, como acabamos de ver. Si Z_n tiende a $\mathcal{N}(0,1)$ en distribución, M_n tiende a $\mathcal{N}(\mu_X, \sigma_X^2/n)$ en distribución. La Figura 55 muestra un buen ejemplo de este tipo de convergencia para la suma de variables de Bernoulli.

82. Ejemplo: Algoritmo Expectation/Maximization (EM)

Sea \underline{X} un vector aleatorio con una distribución de probabilidad determinada por algunos parámetros, $\underline{X} \sim f_{\underline{X}}(\underline{x}; \theta)$. Dado un conjunto de datos observados \underline{x} , la función loglikelihood $L(\theta) = \log(f_{\underline{X}}(\underline{x}; \theta))$ se suele maximizar para estimar los parámetros θ bajo los cuales las observaciones obtenidas sean las más probables –estimación de máxima verosimilitud (Maximum Likelihood, ML)–. Sin embargo, si no se dispone de todos los datos sino que $\underline{x} = [\underline{y} \ \underline{z}]$, donde \underline{y} son los datos observados y \underline{z} son los datos escondidos,

el procedimiento EM propone encontrar el valor esperado de z dados x y un estimado inicial de θ , θ_0 . Se empieza con $z_0 = E[z|x; \theta_0]$, y, con este valor esperado, se encuentra el siguiente estimado de θ , $\theta_1 = \operatorname{argmax}(\log(f_X([y|z_0]; \theta)))$. El proceso se repite ($z_n = E[z|x; \theta_n]$, $\theta_{n+1} = \operatorname{argmax}(\log(f_X([y|z_n]; \theta))$, $n=1,2,\dots$) hasta alcanzar convergencia. Como en el primer paso de cada iteración se calcula el valor esperado de z y en el segundo paso se maximiza el likelihood, el proceso se llama Expectation/Maximization.

Aunque más adelante tendremos un capítulo dedicado al problema de la estimación y la detección, vale la pena introducir aquí este procedimiento algorítmico muy utilizado recientemente para estimación estadística en muchas áreas de la ciencia y la tecnología, pues aplica muchos de los conceptos tratados hasta este capítulo V. Lo plantearémos con un ejemplo sencillo.

Supongamos un espacio de probabilidad con $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathcal{F} = \{0,1\}^\Omega$ y medida de probabilidad dada por $P[\{\omega_1\}] = 1/2$, $P[\{\omega_2\}] = \theta$, $P[\{\omega_3\}] = 2\theta$, $P[\{\omega_4\}] = 1/2 - 3\theta$, siendo θ un parámetro desconocido que queremos estimar a partir de datos experimentales. Ciertamente, el parámetro θ sólo puede tomar valores en el intervalo $[0, 1/6]$ para que la distribución sea válida. Se han hecho algunas repeticiones independientes del experimento obteniendo una observación, y , que consiste de a veces ω_1 , b veces ω_2 , c veces ω_3 y d veces ω_4 . La probabilidad de esta observación es, por supuesto, $P[y|\theta] = (1/2)^a (\theta)^b (2\theta)^c (1/2 - 3\theta)^d$, que es el “likelihood” de y dado θ . Un estimador razonable de θ consiste en escoger el valor de θ que maximice dicho likelihood, para lo cual podemos derivar $L(\theta) = \log(P[y|\theta])$ respecto a θ , igualar a cero y despejar para θ .

$$L(\theta) = \log(P[y|\theta]) = a \cdot \log(1/2) + b \cdot \log(\theta) + c \cdot \log(2\theta) + d \cdot \log(1/2 - 3\theta)$$

$$\frac{d}{d\theta} L(\theta) = \frac{b}{\theta} + \frac{c}{\theta} - \frac{3d}{1/2 - 3\theta} = 0 \Rightarrow \theta = \frac{b+c}{6(b+c+d)}$$

Efectivamente, se trata de un número en el intervalo $[0, 1/6]$ como esperábamos. Hemos resuelto así un problema simple de estimación de máxima verosimilitud, en el que no hubo dificultades porque teníamos información completa: Cuántas veces ocurrió cada posible resultado. Sin embargo, puede que no tengamos información completa. Por ejemplo, puede ser que, en vez de conocer a y b , sólo conozcamos su suma, $h = a+b$. En este caso podemos iniciar con un valor cualquiera de θ , por ejemplo $\theta = 1/12$, y considerar el valor esperado condicional de b , $E[b|\theta] = 2h\theta/(1+2\theta)$. Con esta esperanza condicional podemos maximizar el likelihood mediante el estimador $\theta = (c + E[b|\theta]) / (6(c + d + E[b|\theta]))$. Con este nuevo valor de θ podemos volver a calcular la esperanza condicional de b y repetir hasta convergencia:

1. Haga $n=0$ e inicie con un valor válido $\theta_0 \in [0, 1/6]$
2. Calcule el valor esperado del dato no observado b dado el valor de θ_n , $E[b|\theta_n] = 2h\theta_n / (1 + 2\theta_n)$.
3. Con este valor esperado de b , encuentre el valor de θ_{n+1} que maximiza el likelihood, $\theta_{n+1} = (c + E[b|\theta_n]) / (6(c + d + E[b|\theta_n]))$
4. Repita desde 2 hasta que haya convergencia en el valor de θ .

Así, el ciclo consiste en calcular la esperanza y maximizar el likelihood repetidamente, por lo que el algoritmo se conoce como Expectation/Maximization o, simplemente, algoritmo EM.

Por ejemplo, supongamos que hacemos 20 observaciones y obtenemos los siguientes datos:
 $a = 12, b=1, c=5, d=2$. El valor del parámetro θ que mejor explica estos datos es $(b+c)/[6(b+c+d)] = 1/8$. Sin embargo, supongamos ahora que sólo conocemos $h=a+b=13, c=5, d=2$. No son suficientes datos para estimar θ , por lo que recurrimos al algoritmo E/M empezando con $\theta_0=1/12$:

$$E[b|\theta_0] = 2h\theta_0 / (1+2\theta_0) = 2*13*(1/12) / (1+2/12) = 1.8571$$

$$\theta_1 = (c+E[b|\theta_0]) / (6(c+d+ E[b|\theta_0])) = (5+1.8571) / (6*(5+2+1.8571)) = 0.129$$

$$E[b|\theta_1] = 2h\theta_1 / (1+2\theta_1) = 2.6667$$

$$\theta_2 = (c+E[b|\theta_1]) / (6(c+d+ E[b|\theta_1])) = 0.1322$$

$$E[b|\theta_2] = 2h\theta_2 / (1+2\theta_2) = 2.7182$$

$$\theta_3 = (c+E[b|\theta_2]) / (6(c+d+ E[b|\theta_2])) = 0.1324$$

$$E[b|\theta_3] = 2h\theta_3 / (1+2\theta_3) = 2.7212$$

$$\theta_4 = (c+E[b|\theta_3]) / (6(c+d+ E[b|\theta_3])) = 0.1324$$

$$E[b|\theta_4] = 2h\theta_4 / (1+2\theta_4) = 2.7213$$

$$\theta_5 = (c+E[b|\theta_4]) / (6(c+d+ E[b|\theta_4])) = 0.1324$$

$$E[b|\theta_5] = 2h\theta_5 / (1+2\theta_5) = 2.7213$$

$$\theta_6 = (c+E[b|\theta_5]) / (6(c+d+ E[b|\theta_5])) = 0.1324$$

...

El algoritmo converge al estimador $\theta=0.1324$, que no está lejos del estimador 0.125 obtenido con los datos completos. Por supuesto, entre más datos obtengamos mejor será la estimación. Si en vez de 20 observaciones hacemos 100 y obtenemos $a = 52, b=14, c=22, d=12$, el estimador de máxima verosimilitud es $\theta = (b+c)/[6(b+c+d)] = 0.125$, nuevamente. Pero si sólo contamos con $h=a+b=66, c=22, d=12$ debemos correr el algoritmo E/M. Empecemos otra vez con $\theta_0=1/12$:

$$E[b|\theta_0] = 9.4286$$

$$\theta_1 = 0.1206$$

$$E[b|\theta_1] = 12.8269$$

$$\theta_2 = 0.124$$

$$E[b|\theta_2] = 13.1117$$

$$\theta_3 = 0.1242$$

$$E[b|\theta_3] = 13.1335$$

$$\theta_4 = 0.1242$$

$$E[b|\theta_4] = 13.1354$$

$$\theta_5 = 0.1242$$

$$E[b|\theta_5] = 13.1354$$

$$\theta_6 = 0.1242$$

Con 100 datos observados, el algoritmo converge al estimador $\theta=0.1242$, más cercano al estimador 0.125 obtenido con los datos completos que el estimador obtenido con sólo 10 datos observados.

83. Ejemplo: Canal de comunicaciones con ruido blanco gaussiano

El modelo más común de canal de comunicaciones es el que se muestra en la Figura 93.

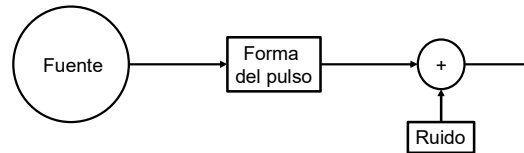


Figura 93. Modelo de un sistema básico de comunicaciones

La fuente genera una secuencia de bits a una tasa de C bits por segundo. Los bits 0 se representan en un pulso eléctrico diferente al de los bits 1 para que puedan ser distinguidos en el receptor. Sin embargo, el canal añade ruido al impulso transmitido, de manera que al receptor llega una versión ruidosa de cada impulso. La tarea del receptor es identificar el bit transmitido.

Para empezar, supongamos que los bits se identifican con un pulso cada uno de amplitud $+A$ ó $-A$ voltios según el valor del bit. Y supongamos que el pulso se contamina con ruido gaussiano de media cero y varianza σ^2 . Si la amplitud del pulso es una variable aleatoria X que toma el valor $+A$ con probabilidad p y toma el valor $-A$ con probabilidad $1-p$, a la salida del canal observamos la variable aleatoria $Y = X + \eta$, donde, como mencionamos, η es una variable aleatoria con media 0 y varianza σ^2 . Primero repliquemos el ejemplo de la definición 64 tomando una sola muestra de Y para detectar el valor de X . Posteriormente veremos la bondad de tomar varias muestras.

La distribución de Y depende del valor de X . Bajo la hipótesis $H_1 : X = +A$, Y será una variable aleatoria Gaussiana con media $+A$ y varianza σ^2 . Bajo la hipótesis $H_0 : X = -A$, Y será una variable aleatoria Gaussiana con media $-A$ y varianza σ^2 :

$$H_0 : X = -A \Rightarrow f_{Y|H} (y|H_0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y+A}{\sigma}\right)^2\right)$$

$$H_1 : X = +A \Rightarrow f_{Y|H} (y|H_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-A}{\sigma}\right)^2\right)$$

La representación de estas dos distribuciones (Figura 94) muestra que entre mayor sea A (entre más separados estén los bits) y entre menor sea σ^2 (entre menos potencia tenga el ruido), será más fácil distinguir cuál de las dos hipótesis es la correcta.

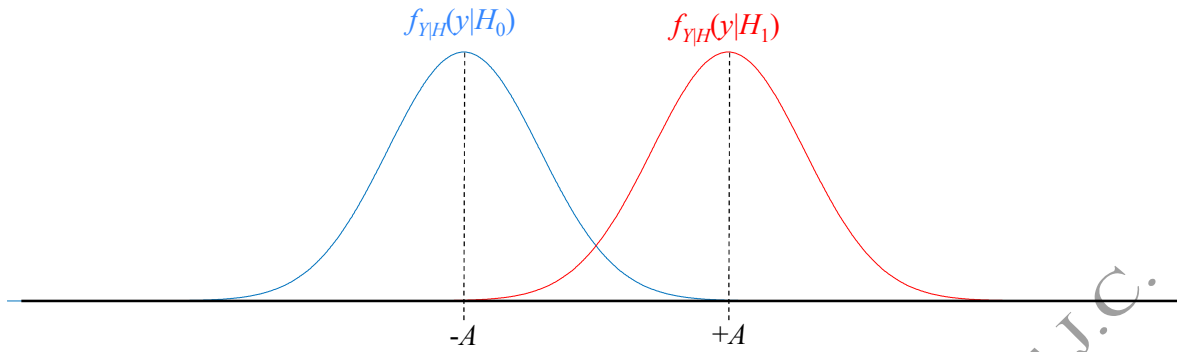


Figura 94. Distribución de la señal recibida según la señal transmitida

La idea es encontrar un valor de comparación con el que se pueda tomar una decisión óptima:

- H_1 Si $y > \gamma$, suponga que H_1 es cierta ($X = +A$);
- $y \geq \gamma$ Si $y < \gamma$, suponga que H_0 es cierta ($X = -A$);
- H_0 Si $y = \gamma$, suponga cualquier hipótesis (lance una moneda balanceada para decidir)

El criterio de optimalidad permitirá determinar el valor apropiado de comparación, γ . En comunicaciones, el criterio más común es minimizar la probabilidad de error:

$$\begin{aligned}
 P_E &= P[\{Y > \gamma\} \cap \{X = -A\}] + P[\{Y \leq \gamma\} \cap \{X = +A\}] \\
 &= P[H_0]P[Y > \gamma | H_0] + P[H_1]P[Y \leq \gamma | H_1] \\
 &= \frac{1-p}{\sigma\sqrt{2\pi}} \int_{\gamma}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{y+A}{\sigma}\right)^2\right) dy + \frac{p}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\gamma} \exp\left(-\frac{1}{2}\left(\frac{y-A}{\sigma}\right)^2\right) dy
 \end{aligned}$$

Nótese que hemos decidido escoger como cierta la hipótesis H_1 cuando $Y = \gamma$, lo cual no afecta la probabilidad de error porque $P[Y = \gamma] = 0$. Considerando la cola de la distribución $\mathcal{N}(0,1)$,

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \exp(-t^2/2) dt$$

Y haciendo la sustituciones $t = (y-A)/\sigma$ y $t = (y+A)/\sigma$ (con $dt = dy/\sigma$ en ambos casos), obtenemos

$$P_E = (1-p)Q\left(\frac{\gamma+A}{\sigma}\right) + p\left[1 - Q\left(\frac{\gamma-A}{\sigma}\right)\right]$$

Tal vez es más común encontrar como las librerías de los lenguajes de programación la función de error

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

de manera que $Q(z) = (1 - erf(t/\sqrt{2}))/2$ y

$$P_E = \frac{1-p}{2} \left[1 - erf\left(\frac{\gamma+A}{\sigma\sqrt{2}}\right)\right] + \frac{p}{2} \left[1 + erf\left(\frac{\gamma-A}{\sigma\sqrt{2}}\right)\right]$$

es una expresión fácil de evaluar, según muestra la Figura 95 para el caso $A = 10$ y $\sigma = 5$, con tres valores distintos de p . Si $\gamma = -\infty$ siempre decidimos que $X = +A$ y la probabilidad de error es $1-p$, que es la probabilidad de que $X = -A$. Si $\gamma = +\infty$ siempre decidimos que $X = -A$ y la probabilidad de error es

p , que es la probabilidad de que $X=+A$. Evidentemente, hay un valor de γ que minimiza la probabilidad de equivocarnos. Veamos cuál es ese valor.

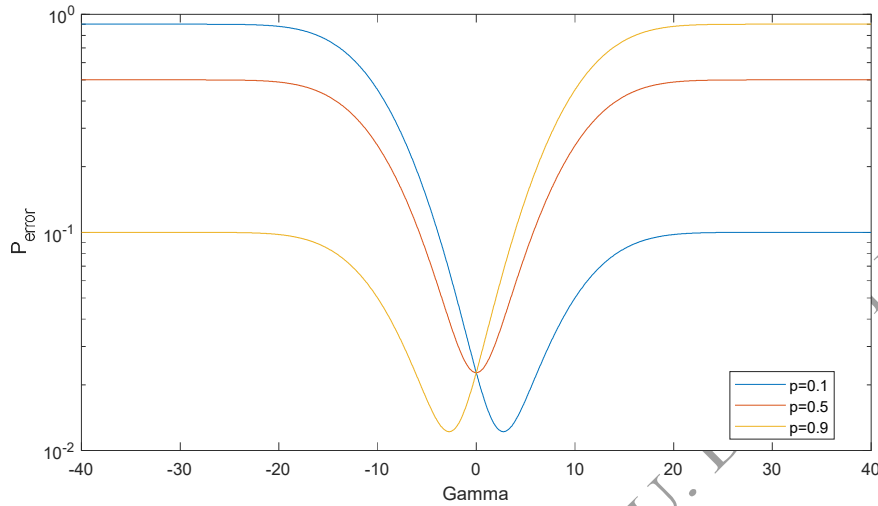


Figura 95. Probabilidad de error en función del nivel de decisión

Para minimizar la probabilidad de error debemos escoger la hipótesis más probable dada la observación:

$$P[H_1 | Y = y] \underset{H_0}{\underset{H_1}{\geq}} P[H_0 | Y = y]$$

Como solo conocemos las probabilidades a priori ($p=P[H_1]$, $1-p=P[H_0]$) y los likelihoods ($Y|H_0 \sim \mathcal{N}(-A, \sigma^2)$, $Y|H_1 \sim \mathcal{N}(+A, \sigma^2)$), debemos usar la regla de Bayes:

$$\begin{aligned} P[H_0 | Y = y] &= \lim_{\Delta y \rightarrow 0} \frac{P[\{X = -A\} \cap \{y < Y \leq y + \Delta y\}]}{P[\{y < Y \leq y + \Delta y\}]} = \dots \\ &= \lim_{\Delta y \rightarrow 0} \frac{P[H_0] P[\{y < Y \leq y + \Delta y\} | H_0]}{P[H_0] P[\{y < Y \leq y + \Delta y\} | H_0] + P[H_1] P[\{y < Y \leq y + \Delta y\} | H_1]} \\ &= \lim_{\Delta y \rightarrow 0} \frac{(1-p) f_{Y|H}(y | H_0) \Delta y}{(1-p) f_{Y|H}(y | H_0) \Delta y + p f_{Y|H}(y | H_1) \Delta y} \\ &= \frac{(1-p) f_{Y|H}(y | H_0)}{(1-p) f_{Y|H}(y | H_0) + p f_{Y|H}(y | H_1)} \end{aligned}$$

Similarmente

$$P[H_1 | Y = y] = \frac{p f_{Y|H}(y | H_1)}{(1-p) f_{Y|H}(y | H_0) + p f_{Y|H}(y | H_1)}$$

Como ambas expresiones tienen el mismo denominador, basta con comparar los numeradores:

$$p \underset{H_0}{f_{Y|H}}(y | H_1) \underset{H_1}{\geq} (1-p) \underset{H_0}{f_{Y|H}}(y | H_0)$$

Lo cual se convierte en una prueba sobre la razón de los likelihoods (LRT –Likelihood Ratio Test–)

$$\frac{\underset{H_1}{f_{Y|H}}(y | H_1)}{\underset{H_0}{f_{Y|H}}(y | H_0)} \underset{H_0}{\geq} \frac{1-p}{p}$$

Nótese que, hasta aquí, todo es idéntico al ejemplo de la definición 64, sólo que usamos las *pdf* condicionales en vez de las *pmf* condicionales por tratarse de una variable continua. En nuestro caso se trata de distribuciones gaussianas, de manera que

$$\frac{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-A}{\sigma}\right)^2\right]}{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y+A}{\sigma}\right)^2\right]} = \exp\left[\frac{1}{2}\left(\left(\frac{y+A}{\sigma}\right)^2 - \left(\frac{y-A}{\sigma}\right)^2\right)\right] \underset{H_0}{\underset{H_1}{\geq}} \frac{1-p}{p}$$

Expandiendo los cuadrados y simplificando,

$$\exp\left[\frac{2Ay}{\sigma^2}\right] \underset{H_0}{\underset{H_1}{\geq}} \frac{1-p}{p}$$

Al tomar logaritmos y multiplicar a ambos lados por $\sigma^2/2A$, obtenemos el criterio de decisión que deseábamos:

$$y \underset{H_0}{\underset{H_1}{\geq}} \gamma = \frac{\sigma^2}{2A} \log\left(\frac{1-p}{p}\right)$$

Efectivamente, sobre la gráfica de la Figura 95 podemos superponer la mínima probabilidad de error que corresponde a cada valor de γ , como muestra la Figura 96.

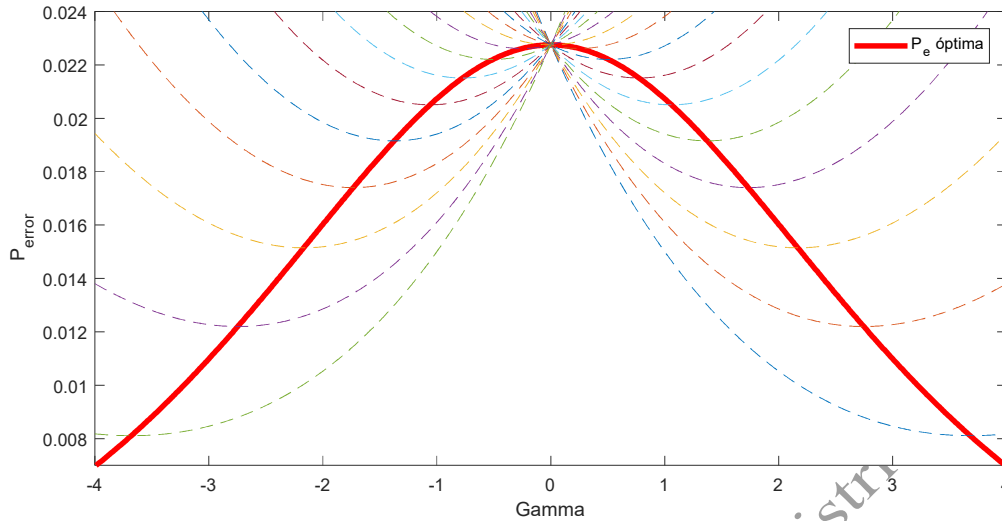


Figura 96. Probabilidad de error óptima

Sin embargo no sinteresa ahora el caso en que podemos disponer de varias observaciones durante el período de un solo bit. Es decir, en vez de ver un valor $Y=y$, vemos un vector de valores $\underline{Y}=\underline{y}=[y_1, y_2, \dots, y_n]^T$. Los likelihoods son ahora $f_{\underline{y}|H_0}(\underline{y}|H_0)$ y $f_{\underline{y}|H_1}(\underline{y}|H_1)$. Si consideramos que cada una de las observaciones es independiente de las demás (que es la razón de hablar de ruido blanco gaussiano, como veremos en la definición 115), estos likelihoods toman la siguiente forma:

$$f_{\underline{y}|H_0}(\underline{y}|H_0) = \prod_{i=1}^n f_{y_i|H_0}(y_i|H_0), \quad f_{\underline{y}|H_1}(\underline{y}|H_1) = \prod_{i=1}^n f_{y_i|H_1}(y_i|H_1)$$

Y como todas las muestras están condicionalmente igualmente distribuidas,

$$f_{\underline{y}|H_0}(\underline{y}|H_0) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\underline{y} + \underline{A})^T(\underline{y} + \underline{A})\right]$$

$$f_{\underline{y}|H_1}(\underline{y}|H_1) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\underline{y} - \underline{A})^T(\underline{y} - \underline{A})\right]$$

Donde \underline{A} es el vector $n \times 1$ $A \cdot [1, 1, \dots, 1]^T$, en el que todos los elementos son el valor constante A .

El likelihood ratio es

$$\frac{f_{\underline{y}|H_1}(\underline{y}|H_1)}{f_{\underline{y}|H_0}(\underline{y}|H_0)} = \exp\left[\frac{1}{2\sigma^2}\left[(\underline{y} + \underline{A})^T(\underline{y} + \underline{A}) - (\underline{y} - \underline{A})^T(\underline{y} - \underline{A})\right]\right] = \exp\left[-\frac{2}{\sigma^2}\underline{y}^T \underline{A}\right] = \exp\left[\frac{2A}{\sigma^2} \sum_{i=1}^n y_i\right]$$

De manera que el criterio de decisión es

$$\exp\left[\frac{2A}{\sigma^2} \sum_{i=1}^n y_i\right] \underset{H_0}{\underset{H_1}{\geq}} \frac{1-p}{p}$$

Tomando logaritmos, multiplicando a ambos lados por $\sigma^2/2A$ y dividiendo por n , obtenemos el siguiente criterio de decisión:

$$\frac{1}{n} \sum_{i=1}^n y_i \underset{H_0}{\overset{H_1}{\geq}} \frac{\sigma^2 / n}{2A} \log \left(\frac{1-p}{p} \right)$$

Nótese que si $n=1$ (una sola muestra) obtenemos el mismo criterio anterior. Pero ahora todas las muestras se resumen en el promedio muestral (una “estadística suficiente”, como veremos más adelante) y el resultado neto es que la varianza del error se ha reducido a una n -ésima parte, por lo que su efecto sobre la estimación es mucho menor (definición 77). Efectivamente, cuando se toman varias muestras independientes y se promedian, la gráfica de la Figura 94 se va estrechando como muestra la Figura 97.

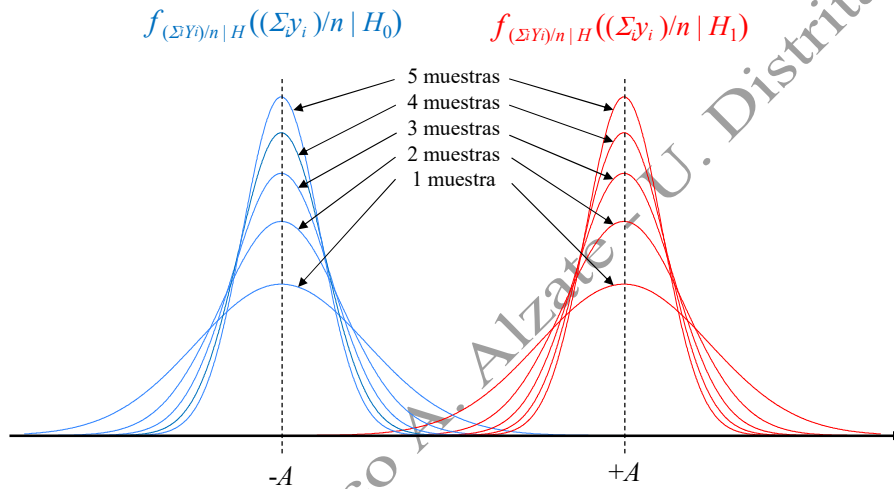


Figura 97. Distribución del promedio muestral de varias mediciones

En consecuencia, la probabilidad de error es

$$P_E = \frac{1-p}{2} \left[1 - \operatorname{erf} \left(\frac{\sqrt{n} \gamma + A}{\sqrt{2} \sigma} \right) \right] + \frac{p}{2} \left[1 + \operatorname{erf} \left(\frac{\sqrt{n} \gamma - A}{\sqrt{2} \sigma} \right) \right]$$

para un nivel de comparación

$$\frac{1}{n} \sum_{i=1}^n y_i \underset{H_0}{\overset{H_1}{\geq}} \gamma,$$

donde el γ óptimo es

$$\gamma^* = \frac{\sigma^2 / n}{2A} \log \left(\frac{1-p}{p} \right),$$

como muestra la Figura 98 cuando $p=0.5$ ($\gamma=0$). Con sólo pasar de 1 a 7 muestras redujimos la tasa de errores de un error cada 44 bits a un error cada 16.5 millones de bits.

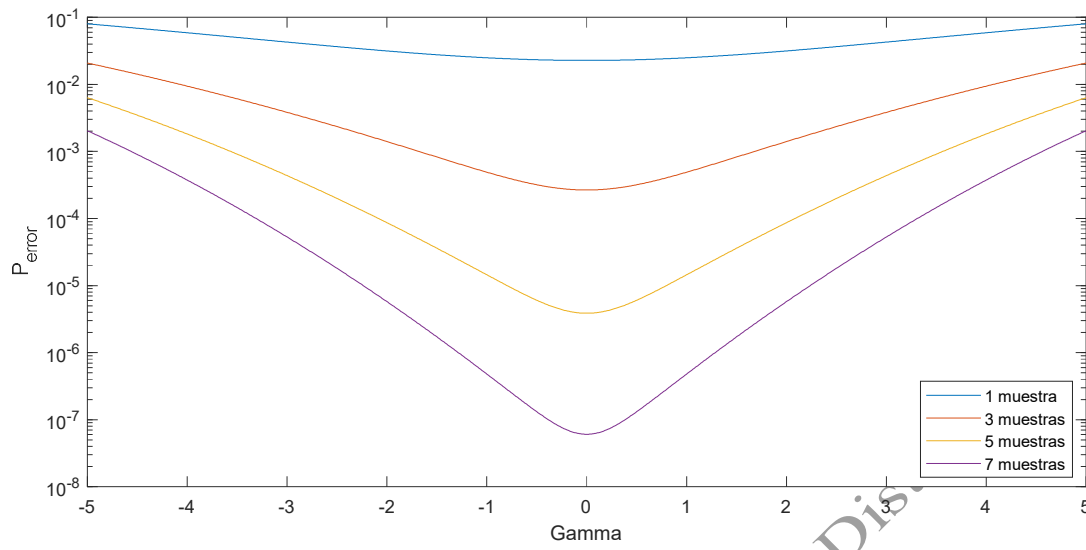


Figura 98. Probabilidad de error con múltiples muestras de cada bit

250 Conceptos - Marco A. Alzate - U. Dis.

VI. Variables aleatorias como vectores en un espacio vectorial

*El conjunto de todas las variables aleatorias con valor esperado cero y varianza finita que se pueden definir en un espacio de probabilidad (Ω, \mathcal{F}, P) dado forman un **espacio vectorial** sobre el campo escalar de los reales.*

El concepto de variable aleatoria como vector (así como el concepto de señal determinística como vector) en un espacio vectorial apropiado es tan importante que tomaremos una larga ruta para explicar de qué se trata esta idea, empezando por algunos repases básicos.

En el curso de álgebra lineal se definió un **vector columna de N dimensiones** como un arreglo de números reales ordenados verticalmente, como el siguiente:

$$\underline{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}, \quad \text{donde } x_i \in \mathbb{R} \text{ para } i = 0, 1, 2, \dots, N-1$$

Para $N=2$ tenemos una pareja ordenada de números reales, que es un elemento del producto cartesiano $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. Para $N=3$ tenemos un trío ordenado de números reales, que es un elemento del producto cartesiano $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$. En general, al conjunto de vectores N -dimensionales le llamamos el **Espacio Euclidiano \mathbb{R}^N** . En los casos $N=1, 2$ ó 3 tenemos una interpretación geométrica muy precisa asociada con las cantidades físicas que tienen una magnitud y una dirección en el espacio (posición, velocidad, aceleración, fuerza, momentum, torque, campo eléctrico, etc.). La Figura 99 muestra algunos ejemplos en \mathbb{R}^2 y \mathbb{R}^3 , donde se aprecia la facilidad de interpretar el concepto de vector desde la geometría.

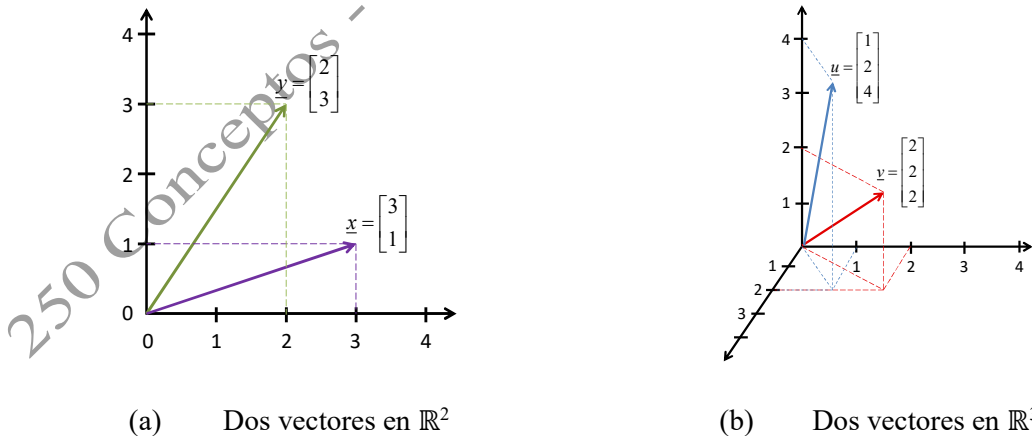


Figura 99. Cantidades con magnitud y dirección en el espacio, o vectores

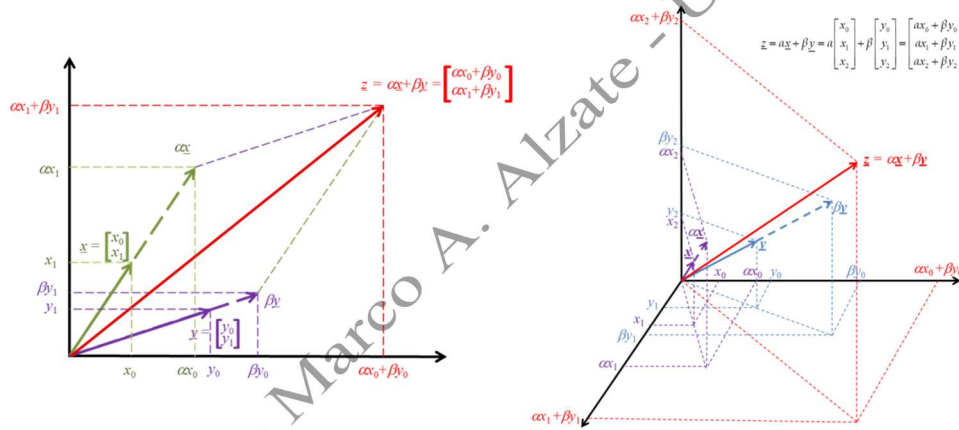
La interpretación geométrica nos permite comprender de manera intuitiva el significado del producto entre un escalar y un vector, la suma entre vectores y la combinación lineal de vectores,

$$\underline{y} = a\underline{x} = \begin{bmatrix} ax_0 \\ ax_1 \\ \vdots \\ ax_n \end{bmatrix} \quad \underline{z} = \underline{x} + \underline{y} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_0 + y_0 \\ x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \underline{z} = a\underline{x} + b\underline{y} = a \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} + b \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} ax_0 + by_0 \\ ax_1 + by_1 \\ \vdots \\ ax_n + by_n \end{bmatrix}$$

según se muestra en la Figura 100, para los casos de \mathbb{R}^2 y \mathbb{R}^3 .

Esas mismas operaciones las podemos extender a dimensiones superiores a 3 pues, aunque la representación geométrica no la podemos dibujar, si podemos extrapolar su significado intuitivo de acuerdo con las respectivas operaciones analíticas que acabamos de definir:

$$\underline{z} = a\underline{x} + \beta\underline{y} = a \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} + \beta \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} ax_0 \\ ax_1 \\ \vdots \\ ax_{N-1} \end{bmatrix} + \begin{bmatrix} \beta y_0 \\ \beta y_1 \\ \vdots \\ \beta y_{N-1} \end{bmatrix} = \begin{bmatrix} ax_0 + \beta y_0 \\ ax_1 + \beta y_1 \\ \vdots \\ ax_{N-1} + \beta y_{N-1} \end{bmatrix}$$



(a) Combinación lineal de vectores en \mathbb{R}^2 (b) Combinación lineal de vectores en \mathbb{R}^3

Figura 100. Combinación lineal de vectores

Estas interpretaciones geométricas de vectores en el espacio euclidiano \mathbb{R}^N se pueden extrapolar a espacios vectoriales más abstractos, como los definidos por variables aleatorias en un espacio de probabilidad dado. Para repasar estos conceptos un poco más lentamente, empezamos por algunos otros conceptos más básicos.

84. Grupo

Un **grupo** (G, Δ) es un conjunto de elementos, G , y una operación definida entre cualquier par de esos elementos, Δ , que satisfacen los siguientes cuatro axiomas:

- G es cerrado para Δ . Esto es, $a, b \in G \Rightarrow a \Delta b \in G$

- La operación Δ es asociativa. Esto es, $a, b, c \in G \Rightarrow a\Delta(b\Delta c) = (a\Delta b)\Delta c$
- La operación Δ tiene un elemento identidad en G . Esto es, $\exists e \in G : \forall a \in G, a\Delta e = e\Delta a = a$
- Cada elemento de G tiene un elemento inverso para Δ en G . Esto es, $\forall a \in G, \exists a^{-1} \in G : a\Delta a^{-1} = a^{-1}\Delta a = e$

Un **grupo abeliano** es un grupo que también satisface la propiedad de conmutatividad. Esto es:

- $a, b \in G \Rightarrow a\Delta b = b\Delta a$

Aunque las anteriores definiciones de grupo y de grupo abeliano parecen muy abstractas, la verdad es que estamos acostumbrados a muchos grupos interesantes con los que tratamos cotidianamente. Por ejemplo, el grupo más común entre nosotros es el grupo de los números reales y la operación de suma entre reales, $(\mathbb{R}, +)$. Otros tres grupos abelianos que usamos cotidianamente son $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ y $(\mathbb{C}, +)$: los enteros, los racionales y los complejos. Otro conjunto que usamos a menudo, $(\mathbb{N}, +)$, no constituye un grupo debido a que los elementos de \mathbb{N} carecen del inverso aditivo. Sin embargo, podemos limitarnos a un rango finito de los enteros, $\mathbb{Z}_N = \{0, 1, 2, \dots, N-1\}$ para algún entero N mayor o igual a 2, y utilizar la suma módulo N , $+_{(N)}$ para establecer la estructura de grupo abeliano. Un caso particular del anterior grupo, de gran interés para la ingeniería electrónica, de computación y de comunicaciones, corresponde al caso $N = 2$, ya que los elementos del grupo son los bits 0 y 1, y la operación de suma módulo 2 ($+_{(2)}$) corresponde con la operación lógica XOR (Figura 101).

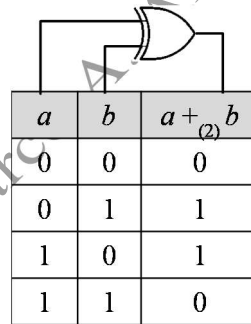


Figura 101. Grupo $(\{0,1\}, \text{XOR})$

El conjunto de los reales con la operación producto, (\mathbb{R}, \cdot) , cumple con todas las propiedades de un grupo abeliano, excepto porque el 0 no tiene inverso multiplicativo y, por tanto, (\mathbb{R}, \cdot) no es un grupo. Sin embargo, basta con sacar el elemento 0 de \mathbb{R} para obtener el grupo abeliano $(\mathbb{R} \setminus \{0\}, \cdot)$. Por supuesto, $(\mathbb{Q} \setminus \{0\}, \cdot)$ y $(\mathbb{C} \setminus \{0\}, \cdot)$ también forman grupos abelianos con elemento identidad 1, donde el inverso multiplicativo de x es $x^{-1} = 1/x$ en todos ellos, pero $(\mathbb{Z} \setminus \{0\}, \cdot)$ no lo es por la ausencia de inversos multiplicativos en \mathbb{Z} . Es posible inventar subconjuntos de los reales que satisfacen las propiedades de grupo con respecto a alguna operación numérica. Por ejemplo, $(G = \{a + b\sqrt{3}, a, b \in \mathbb{Q}\}, +)$ forma un grupo abeliano. También podemos sacudirnos de estereotipos y notar que G no tiene que ser un conjunto de números y que Δ no tiene que ser sumas o multiplicaciones entre números. Por ejemplo (G, Δ) forma un grupo (no-abeliano) si $G =$ Conjunto de permutaciones de n objetos y $\Delta =$ Aplicación de una permutación sobre otra. Por ejemplo, las configuraciones de un cubo de Rubik

forman un grupo con respecto a la rotación de una de sus caras en un ángulo recto. Otro ejemplo de un grupo (abeliano) es el de las matrices rotadoras 2×2 con respecto a la multiplicación matricial:

$$G = \left\{ M(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \theta \in [0, 2\pi) \right\}, \quad M(\theta)M(\alpha) = M(\theta + \alpha)$$

donde la suma entre ángulos se hace módulo 2π .

Los ingenieros podemos llegar muy lejos inventando grupos que satisfagan propiedades útiles para nuestros propósitos. Un ejemplo muy interesante es la criptografía de curva elíptica, donde se considera un conjunto de puntos en el plano cartesiano

$$G = \{(x,y) \in \mathbb{R}^2 : y^2 = x^3 - px - q \text{ para algún par } p,q \in \mathbb{R}\} \cup \{\emptyset\}$$

donde \emptyset es un punto en el infinito. La operación binaria entre dos puntos de G se denota por el símbolo $+$ y se define para dos puntos P y Q en G así: Si X el punto que interseca la curva elíptica al trazar una línea recta entre los puntos P y Q , el punto $P+Q$ es el punto que interseca nuevamente la curva elíptica al trazar una línea recta vertical que pase por X . Si uno de los puntos P o Q hace que la recta que los une sea tangente a la curva, ese punto se repite en la suma. Si la recta es vertical, se introduce el punto en el infinito, que se denota por \emptyset . La Figura 102 muestra un ejemplo de la operación $P+Q=R$. Bajo la operación indicada, los puntos en la curva elíptica forman un grupo abeliano (En criptografía de curva elíptica se restringen los puntos de la curva al conjunto \mathbb{Z}_p , donde p es un primo muy grande, para asegurar así propiedades criptográficas útiles).

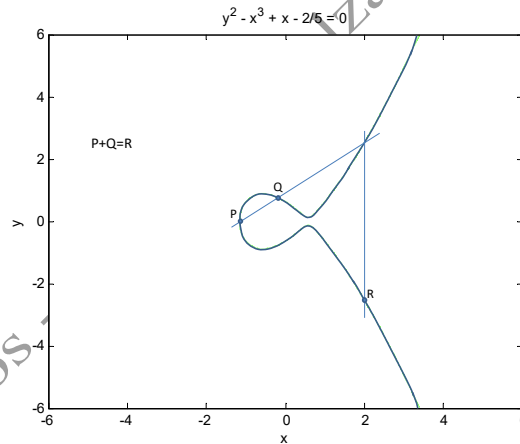


Figura 102. Suma entre dos puntos de una curva elíptica

Un último ejemplo de gran interés para nosotros es el grupo formado por las variables aleatorias definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) , $\mathcal{X} = \{X: \Omega \rightarrow \mathbb{R} \mid A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \forall x \in \mathbb{R}\}$, con respecto a la suma $X+Y = \{(X+Y)(\omega) = X(\omega)+Y(\omega), \omega \in \Omega\} \forall X, Y \in \mathcal{X}$. El par $(\mathcal{X}, +)$ forma un grupo abeliano porque

- Si $X: \Omega \rightarrow \mathbb{R}$ es una función tal que $A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \forall x \in \mathbb{R}$ y $Y: \Omega \rightarrow \mathbb{R}$ es una función tal que $B(y) = \{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F} \forall y \in \mathbb{R}$, entonces $C(z) = \{\omega \in \Omega : X(\omega)+Y(\omega) \leq z\} \in \mathcal{F} \forall z \in \mathbb{R}$ porque, $C(z) = \bigcup_{q \in \mathbb{Q}} \{\omega \in \Omega : X(\omega) \leq q\} \cap \{\omega \in \Omega : Y(\omega) \leq z - q\} \forall z \in \mathbb{R}$.

- $(X(\omega) + Y(\omega)) + Z(\omega) = X(\omega) + (Y(\omega) + Z(\omega)) \forall \omega \in \Omega$.
- Sea $O \in \mathcal{X}$ la variable aleatoria $O = \{O(\omega) = 0, \omega \in \Omega\}$. Entonces $\forall X \in \mathcal{X}, X+O = O+X = X$.
- $\forall X \in \mathcal{X} \exists -X \in \mathcal{X}$ definida así $-X = \{(-X)(\omega) = -X(\omega), \omega \in \Omega : \}$. De esta manera, $X+(-X) = (-X)+X = O$.
- $X(\omega) + Y(\omega) = Y(\omega) + X(\omega) \forall \omega \in \Omega$.

Hay un detalle técnico importante en este ejemplo particular. Decimos que las variables aleatorias X y Y son "iguales con probabilidad 1" si $P[\{\omega \in \Omega : X(\omega) = Y(\omega)\}] = 1$ (pueden haber algunos valores de $\omega \in \Omega$ para los cuales $X(\omega) \neq Y(\omega)$, pero la probabilidad del conjunto de esos valores de ω es cero). Pues bien, los elementos del grupo \mathcal{X} no son tanto variables aleatorias como clases equivalentes de variables aleatorias, donde la relación de equivalencia es la igualdad con probabilidad 1. Esto es, la expresión $X \in \mathcal{X}$ se refiere a todas las variables aleatorias X' que sean iguales a X con probabilidad 1. El elemento identidad, $O \in \mathcal{X}$, es la clase de variables aleatorias equivalentes tales que $P[\{\omega \in \Omega : O(\omega) = 0\}] = 1$. Para todo $X \in \mathcal{X}$ existe la clase equivalente $-X$ tal que $P[\{\omega \in \Omega : (-X)(\omega) = -X(\omega)\}] = 1$, etc.

85. Campo escalar

Un **campo escalar** es una tríada (F, Δ, \circ) compuesta por un conjunto de objetos F y dos operaciones binarias entre ellos, Δ y \circ , que satisfacen las siguientes propiedades

- (1) (F, Δ) forma un grupo abeliano con identidad i_Δ .
- (2) $(F \setminus \{i_\Delta\}, \circ)$ forma un grupo abeliano con identidad i_\circ .
- (3) La operación \circ se distribuye sobre la operación Δ : $\forall a, b, c \in F, a^\circ(b \Delta c) = (a^\circ b) \Delta (a^\circ c)$

Nuevamente, podemos ofrecer muchos ejemplos interesantes de campos escalares, aunque luego nos limitaremos principalmente al campo escalar de los reales con las operaciones de suma y producto $(\mathbb{R}, +, \cdot)$. Otro campo escalar muy interesante pero que casi no usaremos en este libro es el de los números racionales, $(\mathbb{Q}, +, \cdot)$; este es uno de los "más pequeños" campos escalares infinitos que se puedan construir. En cambio, es evidente que los enteros con la suma y el producto entre enteros no forma un campo escalar pues, aunque $(\mathbb{Z}, +)$ es un grupo abeliano con identidad 0, $(\mathbb{Z} \setminus \{0\}, \cdot)$ no constituye un grupo porque no existen los inversos multiplicativos de los enteros diferentes de 1 y -1. Sin embargo, si nos limitamos nuevamente a los enteros no negativos menores a un número primo p , podemos trabajar la suma y el producto módulo p para definir el campo $(\mathbb{Z}_p, +_{(p)}, \cdot_{(p)})$, conocido como "Campo de Galois de orden p ", $GF(p)$. Un ejemplo fascinante de cómo la matemática abstracta de los campos de Galois $GF(p)$ encuentran aplicaciones útiles en el mundo de las comunicaciones se encuentra en el cifrador RSA. Allí el módulo p debe ser el producto de dos números primos muy grandes para asegurar un nivel adecuado de seguridad. Sin embargo, el campo de Galois que mayor impacto ha tenido en la historia de la civilización humana es $GF(2)$, donde la suma es la operación lógica XOR y el producto es la operación lógica AND. Sobre este campo escalar se construye toda la fundamentación de la computación y las comunicaciones digitales de hoy, es decir, ¡sobre este campo escalar se construye la civilización tecnológica que hoy domina el planeta! Este aporte tan

fundamental de Evariste Galois se hace aún más interesante cuando se revisa la biografía de este brillante matemático que murió a los 20 años de edad en un.

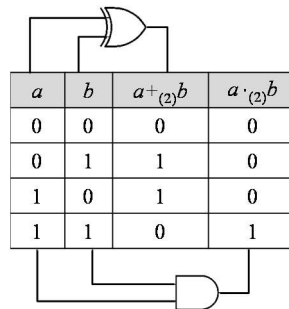


Figura 103. Campo escalar de Galois de orden 2, GF(2).

Existen muchos otros campos escalares interesantes. Por ejemplo, ya vimos que (1) los números racionales forman un campo escalar para la suma y el producto reales. Todos los números racionales son "números construibles", esto es, se pueden obtener a partir de una regla y un compás (dados dos puntos P y Q en el plano, existe una única línea $L(P,Q)$ que los conecta y un único círculo $C(P,Q)$ con centro en P que pasa por Q . Un punto X se puede "construir" a partir de los puntos A, B, C y D si X cae en la intersección de cualquier par de líneas y/o círculos determinados por A, B, C y D). (2) Los números construibles forman un campo escalar con respecto a la suma y la multiplicación reales, de manera que el campo escalar de los números racionales es un sub-campo del campo escalar de los números construibles (los números construibles se forman cerrando los racionales para la raíz cuadrada). Todos los números construibles son números algebraicos, esto es, son las raíces de polinomios finitos con coeficientes enteros. (3) Los números algebraicos forman un campo escalar con respecto a la suma y multiplicación complejos, de manera que el campo escalar de los números construibles son un sub-campo de los números algebraicos. Pero los números algebraicos son un subconjunto de (4) los números complejos que, como ya vimos, forman un campo escalar. Otra vez: los números racionales forman un sub-campo escalar de los números construibles, que forman un sub-campo escalar de los números algebraicos que forman un sub-campo escalar de los números complejos que forman un campo escalar... !Los matemáticos realmente se divierten con estos conceptos!

86. Espacio vectorial

Con los conceptos de grupo y campo escalar podemos repasar el concepto fundamental de **espacio vectorial**. Sea $(F, +, \cdot)$ un campo escalar con una operación binaria de suma entre escalares, $+: F \times F \rightarrow F$, y una operación binaria de producto entre escalares, $\cdot: F \times F \rightarrow F$. Un conjunto de objetos V , a cuyos elementos llamamos vectores, dotado de una suma entre vectores, $+: V \times V \rightarrow V$ y un producto por escalar, $\cdot: F \times V \rightarrow V$, forman el **espacio vectorial $(V, +, \cdot)$ sobre el campo escalar $(F, +, \cdot)$** si se cumplen las siguientes condiciones:

- V es cerrado para $+$. Esto es, $\forall \underline{x}, \underline{y} \in V, \underline{x} + \underline{y} \in V$
- La operación $+$ es asociativa. Esto es, $\forall \underline{x}, \underline{y}, \underline{z} \in V, \underline{x} + (\underline{y} + \underline{z}) = (\underline{x} + \underline{y}) + \underline{z}$

- La operación $+$ tiene un elemento identidad en V . Esto es, $\exists \underline{0} \in V : \forall \underline{x} \in V, \underline{0} + \underline{x} = \underline{x} + \underline{0} = \underline{x}$.
- Cada elemento de V tiene un elemento inverso para $+$ en V . Esto es, $\forall \underline{x} \in V, \exists -\underline{x} \in V$ tal que $\underline{x} + -\underline{x} = -\underline{x} + \underline{x} = \underline{0}$.
- $+$ es conmutativa, esto es, $\forall \underline{x}, \underline{y} \in V, \underline{x} + \underline{y} = \underline{y} + \underline{x}$.
- V es cerrada para el producto por escalar. Esto es, $\forall a \in F, \forall \underline{x} \in V, (a \cdot \underline{x}) \in V$
- El producto por escalar se distribuye entre la suma de escalares. Esto es, $\forall a, b \in F, \forall \underline{x} \in V, (a+b) \cdot \underline{x} = a \cdot \underline{x} + b \cdot \underline{x}$
- El producto por escalar se distribuye entre la suma de vectores. Esto es, $\forall a \in F, \forall \underline{x}, \underline{y} \in V, a \cdot (\underline{x} + \underline{y}) = a \cdot \underline{x} + a \cdot \underline{y}$
- El producto entre escalares y el producto por escalar son asociativos. Esto es, $\forall a, b \in F, \forall \underline{x} \in V, a \cdot (b \cdot \underline{x}) = (a \cdot b) \cdot \underline{x}$
- La identidad del producto entre escalares es la identidad del producto por escalar. Esto es, si 1 es la identidad de $(F \setminus \{0\}, \cdot)$, entonces $\forall \underline{x} \in V, 1 \cdot \underline{x} = \underline{x}$

Nótese el polimorfismo de los símbolos $+$ y \cdot , los cuales se usan para operaciones muy distintas: en $(F, +, \cdot)$ tenemos una suma y un producto de escalares, $+: F \times F \rightarrow F$ y $\cdot: F \times F \rightarrow F$ mientras que en $(V, +, \cdot)$ tenemos una suma entre vectores y un producto por escalar, $+: V \times V \rightarrow V$ y $\cdot: F \times V \rightarrow V$. Normalmente, el contexto dirá fácilmente a qué tipo de suma y a qué tipo de producto se refiere cada símbolo. A manera de ejemplo, nótese que en la expresión $a \cdot (b \cdot \underline{x}) = (a \cdot b) \cdot \underline{x}$ se transforma mágicamente un producto de escalar por vector en un producto entre escalares, mientras que en la expresión $(a+b) \cdot \underline{x} = a \cdot \underline{x} + b \cdot \underline{x}$ se transforma mágicamente una suma entre escalares en una suma entre vectores.

El ejemplo más conocido de espacio vectorial, pues hemos vivido sumergidos en él toda la vida, es el espacio Euclidiano $(\mathbb{R}^3, +, \cdot)$ sobre el campo escalar $(\mathbb{R}, +, \cdot)$. Todas las propiedades geométricas de este espacio surgen de que satisface las 10 condiciones de la definición de un espacio vectorial y, en ese sentido, todos los espacios vectoriales (esto es, todos los conjuntos que satisfagan las mismas diez condiciones) disfrutarán de propiedades geométricas semejantes. Veamos apenas algunos de esos ejemplos.

1. Consideremos el conjunto de señales reales en tiempo continuo definidas en el intervalo unitario $[0, 1]$, $X_{[0,1]} = \{x: [0,1] \rightarrow \mathbb{R}\}$. Ellas forman un espacio vectorial sobre el campo escalar de los reales si definimos el producto por escalar y la suma de vectores de la manera obvia:

$$\forall \underline{x} \in X_{[0,1]}, \forall a \in \mathbb{R}, \text{ si } \underline{z} = a \cdot \underline{x} \text{ entonces } \underline{z} = \{z(t) = (a \cdot x)(t) = a \cdot x(t), t \in [0,1]\}$$

$$\forall \underline{x}, \underline{y} \in X_{[0,1]}, \text{ si } \underline{z} = \underline{x} + \underline{y} \text{ entonces } \underline{z} = \{z(t) = (x+y)(t) = x(t) + y(t), t \in [0,1]\}$$

y, en general, como muestra la Figura 104,

$$\forall \underline{x}, \underline{y} \in X_{[0,1]}, \forall a, b \in \mathbb{R}, \text{ si } \underline{z} = a \cdot \underline{x} + b \cdot \underline{y} \text{ entonces } \underline{z} = \{z(t) = (ax+by)(t) = ax(t) + by(t), t \in [0,1]\}$$

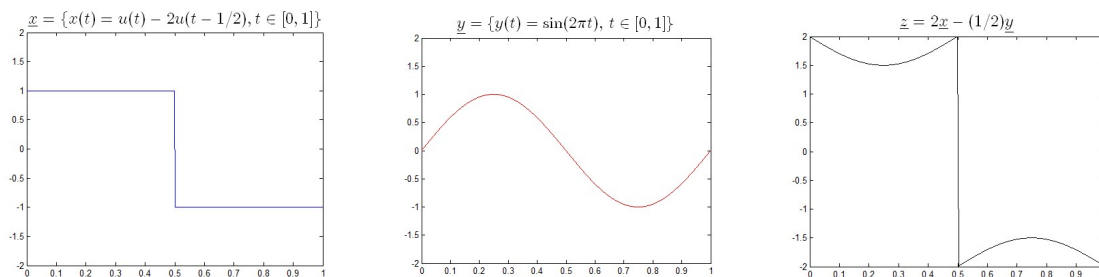


Figura 104. Producto por escalar y suma de vectores en $X_{[0,1]}$

Definiendo así el producto por escalar y la suma de vectores en $X_{[0,1]}$, es fácil ver que todas las propiedades de un espacio vectorial se ven satisfechas, pues todas ellas se reducen a propiedades de los números reales, en este caso los asociados con los valores instantáneos de las señales involucradas. Si el espacio de las señales continuas en el intervalo $[0, 1]$ forma un espacio vectorial, disfruta de todas las características geométricas asociadas a los espacios euclidianos.

2. El conjunto de las matrices reales $M \times N$ forma un espacio vectorial sobre el campo de los reales. Basta con definir de la manera obvia la suma entre matrices y el producto de un escalar por una matriz para verificar fácilmente su estructura de espacio vectorial.

3. El conjunto de polinomios P en una variable real t con coeficientes reales forma un espacio vectorial sobre el campo de los reales. Por ejemplo, el espacio vectorial de los polinomios de grado menor o igual a N se define como $V = \left\{ P: \mathbb{R} \rightarrow \mathbb{R} \text{ tales que } P(t) = \sum_{n=0}^N \alpha_n t^n, N \in \mathbb{N}, \alpha_n \in \mathbb{R}, t \in \mathbb{R} \right\}$. La

suma entre vectores es $P, Q \in V \Rightarrow (P + Q)(t), t \in \mathbb{R} = \{P(t) + Q(t), t \in \mathbb{R}\} \in V$. El producto por escalar

es $P = \left\{ \sum_{n=0}^N \alpha_n t^n, t \in \mathbb{R} \right\} \in V, \beta \in \mathbb{R} \Rightarrow \beta \cdot P = \left\{ (\beta \cdot P)(t) = \beta \cdot P(t) = \sum_{n=0}^N (\beta \alpha_n) t^n, t \in \mathbb{R} \right\} \in V$.

4. El conjunto de las secuencias de N bits, $\{0, 1\}^N$, forma un espacio vectorial sobre el campo escalar $\text{GF}(2)$ si definimos la suma entre vectores como la operación XOR bit a bit, y el producto por escalar como la operación AND entre el escalar y cada uno de los componentes del vector, como muestra la Figura 105.

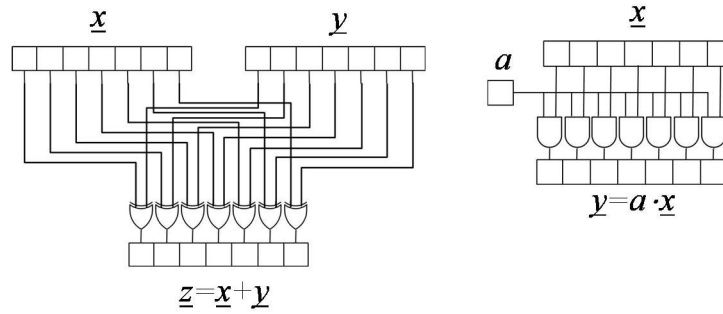


Figura 105. Suma entre vectores y producto por escalar en el espacio vectorial $(\{0,1\}^N, \text{XOR}, \text{AND})$

5. Las señales de energía en tiempo continuo, $L^2(\mathbb{R}) = \left\{ x: \mathbb{R} \rightarrow \mathbb{R} \mid \int_{-\infty}^{\infty} |x(t)|^2 dt < \infty \right\}$, donde la suma entre vectores es $\underline{z} = \underline{x} + \underline{y} = \{z(t) = (x+y)(t) = x(t) + y(t), t \in \mathbb{R}\}$ y el producto por escalar es $\underline{z} = a\underline{x} = \{z(t) = (ax)(t) = ax(t), t \in \mathbb{R}\}$
6. Las señales de energía en tiempo discreto, $l^2(\mathbb{Z}) = \left\{ x: \mathbb{Z} \rightarrow \mathbb{R} \mid \sum_{n=-\infty}^{\infty} |x[n]|^2 < \infty \right\}$, donde la suma entre vectores es $\underline{z} = \underline{x} + \underline{y} = \{z[n] = (x+y)[n] = x[n] + y[n], n \in \mathbb{Z}\}$ y el producto por escalar es $\underline{z} = a\underline{x} = \{z[n] = (ax)[n] = ax[n], n \in \mathbb{Z}\}$
7. Las variables aleatorias con varianza finita y media cero definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) , esto es, $L^2(\Omega, \mathcal{F}, P) = \{X: \Omega \rightarrow \mathbb{R} \mid \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \forall x \in \mathbb{R}, \text{Var}[X] < \infty, E[X] = 0\}$

Como es de esperar, nos detendremos en este último ejemplo. En comparación con el espacio vectorial de las señales de energía, $L^2(\mathbb{R})$, la notación $L^2(\Omega, \mathcal{F}, P)$ para el conjunto de variables aleatorias con varianza finita y media cero definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) sugiere una relación interesante entre la energía de una señal y la varianza de una variable aleatoria. En efecto, como veremos, en ambos casos se trata del cuadrado de la norma de un vector.

En el conjunto $L^2(\Omega, \mathcal{F}, P)$ podemos definir el producto entre una variable aleatoria y un número real (determinístico) y la suma entre variables aleatorias. Si $X \in L^2(\Omega, \mathcal{F}, P)$ tiene una CDF $F_X(x)$, entonces $Z = aX$ es otra variable aleatoria en $L^2(\Omega, \mathcal{F}, P)$ definida por la distribución de X y la probabilidad condicional $P[Z = ax | X = x] = 1$, de manera que³¹

³¹ Como una variable aleatoria no es una variable sino una función, y no es aleatoria sino determinística (definición 29), las definiciones de la suma entre vectores y el producto por escalar son más fáciles desde el espacio de probabilidad subyacente que desde la función de distribución acumulativa: $X+Y = \{(X+Y)(\omega) = X(\omega) + Y(\omega), \omega \in \Omega\}$, $\alpha X = \{(\alpha X)(\omega) = \alpha X(\omega), \omega \in \Omega\}$.

$$F_Z(z) = \begin{cases} F_X(z/a) & a > 0 \\ u(z) & a = 0 \\ 1 - F_X(z/a) + P[X = z/a] & a < 0 \end{cases}$$

Si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ tienen una JCDF $F_{XY}(x, y)$, entonces $Z = X + Y$ es otra variable aleatoria en $L^2(\Omega, \mathcal{F}, \mathbb{P})$ definida por el teorema de la probabilidad total (definición 64), así³¹

$$F_Z(z) = \int_{-\infty}^{\infty} F_{Y|X}(z - x|x) dF_X(x) = \int_{-\infty}^{\infty} F_{X|Y}(z - y|y) dF_Y(y)$$

Con esta definición de suma entre variables aleatorias de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ y producto por escalar es fácil ver que $(L^2(\Omega, \mathcal{F}, \mathbb{P}), +, \cdot)$ constituye un espacio vectorial sobre $(\mathbb{R}, +, \cdot)$ pues satisface las propiedades que definen a cualquier espacio vectorial:

1. $\forall X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}), X + Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$
2. $\forall X, Y, Z \in L^2(\Omega, \mathcal{F}, \mathbb{P}), X + (Y + Z) = (X + Y) + Z,$
3. $\forall X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}), X + Y = Y + X,$
4. Sea la variable aleatoria $O: \Omega \rightarrow \mathbb{R}$ tal que $\int_{o=-\infty}^{\infty} |o|^2 dF_O(o) = 0$ (por ejemplo, $O(\omega) = 0 \forall \omega \in \Omega$ o, con mayor generalidad, $P[O=0]=1$). Entonces $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), X + O = X$
5. $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \exists -X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ tal que $(-X) + X = O$
6. $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \forall a \in \mathbb{R}, aX \in L^2(\Omega, \mathcal{F}, \mathbb{P})$
7. $\forall X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \forall a \in \mathbb{R}, a \cdot (X + Y) = a \cdot X + a \cdot Y$
8. $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \forall a, b \in \mathbb{R}, (a + b) \cdot X = a \cdot X + b \cdot X$
9. $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \forall a, b \in \mathbb{R}, a \cdot (b \cdot X) = (a \cdot b) \cdot X$
10. $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), 1 \cdot X = X$

Como ya mencionamos en el ejemplo de Grupo, para ser un poco más precisos, un vector en el espacio vectorial $L^2(\Omega, \mathcal{F}, \mathbb{P})$ no es exactamente una variable aleatoria sino una clase de variables aleatorias equivalentes. Así, el vector $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ representa a todas las variables aleatorias Y tales que $P[Y=X] = 1$. Por ejemplo, el vector O no es tanto una variable aleatoria como una representación de todas las variables aleatorias que satisfacen la propiedad $P[O=0] = 1$. Lo que ocurre es que, como una variable aleatoria no es ni variable ni aleatoria sino función determinística, hay diferentes maneras como dos funciones pueden ser iguales. Por ejemplo, $\{f(x), x \in \mathbb{R}\} = \{g(x), x \in \mathbb{R}\}$ puede significar que $f(x) = g(x) \forall x \in \mathbb{R}$, o puede significar que $\int_{\mathbb{R}} |f(x) - g(x)|^2 dx = 0$ (en cuyo caso puede haber un número finito (o contable) de valores de x en los que f y g difieren). En nuestro contexto, cuando decimos que dos variables X y Y son "iguales", nos referimos a que $P[\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}] = 1$.

Cuando consideramos un espacio vectorial surgen algunos conceptos básicos como

87. Combinación lineal de vectores

Sea $\{X_1, X_2, \dots\}$ un conjunto de variables aleatorias en $L^2(\Omega, \mathcal{F}, \mathbb{P})$ y $\{\alpha_1, \alpha_2, \dots\}$ un conjunto de números reales. La combinación lineal de $\{X_1, X_2, \dots\}$ mediante los coeficientes $\{\alpha_1, \alpha_2, \dots\}$, $Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots$, es otra variable aleatoria en $L^2(\Omega, \mathcal{F}, \mathbb{P})$. En este caso conviene considerar la interpretación geométrica de la Figura 100.

88. Independencia lineal de un conjunto de vectores³²

Dado un subconjunto de variables aleatorias $\{X_1, X_2, \dots\} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, si ninguna de ellas puede expresarse como una combinación lineal de las otras, decimos que se trata de un conjunto linealmente independiente. Conviene considerar la interpretación geométrica de \mathbb{R}^2 , en la que dos vectores son linealmente independientes si no son colineales (Figura 106). Tres vectores o más en \mathbb{R}^2 no pueden ser linealmente independientes pues al menos uno de ellos es combinación lineal de los otros dos. Igualmente, en \mathbb{R}^3 , tres vectores son linealmente independientes si no son coplanares, pero cuatro vectores o más en \mathbb{R}^3 no pueden ser linealmente independientes.

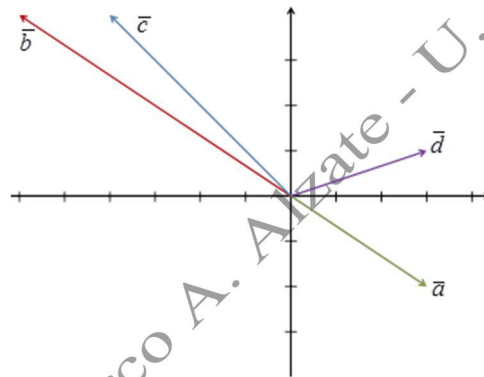


Figura 106. \underline{a} y \underline{b} son linealmente dependientes (colineales), los otros vectores son linealmente independientes

89. Subespacios vectoriales

Cualquier subconjunto $\mathcal{X} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ que sea cerrado para las combinaciones lineales (esto es, tal que $\forall X, Y \in \mathcal{X}, \forall \alpha, \beta \in \mathbb{R}, \alpha X + \beta Y \in \mathcal{X}$) es un espacio vectorial en sí mismo, es decir, es un subespacio vectorial de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Conviene considerar la interpretación geométrica de \mathbb{R}^2 como subespacio vectorial de \mathbb{R}^3 , por ejemplo.

90. Span de un conjunto de vectores

Dado un subconjunto de variables aleatorias $\{X_1, X_2, \dots\} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, podemos formar un subespacio vectorial de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ con todas las combinaciones lineales de $\{X_1, X_2, \dots\}$:

³² No debemos confundir la independencia lineal entre variables aleatorias con la independencia estadística. Si $Y = \sin(2\pi X)$ con $X \sim U(0,1)$, por ejemplo, X y Y son estadísticamente dependientes pero linealmente independientes.

$$\text{span}(X_1, X_2, \dots) = \{\alpha_1 X_1 + \alpha_2 X_2 + \dots, \forall \{\alpha_1, \alpha_2, \dots\} \subset \mathbb{R}\}$$

Conviene considerar la interpretación geométrica de $\mathbb{R}^2 = \text{span}([1 \ 0]^T, [0 \ 1]^T)$, porque $[a, b]^T = a[1 \ 0]^T + b[0 \ 1]^T$. Pero también $\mathbb{R}^2 = \text{span}([x_0 \ x_1]^T, [y_0 \ y_1]^T)$ siempre que $[x_0 \ x_1]^T$ y $[y_0 \ y_1]^T$ sean linealmente independientes. En efecto, $[a, b]^T = \alpha[x_0 \ x_1]^T + \beta[y_0 \ y_1]^T$, donde

$$\begin{bmatrix} a \\ \beta \end{bmatrix} = \begin{bmatrix} x_0 & y_0 \\ x_1 & y_1 \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix}$$

(la inversa de la matriz compuesta por los vectores base siempre existe porque los vectores base son linealmente independientes).

91. Bases y dimensión de un espacio vectorial

Sea $\mathcal{X} \subset L^2(\Omega, \mathcal{F}, P)$ un subespacio vectorial. Cualquier conjunto de variables aleatorias linealmente independientes $\{X_1, X_2, \dots\} \subset \mathcal{X}$ tales que $\mathcal{X} = \text{span}(X_1, X_2, \dots)$ es una base de \mathcal{X} . Todas las bases de \mathcal{X} tienen el mismo número de vectores. A ese número se le conoce como la dimensión de \mathcal{X} . Conviene considerar el ejemplo anterior, en el que $\mathbb{R}^2 = \text{span}([x_0 \ x_1]^T, [y_0 \ y_1]^T)$ para cualquier par de vectores $\underline{x} = [x_0 \ x_1]^T$ y $\underline{y} = [y_0 \ y_1]^T$ que sean linealmente independientes, por lo que la dimensión de \mathbb{R}^2 es 2.

92. Producto interno entre vectores

En cualquier espacio vectorial V se puede definir un producto interno como una función $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$, que satisface la siguientes propiedades:

$$\begin{aligned} \langle \underline{x}, \underline{x} \rangle &\geq 0, \text{ con igualdad sólo si } \underline{x} = \underline{0} \\ \langle \underline{x}, \underline{y} \rangle &= \langle \underline{y}, \underline{x} \rangle \quad (\text{en } \mathbb{C}^n, \langle \underline{x}, \underline{y} \rangle = \langle \underline{y}, \underline{x} \rangle^*) \\ \langle \underline{x} + \underline{z}, \underline{y} \rangle &= \langle \underline{x}, \underline{y} \rangle + \langle \underline{z}, \underline{y} \rangle \end{aligned}$$

$$\text{Para cualquier } \alpha \in \mathbb{R}, \langle \alpha \underline{x}, \underline{y} \rangle = \alpha \langle \underline{x}, \underline{y} \rangle \quad (\text{en } \mathbb{C}^n, \langle \underline{x}, \alpha \underline{y} \rangle = \alpha^* \langle \underline{y}, \underline{x} \rangle^*)$$

En \mathbb{R}^2 , un producto interno muy útil es el producto punto, $\langle [x_0 \ x_1]^T, [y_0 \ y_1]^T \rangle = x_0 y_0 + x_1 y_1$, o correlación. En $L^2(\Omega, \mathcal{F}, P)$, estas cuatro propiedades quedan satisfechas por $\langle X, Y \rangle = E[XY]$, así que la correlación define un posible producto interno en el espacio $L^2(\Omega, \mathcal{F}, P)$. Obsérvese que otro producto interno que satisface casi todas las propiedades de $\langle X, Y \rangle$ es $\text{Cov}(X, Y)$, el cual es idéntico a $E[XY]$ si nos restringimos al espacio compuesto por las variables aleatorias con media cero. Es de notar que, en \mathbb{R}^2 , al producto punto $\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1$ se le conoce como correlación, al igual que a $E[XY]$ en $L^2(\Omega, \mathcal{F}, P)$. En efecto,

$$\begin{aligned} E[XX] &= E[X^2] \geq 0, \text{ con igualdad sólo si } P[X=0]=1 \\ E[XY] &= E[YX] \\ E[(X+Y)Z] &= E[XZ + YZ] = E[XZ] + E[YZ] \\ E[(\alpha X)Y] &= \alpha E[XY] \end{aligned}$$

La covarianza también podría ser un producto interno válido como la correlación, sólo que si el valor medio no es cero, debemos definir el vector cero como el conjunto de variables aleatorias para las cuales $P[X=E[X]]=1$:

$$\begin{aligned} Cov(X,X) &= E[(X-E[X])^2] = Var(X) \geq 0, \text{ con igualdad sólo si } P[X=E[X]]=1 \\ Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] = E[(Y-E[Y])(X-E[X])] = Cov(Y,X) \\ Cov(X+Y,Z) &= E[(X-E[X]+Y-E[Y])(Z-E[Z])] = \\ &= E[(X-E[X])(Z-E[Z]) + (Y-E[Y])(Z-E[Z])] = \\ &= E[(X-E[X])(Z-E[Z])] + E[(Y-E[Y])(Z-E[Z])] = Cov(X,Z) + Cov(Y,Z) \\ Cov(\alpha X,Y) &= E[\alpha(X-E[X])(Y-E[Y])] = \alpha E[(X-E[X])(Y-E[Y])] = \alpha Cov(X,Y) \end{aligned}$$

Como el vector 0 se define mejor con $P[X=0]=1$ que con $P[X=E[X]]=1$, solemos restringirnos al espacio vectorial de las variables aleatorias con media cero.

93. Norma de un vector

En cualquier espacio vectorial V se puede definir una norma como una función $\|\cdot\|: V \rightarrow \mathbb{R}$, que satisface las siguientes propiedades:

$$\forall a \in \mathbb{R}, \forall \underline{u}, \underline{v} \in V,$$

$$\|\underline{u}\| \geq 0, \text{ con igualdad si y sólo si } \underline{u} = \underline{0}.$$

$$\|\alpha \underline{u}\| = |\alpha| \|\underline{u}\|$$

$$\|\underline{u} + \underline{v}\| \leq \|\underline{u}\| + \|\underline{v}\|$$

En \mathbb{R}^2 es común usar la norma Euclidiana, $\|\underline{x}\|^2 = \langle \underline{x}, \underline{x} \rangle = x_0^2 + x_1^2$, directamente relacionada con la longitud del vector. De hecho, la definición $\|\underline{x}\|^2 = x_0^2 + x_1^2$ es sólo una expresión del teorema de pitágoras: la suma del cuadrado de los catetos es el cuadrado de la hipotenusa. Por eso, a la última propiedad de una norma en cualquier espacio vectorial se le conoce como "la desigualdad del triángulo". En $L^2(\Omega, \mathcal{F}, \mathbb{P})$, las cuatro propiedades de una norma también quedan satisfechas por $\|\underline{X}\|^2 = \langle X, X \rangle = E[X^2]$, esto es, el tamaño del vector es la desviación estándar de la variable aleatoria si nos restringimos al espacio compuesto por las variables aleatorias con media cero.

94. Coeficiente de correlación

Nuevamente, las propiedades geométricas del espacio euclidiano \mathbb{R}^N asociado con el producto punto y la norma Euclidiana (o L^2), resultan de las propiedades de cualquier espacio vectorial, dotado de un producto interno en el que la norma esté determinada por ese producto interno. Por ejemplo, consideremos el significado geométrico del producto punto $\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1$, mostrado en la Figura 107. Como $x_0 = \|\underline{x}\| \cos(\theta_x)$, $x_1 = \|\underline{x}\| \sin(\theta_x)$, $y_0 = \|\underline{y}\| \cos(\theta_y)$ y $y_1 = \|\underline{y}\| \sin(\theta_y)$, el producto punto toma la siguiente forma

$$\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1 = \|\underline{x}\| \cos(\theta_x) \|\underline{y}\| \cos(\theta_y) + \|\underline{x}\| \sin(\theta_x) \|\underline{y}\| \sin(\theta_y)$$

$$\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1 = \|\underline{x}\| \cdot \|\underline{y}\| \cdot (\cos(\theta_x) \cos(\theta_y) + \sin(\theta_x) \sin(\theta_y))$$

Haciendo uso de la identidad trigonométrica del coseno de la diferencia de dos ángulos,

$$\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1 = \|\underline{x}\| \cdot \|\underline{y}\| \cdot \cos(\theta_y - \theta_x)$$

pero $\theta = \theta_y - \theta_x$ es el ángulo entre los dos vectores \underline{x} y \underline{y} , de manera que

$$\langle \underline{x}, \underline{y} \rangle = \underline{y}^T \cdot \underline{x} = x_0 y_0 + x_1 y_1 = \|\underline{x}\| \cdot \|\underline{y}\| \cdot \cos(\theta)$$

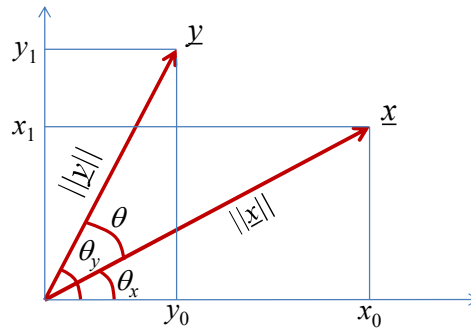


Figura 107. Interpretación geométrica del producto punto: $\langle \underline{x}, \underline{y} \rangle = x_0 y_0 + x_1 y_1 = \|\underline{x}\| \cdot \|\underline{y}\| \cdot \cos(\theta)$

Si calculamos el producto punto de un vector consigo mismo, en cuyo caso $\theta=0$ y $\cos(\theta)=1$, obtenemos el teorema de pitágoras que define la norma euclidiana:

$$\langle \underline{x}, \underline{x} \rangle = |x_0|^2 + |x_1|^2 = \|\underline{x}\|^2$$

Entonces, en cualquier espacio vectorial dotado de un producto interno en el que la norma esté definida por el producto interno, podemos hablar del "coseno del ángulo que forman dos vectores", aunque los vectores sean tan abstractos como dos variables aleatorias:

$$-1 \leq \cos(\theta) = \langle \underline{x}, \underline{y} \rangle / (\|\underline{x}\| \cdot \|\underline{y}\|) \leq 1$$

que es exactamente lo que hemos llamado **coeficiente de correlación**:

$$-1 \leq \rho_{xy} = \frac{\langle \underline{x}, \underline{y} \rangle}{\|\underline{x}\| \cdot \|\underline{y}\|} \leq 1$$

y que, en el espacio $L^2(\Omega, \mathcal{F}, P)$, corresponde a

$$-1 \leq \rho_{XY} = \frac{Cov[X, Y]}{\sigma_X \sigma_Y} \leq 1$$

tal como se había establecido en la definición 69. Por último, como $\langle X, Y \rangle = E[XY]$ es un producto interno válido, $\|X\|^2 = \langle X, X \rangle = E[X^2]$ define una norma válida que coincide con la desviación estándar si $E[X]=0$, en cuyo caso

$$-1 \leq \rho_{XY} = \frac{E[XY]}{\sqrt{E[X^2]E[Y^2}}} \leq 1$$

95. Proyección perpendicular de un vector sobre otro

En \mathbb{R}^n , el componente de \underline{y} en la dirección de \underline{x} tiene una longitud $\|\underline{y}\| \cdot \cos(\theta) = \langle \underline{x}, \underline{y} \rangle / \|\underline{x}\|$, como muestra la Figura 108. De esta manera podemos obtener la proyección ortogonal de \underline{y} sobre \underline{x} si multiplicamos dicha longitud por el vector unitario $\underline{x}/\|\underline{x}\|$ para obtener

$$\hat{\underline{y}} = \frac{\langle \underline{x}, \underline{y} \rangle}{\|\underline{x}\|^2} \underline{x}$$

Esta aproximación minimiza la magnitud del error $\underline{y} - \hat{y}$ entre todas las aproximaciones de \underline{y} en la dirección de \underline{x} , como muestra intuitivamente la geometría de la Figura 108.

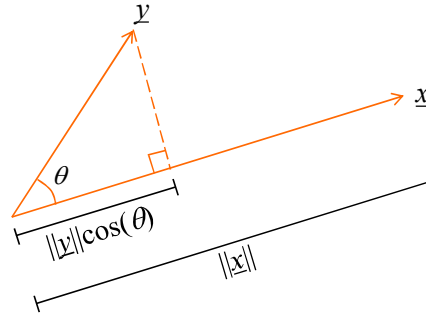


Figura 108. Proyección perpendicular de \underline{y} sobre \underline{x}

Igualmente, si X y Y son dos variables aleatorias con media cero y varianza finita, al remplazar Y por αX cometemos un error cuadrado promedio $E[(Y-\alpha X)^2]$. Para minimizar este error podemos derivar respecto a α e igualar la derivada a cero (después de todo, el error cuadrado promedio forma una parábola con respecto a α). Como el valor esperado es una integral sobre los posibles valores de X y Y y la derivada se hace con respecto a α , podemos intercambiar los operadores: $dE[(Y-\alpha X)^2]/d\alpha = E[2(Y-\alpha X)(-X)] = 2\alpha E[X^2] - 2E[XY]$. Al igualar a cero y despejar α obtenemos exactamente el mismo resultado,

$$\hat{Y} = \frac{E[XY]}{E[X^2]} X = \frac{\langle X, \underline{Y} \rangle}{\|X\|^2} X .$$

96. Ejemplo: Dos variables, dos vectores

Veamos un ejemplo simple pero revelador en el que existe un vector bivaluado con *pdf*

$$f_{XY}(x, y) = 2e^{-(x+y)}u(x)u(y)u(x-y)$$

donde $u: \mathbb{R} \rightarrow \mathbb{R}$ es la función escalón unitario, que vale 0 si la entrada es menor que cero y vale uno si la entrada es mayor o igual a cero. Esto es,

$$f_{XY}(x, y) = \begin{cases} 2e^{-(x+y)} & x \geq y \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Primero que todo nótese que, aunque cada una de las variables X y Y puede tomar cualquier valor no negativo, conjuntamente sólo pueden tomar valores en la región mostrada en la Figura 109.

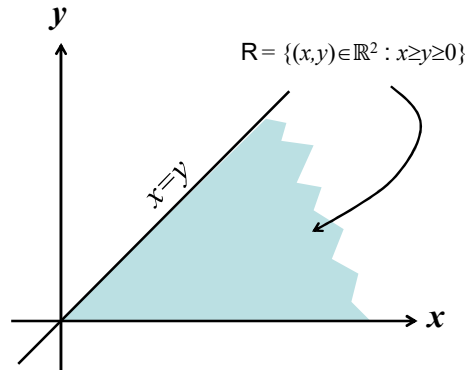


Figura 109. Región de validez de las variables X y Y mencionadas en el texto

Ahora debemos verificar que se trata de una *pdf* válida. Claramente es una función no negativa, pues $e^{-x} \geq 0 \forall x \geq 0$ (en realidad, para todo $x > -\infty$). Además,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 2 \int_0^{\infty} e^{-x} \int_0^x e^{-y} dy dx = 2 \int_0^{\infty} e^{-x} (1 - e^{-x}) dx = e^{-x} (e^{-x} - 2) \Big|_{x=0}^{\infty} = 1$$

Nótese que los rangos de integración aseguran que estemos evaluando las integrales en la región válida, $x \geq y \geq 0$, mostrada en la Figura 109. Una función no negativa cuyo volumen debajo de la superficie sea uno es una *pdf* válida.

El cálculo de las distribuciones marginales es igualmente fácil:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = 2e^{-x} \int_0^x e^{-y} dy = 2e^{-x} (1 - e^{-x}), \quad x \geq 0$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = 2e^{-y} \int_y^{\infty} e^{-x} dx = 2e^{-2y}, \quad y \geq 0$$

Y es una variable aleatoria exponencialmente distribuida con promedio $1/2$, a diferencia de X , que tiene una distribución unimodal sesgada hacia la izquierda, como muestra la Figura 110.

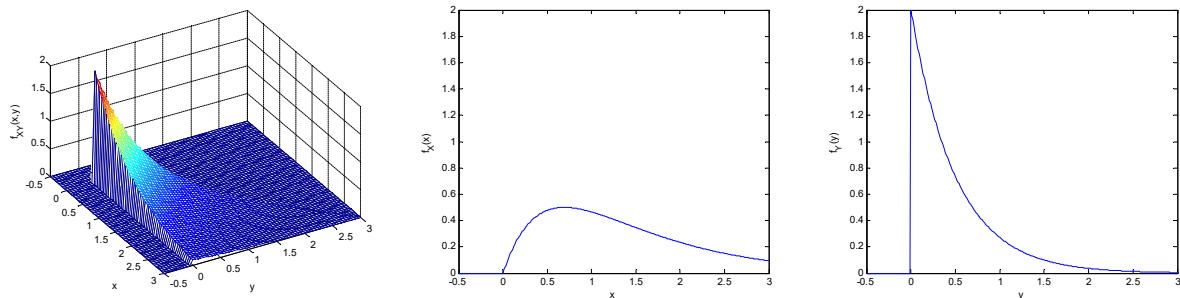


Figura 110. Una función de densidad de probabilidad conjunta y las respectivas marginales

La restricción en los valores posibles de X y Y es la que determina la dependencia entre las dos variables aleatorias. En efecto, $f_{XY}(x,y)$ se puede expresar como el producto entre las funciones $g_1(x) = (\sqrt{2})e^{-x}$ y $g_2(y) = (\sqrt{2})e^{-y}$, lo que sugeriría la independencia entre las variables aleatorias X y Y . Pero la verdad es que dado un valor x de X , Y no puede tomar un valor y mayor a x ; y dado un valor y de Y , X no puede tomar un valor x menor a y . Esto es, las variables aleatorias X y Y no son independientes, como lo refleja el hecho que, en el producto $f_{XY}(x,y) = g_1(x)g_2(y)$, las funciones g_1 y g_2 no corresponden a funciones de densidad de probabilidad válidas, pues el área debajo de cada curva

no es 1 sino $\sqrt{2}$. Más aún, el producto de las marginales es $4e^{-(x+y)}(1-e^{-x})$ para $x \geq y \geq 0$, muy distinto a la verdadera *pdf*. Siendo así, conviene calcular las distribuciones condicionales de X dado Y y de Y dado X :

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} e^{-(x-y)} & x \geq y \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

$$f_{Y|X}(y|x) = \frac{f_{Y,X}(x,y)}{f_X(x)} = \begin{cases} \frac{e^{-y}}{1-e^{-x}} & x \geq y \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Efectivamente, las condicionales son diferentes a las marginales, que es otra demostración de la dependencia entre las dos variables: $f_{X|Y}(x|y)$ depende de y y $f_{Y|X}(y|x)$ depende de x , como muestra la Figura 111.

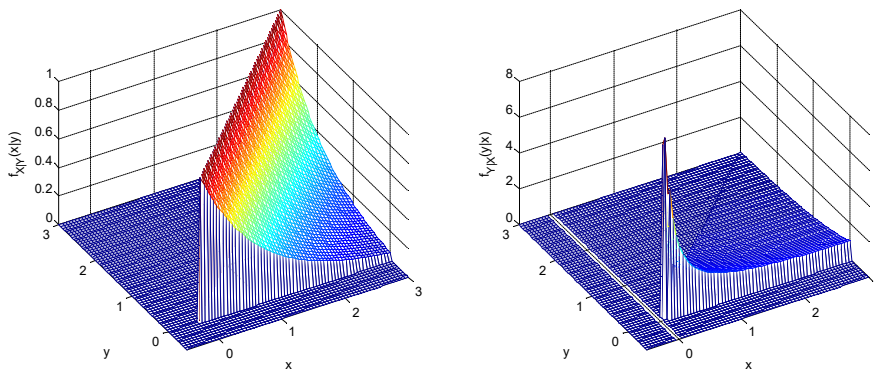


Figura 111. Funciones de densidad de probabilidad condicionales

Podemos calcular ahora algunas estadísticas:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = 2 \int_0^{\infty} x e^{-x} (1 - e^{-x}) dx = e^{-2x} \left(\frac{1}{2} + x \right) - 2e^{-x} (1 + x) \Big|_{x=0}^{\infty} = \frac{3}{2}$$

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = 2 \int_0^{\infty} y e^{-2y} dy = -\frac{1}{2} (1 + 2y) e^{-2y} \Big|_{y=0}^{\infty} = \frac{1}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = 2 \int_0^{\infty} x^2 e^{-x} (1 - e^{-x}) dx = e^{-2x} \left(\frac{1}{2} + x + x^2 \right) - 2e^{-x} (2 + 2x + x^2) \Big|_{x=0}^{\infty} = \frac{7}{2}$$

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y) dy = 2 \int_0^{\infty} y^2 e^{-2y} dy = -e^{-2y} \left(y^2 + y + \frac{1}{2} \right) \Big|_{y=0}^{\infty} = \frac{1}{2}$$

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x,y) dy dx = 2 \int_0^{\infty} x e^{-x} \left(\int_0^x y e^{-y} dy \right) dx = \dots$$

$$\dots = 2 \int_0^{\infty} x e^{-x} (1 - (1+x)e^{-x}) dx = e^{-x} (x+1) ((x+1)e^{-2x} - 2) \Big|_{y=0}^{\infty} = 1$$

de donde

$$V[X] = E[X^2] - E^2[X] = 5/4$$

$$V[Y] = E[Y^2] - E^2[Y] = 1/4$$

$$Cov(X, Y) = E[XY] - E[X]E[Y] = 1/4$$

$$\rho_{XY} = Cov(X, Y) / \sqrt{(V[X]V[Y])} = 1/\sqrt{5}$$

Así, hemos caracterizado completamente las variables X y Y y su variación conjunta. Sin embargo, vale la pena ver su interpretación geométrica. Para esto consideremos las variables de media cero $X_0 = X - E[X] = X - 3/2$ y $Y_0 = Y - E[Y] = Y - 1/2$. Claramente,

$$\langle X_0, Y_0 \rangle = E[X_0 Y_0] = Cov(X, Y) = 1/4$$

$$\|X_0\|^2 = \langle X_0, X_0 \rangle = E[X_0^2] = V[X] = 5/4$$

$$\|Y_0\|^2 = \langle Y_0, Y_0 \rangle = E[Y_0^2] = V[Y] = 1/4$$

$$\cos(\theta) = \frac{\langle X_0, Y_0 \rangle}{\|X_0\| \cdot \|Y_0\|} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1}{\sqrt{5}}$$

Producto punto entre los dos vectores
Norma cuadrada de X_0
Norma cuadrada de Y_0
Coseno del ángulo que forman

De esta manera, podemos representar geoméricamente las dos variables aleatorias como muestra la siguiente figura, en la que hemos incluido la proyección perpendicular de Y_0 sobre X_0 .

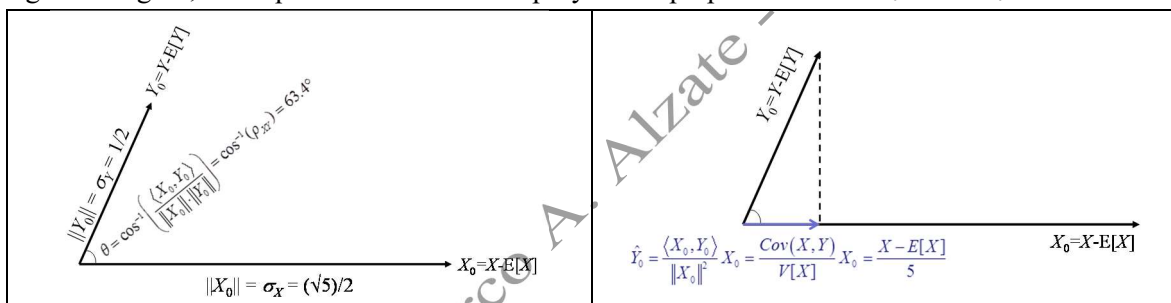


Figura 112. Dos variables aleatorias interpretadas como dos vectores en un espacio vectorial

Una interpretación interesante que explotaremos más adelante es la siguiente. Si no conocemos nada sobre Y_0 pero nos revelan X_0 , también nos revelan algo sobre Y_0 , pues el componente de Y_0 en la dirección de X_0 es $X_0/5$, de manera que nuestra incertidumbre sobre Y_0 se reduce a su componente perpendicular, $Y_0 - X_0/5$. Esta revelación es importante en el problema de la estimación de variables aleatorias a partir de observaciones indirectas, como veremos adelante.

X. Teoría de Colas

132. Teoría de Colas

*La teoría de colas es la rama de las matemáticas que estudia el fenómeno de formación de filas de **clientes** en espera de ser atendidos por uno o varios **servidores** de capacidad finita. El objetivo de la teoría de colas es poder predecir la longitud de la fila, la ocupación de los servidores, los tiempos de espera de los clientes, la tasa de rechazos por falta de cupo, etc., con el fin de proponer mecanismos de análisis, diseño y control.*

Como veíamos en 3, las redes de comunicaciones están conformadas por un conjunto de recursos de capacidad limitada que deben atender las demandas impuestas por los usuarios de la red. Las interacciones dinámicas que se forman cuando los usuarios compiten o colaboran para utilizar los recursos de la red es el tema de estudio de la teoría de colas. El nombre es muy razonable si se considera que, en estos casos, se suele formar una fila de espera, o cola, conformada por los usuarios que encontraron ocupado el recurso solicitado. Por supuesto, la aplicación de esta teoría ha rebasado de lejos el ámbito de las redes de comunicaciones pues es con base en la teoría de colas que se toman decisiones tales como cuántos cajeros ubicar en un supermercado, cuántas pistas de aterrizaje poner en un aeropuerto, cuántos carriles ubicar en los sistemas de entrada y salida de una autopista, a qué velocidad mover una banda transportadora de suministros en una línea de ensamblado, cómo temporizar adecuadamente un semáforo en un cruce de avenidas, etc.

En redes de telecomunicaciones se forman colas al compartir el ancho de banda de enlaces de capacidad finita, al transmitir paquetes en ranuras de tiempo sobre enlaces TDM, al acceder a estructuras de conmutación en enrutadores *cut-through*, al resolver conflictos en protocolos de acceso múltiple sobre enlaces de radio compartidos, etc. A niveles superiores se pueden formar colas en servidores de archivos, servidores web, procesadores en la nube, etc. De forma genérica vamos a representar cualquiera de estos casos en términos de "paquetes" y "enlaces". Esto es, la unidad operacional que fluye a través de un sistema de colas recibirá el nombre genérico de paquete, aunque en realidad puede tratarse de llamadas, sesiones, conexiones, flujos, solicitudes de servicio, demandas, trabajos, clientes, etc. Igualmente, diremos que el servicio solicitado lo ofrece, en términos genéricos, un "enlace", aunque en realidad puede ser un servidor de archivos, una ranura en un enlace TDM, un servidor web, una banda de frecuencia en un ambiente de radio, etc.

En todos estos ejemplos podemos detectar la misma estructura (**Figura 114**): una secuencia de paquetes llegan a un centro de servicio, donde existe cierto número de enlaces. Los paquetes desean ser transmitidos por al menos uno de los enlaces pero, si un paquete no encuentra enlaces disponibles al momento de su llegada, debe unirse a una cola a esperar que llegue su turno de transmisión. Cuando el paquete sea seleccionado para ser transmitido, se le asignará uno o varios enlaces, o una fracción de la capacidad de un enlace (de acuerdo con unas reglas conocidas como "disciplina de servicio"). El paquete ocupará los recursos asignados durante un tiempo específico hasta que termine su transmisión. Entonces diremos que el paquete salió del sistema.

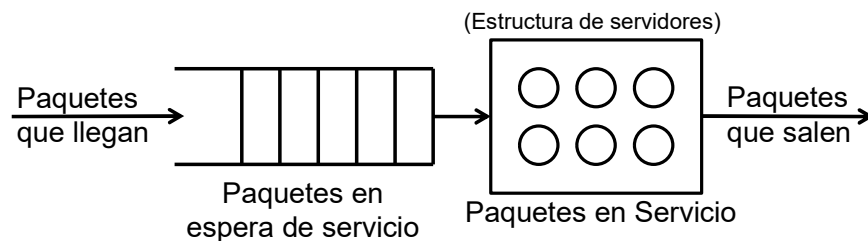


Figura 114. Estructura de un sistema de cola

133. Estructura Básica de un Modelo de Colas

Un modelo de colas se caracteriza por los siguientes elementos: (1) un proceso estocástico **A** que describe el **proceso de llegada** de los clientes; (2) un proceso estocástico **B** que describe el **proceso de servicio** de los clientes; (3) un número entero **C** que representa el **número de servidores**, (4) un número entero **D** que representa el **cupo** o máximo número de clientes que pueden permanecer dentro del sistema; (5) un número entero **E** que representa el **número de clientes potenciales** que podrían solicitar un servicio; (6) un mecanismo **F** de atención de clientes o **disciplina de servicio**; y (7) un mecanismo **G** de gestión de los cupos en el sistema o **administración de memoria**. La **notación de Kendall** se refiere a la especificación de las primeras tres características en el formato $A/B/C$, en cuyo caso se supone que el cupo en el sistema es infinito, que hay un número infinito de posibles clientes, que se atiende primero al primero que llega y que los cupos en el sistema se asignan por orden de llegada hasta que se agotan. Si alguna de estas características es diferente, se usa la notación extendida de Kendall $A/B/C/D/E/F/G$.

Se pueden hacer muchas suposiciones alternativas con respecto al comportamiento de los elementos descritos en la estructura de la **Figura 114**. Siete características relevantes por especificar se mencionan en la **Figura 115**, incluyendo una descripción completa del **proceso de llegada** de paquetes (instantes de llegada, tiempos entre llegadas o número de llegadas en un período de tiempo, por ejemplo) y del **proceso de servicio** (tiempo que toma la transmisión de cada paquete, el cual dependen a su vez de la longitud de cada paquete y del ancho de banda asignado a cada paquete). También se debe especificar el número de enlaces disponibles para atender a los paquetes que lleguen (**número de servidores** –o número de canales TDM, FDM o CDMA disponibles, por ejemplo–), así como la disponibilidad de memoria para almacenar tanto los paquetes en servicio como los paquetes que fueron admitidos al sistema pero que están esperando en la cola su turno de transmisión –**cupos en el sistema**–. En algunos casos el número de potenciales usuarios del sistema puede ser finito, de manera que si un usuario tiene una solicitud de servicio pendiente dentro del sistema puede quedar en espera, en cuyo caso la tasa de llegadas dependerá del número de paquetes en el sistema. Sin embargo, si suponemos un número infinito de usuarios, cada uno participando con un infinitésimo del tráfico, el análisis puede incluir que la tasa de llegadas sea independiente de la ocupación del sistema. Por eso es importante especificar también el **número de clientes**. Generalmente los paquetes

se suelen atender en orden de llegada (FIFO –First-in, first-out–), aunque a veces es importante asignar algún tipo de prioridad o buscar alguna manera más equitativa de asignar el ancho de banda, lo cual se traduce en diferentes **disciplinas de servicio**. Por último, al almacenar los paquetes en la cola podemos hacerlo en orden de llegada hasta que se agote el cupo –drop-tail–, o podemos asignar algún mecanismo activo de **gestión de memoria** que nos permita señalar el comienzo de la congestión antes de que se presente.

La **notación de Kendall** se usa para especificar las primeras tres características de un sistema de colas particular mediante tres parámetros A/B/C, donde A especifica el proceso de llegadas, B especifica el proceso de servicios y C especifica el número de servidores. En ese caso, los cuatro parámetros restantes toman los valores especificados en la **Figura 115**. Si se especifica más que los tres primeros parámetros, hablamos de la notación extendida de Kendall.

Proceso de llegadas	Proceso de servicios	Número de servidores	Cupos en el sistema (∞)	Número de clientes (∞)	Disciplina de servicio (FIFO)	Gestión de memoria (Droptail)
---------------------	----------------------	----------------------	----------------------------------	---------------------------------	-------------------------------	-------------------------------

Figura 115. Características por especificar en un sistema de cola

El proceso de llegadas se refiere al mecanismo mediante el cual los paquetes se generan a la entrada del sistema. Una alternativa es que los paquetes lleguen en instantes igualmente espaciados en el tiempo, un paquete cada τ segundos, en cuyo caso la tasa de llegadas es constante y corresponde a $\lambda=1/\tau$ paquetes por segundo. Esta suposición se representa por la letra **D** de **Determinístico**. Por supuesto, ésta no es una suposición muy realista en redes modernas de comunicaciones, porque la llegada de paquetes se suele dar en instantes aleatorios. Entonces se hace necesario describir la distribución de probabilidad conjunta de los instantes de llegadas.

Si las llegadas se suponen completamente aleatorias, los tiempos entre llegadas serán independientes e idénticamente distribuidos de acuerdo con una distribución exponencial, generando un proceso de llegadas tipo Poisson. En este caso, como los tiempos entre llegadas son exponenciales e independientes, esto es, sin memoria (definición 51), se habla de procesos **M** (Memoryless o **sin-Memoria**).

En otros casos, los tiempos entre llegadas puede seguir siendo independientes e idénticamente distribuidos (iid), pero no necesariamente exponenciales. Podemos hablar de procesos **GI**, distribución **general e independiente**, si sólo conocemos algunas estadísticas particulares tales como el valor medio y la varianza de los tiempos entre llegadas, pero no su distribución precisa. O podemos modelar específicamente una distribución de probabilidad no exponencial como la distribución **Erlang de k etapas**, o proceso E_k , o la distribución **hiper-exponencial de k etapas**, o proceso H_k .

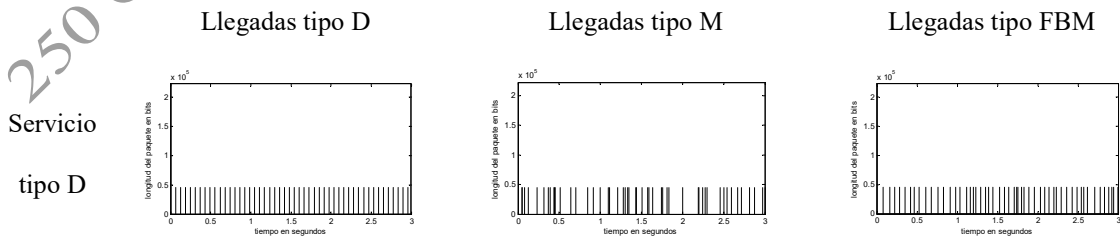
Desafortunadamente, la suposición de independencia entre los tiempos de llegada está lejos de ser apropiada para el tráfico en redes modernas de comunicaciones, pues se han detectado importantes correlaciones entre los tiempos entre llegadas de paquetes. Los primeros modelos intentan capturar correlaciones que decaen exponencialmente con el tiempo, o modelos **SRD –short range**

dependence-. Uno de los más comunes es el proceso de **Poisson Markovianamente Modulado (MMPP)**, en el que la intensidad de un proceso de Poisson varía según el estado de una cadena de Markov subyacente. En los procesos tipo **fase (PH)** los tiempos entre llegadas están dados por los instantes de absorción en una cadena de Markov finita con algunos estados transientes y un estado absorbente. Aunque este proceso es realmente un modelo no correlacionado porque después de la absorción la cadena se renueva (proceso de renovación), inspira el Proceso de Llegadas Markovianas (**MAP -Markovian Arrival Process-**): una cadena de Markov finita con varios estados absorbentes se reinicia en un estado transiente que depende del estado absorbente que alcanzó la última vez. El modelo MAP se puede extender al Proceso MAP por Lotes (**BMAP -Batch Markovian Arrival Process-**), en el que cada evento MAP genera un lote de llegadas (el tamaño del lote se puede modelar de diferentes maneras).

Otro modelo SRD de interés es el **Proceso Autoregresivo (AR)**, en el que el siguiente tiempo entre llegadas depende explícitamente de los anteriores tiempos entre llegadas dentro de una ventana de tiempo de longitud determinada. Si además se considera una combinación de una secuencia de entrada completamente aleatoria (modelo de **promedios móviles** o **MA -moving average-**), se constituye un proceso **ARMA** o modelo autoregresivo de promedios móviles. Si los tiempos entre llegadas corresponden a un modelo ARMA, los instantes de llegada se consiguen acumulando (o integrando) los tiempos entre llegadas, a los que se denomina modelo **ARIMA** (auto-regressive integrated moving average).

Sin embargo, mediciones detalladas de tráfico realizadas en las últimas dos décadas revelan una estructura de correlación mucho más rica y compleja en casi todos los tipos de tráfico sobre las modernas redes de comunicaciones, que se extiende a muchas escalas de tiempo, en lo que se conoce como "dependencia de rango largo", **LRD**. El fenómeno de la dependencia de rango largo hace necesario desarrollar modelos adicionales de tráfico capaces de representar estas correlaciones. A pesar de sutiles pero definitivas diferencias conceptuales, en general se habla de modelos de **tráfico fractal** cuando se representan las correlaciones mediante dependencia de rango largo. Un modelo fractal ampliamente usado para representar las secuencias de instantes de llegadas es el **FBM -Movimiento Browniano Fraccional-**, en cuyo caso al proceso de tiempos entre llegadas se le conoce como **FGN -Ruido Gaussiano Fraccional-**. En estos modelos, los tiempos entre llegadas son variables gaussianas cuya correlación decae hiperbólicamente con el tiempo. Otro modelo ampliamente utilizado es el modelo ARIMA en el que el proceso de integración se hace en una dimensión fraccional (**FARIMA -fractional auto-regressive integrated moving average-**). Otro modelo muy usado es el $M/G/\infty$, que se asemeja a MAP, excepto por que la tasa de llegadas está dada por el estado de ocupación de un sistema de colas con infinitos servidores con distribuciones de cola pesada en sus tiempos de servicio. En estos modelos (FBM, FARIMA y $M/G/\infty$), la correlación de largo plazo se consigue sumando pequeños componentes a diferentes escalas, en lo que se denomina una cascada aditiva. Pero algunos modelos utilizan una cascada multiplicativa para construir y descomponer el proceso, tal como ocurre con el modelo **MWM (multifractal wavelet model)**, en el que los detalles de una escala a otra se obtienen multiplicando por fracciones aleatorias.

Si consideramos que todos los modelos anteriores se referían al proceso de tiempos entre llegadas de paquetes, es claro que al describir el proceso de servicio mediante el modelado probabilístico de los tiempos de transmisión tendremos las mismas opciones (modelos D , M , E_k , H_k , PH , MAP , AR , $MMPP$, FBM , $FARIMA$, MWM , etc.). Cuando los tiempos de servicio son los tiempos de transmisión de los paquetes sobre un enlace de transmisión de capacidad fija, la variabilidad aleatoria de los tiempos se debe a la variabilidad aleatoria de las longitudes de los paquetes. Por eso, una gráfica de la longitud de los paquetes que llegan a la cola en función del instante en que llegan es una buena representación de los dos primeros componentes de la estructura de la **Figura 115**. La Figura 116 muestra algunas trazas correspondientes a tres tipos de procesos aplicados tanto al modelado de los tiempos de llegada como al proceso de los tiempos de servicio (tamaño de paquetes). El eje horizontal de cada traza representa el tiempo en segundos y en él ubicamos una línea vertical en el instante en que llega cada paquete. El eje vertical es la longitud de los paquetes en bits, de manera que cada línea vertical tiene una altura correspondiente a la longitud del paquete que llegó en ese instante. Si consideramos un enlace de capacidad fija C bps, el eje vertical es C veces el tiempo de transmisión del paquete, de manera que las gráficas representan tiempos de servicio vs tiempos de llegada, caracterizando así los dos primeros aspectos de la estructura de un sistema de colas de la **Figura 115**. En la primera fila se observan distintas trazas en las que los tiempos de servicio son constantes, es decir, son tipo D (determinísticos). En la segunda fila los tiempos de servicio están exponencialmente distribuidos y son independientes entre ellos, es decir, forman un proceso tipo M (memoryless –sin memoria–). En la tercera fila, los tiempos de servicio están correlacionados entre sí, con una función de correlación que decae hiperbólicamente con el tiempo. Los incrementos son estacionarios y Gaussianos, de manera que los tiempos de servicio forman un proceso FBM –fractional Brownian motion–. De la misma manera, las columnas de la Figura 116 representan procesos tipo D , M y FBM para los tiempos de llegada: En la primera columna los tiempos entre llegadas son constantes, en la segunda columna son exponenciales e independientes y en la tercera columna los tiempos entre llegadas son gaussianos con dependencia de rango largo. Usando los primeros dos términos de la notación de Kendall, las trazas se describen, yendo de izquierda a derecha y de arriba a abajo, como D/D , M/D , FBM/D , D/M , M/M , FBM/M , D/FBM , M/FBM y FBM/FBM .



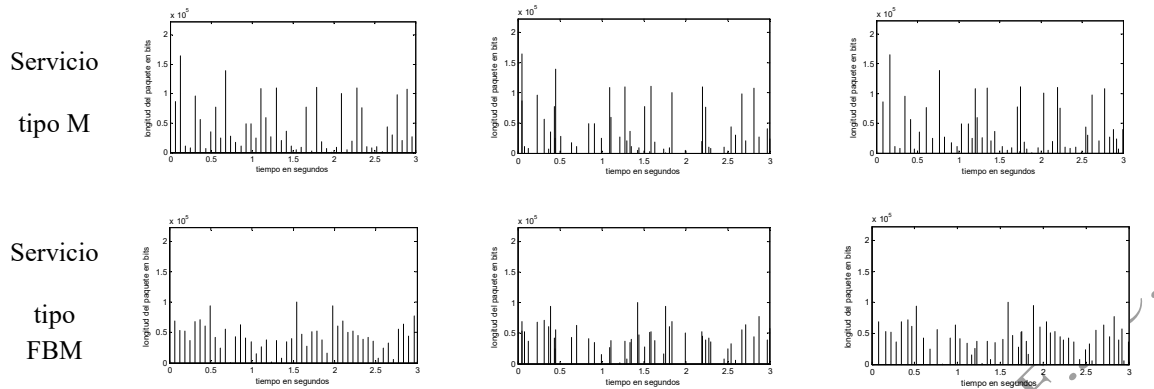


Figura 116. Combinaciones de tres tipos de procesos para modelar las llegadas y los servicios

El tercer parámetro, número de servidores, puede referirse al número de enlaces, el número de canales TDM o FDM, o el número máximo de usuarios admisibles en una celda CDMA para que la relación entre la potencia de la señal y la potencia de la interferencia permita garantizar una tasa de error específica. Es un número entero que siempre debe especificarse. Los parámetros cuatro y cinco son simplemente dos números naturales que, por defecto, se consideran infinitos. Aunque la memoria de un enrutador no puede ser infinita, si puede ser arbitrariamente grande, dado el bajo precio de las memorias de estado sólido, por lo que la suposición de memoria infinita (cupó infinito) no es tan irreal como parece. Igualmente, aunque el número de enlaces de entrada de un enrutador sea pequeño, cada uno de ellos puede traer el tráfico compuesto de un gran número de usuarios. La suposición de un número infinito de potenciales clientes se refiere a esta condición, en la que el tráfico generado por un usuario individual es despreciable comparado con el tráfico total de todos los usuarios. Si éste no es el caso, la activación o desactivación de un usuario puede afectar la tasa total de llegadas y, entonces, será conveniente considerar un número finito de potenciales clientes.

La disciplina de servicio y la gestión de memoria están íntimamente relacionadas con los mecanismos de **Calidad de Servicio (QoS)** en redes modernas de comunicaciones. La disciplina de servicio se refiere a cómo se va a signar el enlace (el recurso de transmisión) entre los diferentes flujos que componen el tráfico de entrada. La disciplina más obvia y simple de entender es **FIFO (Firs In, First Out)** o FCFS (First Come, First Served), según la cual los paquetes se atienden por orden de llegada, sin considerar ningún criterio adicional como tamaño, origen o prioridad. Su gran ventaja es que es muy simple de entender y de implementar, pero el desempeño general de la red se puede ver impactado de manera negativa. Por ejemplo, si un paquete muy grande (que ocupa el canal durante 10 ms) llega inmediatamente antes que un paquete muy pequeño (que ocupa el canal durante 0.5 ms), la disciplina FIFO hará que el primer paquete tarde 10 ms y el segundo tarde 10.5 ms para un promedio de 10.25 ms. Si se atendiera primero al más pequeño, los retardos serían 0.5 ms y 10.5 ms, respectivamente, para un promedio de 5.5 ms, una reducción de más del 46%. Por este motivo se proponen diferentes mecanismos, como **PQ (Priority Queueing)**, según el cual los paquetes se atienden de acuerdo con una medida de prioridad asignada externamente. Por supuesto, esto mejora el desempeño para el tráfico de alta prioridad a costa del deterioro del desempeño para el tráfico de baja prioridad, pero le da libertad al administrador de la red para asignar los recursos más inteligentemente. Típicamente, su implementación incluye mantener una cola para cada clase

prioridad, de manera que una cola particular se atiende sólo cuando todas las colas de prioridad superior se encuentran desocupadas. Claro, esto puede resultar injusto en períodos de congestión durante los cuales los flujos de mínima prioridad pueden dejar de ser atendidos. Para ser más equitativos, se puede usar la disciplina **FQ (Fair Queueing)**, en la que los distintos flujos ocupan colas diferentes, las cuales se atienden en un orden cíclico (**round robin**). Esto es, se atiende un paquete de cada cola, en estricto orden cíclico, saltando las colas que se encuentran desocupadas en su respectivo turno. De esta manera los flujos se desacoplan entre ellos, ya que un flujo muy intenso no afectará el desempeño de flujos más ligeros (si usan paquetes de longitudes semejantes). Sin embargo, es suficiente que un flujo use paquetes más grandes que los demás para ganar una fracción superior del ancho de banda. Este problema se resolvería de la manera más equitativa si los turnos se tomaran bit-por-bit, no paquete-por-paquete. Este mecanismo se denomina **PS (Processor Sharing)** en el mundo de los sistemas operativos, pero resulta imposible de implementar en un sistema de comunicaciones. Por eso se trata de aproximar mediante un mecanismo conocido como **BRFQ (Bit Round Fair Queueing)**, en el que primero se calcula los tiempos de salida de los paquetes si se usara PS, y luego se atienden los paquetes en el correspondiente orden de salida. Esto elimina la ventaja inadecuada de los flujos que envían paquetes más grandes cuando se usa la disciplina FQ. Aunque BRFQ es la disciplina de servicio más equitativa, no permite al administrador de la red proporcionar diferentes capacidades a diferentes flujos. Por eso es muy típico en redes modernas usar una combinación de PQ y FQ conocida como **WFQ (Weighted Fair Queueing)**, que también se puede ver como una forma generalizada de PS (**GPS –Generalized Processor Sharing–**). Simplemente, en cada ciclo FQ los flujos de mayor prioridad reciben atención de más de un bit. Las prioridades se refieren al número de bits atendidos en cada ciclo. Si hay tres flujos, uno de los cuales transmite 1 bit por ciclo, otro 2 bits por ciclo y un tercero que transmite 3 bits por ciclo, se garantiza una asignación del 17%, el 33% y el 50% de la capacidad del canal, respectivamente. WFQ es la versión paquetizada de GPS, como BRFQ es la versión paquetizada de PS. WFQ calcula y asigna un tiempo virtual de finalización a cada paquete y luego transmite los paquetes de acuerdo con este tiempo calculado. Desafortunadamente, estos cálculos hacen de WFQ un algoritmo computacionalmente complejo y no resulta fácilmente escalizable cuando haya miles o decenas de miles de flujos, por lo que existen versiones donde el cálculo es sólo aproximado, aunque se logran las mismas ventajas de garantizar una capacidad determinada con equidad (por ejemplo, el mecanismo **SCFQ –Self-Clocked Fair Queueing–**).

El último parámetro por especificar es el mecanismo de gestión de memoria en el buffer. Normalmente los paquetes se aceptan mientras haya espacio donde almacenarlos y sólo cuando se agota el espacio se produce el primer descarte. Estos descartes se seguirán produciendo hasta que se desocupe el espacio de memoria suficiente para aceptar un nuevo paquete. A dicho mecanismo se le denomina **DropTail –descarte al final de la cola–**. Con DropTail, la congestión se notifica a los usuarios mediante descarte de paquetes cuando ya los recursos de almacenamiento se han agotado completamente, esto es, la congestión se notifica por primera vez cuando la red ya está completamente congestionada. Por eso una alternativa frecuentemente usada en internet es **RED (Random Early Detection)**, donde la longitud de la cola se va promediando en línea y, si el promedio está por debajo de un nivel mínimo, se aceptan todos los paquetes que lleguen; si el promedio supera un nivel máximo, se descartan todos los paquetes que lleguen; y si el promedio está en el rango

intermedio, se escogen algunos paquetes al azar para ser descartados, donde la probabilidad de descarte aumenta linealmente con la ocupación promedio de la cola. Aunque RED avisa a los usuarios sobre el comienzo de la congestión antes de que se agoten los recursos, tiene algunas dificultades, especialmente con la forma en que se seleccionan los paquetes para descartar, sin considerar la fuente de los paquetes. Por eso se implementan variantes como **WRED (weighted RED)**, en el que la probabilidad de descarte se asigna de manera diferente para cada tipo de flujo, o **ARED (adaptive RED)**, en el que la probabilidad de descarte se ajusta automáticamente de acuerdo con la dinámica de la longitud de la cola. Otro esquema conocido como **BLUE** (no parece ser ningún acrónimo, sólo un contraste de RED) mantiene una probabilidad de pérdida p que se incrementa cada vez que la cola supera un nivel superior y se reduce cada vez que la cola se desocupa. **REM (Random Exponential Marking)** intenta estabilizar la longitud de la cola alrededor de un punto de equilibrio, ajustando la tasa de llegadas a la capacidad del enlace. Si la diferencia es positiva, aumenta la probabilidad de indicar la congestión; si la diferencia es negativa, la probabilidad disminuye. En general, a todos estos mecanismos que contrastan con DropTail se les conoce con el nombre genérico de **AQM (Active Queue Management)**.

134. Colas determinísticas

En una cola D/D/n/k los paquetes llegan en intervalos constantes de tiempo, $1/\lambda$, y ocupan su servidor durante un tiempo constante de servicio, $1/\mu$. Mientras $\lambda \leq n\mu$, donde n es el número de servidores, el número promedio de paquetes en el sistema es λ/μ . En otro caso, la cola crece sin límite (si el cupo es $k=\infty$) o tendrá pérdidas periódicas (si el cupo es $k<\infty$). En este último caso, la traza del número de paquetes en el sistema en cada instante, $\{N(t), t \geq 0\}$, se repetirá periódicamente entre cada par de pérdidas consecutivas y los promedios temporales en dicho período determinarán las medidas de desempeño del sistema.

Un ejemplo que revela muy bien muchos aspectos de la dinámica de los sistemas de colas es el de la cola D/D/1, mostrado en la Figura 117. Los paquetes llegan a una tasa constante de λ paquetes por segundo (exactamente un paquete cada $1/\lambda$ segundos) y todos tienen un tiempo de servicio constante de $1/\mu$ segundos (mientras el servidor esté ocupado, atiende μ paquetes por segundo).

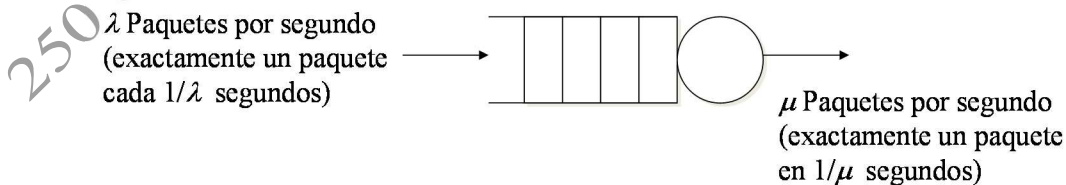


Figura 117. Modelo de cola D/D/1

Si $\mu > \lambda$, el tiempo de servicio de cada paquete se cumple antes de la llegada del próximo paquete, de manera que cada paquete que llega encuentra el servidor desocupado. En consecuencia, el número de

paquetes en el sistema en el instante t , $N(t)$, será una onda cuadrada con ciclo útil $(1/\mu)/(1/\lambda) = \lambda/\mu \equiv \rho < 1$, como muestra la Figura 118.

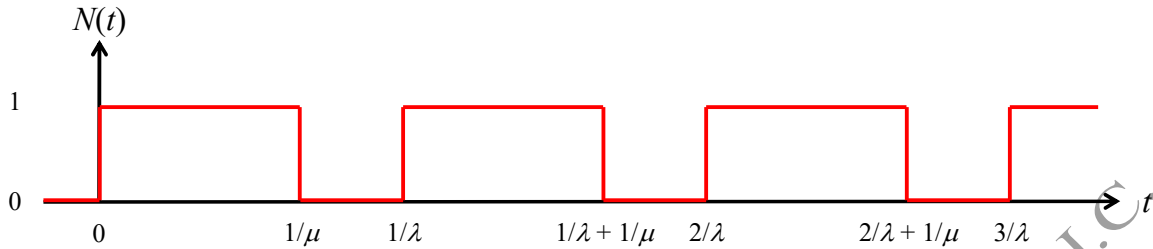


Figura 118. Número de paquetes en el sistema D/D/1, $N(t)$, cuando $\mu > \lambda$

Nótense los múltiples significados del ciclo útil de dicha curva: (1) ρ es la utilización del enlace (la fracción de tiempo que permanece ocupado), (2) ρ es la intensidad de tráfico del sistema (la relación entre la demanda y la capacidad), (3) ρ es el número promedio de paquetes en el sistema. En este último caso, la interpretación del promedio es determinística:

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt = \frac{1}{1/\lambda} \int_0^{1/\mu} dt = \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu} \equiv \rho$$

Pero, si decidimos ver el sistema en un instante aleatorio, t_0 , podemos considerar una interpretación probabilística de la misma estadística:

$$P[N(t_0) = 1] = \frac{1/\mu}{1/\lambda} = \rho, \quad P[N(t_0) = 0] = \frac{1/\lambda - 1/\mu}{1/\lambda} = 1 - \rho \Rightarrow E[N(t_0)] = (1)\rho + (0)(1 - \rho) = \rho$$

Como cada paquete que llega encuentra el enlace desocupado, el tiempo que permanece en el sistema es exactamente su tiempo de transmisión, $E[T] = 1/\mu$, de manera que nunca se forma una cola de paquetes en espera de ser transmitidos: $E[Q] = 0$ y $E[W] = 0$.

En este caso aplican principios de estacionariedad y ergodicidad, gracias a los cuales pudimos calcular las medidas probabilísticas como las medias temporales a partir de una sola traza. En efecto, el número de paquetes en el sistema permanece acotado entre 0 y 1, por lo que ni siquiera necesitamos disponer de un *buffer* para los paquetes que encuentren el enlace ocupado, pues cada paquete encontrará el enlace disponible al momento de su llegada. Es decir, la respuesta de la Figura 118 es igualmente válida para los sistemas D/D/1, D/D/1/1 o, en general, D/D/1/k con $k > 0$, siempre que $\mu > \lambda$.

Otro caso diferente ocurre si la demanda supera a la capacidad, en cuyo caso la cola crece sin límite a medida que transcurre el tiempo. La Figura 119 muestra el comportamiento del número de paquetes en el sistema cuando la tasa de llegadas, $\lambda = 1.5$ paquetes/segundo, supera a la capacidad, $\mu = 1$ paquete/segundo. En la figura hemos supuesto que, si una llegada y una salida ocurren simultáneamente, primero se procesa la salida y después se procesa la llegada, aunque ambas en el mismo instante.

La forma de onda entre 0 y 2 segundos se repite entre 2 y 4, pero con un paquete más en el sistema; después se vuelve a repetir entre 4 y 6 segundos, pero con dos paquetes más en el sistema. En general,

la forma de onda se repite entre $2n$ y $2n+2$ segundos, pero sobre un piso de n paquetes más en el sistema. Este sistema no es estacionario porque $E[N(t)]$ toma valores crecientes con t . Como el servidor siempre está ocupado, la utilización es igual a 1.

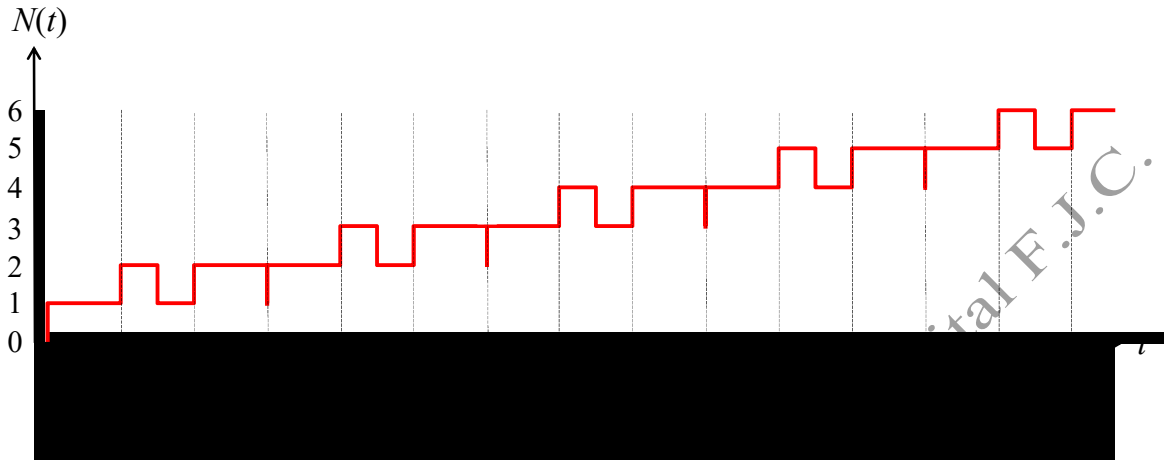


Figura 119. Número de paquetes en el sistema D/D/1, $N(t)$, cuando $\lambda > \mu$ (en este caso, $\mu=1$ y $\lambda=1.5$)

Una manera de estabilizar el sistema es limitando el cupo de paquetes que pueden estar dentro del sistema. En el caso anterior, si limitamos a 3 paquetes el cupo en el sistema (uno siendo atendido y dos esperando en cola), el número de paquetes en el sistema, iniciando desocupado en el instante 0, en que se presenta la primera llegada, sería como muestra la Figura 120.

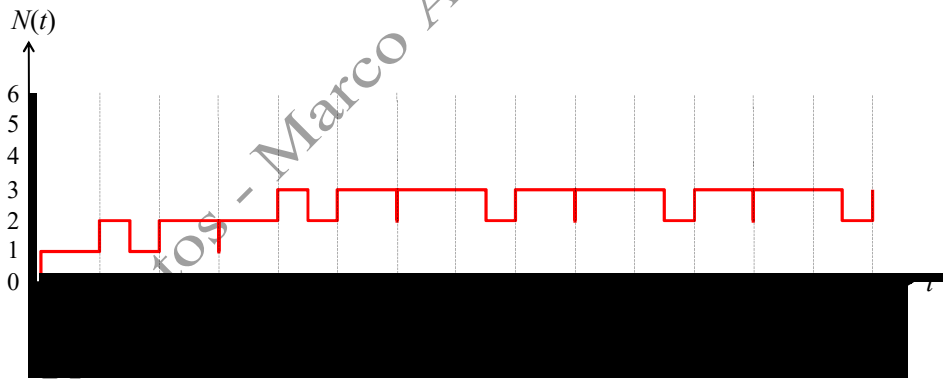


Figura 120. Número de paquetes en el sistema D/D/1/3, $N(t)$, cuando $\mu=1$ y $\lambda=1.5$

Después de un transiente de $8/3$ de segundo viene una secuencia cíclica que se repite cada dos segundos. En cada período de dos segundos en el sistema permanecen 2 paquetes durante $1/3$ de segundo y 3 paquetes durante $5/3$ de segundo. En consecuencia, el número promedio de paquetes en el sistema es

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt = \frac{1}{2} \int_{8/3}^{14/3} N(t) dt = \frac{1}{2} \left[2 \left(\frac{1}{3} \right) + 3 \left(\frac{5}{3} \right) \right] = \frac{17}{6}$$

En cada período de dos segundos se pierde un paquete de los tres que llegan, debido a que no encuentra cupo en la cola (en la Figura 120 los paquetes rechazados llegan en los instantes $14/3$, $20/3$ y $26/3$), por lo que en este sistema D/D/1/3 la tasa de pérdidas es $P_B=1/3$. La tasa efectiva de llegadas

es $\Gamma = \lambda(1 - P_B)$, igual a un paquete por segundo, que es la misma tasa de servicios, μ , ya que el servidor permanece ocupado el 100% del tiempo, transmitiendo exactamente un paquete por segundo. Con respecto al retardo, uno de los tres paquetes que llegan en un período es rechazado, otro tarda $8/3$ de segundo en el sistema y otro tarda 3 segundos en el sistema, para un promedio de $E[T] = (3 + 8/3)/2 = 17/6$ de segundo. Aunque ya lo veremos con más detalle, es interesante notar que $E[N] = \Gamma \cdot E[T]$. Por último, el tiempo que cada paquete permanece en promedio en la cola corresponde a los $17/6$ de segundo que permanece en el sistema menos el segundo que permanece en el servidor, $E[W] = 11/6$. El número promedio de paquetes en la cola será $E[Q] = \Gamma \cdot E[W] = 11/6$. En efecto, durante un período en la cola permanece 1 paquetes durante $1/3$ de segundo y 2 paquetes durante $5/3$ de segundo, para un promedio de $(1/3 + 10/3)/2 = 11/6$ paquetes.

Otra manera de estabilizar el sistema cuando $\lambda > \mu$ es añadiendo más servidores (que es una forma de aumentar la capacidad). Por ejemplo, consideremos el caso de la cola D/D/2. Si $\lambda < \mu$, uno de los servidores permanecerá desocupado y el otro atenderá los paquetes a una tasa mayor a la tasa con que llegan, de manera que este sistema es idéntico al D/D/1 con $\lambda < \mu$. Si $\lambda > 2\mu$, aún con los dos servidores ocupados permanentemente, el sistema no alcanzará a atender a los paquetes a la velocidad con que llegan, de manera que la cola crecerá constantemente a una tasa $\lambda - 2\mu$ paquetes por segundo, esto es, el sistema será inestable. Así pues, el caso de interés para el sistema D/D/2 es cuando $\mu < \lambda < 2\mu$. Para este caso, la Figura 121 muestra una traza del número de paquetes en el sistema en cada instante. Se observa nuevamente una señal periódica, precedida por un pequeño transiente entre 0 y $1/\lambda$, por lo que el número promedio de paquetes en el sistema se calcula fácilmente:

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt = \frac{1}{1/\lambda} \int_{1/\lambda}^{2/\lambda} N(t) dt = \lambda \left[2 \left(\frac{1}{\mu} - \frac{1}{\lambda} \right) + 1 \left(\frac{2}{\lambda} - \frac{1}{\mu} \right) \right] = \frac{\lambda}{\mu} \equiv \rho$$

Entonces, en el sistema D/D/2 con $\mu < \lambda < 2\mu$, se obtiene que $1 < E[N] = \rho < 2$. Si consideramos disponer de tres servidores (el sistema D/D/3), encontramos que con $0 < \lambda < \mu$, dos de los tres servidores permanecerán desocupados y el sistema se comportará de manera idéntica a D/D/1; con $\mu < \lambda < 2\mu$, un servidor permanecerá desocupado y el sistema se comportará de manera idéntica a D/D/2; con $\lambda > 3\mu$, los tres servidores permanecerán ocupados y la cola crecerá constantemente a una tasa $\lambda - 3\mu$, paquetes por segundo, haciendo inestable al sistema. El caso de interés es, entonces, cuando $2\mu < \lambda < 3\mu$, como muestra la traza de la Figura 122.

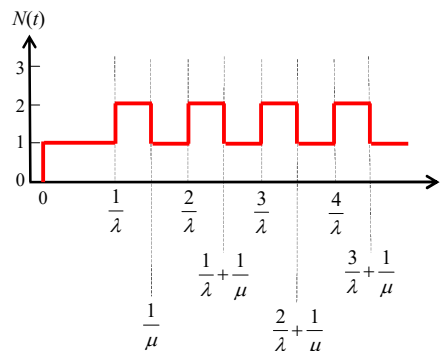


Figura 121. Número de paquetes en un sistema D/D/2 con $\mu < \lambda < 2\mu$

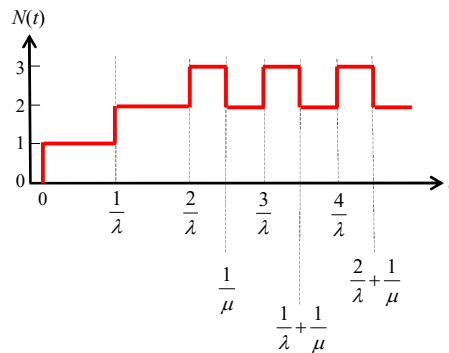


Figura 122. Número de paquetes en un sistema D/D/3 con $2\mu < \lambda < 3\mu$

Si el sistema empieza desocupado en el momento de la primera llegada, habrá dos llegadas adicionales antes de que ocurra la primera salida, cada una de ellas iniciando su transmisión inmediatamente, pues encuentran un servidor desocupado. Desde entonces las llegadas y las salidas se dan intercaladamente en períodos de $1/\lambda$ segundos, donde los tres enlaces permanecen ocupados durante $(1/\mu - 2/\lambda)$ segundos, y sólo dos permanecen ocupados durante $(3/\lambda - 1/\mu)$ segundos en cada período. De esta manera, el número promedio de paquetes en el sistema es

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt = \frac{1}{1/\lambda} \int_{2/\lambda}^{3/\lambda} N(t) dt = \lambda \left[3 \left(\frac{1}{\mu} - \frac{2}{\lambda} \right) + 2 \left(\frac{3}{\lambda} - \frac{1}{\mu} \right) \right] = \frac{\lambda}{\mu} \equiv \rho$$

Entonces, para el sistema D/D/3 con $2\mu < \lambda < 3\mu$, se obtiene que $2 < E[N] = \rho < 3$.

En general, si consideramos disponer de k servidores (sistema D/D/ k), encontramos que si $\lambda < (k-1)\mu$, el sistema es idéntico a D/D/ n , con $n = \lceil \lambda/\mu \rceil < k$. Si $\lambda > k\mu$, el sistema es inestable. Pero si $(k-1)\mu < \lambda < k\mu$, el sistema se comportará como muestra la Figura 123. Con la k -ésima llegada termina un período transiente e inicia un ciclo periódico que se repite cada $1/\lambda$ segundos, en el que el sistema permanece con todos los k servidores ocupados durante $(1/\mu - (k-1)/\lambda)$ segundos, y sólo $k-1$ permanecen ocupados durante el resto del período, $(k/\lambda - 1/\mu)$ segundos. El número promedio de paquetes en el sistema es, entonces, $E[N] = \frac{k(1/\mu - (k-1)/\lambda) + (k-1)(k/\lambda - 1/\mu)}{1/\lambda} = \frac{\lambda}{\mu} = \rho$

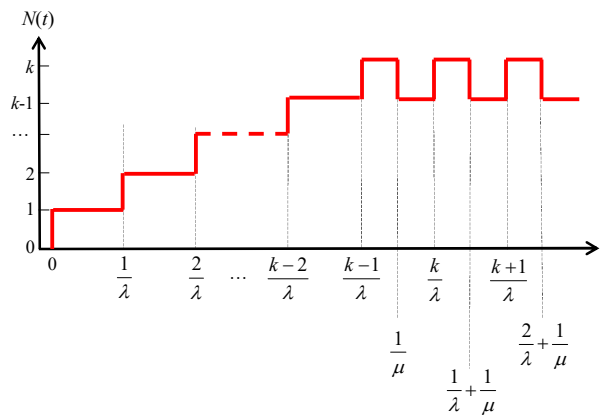


Figura 123. Número de paquetes en un sistema D/D/ k con $(k-1)\mu < \lambda < k\mu$

Si logramos construir una curva de $N(t)$ vs t podemos determinar el comportamiento del sistema: ¿Es estable? Si sí, ¿Cómo es un período en régimen estable? Así obtenemos $E[N]$ y todas las demás medidas de desempeño. Lo interesante es que esa curva, en el caso D/D/..., se puede construir de manera tan sistemática que podemos hacer que el computador la construya reproduciendo los instantes de llegada y salida y su interacción entre ellos, como se muestra a continuación.

135. Simulación de un Sistema de Colas

La simulación consiste en estudiar un sistema mediante la ejecución de un programa de computador que reproduzca las interacciones entre los componentes del sistema a medida que transcurre un tiempo virtual en el computador. En sistemas de cola conviene usar el enfoque de los eventos discretos (los estados del sistema cambian de manera discreta pero en instantes continuos de tiempo), según el cual se mantiene una lista de eventos futuros en la cual se busca el evento inminente para avanzar el reloj a ese instante y procesar el evento correspondiente. A medida que se van actualizando las variables de estado se pueden ir actualizando sus estadísticas. La aleatoriedad se obtiene mediante la generación (determinística) de secuencias pseudo-aleatorias. Como una simulación que incluya estas secuencias es sólo un experimento aleatorio, se debe hacer una planeación juiciosa de experimentos para determinar el tiempo de simulación y el número de repeticiones que aseguren la significancia estadística de los resultados.

Como se mencionó en la definición 1, la construcción del modelo matemático de las colas determinísticas se basó en

1. Identificar los **componentes** que intervienen en el sistema (enlaces, buffer y paquetes)
2. Determinar las **variables descriptivas** de esos componentes (estado de ocupación de los enlaces, número de paquetes en el buffer, tiempo de llegada y tiempo de servicio de cada paquete)
3. Identificar las **reglas de interacción** entre los componentes (cuando un paquete llega, si encuentra un enlace desocupado, lo ocupa durante su tiempo de servicio. Si no, se une a la cola en espera de su turno de servicio. Si no hay cupo en la cola para él, es rechazado).
4. Determinar los **parámetros descriptivos** de esas interacciones (número de enlaces, cupo en el sistema, tasa del proceso determinístico de llegadas, tasa del proceso determinístico de servicios)

Formulado de esta manera, hemos expresado el sistema mediante un modelo matemático. La solución analítica, llevada a cabo en la definición 134, consistió en verificar cuándo se forma un ciclo periódico en la ocupación del sistema. Si no se alcanzaba dicho ciclo, el sistema era inestable. En otro caso, bastaba con evaluar las estadísticas temporales durante un período para obtener las medidas de desempeño del sistema. Por supuesto, el mismo modelo matemático se puede evaluar también mediante simulación, como haremos en este apartado.

La simulación consiste en estudiar el sistema mediante la ejecución juiciosa de **un programa de computador que reproduzca las interacciones entre los componentes** a medida que transcurre un

tiempo virtual en el computador. Una secuencia de estados producida por el simulador es una posible traza muestral del sistema bajo estudio, de manera que de ella podremos tomar posibles medidas de desempeño de dicho sistema. Es importante notar que, si hay algún componente de aleatoriedad en el modelo, **cada ejecución del programa es sólo un experimento aleatorio** y hay que tratarlo como tal, esto es, el estudio mediante simulación debe corresponder al análisis estadístico de un número cuidadosamente calculado de experimentos cuidadosamente planeados.

Un aspecto importante en la simulación es, entonces, el avance del tiempo virtual. Consideremos dos alternativas.

a. Simulación de tiempo discreto

La primera alternativa es considerar un incremento fijo de tiempo Δt , haciendo que las interacciones entre los componentes del sistema se evalúen cada Δt segundos y avanzando el reloj de simulación en pasos de Δt segundos. Este mecanismo es útil para simular sistemas cuya naturaleza sea de tiempo discreto (como circuitos digitales sincrónicos, por ejemplo) o para simular sistemas representados mediante ecuaciones diferenciales, en cuyo caso los incrementos infinitesimales de tiempo se reemplazan por incrementos discretos suficientemente pequeños, Δt . En la simulación del circuito RC de la definición 1, por ejemplo, usamos un incremento $\Delta t = RC/25$, que resultó un paso suficientemente pequeño para obtener resultados aproximadamente correctos sin hacer demasiadas iteraciones innecesarias. En general, ése es el compromiso en la selección de Δt : Si se escoge demasiado pequeño, se desperdiciará tiempo de computación haciendo cálculos que no significarán un mejor conocimiento del sistema. Si se escoge demasiado grande, se puede perder información sobre el comportamiento dinámico del sistema por la acumulación de errores. Desafortunadamente, si el estado del sistema cambia de manera continua con el tiempo, no hay otra alternativa para simularlo en un computador digital.

Por ejemplo, los movimientos relativos de la luna y la tierra se dan en tiempo continuo y no hay alternativa para su simulación por computador distinta a calcular las pequeñas variaciones de posición en pequeños incrementos de tiempo. En este caso, los componentes son la luna y la tierra, sus variables descriptivas dinámicas son la posición, la velocidad y la aceleración. La regla de interacción la determina la ley universal de la gravitación, cuyos parámetros son las masas de los cuerpos y la constante gravitacional: Las dos se atraen una a la otra con una fuerza de magnitud $F = G \cdot M \cdot m / r^2$, donde G es la constante gravitacional (la fuerza de atracción entre dos masas de 1 Kg a 1 m de distancia), M es la masa de la tierra, m es la masa de la luna y r es la distancia entre ellas. La aceleración que experimenta la tierra en dirección a la luna es $a_T = G \cdot m / r^2$, mientras la aceleración que experimenta la luna en dirección a la tierra es $a_L = G \cdot M / r^2$. En un intervalo pequeño de tiempo Δt , estas aceleraciones implican un cambio de las respectivas velocidades en una cantidad $a_T \Delta t$ y $a_L \Delta t$, respectivamente. Con las nuevas velocidades v_T y v_L , en el mismo intervalo de tiempo las posiciones cambiarán en una cantidad $v_T \Delta t$ y $v_L \Delta t$, respectivamente. Si calculamos estas variaciones iterativamente cada Δt segundos, podremos construir las trayectorias de los dos cuerpos. El siguiente

programa en matlab® lleva a cabo esta simulación, donde el incremento de tiempo es de un cuarto de hora. Los resultados se observan en la Figura 124.

```
m1 = 5.97e24; % Masa de la tierra (kg)
m2 = 7.35e22; % Masa de la luna (kg)
G = 6.67e-11; % Constante gravitacional (m³kg⁻¹s⁻²)
x1 = [0 0]; % Posición original de la tierra
x2 = [4.0561e8 0]; % Posición original de la luna (m). En promedio
% son 384403 km, pero en el apogeo está a 405610 km y en el perigeo a 363026 km
v1 = [0 0]; % Velocidad original de la tierra
v2 = [0 970]; % Velocidad original de la luna (ms⁻¹). En promedio.
% son 1022 m/s, pero en el el apogeo va a 970 m/s y en el perigeo a 1082 m/s
t = 0; % Reloj de simulación
dt = 900; % Evalúa cada cuarto de hora
T = zeros(1500,1); % Almacena las posiciones de la tierra
XT = T; YT = T; XL = T; YL = T; % y de la luna en función del tiempo
for n=1:4238
    r2 = (x2 - x1)*(x2 - x1)'; % Distancia (al cuadrado) entre los dos cuerpos
    F = G*m1*m2/r2; % Fuerza de atracción entre los dos cuerpos
    a1 = (F/m1)*(x2-x1)/sqrt(r2); % Aceleración que la luna impone a la tierra
    a2 = (F/m2)*(x1-x2)/sqrt(r2); % Aceleración que la tierra impone a la luna
    v1 = v1 + a1*dt; % Actualiza la velocidad de la tierra
    v2 = v2 + a2*dt; % Actualiza la velocidad de la luna
    x1 = x1 + v1*dt; % Actualiza la posición de la tierra
    x2 = x2 + v2*dt; % Actualiza la posición de la luna
    t = t+dt/3600/24; % Tiempo (en días)
    XT(n) = x1(1); YT(n) = x1(2); T(n) = t; % Traza de las posiciones de la tierra
    XL(n) = x2(1); YL(n) = x2(2); % y de la luna
end
figure
subplot(221); scatter(XT,YT,'bo'); hold on; scatter(XL,YL,'ko'); axis equal
title('Posiciones de la tierra y la luna')
subplot(222); scatter(0,0,'bo'); hold on; scatter(XL-XT,YL-YT,'ko'); axis equal
title('Posiciones de la luna respecto a la tierra')
subplot(223); plot(T,XT,'b-',T,YT,'r-'); axis tight
title('Posición de la tierra'); legend('posición horizontal','posición vertical')
xlabel('tiempo (días)'); ylabel('Posición (metros)')
subplot(224); plot(T,XL-XT,'b-',T,YL-YT,'r-'); axis tight
title('Posición de la luna respecto a la tierra');
legend('posición horizontal','posición vertical')
xlabel('tiempo (días)'); ylabel('Posición (metros)')
```

Listado 2. Simulación de los movimientos de la tierra y la luna

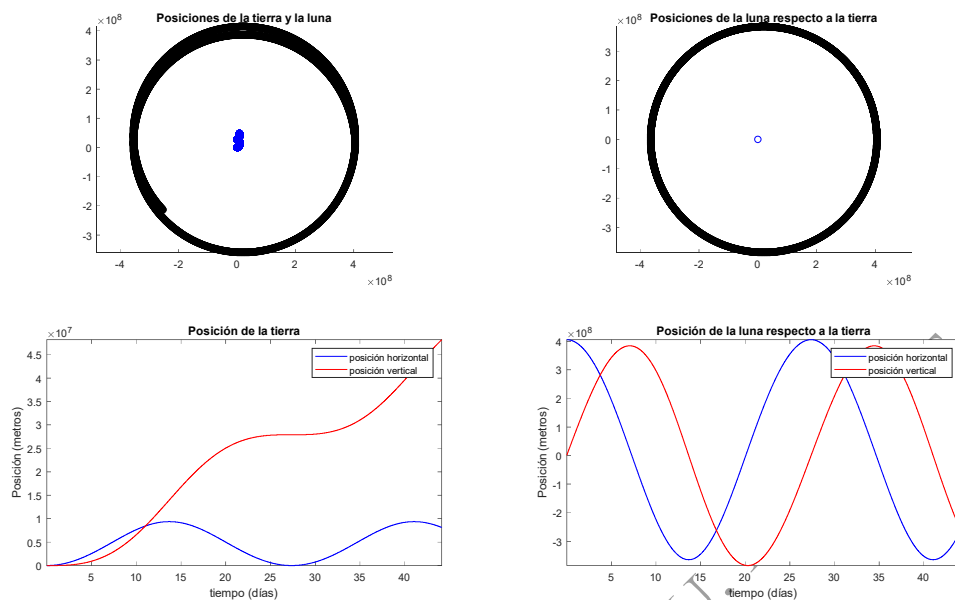


Figura 124. Simulación básica de los movimientos de la tierra y la luna

Nótese que este programa tan simple predice un período de 27.37 días para la órbita de la luna, lo cual no está nada mal comparado con los 27.32 días que realmente dura. Al estudiante le gustará jugar con diferentes masas, posiciones iniciales y velocidades, para observar muchos movimientos interesantes que pueden surgir de esta interacción entre dos objetos. Una pequeña modificación permitirá incluir n objetos y descubrir fenómenos fascinantes (basta construir un arreglo de masas, posiciones y velocidades y sumar los efectos de cada uno de los objetos sobre cada uno de los demás).

b. Simulación de eventos discretos (DES)

Escoger un incremento fijo de tiempo no es apropiado para el tipo de simulaciones que queremos hacer, en las que los paquetes pueden llegar o salir en cualquier momento (tiempo continuo) pero cada llegada o cada salida altera el estado del sistema en pasos discretos (un paquete más o un paquete menos). En estos modelos, una llegada o una salida es un "evento", una acción instantánea que altera el estado del sistema, de manera que este estado permanece invariable durante el tiempo comprendido entre dos eventos consecutivos. Por esto, a este tipo de modelos se les denomina "de **Eventos Discretos**". En estos casos, si hacemos que el reloj de simulación salte directamente de un evento al siguiente, logramos la máxima eficiencia y la máxima precisión con respecto a los modelos de tiempo discreto: Se trata de evaluar las reglas de interacción solamente en los instantes en que cambia el estado del sistema. Para implementar este concepto, el simulador debe mantener una **lista de eventos futuros**, una estructura de datos donde se almacena los eventos que ocurrirán en el futuro y los instantes en que habrán de ocurrir, según haya sido determinado por eventos anteriores, además de algunos otros parámetros que describan el evento por ocurrir. Por ejemplo, en la cola D/D/1, los eventos son las llegadas y las salidas, de manera que la lista de eventos futuros se reduce a dos variables: el instante de la próxima llegada y el instante de la próxima salida. En otros casos, se puede tratar de estructuras muy elaboradas asociadas con algoritmos muy eficientes para su actualización

(inclusión y eliminación de campos en la lista). De cualquier manera, los eventos se programan dinámicamente a medida que la simulación avanza. Por ejemplo, la llegada de un paquete puede programar la siguiente llegada dentro de $1/\lambda$ segundos (si las llegadas son determinísticas) o dentro de un tiempo aleatorio generado como una muestra de una distribución adecuada (una variable exponencialmente distribuida con promedio $1/\lambda$ segundos, por ejemplo). De la misma manera, cuando un paquete inicia su servicio, se puede programar su salida dentro de un tiempo (determinístico, uniforme, exponencial, etc.) con promedio $1/\mu$.

Así pues, después de inicializar algunas variables y programar los primeros eventos futuros (por ejemplo, la primera llegada), el proceso avanza como muestra la Figura 125: Se busca en la lista de eventos futuros el evento más próximo en el tiempo. Como nada ocurrirá con el sistema entre el instante actual y el instante de ese próximo evento, se adelanta el reloj de simulación al instante de ocurrencia de ese próximo evento y se actualiza el estado del sistema de acuerdo con el tipo de evento correspondiente (llega un nuevo paquete, sale un paquete o un paquete pasa de un sitio a otro, por ejemplo). Con el nuevo estado y el nuevo valor del reloj de simulación se actualizan las estadísticas. De otro lado, de acuerdo con el tipo de evento que haya ocurrido, y con la ayuda de un generador de muestras de variables aleatorias, se podrían programar más eventos futuros. Mientras dure la simulación, este proceso se repite constantemente. La simulación terminará cuando se haya alcanzado un tiempo de simulación suficientemente largo como para que el sistema haya alcanzado la estabilidad en la distribución de sus estados y haya permanecido en esa condición suficientemente como para que las estadísticas no se vean afectadas por los datos del período de distribución transiente. Al final de la simulación se presentan los reportes necesarios, como las estadísticas finales, por ejemplo.

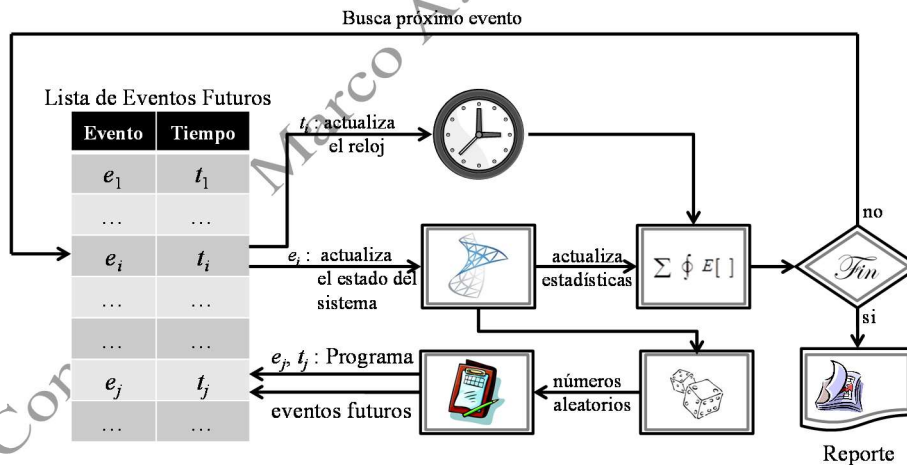


Figura 125. Simulación de Eventos Discretos

En términos de un proceso secuencial, la simulación de eventos discretos sigue el siguiente flujo:

1. Inicialice las variables de estado del sistema (típicamente, el sistema empieza vacío)
2. Asigne un valor inicial al reloj de simulación (típicamente cero)
3. Programe un evento inicial (típicamente, la primera llegada en el instante cero)

4. Mientras la condición de terminación sea falsa (típicamente, mientras el reloj de simulación sea menor que un tiempo máximo o mientras las estadísticas temporales no alcancen un valor convergente), repita el proceso de los pasos 5 a 9:
5. Avance el reloj de simulación al tiempo de ocurrencia del siguiente evento
6. Ejecute las actualizaciones de estado correspondiente a ese siguiente evento
7. Retire este evento de la lista
8. Programe algunos eventos futuros, si es oportuno
9. Actualice las estadísticas de acuerdo con el nuevo estado y el nuevo valor del reloj de simulación
10. Después que se cumpla la condición de terminación, genere el reporte de las estadísticas

Consideremos, por ejemplo, la cola D/D/1 analizada en la definición **¡Error! No se encuentra el origen de la referencia.** El mismo modelo matemático que analizamos en esa sección se simula con el programa en matlab® del Listado 3.

```
1 function ColaDD1(lambda,mu,tiempoSimulacion)
2   reloj = 0; % Reloj de la simulación
3   proximaLlegada = 0; % Tiempo de la siguiente llegada
4   proximaSalida = inf; % Tiempo de la siguiente salida
5   paquetesSistema = 0; % Número de paquetes en el sistema [N(t)]
6   while reloj < tiempoSimulacion
7       if proximaLlegada < proximaSalida % Procesa una llegada
8           reloj = proximaLlegada; % Actualiza el reloj de simulación
9           proximaLlegada = reloj + 1/lambda; % Programa la próxima llegada.
10          paquetesSistema = paquetesSistema+1; % Hay un paquete más en el sistema
11          if paquetesSistema == 1 % Si es el único, inicia su transmisión inmediatamente
12              proximaSalida = reloj + 1/mu; % Programa la próxima salida
13          end
14       else % Procesa una salida
15           reloj = proximaSalida; % Actualiza el reloj de simulación
16           paquetesSistema = paquetesSistema-1; % Hay un paquete menos en el sistema
17           proximaSalida = inf; % Si el sistema queda vacío, no se espera una salida
18           if paquetesSistema > 0 % En otro caso,
19               proximaSalida = reloj + 1/mu; % Programa la próxima salida
20           end
21       end
22   end
```

Listado 3. Primera versión del simulador de la cola D/D/1, en matlab®

El programa recibe como parámetros de entrada la tasa de llegadas (λ), la tasa de servicio (μ) y el tiempo que debe durar la simulación ($t_{\text{simulación}}$). La simulación se inicializa en las líneas 2 a 5. Primero se inicializa en cero la variable `reloj`, que lleva el registro del tiempo durante la simulación. Como hay sólo dos clases de eventos, las llegadas y las salidas, la lista de eventos futuros (¡la estructura de datos más importante en un programa de simulación de eventos discretos!) se reduce a dos variables: El tiempo de la próxima llegada (`proximaLlegada`) y el tiempo de la próxima salida (`proximaSalida`). La línea 3 programa la primera llegada para el instante cero, sin esperar ninguna salida (línea 4), ya que el sistema empieza vacío (línea 5). La simulación dura mientras no se haya alcanzado el máximo tiempo de simulación (línea 6). Durante la simulación se escoge el evento más próximo (líneas 7 y 14). Si es una llegada, se actualiza el reloj al tiempo de la próxima llegada (línea 8), se programa la próxima llegada dentro de $1/\lambda$ segundos (línea 9) y se incrementa el número de paquetes en el sistema (línea 10). Si el paquete que acaba de llegar es el único paquete en el sistema, ocupa el enlace durante $1/\mu$ segundos (líneas 11 a 13). Por otro lado, si el evento más próximo es una salida, se actualiza el reloj al tiempo de la próxima salida (línea 15), se

decrementa el número de paquetes en el sistema (línea 16) y se desocupa el enlace (línea 17). Sin embargo, si aún quedan paquetes en la cola, el enlace se asigna al primer paquete en la cola y se programa la próxima salida para dentro de $1/\mu$ segundos (líneas 18 a 20). Si se observa bien, este programa es sólo una descripción formal de las reglas de interacción entre los componentes del sistema.

Desafortunadamente el programa anterior no produce ningún resultado visible. Sería interesante ver, por ejemplo, una traza del número de paquetes en el sistema a medida que transcurre el tiempo, $\{N(t), t \geq 0\}$. Para esto debemos registrar los instantes en que cambia el estado del sistema y el valor que asume en cada instante, por lo que un arreglo de dos columnas estará bien, como muestran las líneas 5a, 5b, 6a, 6b, 23 y 24 que se insertarían en el Listado 3, así:

```
...
5 paquetesSistema = 0;           % Número de paquetes en el sistema [N(t)]
5a traza = zeros(ceil(2*tiempoSimulacion*lambda),2); % El evento i ocurrió en el instante
5b numeroEvento = 0;           % traza(i,1) y dejó traza(i,2) paquetes en el sistema
6 while reloj < tiempoSimulacion
6a     numeroEvento = numeroEvento+1; % Un evento más
6b     traza(numeroEvento,:) = [reloj paquetesSistema]; % Registra (tiempo,estado) en cada evento
7     if proximaLlegada < proximaSalida % Procesa una llegada
...
23 traza = traza(1:numeroEvento,:);
24 stairs(traza(:,1),traza(:,2))
```

Listado 4. Modificaciones al programa del Listado 3 para obtener una traza de $\{N(t), t \geq 0\}$

El arreglo `traza` se declara con suficiente espacio para almacenar 2 eventos por cada paquete que alcance a llegar (suponiendo que al final de la simulación el sistema quede vacío, esto es, que cada paquete que haya llegado durante la simulación también haya alcanzado a irse antes de que culmine la simulación). La primera columna guarda el instante en que ocurrió un evento y la segunda columna guarda el valor que el estado del sistema asumió después de dicho evento (número de paquetes en el sistema). Por eso se usa una variable `numeroEvento` para subindicar el arreglo, la cual se incrementa con cada evento (cada ciclo del `while`) para almacenar los datos necesarios (`reloj` y `paquetesSistema`). Al final de la simulación se trunca el arreglo a los eventos almacenados y se grafica el número de paquetes en el sistema en función del tiempo. La Figura 126 muestra la traza generada cuando la capacidad es $\mu=1$ paquete por segundo y la demanda crece desde $\lambda=0.25$ hasta $\lambda=1.5$ en pasos de 0.25 paquetes por segundo. Mientras λ sea menor o igual a μ , el sistema alcanza un ciclo periódico. De otra manera, el sistema es inestable porque la cola crece sin límite a medida que transcurre el tiempo.

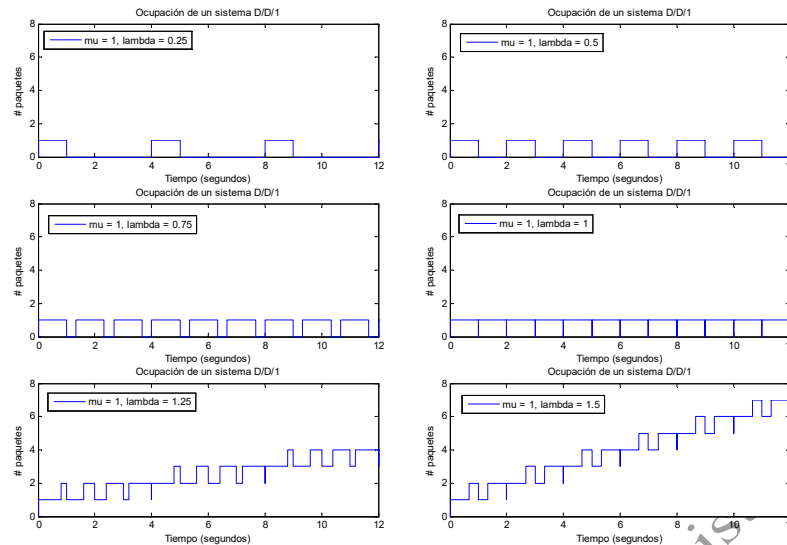


Figura 126. Trazas obtenidas del programa simulador de la cola D/D/1

Limitar el cupo en el sistema resulta muy fácil. Basta introducir el cupo en el sistema como un parámetro adicional (`cupo`) y, al momento de la llegada de un paquete, verificar si ya existen `cupo` paquetes en el sistema, en cuyo caso el nuevo paquete se retira sin ser admitido en el sistema. Basta insertar las líneas 9a y 13a y cambiar la línea 1 como muestra el siguiente listado:

```
1 function ColaDD1k(lambda,mu,cupo,tiempoSimulacion)
9a  if paquetesSistema < cupo
13a end
```

Listado 5. Modificaciones al programa del Listado 3 para simular una cola D/D/1/k

Resulta ligeramente más elaborado añadir varios enlaces ya que en este caso debemos considerar la próxima salida en cada enlace, para lo cual convendría, por ejemplo, disponer de un arreglo de próximas salidas, con un elemento por cada enlace en el sistema. Si `proximaSalida(i)` es infinito, significa que el enlace `i` está desocupado. Cuando un paquete llega, mira si alguno de los enlaces está desocupado y, si es así, lo ocupa y programa su próxima salida para dentro de $1/\mu$ segundos. Cuando un paquete sale, si deja en el sistema menos paquetes que el número de servidores es porque no queda ninguno en espera de servicio (todos los paquetes que haya, si es que hay alguno, están siendo atendidos). En ese caso, la próxima salida por el enlace que se acaba de desocupar se programa para dentro de un tiempo infinito. De otra manera, dicho enlace se asigna inmediatamente al primer paquete en la cola, programando su próxima salida para dentro de $1/\mu$ segundos. Con n servidores, ahora la lista de eventos futuros tiene $n+1$ variables: La próxima llegada y cada una de las próximas salidas por cada uno de los n servidores. El listado del programa que devuelve la traza de paquetes en el sistema con n servidores y k cupos en el sistema se muestra en el Listado 6.

```
1 function traza = ColaDDnk(lambda,mu,servidores,cupo,tiempoSimulacion)
2  reloj = 0; % reloj de la simulación
3  proximaLlegada = 0; % Tiempo de la siguiente llegada
4  proximaSalida = inf*ones(servidores,1); % Tiempo de la siguiente salida en cada servidor
5  paquetesSistema = 0; % Número de paquetes en el sistema [N(t)]
6  traza = zeros(ceil(2*tiempoSimulacion*lambda),2); % El evento i ocurrió en el instante
7  numeroEvento = 0; % traza(i,1) y dejó traza(i,2) paquetes en el sistema
8  while reloj < tiempoSimulacion % Inicia el ciclo de simulación
9  numeroEvento = numeroEvento+1; % Un evento más
10 traza(numeroEvento,:) = [reloj paquetesSistema]; % Registra (tiempo,estado) en cada evento
11 if proximaLlegada < min(proximaSalida) % Procesa una llegada
12  reloj = proximaLlegada; % Actualiza el reloj de simulación
```

```

13     proximaLlegada = reloj + 1/lambda; % Programa la próxima llegada
14     if paquetesSistema < cupo           % El sistema puede aceptar este paquete
15         paquetesSistema = paquetesSistema+1;
16         if paquetesSistema <= servidores % De hecho, lo puede empezar a transmitir ya
17             ss = find(proximaSalida==inf); % Busca el primer servidor desocupado
18             ss = ss(1);
19             proximaSalida(ss) = reloj + 1/mu; % Programa la próxima salida de este servidor
20         end
21     end
22 else % Procesa una salida
23     [reloj,k] = min(proximaSalida); % Actualiza el reloj de simulación e indica el servidor
                                     % que termina el servicio
24     paquetesSistema = paquetesSistema - 1;
25     if paquetesSistema < servidores
26         proximaSalida(k) = inf; % Este servidor queda desocupado
27     else
28         proximaSalida(k) = reloj + 1/mu; % Programa la próxima salida de este servidor
29     end
30 end
31 end
32 traza = traza(1:numeroEvento,:); % Fin de la simulación: Grafica la ocupación del sistema
33 stairs(traza(:,1),traza(:,2)) % (número de paquetes en el sistema como función del tiempo)
34 end

```

Listado 6. Simulador de la cola D/D/n/k

La Figura 127 muestra las trazas generadas por el programa anterior cuando (1) se tienen 3 servidores cada uno con una capacidad de $\mu=1$ paquete por segundo, (2) en el sistema se pueden admitir hasta 10 paquetes (3 en servicio y 7 en espera), y (3) la demanda crece desde $\lambda=0.5$ hasta $\lambda=5.5$ en pasos de 1 paquete por segundo. Mientras λ sea menor o igual a μ , el sistema alcanza el ciclo periódico de la cola D/D/1. Mientras λ este entre μ y 2μ , el sistema alcanza el ciclo periódico de la cola D/D/2. Mientras λ este entre 2μ y 3μ , el sistema alcanza el ciclo periódico de la cola D/D/3. De otra manera, el sistema crecerá hasta llenar el espacio disponible, en cuyo caso alcanzará un período cíclico en el que la ocupación del sistema alterna entre 9 y 10 paquetes, con un ciclo útil que aumenta con λ .

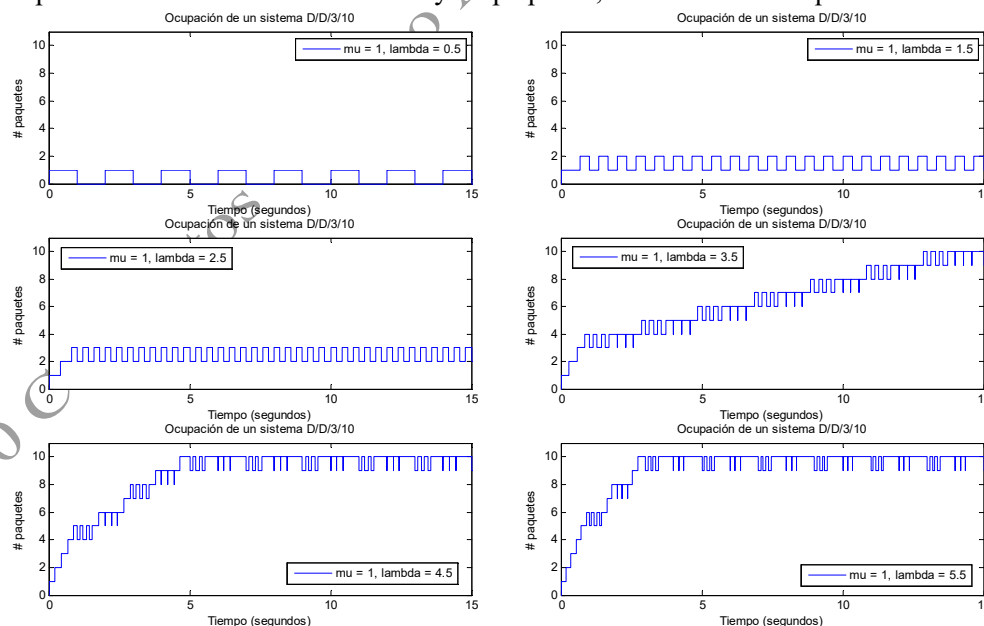


Figura 127. Trazas obtenidas del programa simulador de la cola D/D/3/10

c. Adquisición de Estadísticas

En los programas anteriores, tratándose de colas D/D/..., todas las estadísticas de interés se pueden obtener a partir de la traza del número de paquetes en el sistema, $\{N(t), t \geq 0\}$, como se hizo en el tratamiento analítico del modelo (en la definición anterior). Sin embargo, podemos utilizar el mismo programa de simulación para hallar esas estadísticas temporales de interés.

En primer lugar, para hallar el promedio del número de paquetes en el sistema,

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt$$

debemos calcular el área debajo de la curva de $N(t)$ durante el tiempo de simulación y dividirla por el tiempo de simulación. Obsérvese que, dependiendo del estado inicial del sistema, dicha traza puede pasar por un período inicial transiente antes de alcanzar su ciclo periódico. Con el modelo analítico buscábamos promediar durante un ciclo periódico, lo que equivalía al límite $T \rightarrow \infty$. Con el modelo de simulación debemos (1) empezar a calcular el área debajo de la curva una vez alcanzado el ciclo periódico o (2) simular durante un tiempo lo suficientemente largo como para que el efecto del ciclo transiente sea despreciable. En esta introducción usaremos la estrategia (2).

Como el estado del sistema permanece constante entre eventos, el cálculo del área debajo de la curva $N(t)$ se reduce a la acumulación de las áreas de los rectángulos formados entre eventos consecutivos, como muestra la Figura 128.

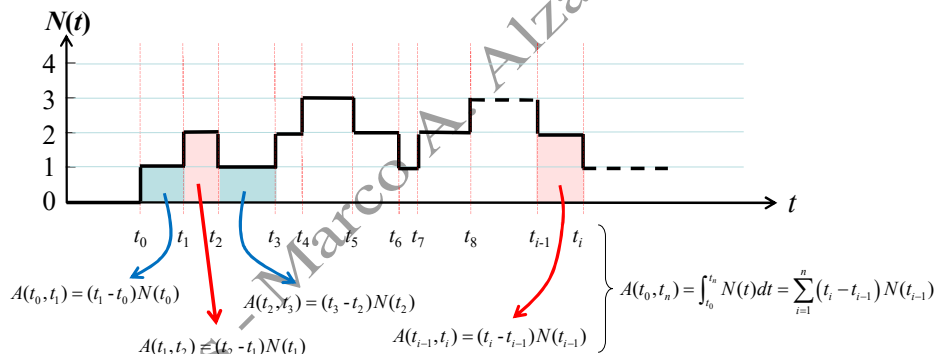


Figura 128. Cálculo del área debajo de la curva de $N(t)$

Así pues, el cálculo del área debajo de la curva consiste en recordar el tiempo del último evento (t_{ue}) y, antes de actualizar `paquetesSistema`, añadir el término `paquetesSistema*(reloj-tue)` a la suma acumulada del área. Al final de la simulación dividimos esta suma acumulada por el tiempo de simulación para obtener el número promedio de paquetes que permanecieron en el sistema (ver las líneas 23, 29, 52 y 70 del Listado 7). De idéntica manera podemos calcular el número promedio de paquetes en la cola, $E[Q]$, o el número promedio de paquetes en servicio, $E[X]$. Más aún, si además de acumular el área bajo la curva de $N(t)$ acumulamos también el área bajo $N^2(t)$, podemos calcular el segundo momento $E[N^2]$ para evaluar también la varianza, $V[N] = E[N^2] - E^2[N]$. Si quisiéramos evaluar estadísticas de orden superior, seguiríamos un procedimiento semejante.

Pero otras estadísticas de interés están relacionadas con el tiempo que pasa un paquete en el sistema, en la cola o en servicio. En este caso la forma de las estadísticas varía pues, si se han ido $n(t)$ paquetes

del sistema hasta el instante t , donde el i -ésimo paquete llegó en el instante t_{Li} y salió en el instante t_{Si} , el tiempo promedio de permanencia en el sistema es

$$E[\tau] = \lim_{T \rightarrow \infty} \frac{1}{n(T)} \sum_{i=1}^{n(T)} (t_{Si} - t_{Li})$$

Nótese que para calcular estos promedios se hace necesario recordar el instante en que llegó cada paquete que sale del sistema, para lo cual no basta mantener la cuenta del número de paquetes en el sistema sino que hace falta llevar un registro de los instantes de llegada de los paquetes que se encuentran en el sistema. Para esto, necesitamos implementar la cola del enrutador en el programa de simulación, de manera que `Cola(i)` contenga el instante de llegada del paquete que ocupa la posición i en la cola. Una primera opción es una lista ordenada en el que el primer elemento del arreglo corresponde al primer paquete de la cola, pero entonces sería necesario desplazar todos los paquetes del arreglo una posición cada vez que el primer paquete en la cola inicia su servicio. Como este desplazamiento es computacionalmente costoso, se prefiere usar un "arreglo circular" con dos punteros, uno al primer elemento de la cola y otro al último elemento de la cola, de manera que las operaciones "Entrar a la cola" y "salir de la cola" se hacen muy sencillas. Para introducir un paquete a la cola se almacena su tiempo de llegada en la posición del último elemento de la cola y se incrementa el puntero circularmente. Para sacar un paquete de la cola se lee su tiempo de llegada en la posición del primer elemento de la cola y se incrementa el puntero circularmente. Si los dos punteros coinciden en la misma posición es porque la cola se encuentra desocupada. Así, el puntero al último elemento de la cola indica la posición donde se insertará el próximo paquete que llegue al buffer y el puntero al primer elemento de la cola indica la posición de donde saldrá el siguiente paquete en dejar el buffer, como muestra la Figura 129 (ver líneas 14 a 21, 39 a 45 y 60 a 64 del Listado 7).

Con este arreglo, podemos saber cuándo llegó al sistema cada paquete que sale de él y, así, podemos calcular el tiempo que permaneció cada paquete dentro del sistema para ir actualizando la suma de los tiempos de permanencia y poder calcular el promedio al final de la simulación.

Otras dos medidas de interés son la tasa de pérdidas y el caudal. Para el primer caso es tan simple como contar el número de paquetes que han llegado, N_A , independientemente de si fueron aceptados o no, y el número de paquetes que se han rechazado por falta de cupo en el sistema, N_B . La probabilidad de pérdida se puede interpretar como la fracción de paquetes rechazados, $P_B = N_B/N_A$ (ver líneas 10, 11, 27, 47 y 76 del Listado 7). Para el segundo caso, hay dos formas teóricamente equivalentes de describir el caudal: Una es la tasa de paquetes que efectivamente se aceptan por segundo, y otra es la tasa de paquetes que efectivamente se transmiten por segundo. Si $N_{in}(t)$ es el número de paquetes que han sido aceptados hasta el instante t , $N_{out}(t)$ es el número de paquetes que han sido atendidos hasta el instante t , y $N(t)$ es el número de paquetes en el sistema en el instante t , tenemos que

$$\Gamma = \lim_{T \rightarrow \infty} \frac{N_{in}(T)}{T} = \lim_{T \rightarrow \infty} \frac{N_{out}(T) + N(T)}{T} = \lim_{T \rightarrow \infty} \frac{N_{out}(T)}{T} + \lim_{T \rightarrow \infty} \frac{N(T)}{T} = \lim_{T \rightarrow \infty} \frac{N_{out}(T)}{T}$$

de manera que las dos definiciones son equivalentes si $\lim_{T \rightarrow \infty} N(T) < \infty$. En la simulación se tiene un tiempo de simulación finito, así que puede haber alguna diferencia entre ambas medidas (ver líneas 12, 50 y 75 del Listado 7).

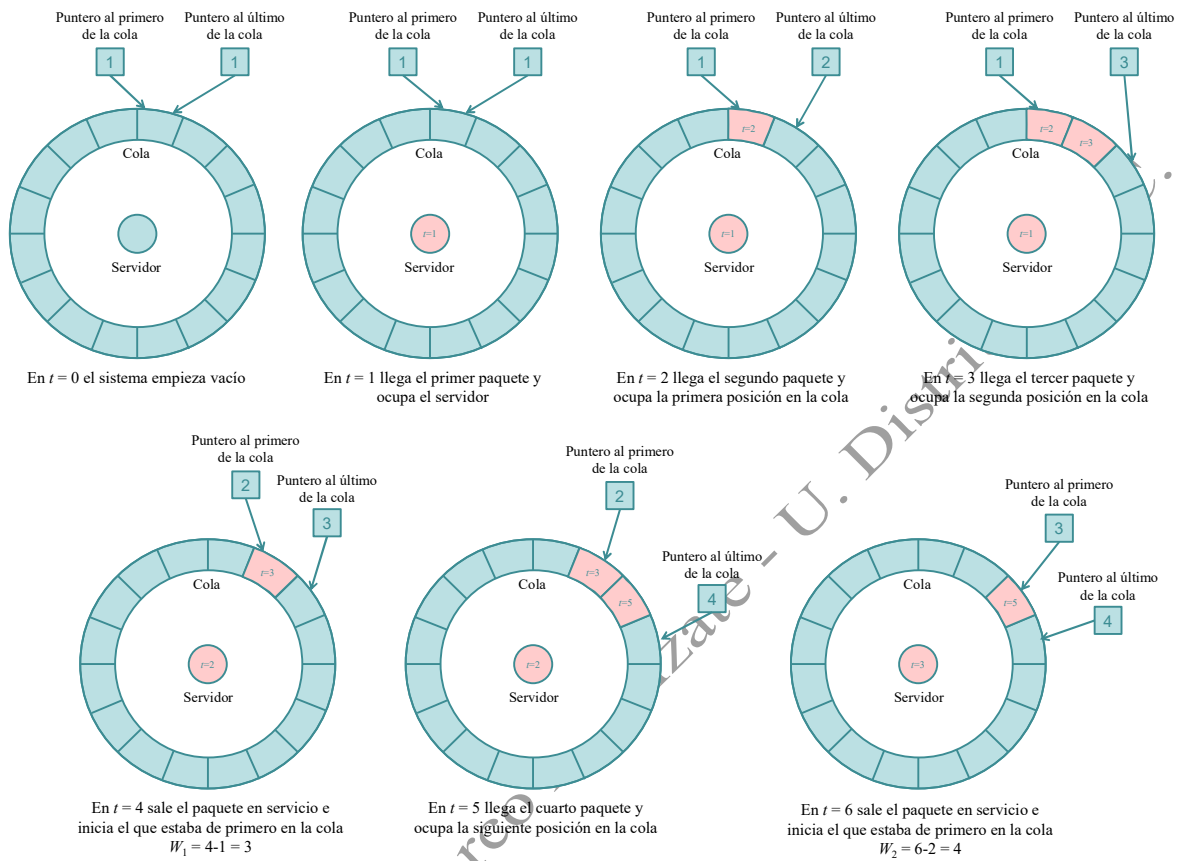


Figura 129. Buffer circular para llevar el registro del tiempo de llegada de cada paquete en el sistema

El Listado 7 muestra un programa en matlab® que simula la cola D/D/n/k, generando la traza de $N(t)$ vs t , y retornando las medidas de $E[N]$ –número promedio de paquetes en el sistema–, $E[Q]$ –número promedio de paquetes en la cola–, $E[T]$ –tiempo promedio de permanencia en el sistema–, $E[W]$ –tiempo promedio de espera en la cola–, P_B –probabilidad de bloqueo– y Γ –caudal–.

```

1 function [EN, EQ, ET, EW, PB, G] = ColaDDnk(lambda,mu,servidores,cupo,tiempoSimulacion)
% [EN, EQ, ET, EW, PB, G] = ColaDDnk(lambda,mu,servidores,cupo,tiempoSimulacion)
% Simula durante TIEMPOSIMULACION segundos un sistema de cola FIFO con un numero dado de SERVIDORES
% y un CUPO máximo en el sistema (CUPO >= SERVIDORES), cuando la tasa de llegadas es constante
% (LAMBDA por segundo) y la tasa de servicio de cada servidor es constante e igual para cada
% servidor (MU por segundo) -Cola D/D/SERVIDORES/CUPO-
%
% Marco Aurelio Alzate Monroy, Univesidad Distrital, 2014
%
2 reloj = 0; % reloj de la simulación
3 proximaLlegada = 0; % Tiempo de la siguiente llegada
4 proximaSalida = inf*ones(servidores,1); % Tiempo de la siguiente salida en cada servidor
5 paquetesSistema = 0; % Número de paquetes en el sistema [N(t)]
6 paquetesCola = 0; % Número de paquetes en la cola [Q(t)]
7 traza = zeros(floor(3*tiempoSimulacion*lambda),2); % El evento i ocurrió en el instante
8 numeroEvento = 0; % traza(i,1) y dejó traza(i,2) paquetes en el sistema
    
```

```

9     EN = 0; % Promedio de paquetes en el sistema (durante la
10    NB = 0; % simulación, área debajo de la curva N(0:reloj))
11    NA = 0; % Número de paquetes bloqueados
12    ND = 0; % Número de paquetes que han llegado
13    EQ = 0; % Número de paquetes que han salido
simulación % Promedio de paquetes en la cola (durante la
14    EW = 0; % área debajo de la curva Q(0:reloj))
de % Suma de los retardos de los paquetes que han salido
15    tamañoCola = cupo - servidores; % la cola
16    if tamañoCola==inf % Calcula el tamaño de la cola
17        tamañoCola=ceil(lambda*tiempoSimulacion); % La cola será un arreglo circular donde se almacena
18    end % el instante de llegada de cada paquete
19    cola = zeros(tamañoCola,1); % Se usa un puntero ULTIMOCOLA donde se almacenará el
20    primeroCola=1; % próximo paquete que llegue a la cola,
% y un puntero PRIMEROCOLA de donde se tomará el
21    ultimoCola=1; % paquete que salga de ella (FIFO)
22    while reloj<tiempoSimulacion % Estos punteros se incrementarán circularmente
23        tue = reloj; % Inicia el ciclo de simulación
24        numeroEvento = numeroEvento+1; % tiempo del último evento
25        traza(numeroEvento,:) = [reloj paquetesSistema]; % Un evento más
26        if proximaLlegada<min(proximaSalida) % Registra (tiempo,estado) en cada evento
27            NA = NA + 1; % Procesa una llegada
28            NA = NA + 1; % Una llegada más
29            reloj = proximaLlegada; % Actualiza el reloj de simulación
30            EN = EN + paquetesSistema*(reloj - tue); % Actualiza el área bajo la curva N(t)
31            EQ = EQ + paquetesCola*(reloj - tue); % Actualiza el área bajo la curva de Q(t)
32            proximaLlegada = reloj + 1/lambda; % Programa la próxima llegada
33            if paquetesSistema<cupo % El sistema puede aceptar este paquete
34                paquetesSistema = paquetesSistema+1;
35                if paquetesSistema <= servidores % De hecho, lo puede empezar a transmitir ya
36                    ss = find(proximaSalida==inf ); % Busca el primer servidor desocupado
37                    ss = ss(1);
38                    proximaSalida(ss) = reloj + 1/mu; % Programa la próxima salida de este servidor
39                else % Este paquete debe esperar en cola
40                    paquetesCola = paquetesCola + 1; % Lo almacena en la última posición de la
41                    cola
42                    cola(ultimoCola)=reloj;
43                    ultimoCola = ultimoCola+1; % (y actualiza circularmente dicha posición)
44                    if ultimoCola>tamañoCola
45                        ultimoCola=1;
46                    end
47                end
48            else
49                NB = NB + 1; % Este paquete es rechazado porque no hay dónde
50                ubicarlo
51            end
52            else % Procesa una salida
53                ND = ND + 1; % Una salida más
54                [reloj,k] = min(proximaSalida); % Actualiza el reloj de simulación e indica el
55                servidor
56                % que termina el servicio
57                EN = EN + paquetesSistema*(reloj - tue); % Actualiza el área bajo la curva N(t)
58                EQ = EQ + paquetesCola*(reloj - tue); % Actualiza el área bajo la curva de Q(t)
59                paquetesSistema = paquetesSistema - 1;
60                if paquetesSistema < servidores
61                    proximaSalida(k) = inf ; % Este servidor queda desocupado
62                else
63                    paquetesCola = paquetesCola - 1; % Este servidor empieza a atender al primer
64                    paquete
65                    % de la cola
66                    proximaSalida(k) = reloj + 1/mu; % Programa la próxima salida de este servidor
67                    EW = EW + (reloj - cola(primerocola));
68                    primeroCola = primeroCola+1; % (y actualiza circularmente dicha
69                    posición)
70                end
71            if primeroCola>tamañoCola

```

```

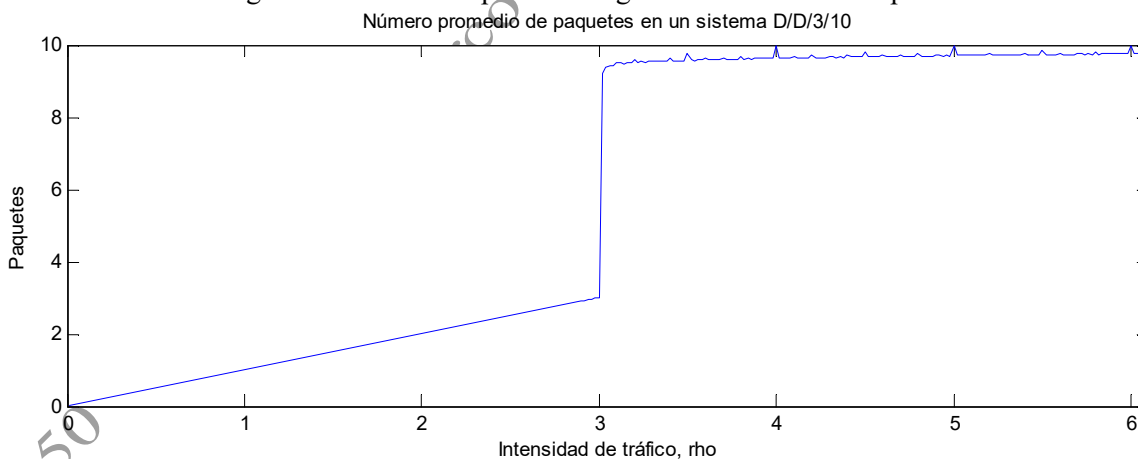
63         primeroCola=1;
64     end
65 end
66 end
67 end
68 traza = traza(1:numeroEvento,:); % Fin de la simulación: Grafica la ocupación del sistema
69 stairs(traza(:,1),traza(:,2)) % (número de paquetes en el sistema como función del tiempo)
70 EN = EN/reloj; % Promedio del número de paquetes que permanecieron en el sistema
71 EQ = EQ/reloj; % Promedio del número de paquetes que permanecieron en la cola
72 EW = EW/ND; % Retardo promedio en la cola
73 ET = EW + 1/mu; % Retardo promedio en el sistema
74 G1 = (NA - NB)/reloj; % Número de paquetes aceptados por segundo
75 G2 = ND/reloj; % Número de paquetes transmitidos por segundo
76 PB = NB/NA; % Fracción de paquetes que fueron rechazados
77 G = (G1+G2)/2; % ¿Verdadero caudal? G1, G2 y G son iguales si la cola queda vacía
78 end

```

Listado 7. Programa completo para simulación de la cola D/D/n/k

El programa anterior se usó para medir el número promedio de paquetes en un sistema D/D/3/10 donde cada servidor tiene una capacidad de $\mu=1$ paquete por segundo y se varía la tasa de llegadas entre $\lambda=0$ y $\lambda=6.1$. La gráfica $E[N]$ vs ρ , donde ρ se define como λ/μ se muestra en la Figura 130 (¿por qué se alcanza una ocupación del 100% para valores enteros de ρ mayores a 3?).

Es importante hacer notar que, aunque hemos hablado de medias y varianzas, no hay nada aleatorio en estos sistemas, por lo que en realidad estamos haciendo referencia a promedios temporales de señales completamente determinísticas. Como vimos, la principal fuente de aleatoriedad en los sistemas de colas está en los tiempos entre llegadas de paquetes consecutivos y en los tiempos de servicio que cada paquete requiere para su transmisión. En estos casos, si conocemos la función de densidad de probabilidad de dichas cantidades, será suficiente con usar muestras de dichas distribuciones al escoger el instante de la próxima llegada o el instante de la próxima salida.

Figura 130. $E[N]$ en función de ρ para una cola D/D/3/10 con $\mu=1$

d. Generación de Muestras de Variables Aleatorias

Hemos insistido en que la aleatoriedad de un experimento se refiere a la imposibilidad de predecir su resultado antes de ejecutar el experimento. La característica más sobresaliente de un programa de computador es, precisamente, que los resultados son perfectamente predecibles, pues resultan de la realización de un algoritmo específico ¿Cómo puede una máquina así generar muestras de variables

aleatorias? La verdad es que no lo hace, pues no puede lanzar dados o monedas. En vez de esto, se usa un algoritmo determinístico para generar secuencias de números completamente determinísticas que "se comportan" como si fueran aleatorias³³. Por ejemplo, considere los primeros 100000 dígitos de la expansión decimal de π (3141592653 5897932384 6264338327 9502884197 1693993751 0582097494 4592307816 4062862089 9862803482 5342117067 9821480865 1328230664 7093844609 5505822317 2535940812 8481117450 2841027019 3852110555 9644622948 9549303819 6442881097 5665933446 1284756482 3378678316 5271201909 1456485669 2346034861 0454326648 2133936072 6024914127 3724587006 6063155881 7488152092 0962829254 etc.). La gráfica izquierda de la Figura 131 muestra cómo la distribución de cada dígito es perfectamente uniforme (cada dígito aparece cerca de 10000 veces). Si condicionamos en el dígito anterior, por ejemplo si calculamos la frecuencia relativa de los dígitos precedidos por un 0, como muestra la gráfica intermedia de la Figura 131, la distribución sigue siendo uniforme (cada dígito aparece cerca de 1000 veces después de un 0). Más aún, la función de autocovarianza normalizada es un impulso unitario, lo cual enfatiza la sugerencia de independencia, pues ninguna sub-secuencia de dígitos parece contener información sobre cuál será el siguiente dígito, como muestra la gráfica de la derecha de la Figura 131. Otra característica interesante es que una secuencia de N dígitos sólo se repetirá cerca de 10^{5-N} veces. Por ejemplo, la secuencia '3' aparece 10026 veces, la secuencia '31' aparece 975 veces, la secuencia '314' aparece 95 veces, la secuencia '3141' aparece 9 veces y la secuencia '31415' aparece 2 veces, pero la secuencia '314159' sólo aparece una vez (de hecho aparecen todas las secuencias de 4 dígitos o menos, pero sólo aparece el 37% de las secuencias de 5 dígitos). Si pudiésemos generar 100000 dígitos decimales realmente aleatorios, todas estas características también se cumplirían.

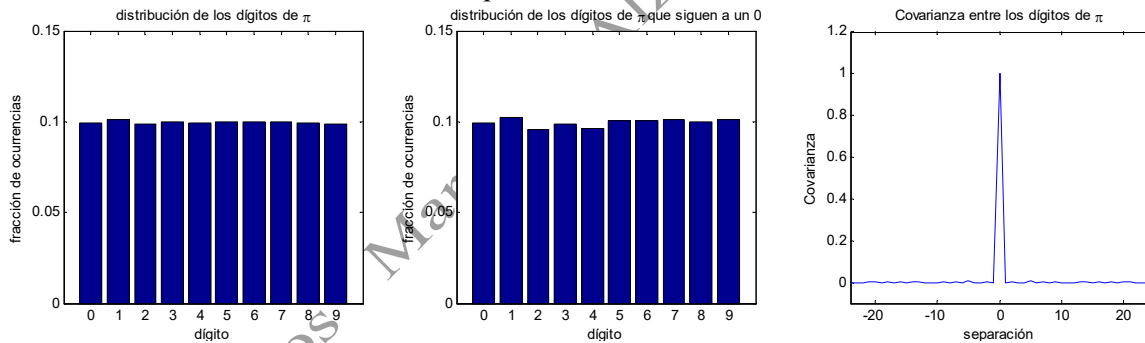


Figura 131. Características estadísticas de los primeros 100000 dígitos de la expansión decimal de π

Así como existen algoritmos para generar la secuencia de dígitos de π , que es una secuencia determinística que satisface todos los criterios de aleatoriedad, existen diferentes algoritmos para generar secuencias de números enteros que satisfagan criterios de aleatoriedad como los mencionados anteriormente, aunque en realidad sean completamente determinísticos. Por esa razón se conocen como secuencias pseudo-aleatorias. La idea es que un generador de números enteros entre 0 y $m-1$, que satisfaga criterios como los anteriores, puede usarse para generar números racionales en el intervalo $[0,1)$ si dividimos cada entero de la secuencia entre m ; estos racionales podrían interpretarse como probabilidades en el programa simulador.

³³ Si hace falta verdadera aleatoriedad, se puede usar un servicio como el de hotbits, de fourmilab (<http://www.fourmilab.ch/hotbits/>), que usa la temporización de eventos espontáneos de desintegración de núcleos atómicos (decaimiento radioactivo).

El área de la generación de números pseudo-aleatorios y la generación de muestras con otras distribuciones distintas a la uniforme ha sido ampliamente estudiada desde la teoría de números y ha arrojado muy interesantes algoritmos. Aquí mencionaremos unos pocos aspectos introductorios para que el estudiante tenga una idea muy general y pueda ser, al menos, un buen usuario de las funciones disponibles en las bibliotecas de los lenguajes de programación que utiliza.

Uno de los algoritmos más usados en la generación de secuencias uniformemente distribuidas de números enteros es el generador congruencial lineal que, partiendo de una semilla Z_0 , genera los términos de la secuencia así:

$$Z_n = (\alpha Z_{n-1} + \beta) \bmod m, \quad n \geq 1$$

donde α , β y m son parámetros del generador. Este es un algoritmo rápido que requiere poca memoria. Como los números obtenidos se evalúan módulo m , el conjunto de números que puede generar es $\mathcal{Z} \subseteq \{0, 1, \dots, m-1\}$. Si la semilla Z_0 se repite después de K iteraciones, $Z_0 = Z_K$, toda la secuencia se repetirá a partir de ahí y, por consiguiente, el período del generador será $|\mathcal{Z}| = K$. La teoría de números describe condiciones en α , β , m y Z_0 para que (1) la secuencia esté uniformemente distribuida en \mathcal{Z} , (2) los números de la secuencia sean estadísticamente independientes, y (3) tengan un período tan grande como sea posible (ojalá m) —en efecto, si un programa de simulación necesita 10000 números aleatorios, no sirve un generador que repita la secuencia después de 5000 muestras—. Obsérvese que, si hace falta, cualquier secuencia puede ser reproducible inicializando Z_0 adecuadamente; esto es de gran importancia para comparar resultados de diferentes modelos de simulación bajo idénticas condiciones.

Por ejemplo, empezando con $Z_0 = 0$, el generador lineal congruencial $\{Z_n = (5Z_{n-1} + 1) \bmod 16\}$ genera la secuencia $\{1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5, 10, 3, 0\}$ que es perfectamente uniforme y tiene máximo período, aunque se pueden ver fácilmente ciertas correlaciones (en cuatro ocasiones ocurre el par $(n, n+1)$, pero nunca ocurre el par $(n, n+2)$, por ejemplo). Un generador útil debe tener un valor alto de m (para un período suficientemente grande) y un valor alto de α (para proporcionar suficiente aleatoriedad). Algunos generadores lineales congruenciales usados en diferentes lenguajes de programación son los siguientes:

$$\begin{aligned} Z_n &= (1103515245 Z_{n-1} + 12345) \bmod 2^{31} \text{ (la librería de GNU para el lenguaje C)} \\ Z_n &= (22695477 Z_{n-1} + 1) \bmod 2^{32} \text{ (C/C++ de Borland)} \\ Z_n &= (6364136223846793005 Z_{n-1} + 1) \bmod 2^{64} \text{ (la librería estándar C de newlib)} \\ Z_n &= (16807 Z_{n-1} + 0) \bmod 2^{31}-1 \text{ (original de matlab®)} \\ &\text{etc.} \end{aligned}$$

Sin embargo, existen otros algoritmos recientemente desarrollados con mejores características. Por ejemplo, matlab permite escoger entre el generador multiplicativo congruencial anterior y otros métodos como el siguiente generador multiplicativo retardado de Fibonacci,

$$Z_n = (Z_{n-31} * Z_{n-63}) \bmod 2^{64}$$

que tiene un período de 2^{124} , o el más aceptado en la actualidad, el "Mersene twister", que tiene un increíble período de $2^{19937}-1$ y mucho mejores características estadísticas, aunque requiere más recursos de cómputo y de memoria para su evaluación. Para calcular Z_n , el algoritmo Mersene twister usa una forma generalizada de generadores por registros de desplazamiento y generadores retardados de Fibonacci. Matlab usa, por defecto, el generador Mersene twister y le da al usuario la opción de cambiarlo por otros de su preferencia.

Con cualquiera de estos generadores, la secuencia $r_n = Z_n/m$ corresponde a una serie de números uniformemente distribuidos en el intervalo $[0, 1)$, independientes entre ellos, de manera que podemos usar la secuencia como una generadora de probabilidades. En la mayoría de lenguajes de programación, existe la función `rand()` para generar el siguiente número de la secuencia (como en matlab). Así, por ejemplo, para simular la lanzada de una moneda, asignamos los valores 0 y 1 a los elementos del espacio muestral cara y sello, respectivamente, y escogemos cada uno de los posibles valores con probabilidad $1/2$. Esto es, evaluamos $r = \text{rand}()$ y, si r cae en el intervalo $[0,0.5)$, escogemos 0 y, si r cae en el intervalo $[0.5,1)$, escogemos 1.

```
>> lado = {'cara','sello'};
>> moneda = (rand()>=0.5); disp(lado{moneda+1})
```

Cada vez que repetamos la última línea obtendremos el resultado 'cara' o 'sello' con idéntica probabilidad, independientemente de anteriores resultados. Este algoritmo equivale a invertir la función de distribución acumulativa de Bernoulli, como muestra la Figura 132.

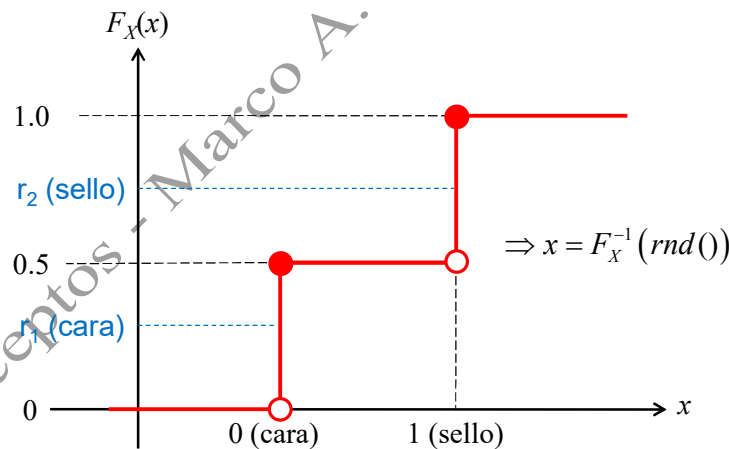


Figura 132. Obtención de muestras de variables de Bernoulli mediante inversión de la CDF

Si tenemos un enlace que permanece ocupado el 80% del tiempo y queremos generar una muestra de cuántos enlaces están ocupados en un instante dado usamos la siguiente línea:

```
>> EnlacesOcupados = (rand()>=0.2)
```

que retorna 1 con probabilidad 0.8 y 0 con probabilidad 0.2. Si transmitimos un bit sobre un canal con $\text{BER}=10^{-5}$ y queremos generar una muestra de cuántos bits se dañaron durante la transmisión, usamos la siguiente línea

```
>> BitsErrados = (rand()<1e-5)
```

que retorna 1 con probabilidad 10^{-5} y 0 con probabilidad $1-10^{-5}$. En cada caso, hemos invertido la CDF de la variable aleatoria de Bernoulli correspondiente. Si deseamos generar el estado de ocupación de 10 enlaces o el estado de error de 100000 bits usaríamos líneas como las siguientes

```
>> EnlacesOcupados = (rand(10,1)>=0.2);
>> BitsErrados = (rand(100000,1)<1e-5);
```

Igual podemos hacer, por ejemplo, para lanzar cien dados:

```
>> r = rand(100,1); dado = (r>=0)+(r>=1/6)+(r>=1/3)+(r>=1/2)+(r>=2/3)+(r>=5/6);
```

Con la anterior instrucción, el vector `dado` será un arreglo de 100 números uniformemente distribuidos en $\{1,2,3,4,5,6\}$ e independientes entre ellos pues, efectivamente, hemos hecho la inversión de la CDF, $\text{dado} = F^{-1}(\text{rnd}())$, como muestra la Figura 133.

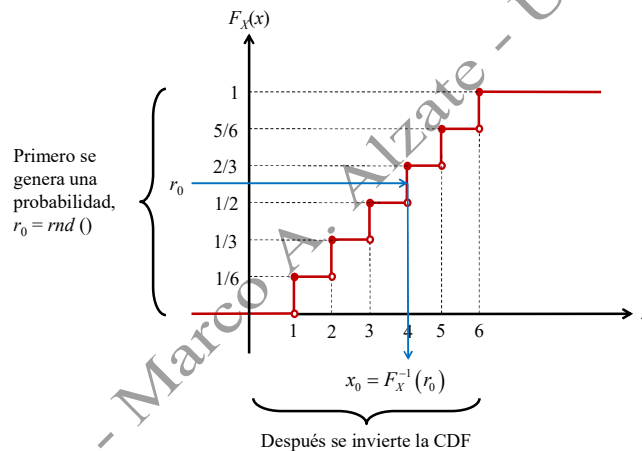


Figura 133. Simulación de lanzar un dado mediante inversión de la CDF

Supongamos que queremos generar una muestra de una variable aleatoria geoméricamente distribuida en $\{0,1,2,\dots\}$ con parámetro p , esto es, $P[X=k] = p^k(1-p)$. La CDF de esta variable es

$$F_X(n) = \sum_{j=0}^n p^j (1-p) = 1 - p^{n+1}$$

de manera que, si hacemos $\text{rand}() = F_X(n)$, obtenemos $n = \log(1 -$

$\text{rand}()) / \log(p) - 1$. Como muestra la Figura 134, debemos seleccionar el mínimo entero no menor a n y, como $1 - \text{rand}()$ es otra probabilidad, podemos usar $\text{rand}()$ en vez de $1 - \text{rand}()$. Por ejemplo, las siguientes líneas en matlab generan 10000 muestras de una variable aleatoria geométrica y comparan su histograma con la pmf $P[X=k] = p^k(1-p)$, como muestra la gráfica izquierda de la Figura 135.

```
>> r = rand(10000,1);
>> n = ceil(log(r)/log(0.8) - 1); mx = max(n);
>> h = hist(n,0:mx); h = h/sum(h);
>> stem(0:mx,h,'bo'); hold on; stem(0:mx,0.2*0.8.^(0:mx),'rx')
```

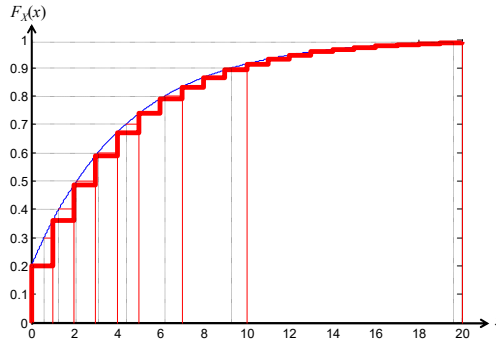



Figura 134. Generación de muestras de una variable aleatoria geométrica

Como último ejemplo, consideremos una variable aleatoria T exponencialmente distribuida con promedio $1/\lambda$, $F_T(t) = 1 - e^{-\lambda t}$. Usando como probabilidad $1 - \text{rand}()$ obtenemos $t = -\log(\text{rand}())/\lambda$. La siguiente secuencia de instrucciones de matlab® genera los tiempos de servicio de 10000 paquetes distribuidos exponencialmente con promedio 2 ms y estima la pdf mediante el histograma, como muestra la gráfica derecha de la Figura 135.

```
>> t = -log(rand(10000,1))/500;
>> [h,s] = hist(t,20);
>> d = s(2)-s(1);
>> f = h/sum(h)/d;
>> g = 500*exp(-500*s);
>> plot(s,f,'b-',s,g,'r--')
```

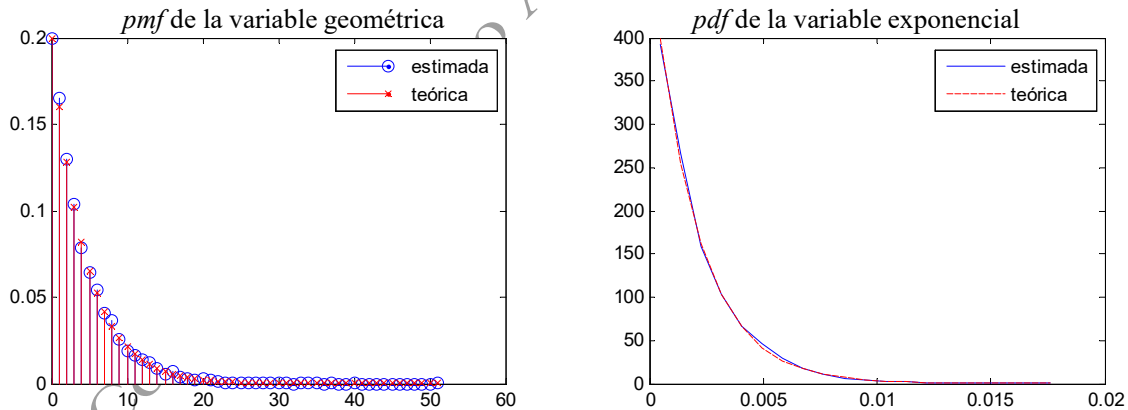


Figura 135. Distribuciones teóricas y empíricas, estimadas de 10000 muestras pseudo-aleatorias

Desafortunadamente no todas las funciones acumulativas de distribución son invertibles. Por ejemplo, la Gaussiana ni siquiera tiene una expresión en forma cerrada. Entonces se usan otras propiedades estadísticas como, por ejemplo, el hecho de que un punto al azar en un cuadrado unitario, (u_1, u_2) , puede transformarse en dos muestras independientes de variables aleatorias gaussianas con media cero y varianza 1 mediante la transformación

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \sqrt{-2 \log(u_1)} \begin{bmatrix} \cos(2\pi u_2) \\ \sin(2\pi u_2) \end{bmatrix}$$

que en matlab se consigue así (por ejemplo)

```
>> u = rand(2,1);
>> x1 = sqrt(-2*log(u(1)))*cos(2*pi*u(2));
>> x2 = sqrt(-2*log(u(1)))*sin(2*pi*u(2));
```

En efecto, despejando para u_1 y u_2 obtenemos

$$u_1 = e^{-(x_1^2+x_2^2)/2} \quad u_2 = \frac{1}{2\pi} \tan^{-1}\left(\frac{x_2}{x_1}\right)$$

cuyo jacobiano es la es la función de densidad de probabilidad conjunta de dos variables normales independientes,

$$\begin{vmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} \end{vmatrix} = -\frac{1}{2\pi} e^{-(x_1^2+x_2^2)/2} = -\left(\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-x_2^2/2}\right)$$

Usando este *método Box-Muller*, como se le conoce, podemos verificar la función de densidad de probabilidad de las muestras obtenidas (ver Figura 136):

```
>> u = rand(10000,2);
>> x1 = sqrt(-2*log(u(:,1)))*cos(2*pi*u(:,2));
>> x2 = sqrt(-2*log(u(:,1)))*sin(2*pi*u(:,2));
>> [h,s] = hist([x1; x2],20);
>> d = s(2)-s(1);
>> f = h/sum(h)/d;
>> g = exp(-s.*s/2)/sqrt(2*pi);
>> plot(s,f,'b-',s,g,'r--')
>> legend('estimada','teórica')
```

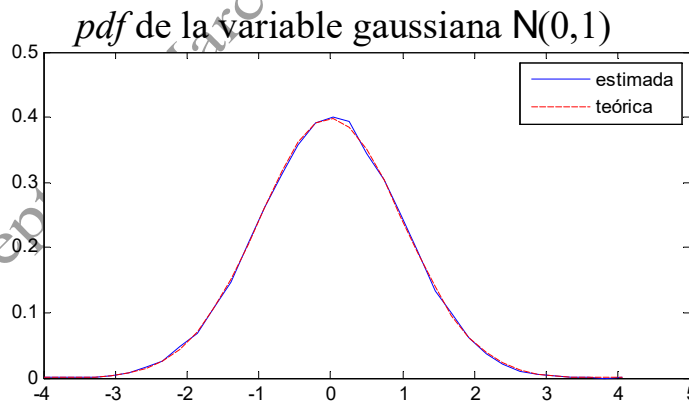


Figura 136. Estimación de la pdf $N(0,1)$ a partir de muestras generadas mediante Box-Muller

Como la CDF en un punto es el área debajo de la pdf desde $-\infty$ hasta ese punto, otro algoritmo típico consiste en generar puntos uniformemente distribuidos en un rectángulo que contenga toda el área debajo de la *pdf*. Si el punto cae debajo de la curva, su coordenada x es una muestra de la variable aleatoria. Si no, se rechaza el punto seleccionado y se repite el procedimiento hasta obtener una

muestra válida. Este *método de rechazo*, como se le conoce, tiene algunas variantes interesantes que buscan mayor eficiencia o mayor exactitud, como el método ziggurat.

En general, los métodos de generación de números pseudo-aleatorios y de muestras de variables aleatorias con distintas distribuciones es un campo de la teoría de números que avanza aceleradamente y que tiene importantes aplicaciones no sólo en simulación sino en criptografía, por ejemplo (aunque en esa área existen requerimientos adicionales de seguridad). En este aparte sólo queríamos mostrar una breve introducción a algunos de los métodos utilizados, en especial si el estudiante requiere generar muestras de sus propias distribuciones empíricas o teóricas.

Matlab, en particular, no sólo permite seleccionar el algoritmo con que se obtienen muestras uniformes en $[0,1)$ sino que también permite escoger el algoritmo con que se obtienen muestras de la variable normal a partir de las uniformes (ver en la documentación del matlab los detalles de la instrucción `RandStream()`). Excepto por razones de compatibilidad con programas existentes, o por la necesidad de usar múltiples flujos, es adecuado usar los algoritmos por defecto que matlab propone (Mersenne twister para números uniformes en $[0,1)$ y Ziggurat para números gaussianos con media cero y norma 1) pues, hasta ahora, son los que mejores características presentan. De otro lado, matlab incluye rutinas para generar muestras de una gran cantidad de distribuciones mediante la instrucción `random()` –Beta, Binomial, χ^2 , F , Exponencial, Gamma, Gaussiana, Geométrica, hipergeométrica, lognormal, Pareto, Poisson, Rayleigh, t de student, uniforme continua o discreta, Valor Extremo, Weibull, etc.–.

Por ejemplo, si en el programa simulador del Listado 7 cambiamos los tiempos entre llegadas constantes por tiempos exponencialmente distribuidos con promedio $1/\lambda$ y cambiamos los tiempos de servicio constantes por tiempos exponencialmente distribuidos con promedio $1/\mu$, habremos pasado de simular la cola $D/D/n/k$ a simular la cola $M/M/n/k$. Aunque el cambio es mínimo en cuanto a la modificación del listado del programa simulador, el cambio en el análisis de los resultados de simulación es monumental, como veremos en el siguiente subtítulo.

e. Simulación de colas aleatorias: Análisis de resultados

Como mencionamos, para convertir el programa simulador de la cola $D/D/n/k$ a un programa que simule la cola $M/M/n/k$ basta con cambiar la distribución de los tiempos entre llegadas y tiempos de servicio, de determinística a exponencial. Las tres líneas relevantes del programa del Listado 7 son las siguientes

```
31 proximaLlegada = reloj + 1/lambda; % Programa la próxima llegada
37 proximaSalida(ss) = reloj + 1/mu; % Programa la próxima salida de este servidor
59 proximaSalida(k) = reloj + 1/mu; % Programa la próxima salida de este servidor
```

que ahora deberán expresar lo siguiente:

```
31 proximaLlegada = reloj + random('exp',1/lambda); % Programa la próxima llegada
37 proximaSalida(ss) = reloj + random('exp',1/mu); % Programa la próxima salida de este servidor
59 proximaSalida(k) = reloj + random('exp',1/mu); % Programa la próxima salida de este servidor
```

Igualmente valdría la pena cambiar ahora la primera línea para indicar que es la cola M/M y no la cola D/D:

```
1 function [EN, EQ, ET, EW, PB, G] = ColaMMnk(lambda,mu,servidores,cupo,tiempoSimulacion)
```

Mientras que al simular la cola D/D/n/k basta con obtener una traza y tomar las estadísticas temporales de esa traza, pues todas las simulaciones arrojarán la misma traza ya que no hay ningún componente aleatorio en ellas, cada ejecución del programa simulador de la cola M/M/n/k arrojará una traza diferente debido a la inclusión de muestras de variables aleatorias. Esto genera una incertidumbre en los resultados de la simulación de la cola M/M/n/k que no existían con la simulación de la cola D/D/n/k. Por ejemplo, si ejecutamos tres veces el programa simulador de la cola D/D/1 con $\lambda=0.8$ y $\mu=1$, obtendremos tres veces la traza observada en la tercera gráfica de la Figura 126, pero si ejecutamos tres veces el programa simulador de la cola M/M/1 con los mismos parámetros, obtendremos tres trazas diferentes, como muestra la Figura 137.

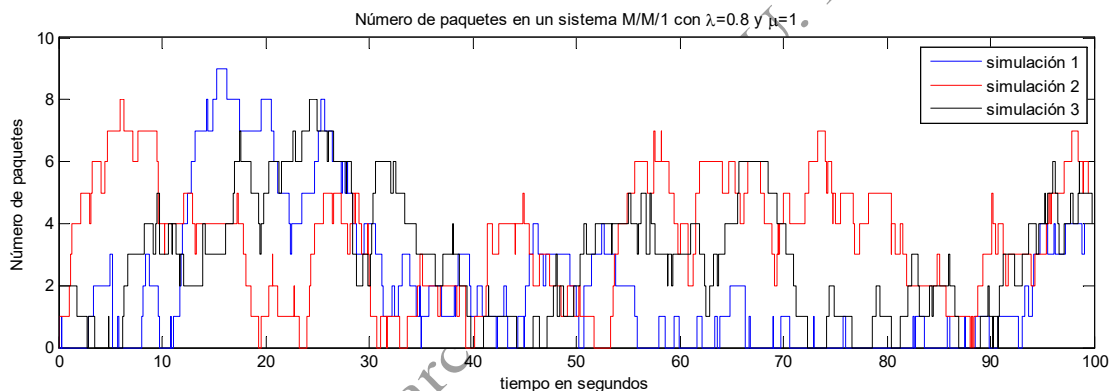


Figura 137. Tres simulaciones distintas de un mismo sistema M/M/1

Este ejemplo ilustra muy bien el concepto de proceso estocástico: Cada ejecución del programa simulador es un experimento aleatorio que genera como resultado una traza del número de paquetes en función del tiempo. Con n simulaciones tenemos n trazas, $\{N_i(t), t \geq 0\}, i=1,2,\dots,n$. Si hacemos n simulaciones, podemos estimar el número promedio de paquetes en el sistema en un instante t_0 cualquiera mediante el estimador

$$\hat{E}[N(t_0)] = \frac{1}{n} \sum_{i=1}^n N_i(t_0)$$

Si suponemos que el sistema alcanza la estacionariedad, podríamos considerar que la función de valor medio llega a dejar de depender del tiempo:

$$\hat{E}[N] = \lim_{t_0 \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n N_i(t_0)$$

Y, si la varianza es finita, al tomar más y más muestras se va reduciendo la incertidumbre de la estimación:

$$E[N] = \lim_{n \rightarrow \infty} \lim_{t_0 \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n N_i(t_0)$$

Por otro lado, con una sola traza, las propiedades de ergodicidad y estacionariedad nos permiten saber que basta con simular durante un tiempo infinito para calcular

$$E[N] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt$$

tal como lo hacíamos con la cola D/D/n/k. Claro, con la cola D/D/n/k no queremos estimar la media de un proceso estocástico estacionario y ergódico, como con la cola M/M/n/k, sino que sólo queremos minimizar el efecto del período transiente.

Así pues, realicemos veinte simulaciones para generar veinte trazas como las de la Figura 137 $\{N_i(t), t \geq 0\}, i=1,2,\dots,20$, y calculemos el promedio del número de paquetes en cada instante:

```
for i=1:20
    [traza,EN,EQ,ET,EW,G,PB] = ColaMMnk(0.75,1,1,inf,1500); % Corre 20 simulaciones
    dt = traza(2:end,1) - traza(1:end-1,1); % genera una traza
    areas = traza(1:end-1,2).*dt; % tiempo entre eventos
    EN = cumsum(areas)./traza(2:end,1); % Áreas de los rectángulos
    plot(traza(2:end,1),EN); % Promedio acumulado hasta
    hold on % cada instante
end
```

Si tomamos muestras de EN cada segundo, podemos promediar los veinte promedios temporales en instantes enteros para considerar el promedio temporal (en la variable tiempo) y el promedio probabilístico (en la variable del espacio muestral). El resultado se aprecia en la Figura 138 para $\lambda=0.75$ y $\mu=1$.

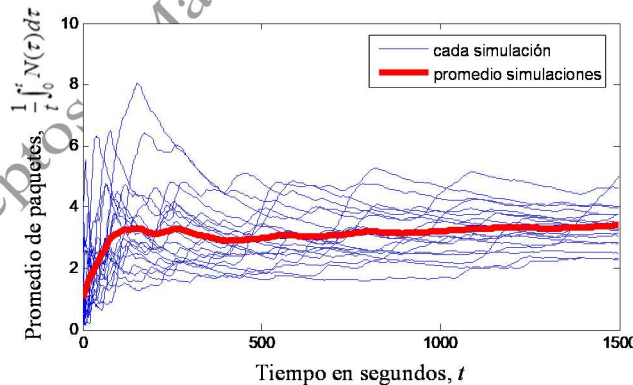


Figura 138. Trazas obtenidas de 20 simulaciones y función de valor medio

Como podemos notar, cada traza (y el promedio estadístico de las trazas) sufre un transiente antes de alcanzar la estabilidad, aunque el promedio estadístico se estabiliza más rápido, tendiendo más claramente a lo que parece ser el verdadero valor esperado del número de paquetes en el sistema. El tiempo de simulación debe ser tan pequeño como se pueda por razones de eficiencia, pero también debe ser tan grande como se pueda para asegurar que el sistema alcanza el estado estable en el que queremos medir los promedios. De otro lado, a menos que empecemos a tomar estadísticas una vez

haya transcurrido el período transiente, debemos alargar el tiempo de simulación para que el efecto de este período transiente sobre las estadísticas temporales se haga despreciable. En general, el tiempo de simulación se escoge de acuerdo con algunas pruebas piloto como las de la Figura 138.

El otro aspecto importante en la planeación de experimentos consiste en la selección del número de repeticiones que debemos hacer para asegurar cierto nivel de confianza en los resultados de la simulación. Una alternativa es hacer una única simulación durante un tiempo increíblemente largo, pues la ergodicidad nos asegura que así nos acercaremos indefectiblemente al verdadero valor medio. El problema es que puede que el tiempo que se requiera sea excesivamente largo y, aun así, no tendríamos manera de medir si ya estamos suficientemente cerca del verdadero promedio o si debemos alargar el tiempo de simulación para acercarnos con un nivel de confianza suficiente.

Supongamos, entonces, que escogimos un tiempo total de simulación T y repetimos la simulación n veces para obtener n trazas $\{N_i(t), 0 \leq t \leq T\}, i=1,2,\dots,n$. Con cada simulación calculamos el promedio temporal $\left\{N_i = \frac{1}{T} \int_0^T N_i(t) dt, i=1,2,\dots,n\right\}$ y finalmente calculamos el promedio estadístico de los n resultados, $\hat{N} = \frac{1}{n} \sum_{i=1}^n N_i$. Este valor será nuestro estimador del número promedio de paquetes en el sistema, pero ¿qué tan cerca o qué tan lejos podemos estar del verdadero valor esperado?

Consideremos cada N_i (el promedio temporal de cada simulación) como una variable aleatoria cuyo valor esperado es el verdadero promedio que queremos encontrar, $m = E[N]$, y que tiene una varianza finita, σ^2 . Sabemos entonces que el promedio \hat{N} tiende a tener una distribución normal con media m y varianza σ^2/n , y la variable normalizada $Z = (\hat{N} - m)/(\sigma/\sqrt{n})$ tiende a tener una distribución normal con media cero y varianza 1, $\mathcal{N}(0,1)$, a medida que aumenta el número de simulaciones, n (ver definición 81). El $\alpha \cdot 100\%$ percentil superior de Z es el valor z_α para el cual $P[Z > z_\alpha] = \alpha$, de manera que $P[|Z| \leq z_{\alpha/2}] = 1 - \alpha$, como muestra la Figura 139.

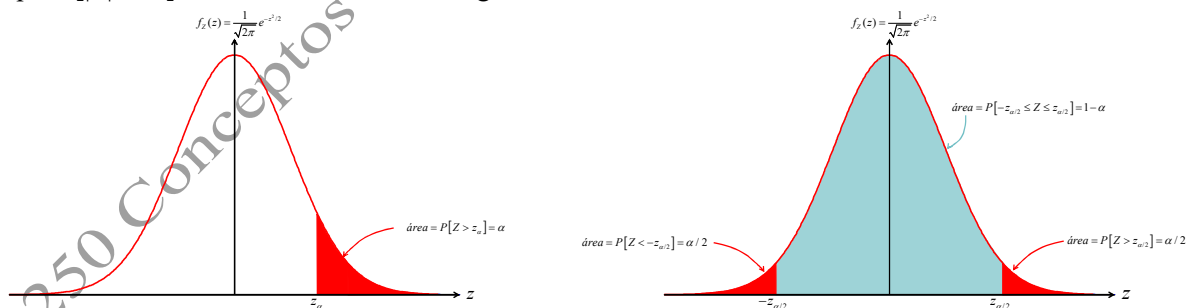


Figura 139. Conceptos de percentil (a) e intervalo de confianza (b)

En matlab® resulta muy fácil hallar los percentiles de la distribución normal. Por ejemplo, el valor de z que hace $P[Z > z] = 0.05$ (el 5% percentil) se calcula así:

```
>> z_05 = norminv(1 - 0.05, 0, 1)
```

pues la instrucción `norminv` invierte la CDF de la distribución normal, de manera que `norminv(p, μ, σ)` retorna el valor de x que hace $F_X(x) = P[X \leq x] = p$ si $X \sim \mathcal{N}(\mu, \sigma)$.

Tomemos la última expresión y remplacemos Z por $(\hat{N} - m)/(\sigma/\sqrt{n})$ para multiplicar por σ/\sqrt{n} en los tres términos dentro del argumento de la probabilidad:

$$P[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}] = P\left[-z_{\alpha/2} \leq \frac{\hat{N} - m}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] = P\left[-\frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq \hat{N} - m \leq \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Restando \hat{N} en los tres términos y multiplicando por -1 , obtenemos la siguiente expresión. Ella determina un intervalo alrededor de \hat{N} donde se encuentra la verdadera media, m , con una confianza de $1 - \alpha$:

$$P\left[\hat{N} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq m \leq \hat{N} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

El intervalo que delimita m , $\hat{N} \pm z_{\alpha/2}\sigma/\sqrt{n}$, se denomina "intervalo de confianza" y la cantidad $1 - \alpha$ es el "nivel de confianza", pues $1 - \alpha$ es la probabilidad con que podemos asegurar que m se encuentra en dicho intervalo. Por ejemplo, si queremos un nivel de confianza del 95% ($\alpha = 0.05$), usamos el percentil 2.5%, que es $z_{0.025} = 1.96$ ($\text{norminv}(1 - 0.025, 0, 1) = 1.96$) y, con él, calculamos el intervalo del 95% de confianza en términos de \hat{N} , σ^2 y n .

Desafortunadamente la anterior expresión se basa en el conocimiento de la varianza, σ^2 , cuando apenas estamos tratando de estimar la media, m . Una alternativa es estimar σ^2 a partir de las muestras de las simulaciones mediante la expresión

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (N_i - \hat{N})^2$$

pero, en este caso, la variable normalizada $T = (\hat{N} - m)/(S/\sqrt{n})$ no es exactamente normal, sino que tiene una distribución t de student con $n-1$ grados de libertad. La Figura 140 muestra la distribución t de student con 1, 2, 5 e ∞ grados de libertad, donde el último caso corresponde exactamente a la variable $\mathcal{N}(0,1)$. Como era de esperarse, la distribución t es más dispersa que la distribución normal debido a que el uso del estimador S^2 en vez de la varianza σ^2 introduce mayor incertidumbre.

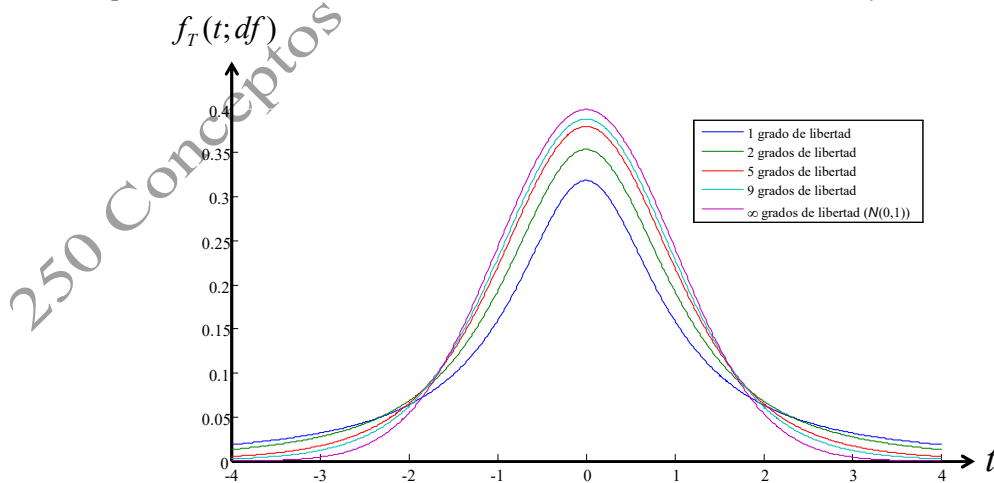


Figura 140. Comparación entre la distribución t de student y la distribución gaussiana

Procediendo de la misma manera que cuando se conocía la varianza,

$$P\left[\hat{N} - \frac{t_{\alpha/2; n-1} S}{\sqrt{n}} \leq m \leq \hat{N} + \frac{t_{\alpha/2; n-1} S}{\sqrt{n}}\right] = 1 - \alpha$$

donde $t_{\alpha, df}$ es el α -100% percentil de la distribución t de student con df grados de libertad, esto es, $P[T > t_{\alpha, df}] = \alpha$ si T es una variable aleatoria con distribución t de student con df grados de libertad. El percentil $t_{\alpha, df}$ es igualmente fácil de calcular con matlab:

```
>> t_alfa_df = tinv(1-alfa,df);
```

El programa del Listado 8 realiza tantas simulaciones como sean necesarias para satisfacer unos requerimientos de confianza específicos. Por ejemplo, supongamos que queremos obtener un intervalo del 95% de confianza para el número promedio de paquetes en una cola M/M/1 con $\lambda=0.75$ y $\mu=1$, que sea $\pm 5\%$ del promedio estimado. La manera de invocar el programa sería así:

```
>> [m,v,delta]=SimulacionesMMnk(0.75,1,1,inf,8000,0.05,0.95)
```

Un posible resultado sería el siguiente, en el que se requirieron 18 simulaciones:

```
Con 2 simulaciones, el número promedio de paquetes es 2.8562 más o menos 243.646%
Con 3 simulaciones, el número promedio de paquetes es 2.9041 más o menos 47.3829%
Con 4 simulaciones, el número promedio de paquetes es 2.9323 más o menos 24.7332%
Con 5 simulaciones, el número promedio de paquetes es 2.9419 más o menos 16.6842%
Con 6 simulaciones, el número promedio de paquetes es 2.9330 más o menos 12.6756%
Con 7 simulaciones, el número promedio de paquetes es 2.8847 más o menos 11.1464%
Con 8 simulaciones, el número promedio de paquetes es 2.8981 más o menos 9.3492%
Con 9 simulaciones, el número promedio de paquetes es 2.9213 más o menos 8.1855%
Con 10 simulaciones, el número promedio de paquetes es 2.9673 más o menos 7.8915%
Con 11 simulaciones, el número promedio de paquetes es 2.9848 más o menos 7.1108%
Con 12 simulaciones, el número promedio de paquetes es 2.9848 más o menos 6.4123%
Con 13 simulaciones, el número promedio de paquetes es 3.0277 más o menos 6.5325%
Con 14 simulaciones, el número promedio de paquetes es 3.0239 más o menos 6.0101%
Con 15 simulaciones, el número promedio de paquetes es 3.0022 más o menos 5.8067%
Con 16 simulaciones, el número promedio de paquetes es 3.0099 más o menos 5.4117%
Con 17 simulaciones, el número promedio de paquetes es 3.0000 más o menos 5.1205%
Con 18 simulaciones, el número promedio de paquetes es 2.9862 más o menos 4.9235%
```

```
function [m,v,delta] = SimulacionesMMnk(lambda, mu, servidores, cupo, ...
    tiempoSimulacion, intervalo, confianza)
% [m,v,delta]=N_Simulaciones_MMnk(0.75,1,1,inf,8000,0.05,0.95)
% Ejecuta no menos de 5 ni más de 50 simulaciones de una cola
M/M/servidores/cupo
% con tasa de llegadas lambda y tasa de servicio mu. Cada simulación se ejecuta
% durante tiempoSimulación segundos. El número de simulaciones se detiene si
los
% intervalos del 95% de confianza se alcanzan a estimar dentro de más o menos
5%
% del promedio estimado. Devuelve la media (m), la varianza (v) y el intervalo
de
% E[N], E[Q], E[T], E[W], Gamma y Pb
%
    suma1 = [0 0 0 0 0 0]; % Acumula las muestras de las 6 variables
```



```

suma2 = [0 0 0 0 0 0];           % Acumula los cuadrados de las muestras
numeroSimulaciones = 0;         % Cuenta las simulaciones que debe hacer
porcentaje = 100;               % Nivel de confianza
while ((porcentaje/100 > intervalo) * (numeroSimulaciones<50) + ...
      (numeroSimulaciones<2))
    % no menos de dos ni más de 50 simulaciones, esperando satisfacer el
    % intervalo de confianza solicitado
    numeroSimulaciones = numeroSimulaciones + 1; % Una simulación más
    [traza,EN,EQ,ET,EW,G,PB] = ColaMMnk(lambda, mu, servidores, cupo, ...
                                         tiempoSimulacion);
    suma1 = suma1 + [EN EQ ET EW G PB]; % Acumula las medidas
    suma2 = suma2 + [EN^2 EQ^2 ET^2 EW^2 G^2 PB^2]; % y sus cuadrados
    if numeroSimulaciones > 1
        m = suma1/numeroSimulaciones; % Estima los promedios y las
varianzas
        v = (suma2 - suma1.*suma1/numeroSimulaciones)/(numeroSimulaciones-
1);

        percentil = tinv(0.5+confianza/2,numeroSimulaciones-1);
        delta = percentil*sqrt(v/numeroSimulaciones);
        porcentaje = 100*delta(1)/m(1);
        disp(['Con ' num2str(numeroSimulaciones) ' simulaciones, el ' ...
              'número promedio de paquetes es ' num2str(m(1)) ' más o ' ...
              'menos ' num2str(porcentaje) '%']);
    end
end

```

Listado 8. Este programa hace tantas simulaciones como sean necesarias para obtener un intervalo de confianza y un nivel de confianza específicos

Es interesante notar que un estudio de simulación que involucre aleatoriedad jamás podrá arrojar resultados como "el número promedio de paquetes es 3", sino que sus resultados deben tener la siguiente forma: "Según el estudio de simulación, con un nivel de confianza del 95% el número promedio de paquetes se encuentra entre 2.8369 y 3.1355".

Por último, llevaremos a cabo un estudio de simulación que permita ver cómo varía el número promedio de paquetes en el sistema con la intensidad de tráfico, $\rho = \lambda/\mu$. El programa del Listado 9 recorre diferentes valores de ρ y, para cada uno de ellos, repite tantas simulaciones como sean necesarias para lograr intervalos de $\pm 5\%$ con el 95% de confianza (a menos que se requieran más de 50 simulaciones). La gráfica final presenta los intervalos del 95% de confianza en función de ρ , como muestra la Figura 141. En total se hicieron 174 simulaciones y todos los puntos alcanzaron intervalos del 95% de confianza menores al 5%, excepto $\rho=0.98$ que, después de 50 simulaciones, sólo había logrado un intervalo del 95% de confianza de $\pm 7.2\%$.

```

function [EN, ENSup, ENinf] = DesempenhoMMnk(servidores,cupo)
% function [EN, ENSup, ENinf] = DesempenhoMMnk(1,inf)
% Encuentra mediante un estudio de simulación el número promedio de paquetes en
% el sistema M/M/servidores/cupo para diferentes valores de la intensidad de
% tráfico en el intervalo [0,servidores]. Determina intervalos del 95% de
% confianza, intentando que estén dentro del 5% alrededor de la media estimada.
%

```

```

mu = 10; % Tasa de servicio de c/servidor
rho = servidores*[0.02 2 4 5 6 7 7.5 8 8.5 9 9.2 9.5 9.8]/10; %
intensidades
for i = 1:13 % Para cada intensidad de tráfico
    lambda = rho(i)*mu; % calcula la tasa de llegadas
    disp('=====') % y ejecuta el número de
simulaciones
    disp(['rho = ' num2str(rho(i)) ' : ']) % necesarias (hasta 50) para
    disp('=====') % obtener intervalos de más o menos
    % 5% con 95% de confianza
    [m,v,delta]=SimulacionesMMnk(lambda,mu,servidores,cupo,8000,0.05,0.95);
    EN(i) = m(1); % Almacena el promedio de paquetes
    ENSup(i) = m(1) + delta(1); % en el sistema y sus intervalos
    ENinf(i) = m(1) - delta(1);
end
plot(rho,EN,'r-',rho,rho./(1-rho),'k-.',rho,ENsup,'b--',rho,ENinf,'b--')
axis([0 1 0 60]) % Grafica los resultados obtenidos
xlabel('Intensidad de tráfico') % (incluyendo la comparación con la
ylabel('Número de paquetes') % solución teórica)
legend('promedio','Teórico','Intervalo del 95% de confianza')
end

```

Listado 9. Estudio de la cola M/M/1 mediante simulación

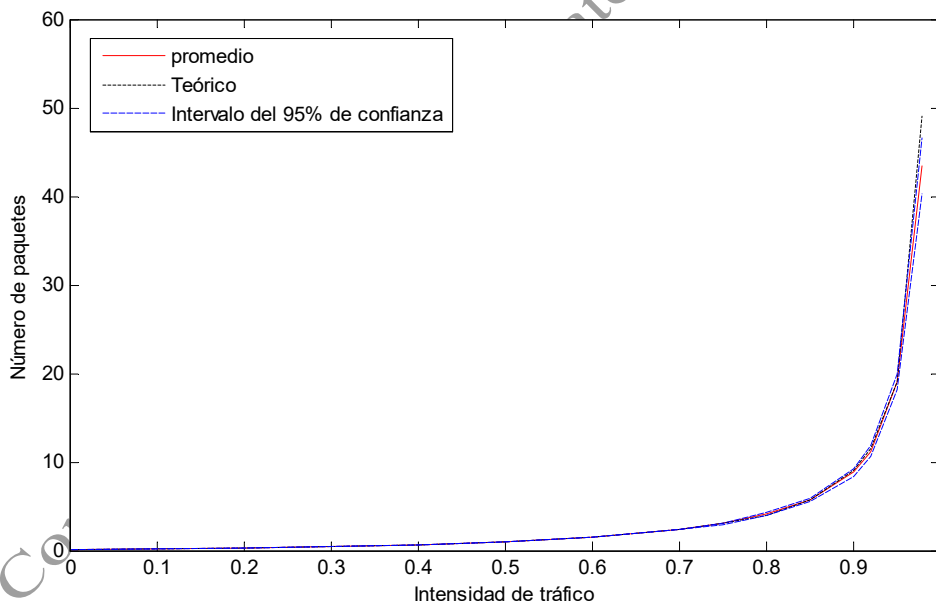


Figura 141. Estudio de simulación de la cola M/M/1

Un aspecto novedoso de la Figura 141 es que también se compara con un resultado teórico, $E[N]=\rho/(1-\rho)$, el cual deduciremos en la siguiente definición. En efecto, un estudio de simulación siempre necesita una validación que les permita, tanto al investigador como al lector del estudio, ganar cierta confianza en la validez estadística de los resultados. Un método interesante es comparar con resultados teóricos cuando eso es posible (como en el caso de la cola M/M/1). De hecho, como ya mencionamos, cada una de las soluciones analíticas, experimentales y de simulación de un modelo matemático se puede usar para validar otras soluciones. Por ejemplo, ahora que el modelo de

simulación fue capaz de reproducir un resultado teórico de la cola M/M/1, parece haber quedado validado para ser usado en el estudio de sistema ligeramente más complicados cuya solución analítica sea muy difícil. Sin embargo, vale la pena hacer notar que en el punto $\rho=0.98$, cuando llegamos a las 50 simulaciones sin haber alcanzado el intervalo del 5% que deseábamos, el resultado teórico está ligeramente por encima de la cota superior del correspondiente intervalo del 95% de confianza.

Definitivamente, parece oportuno empezar el tratamiento analítico de las colas M/M/n/k.

136. La cola M/M/1

Partimos de la suposición de que las llegadas forman un proceso independiente de Poisson con intensidad λ paquetes por segundo. Como vimos en la definición (d), esta propiedad consiste en que, en un intervalo de tiempo infinitesimal, el número de llegadas es proporcional a la longitud del intervalo y las llegadas simultáneas son improbables:

$$P[A(t, t + \Delta t) = n] = \begin{cases} \lambda\Delta t + o(\Delta t) & n = 1 \\ 1 - \lambda\Delta t + o(\Delta t) & n = 0 \\ o(\Delta t) & n > 1 \end{cases}$$

donde $o(\Delta t)$ es cualquier función que tienda a 0 más rápidamente que Δt .

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$$

La anterior descripción del proceso de Poisson conducía a que los tiempos entre llegadas son exponenciales e independientes, como se mostró en la definición 44(b) (de ahí la M de *Memoryless* en la especificación del proceso de llegadas, pues la distribución exponencial es la única distribución en tiempo continuo que no tiene memoria, como se mostró en la definición 51). Consecuentemente, como los tiempos de servicio son exponenciales, mientras el servidor esté ocupado el número de paquetes que salen en un intervalo infinitesimal de tiempo es proporcional a la longitud del intervalo y las salidas simultáneas son improbables

$$P[D(t, t + \Delta t) = n] = \begin{cases} \mu\Delta t + o(\Delta t) & n = 1 \\ 1 - \mu\Delta t + o(\Delta t) & n = 0 \\ o(\Delta t) & n > 1 \end{cases}$$

Ahora queremos describir la distribución de probabilidad del proceso $\{N(t), t \geq 0\}$, donde $N(t)$ es la variable aleatoria que representa el número de paquetes en el sistema en el instante t . Por simplicidad, denotaremos su distribución mediante $p_n(t) = P[N(t)=n]$. Tenemos, pues, una variable continua (el tiempo) y una variable discreta (el estado) y, como de costumbre en el modelamiento probabilístico de sistemas, queremos expresar probabilidades que involucren incrementos infinitesimales de tiempo e incrementos unitarios de estado como suma de productos de probabilidades más simples, para lo cual el teorema de la probabilidad total resulta la herramienta ideal (definición 19). Sólo necesitamos encontrar una partición adecuada del espacio de probabilidad. Si consideramos las probabilidades de estado dentro de un infinitésimo de tiempo, la partición adecuada para relacionar las variaciones infinitesimales de probabilidad será el estado actual, pues la transición del estado k al estado n ocurrirá si la suma de llegadas menos la suma de salidas es $n-k$:

$$p_n(t + \Delta t) = \sum_{k=0}^{\infty} p_k(t) P[A(t, t + \Delta t) - D(t, t + \Delta t) = n - k | N(t) = k]$$

Pero las llegadas y salidas son independientes entre sí e independientes del estado del sistema, así que

$$P[A(t, t + \Delta t) - D(t, t + \Delta t) = n - k | N(t) = k] = P[A(t, t + \Delta t) - D(t, t + \Delta t) = n - k] = \\ = \sum_{j=0}^{\infty} P[A(t, t + \Delta t) = j] P[D(t, t + \Delta t) = j - n + k]$$

Como casi todos los términos de la suma anterior son $o(\Delta t)$, los podemos agrupar en un solo $o(\Delta t)$ y así quedan sólo los incrementos y decrementos unitarios en la primera relación de recurrencia. Para $n \geq 1$,

$$p_n(t + \Delta t) = p_{n-1}(t)P[A(t, t + \Delta t) = 1]P[D(t, t + \Delta t) = 0] + \\ p_{n+1}(t)P[A(t, t + \Delta t) = 0]P[D(t, t + \Delta t) = 1] + \\ p_n(t)P[A(t, t + \Delta t) = 0]P[D(t, t + \Delta t) = 0] + \\ o(\Delta t)$$

Al remplazar las expresiones originales para una llegada, una salida, cero llegadas y cero salidas obtenemos

$$p_n(t + \Delta t) = p_{n-1}(t)(\lambda \Delta t + o(\Delta t))(1 - \mu \Delta t + o(\Delta t)) + \\ p_{n+1}(t)(1 - \lambda \Delta t + o(\Delta t))(\mu \Delta t + o(\Delta t)) + \\ p_n(t)(1 - \lambda \Delta t + o(\Delta t))(1 - \mu \Delta t + o(\Delta t)) + \\ o(\Delta t), \quad n \geq 1$$

Podemos agrupar la larga lista de funciones $o(\Delta t)$ en un solo $o(\Delta t)$ — esa función ómicron donde agrupamos todas las demás es como nuestro tarro de la basura— para obtener

$$p_n(t + \Delta t) = p_{n-1}(t)(\lambda \Delta t)(1 - \mu \Delta t) + p_{n+1}(t)(1 - \lambda \Delta t)(\mu \Delta t) + p_n(t)(1 - \lambda \Delta t)(1 - \mu \Delta t) + o(\Delta t)$$

Si expandimos los productos y notamos que $\alpha \Delta t^2$ es otra función $o(\Delta t)$ que podemos depositar en el tarro de la basura, obtenemos

$$p_n(t + \Delta t) = p_{n-1}(t)(\lambda \Delta t) + p_{n+1}(t)(\mu \Delta t) + p_n(t)(1 - \lambda \Delta t - \mu \Delta t) + o(\Delta t)$$

Ahora basta con restar a ambos lados $p_n(t)$ y dividir por Δt para mostrar claramente a donde queremos llegar:

$$\frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} = \lambda p_{n-1}(t) + \mu p_{n+1}(t) - (\lambda + \mu) p_n(t) + \frac{o(\Delta t)}{\Delta t}$$

pues, tomando el límite cuando Δt tiende a cero, obtenemos la ecuación diferencial, de diferencias, lineal y con coeficientes constantes que regula la evolución temporal de las probabilidades de estado del sistema M/M/1:

$$\frac{d}{dt} p_n(t) = \lambda p_{n-1}(t) + \mu p_{n+1}(t) - (\lambda + \mu) p_n(t), \quad n \geq 1$$

Podemos suponer condiciones iniciales apropiadas para resolver el anterior sistema de ecuaciones diferenciales de diferencias. Por ejemplo, podemos suponer que el sistema empieza desocupado ($p_0(0) = 1$ y $p_n(0) = 0$ para $n > 0$) y tener en cuenta que en el estado cero no se producen salidas ($dp_0(t)/dt = \mu p_1(t) - \lambda p_0(t)$). La solución del sistema se puede obtener de manera más o menos directa combinando juiciosamente la transformada de Laplace para el tiempo continuo y la transformada Z para el estado discreto, llegando a expresiones de las probabilidades basadas en funciones modificadas de Bessel. Bajo cierta condición muy particular que veremos más adelante, esas soluciones tienen un período transiente que se desvanece y una solución de estado estable en la que las probabilidades permanecen constantes. En vez de pasar por este proceso de solución del sistema de ecuaciones diferenciales para buscar después su solución de estado estable con probabilidades constantes (si existe), vamos a partir

de suponer la existencia de dicha estabilidad para encontrar las probabilidades de estado "a la larga", cuando ya no dependen del tiempo, esto es, supondremos la siguiente propiedad de estabilidad

$$\lim_{t \rightarrow \infty} p_n(t) = p_n \Rightarrow \lim_{t \rightarrow \infty} \frac{d}{dt} p_n(t) = 0$$

y después veremos bajo qué condiciones se puede dar dicha estabilidad. Así, si las probabilidades no dependen del tiempo y, por consiguiente, su derivada es cero, la expresión que surge del sistema de ecuaciones diferenciales de diferencia es la siguiente relación de recurrencia:

$$\begin{aligned} (\lambda + \mu) p_n &= \lambda p_{n-1} + \mu p_{n+1}, \quad n \geq 1 \\ \lambda p_0 &= \mu p_1, \quad n = 0 \end{aligned}$$

Bajo la suposición de que la anterior recurrencia es cierta, no sólo llegaremos a una expresión cerrada para las probabilidades estables del estado del sistema, sino que encontraremos la condición necesaria y suficiente para que se dé la estabilidad que suponemos.

Primero es conveniente darle una interpretación física a la relación de recurrencia entre las probabilidades de estado del sistema, una vez se ha alcanzado la estabilidad. Si hay un flujo estable de algún fluido, por ejemplo agua, a través de una tubería, sobre cualquier superficie cerrada imaginaria que pongamos a lo largo del flujo se debe satisfacer el principio de continuidad o conservación de flujo: El flujo total de agua que entra en la superficie debe ser igual al flujo total de agua que sale de la superficie. Si entra más de la que sale, el flujo no es estable porque adentro de la superficie cerrada algo se está llenando y terminará desbordándose. Si sale más de la que entra, el flujo no es estable porque debe haber una tanque adentro que se está desocupando. Pues bien, gracias a esta analogía, a las relaciones de recurrencia que acabamos de ver se les denomina ecuaciones de balance de flujo, pues se trata de considerar la cantidad de paquetes que pasan por segundo de un estado a otro. Por ejemplo, del estado 0, en el que se permanece con probabilidad p_0 , se pasa al estado 1 con una tasa λ ; y del estado 1, en el que se permanece con probabilidad p_1 , se pasa al estado 0 con una tasa μ . La relación $\lambda p_0 = \mu p_1$ refleja la igualdad de estos dos flujos, los cuales salen y entran, respectivamente, de una superficie cerrada que bordea el estado cero, como muestra la Figura 142. La misma figura muestra una superficie cerrada que bordea el estado n : el flujo que entra es $\lambda p_{n-1} + \mu p_{n+1}$ y el flujo que sale es $(\lambda + \mu) p_n$, cuyo balance conduce a la ecuación de recurrencia obtenida de suponer estabilidad en la ecuación diferencial de diferencias: $(\lambda + \mu) p_n = \lambda p_{n-1} + \mu p_{n+1}$.

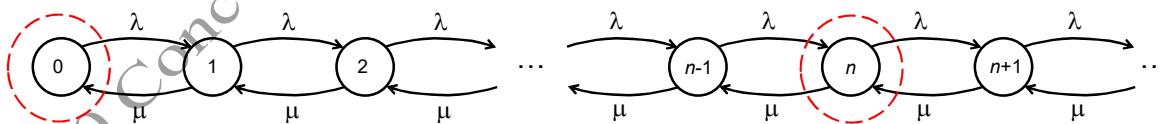


Figura 142. Diagrama de transición entre estados y superficies para balance de flujo

Usando las superficies de la Figura 143, y definiendo la intensidad de tráfico como $\rho \triangleq \lambda / \mu$, podemos establecer las siguientes relaciones entre las probabilidades de estado p_n y la probabilidad de que el sistema esté desocupado, p_0 :

$$\begin{aligned} \lambda p_0 &= \mu p_1 \Rightarrow p_1 = \rho p_0 \\ \lambda p_1 &= \mu p_2 \Rightarrow p_2 = \rho p_1 = \rho^2 p_0 \\ \lambda p_2 &= \mu p_3 \Rightarrow p_3 = \rho p_2 = \rho^3 p_0 \\ \lambda p_{n-1} &= \mu p_n \Rightarrow p_n = \rho p_{n-1} = \rho^n p_0, \quad n > 0 \end{aligned}$$

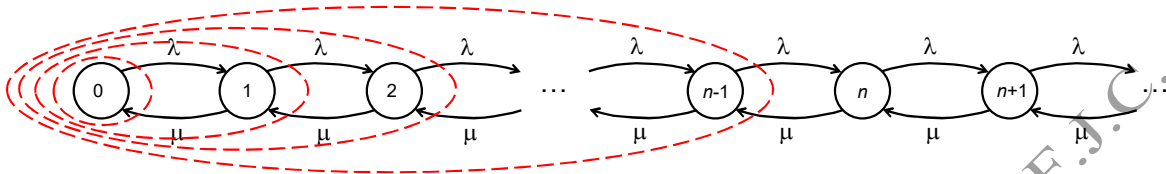


Figura 143. Algunas superficies para establecer ecuaciones de balance de flujo

Ahora que tenemos una expresión general para las probabilidades de estado en función de la probabilidad de que el sistema esté vacío, $p_n = \rho^n p_0$, podemos normalizar la suma de todas las probabilidades (el primer teorema de la probabilidad, $P[\Omega]=1$) para despejar p_0 :

$$\sum_{n=0}^{\infty} p_n = p_0 \sum_{n=0}^{\infty} \rho^n = 1$$

Esta relación sólo es posible si la suma converge, lo cual ocurre únicamente si $\rho < 1$. Esta es la condición de estabilidad que estábamos buscando: Si $\lambda < \mu$, la suma converge y p_0 tomara un valor mayor a cero, indicando que el sistema visitará el estado vacío recurrentemente, por lo que la cola no crecerá indefinidamente. Pero si $\rho \geq 1$ ($\lambda \geq \mu$), la suma no converge y, por consiguiente, la probabilidad de que el sistema esté desocupado será $p_0=0$. Esto implica que la suposición de estabilidad no fue correcta y, para conocer la probabilidad de estado en cada instante, toca buscar la solución en la ecuación diferencial de diferencias original.

Bajo la condición de estabilidad, $\rho < 1$, la suma de la serie geométrica conduce a $1/(1-\rho)$,

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho}$$

de manera que

$$p_0 = 1 - \rho \Rightarrow p_n = \rho^n (1 - \rho)$$

Por un lado, hemos encontrado que si el sistema se llega a estabilizar, el estado del sistema (el número de paquetes en el sistema después de un tiempo suficientemente grande como para considerar que ya se alcanzó la estabilidad) está distribuido geoméricamente, con parámetro $\rho = \lambda/\mu$. Pero también hemos encontrado otra interpretación muy interesante para la intensidad de tráfico: $\rho = \lambda/\mu$ es la probabilidad de que el sistema esté ocupado:

$$\rho \triangleq \lambda / \mu = 1 - p_0$$

Esto es, podemos interpretar la intensidad de tráfico también como la utilización del enlace (la fracción de tiempo que el enlace permanece ocupado).

Conociendo la distribución de probabilidad del número de paquetes en el sistema, N , es fácil contestar cualquier pregunta que podamos hacernos sobre la variable aleatoria N . Por ejemplo, ¿cuál es el número promedio de paquetes que permanecen en el sistema?

$$E[N] = \sum_{n=0}^{\infty} n p_n = (1-\rho) \rho \sum_{n=0}^{\infty} n \rho^{n-1} = (1-\rho) \rho \frac{d}{d\rho} \frac{1}{1-\rho} = \frac{\rho}{1-\rho}$$

Este es exactamente el resultado que incluimos en la Figura 141 como "resultado teórico" para validar los resultados del estudio de simulación.

Para calcular otras estadísticas de interés conviene tener en cuenta el principio PASTA (Poisson Arrivals See Time Averages), que veremos en la siguiente definición. De acuerdo con este principio, los paquetes que llegan según un proceso de Poisson encuentran el sistema en estados típicos, de manera que la probabilidad de que haya n paquetes en el sistema, $p_n = P[N(t)=n]$ para todo t suficientemente grande, es también la probabilidad de que un paquete encuentre n paquetes en el sistema en el instante de su llegada, $p_n = P[N(t_A) = n]$ donde t_A es el instante de llegada de un paquete (una vez alcanzada la estabilidad, claro). Esta es una propiedad muy sorprendente de las llegadas Poisson si se tiene en cuenta que, por ejemplo, la probabilidad de que haya cero paquetes en un sistema D/D/1 estable es $p_0 = 1-\rho$ pero la probabilidad de que un paquete que llegue encuentre cero paquetes en el sistema es $\hat{p}_0 = 1$ (todos los paquetes encuentran el sistema desocupado).

Pues bien, de acuerdo con el principio PASTA, los paquetes encuentran en promedio $E[N]$ paquetes en el sistema. Pero, como los tiempos de servicio están exponencialmente distribuidos y la distribución exponencial no tiene memoria, al servidor le toma, en promedio, $E[N]/\mu$ segundos para atender a todos los paquetes que había en el momento de la llegada de un nuevo paquete, de manera que el tiempo promedio que un paquete permanece en la cola esperando su turno de transmisión es $E[W] = E[N]/\mu = \rho/(\mu-\lambda)$ segundos. Después de esperar en cola, cada paquete permanece en el sistema otros $1/\mu$ segundos en promedio, de manera que el tiempo total de permanencia en el sistema es $E[T] = E[W] + 1/\mu = 1/(\mu-\lambda)$ segundos. Por otro lado, vimos que la probabilidad de que el sistema esté vacío es $p_0 = 1-\rho$, de manera que la probabilidad de que haya al menos un paquete en el sistema, esto es, la probabilidad de que haya un paquete en servicio, es $1-p_0 = \rho$. El número promedio de paquetes en servicio es, entonces, $E[X] = 0(1-\rho) + 1(\rho) = \rho$. El número total de paquetes en el sistema incluye los que están en servicio y los que están en cola esperando su turno de transmisión, por lo que el número promedio de paquetes en la cola será $E[Q] = E[N] - E[X] = \rho/(1-\rho) - \rho = \rho^2/(1-\rho)$. Como el sistema tiene cupo infinito, no sucede que un paquete sea rechazado por el sistema, de manera que la probabilidad de rechazo, o de bloqueo, de un paquete es $P_B = 0$. Por último, como todos los paquetes se aceptan y el sistema es estable (se desocupa eventualmente porque $p_0 = 1-\rho > 0$), entonces la tasa efectiva de paquetes que se transmiten por segundo, el caudal o *throughput*, es $I = \lambda$. Otra forma de ver el caudal es que, mientras el enlace esté ocupado, transmite μ paquetes por segundo, de manera que $I = \mu(1-p_0) = \mu\rho = \lambda$. En resumen, la siguiente tabla presenta las medidas de desempeño que acabamos de encontrar para la cola M/M/1 con tasa de llegadas λ y tasa de servicio μ :

Tabla 1. Medidas de desempeño de la cola M/M/1

Medida	Significado
$E[N] = \frac{\rho}{1-\rho}$	Número promedio de paquetes en el sistema
$E[Q] = \frac{\rho^2}{1-\rho}$	Número promedio de paquetes en la cola
$E[X] = \rho$	Número promedio de paquetes en servicio
$E[T] = \frac{1}{\mu - \lambda}$	Tiempo promedio de permanencia en el sistema
$E[W] = \frac{\rho}{\mu - \lambda}$	Tiempo promedio de espera en la cola
$P_B = 0$	Probabilidad de que un paquete sea rechazado
$I = \lambda$	Caudal (tasa efectiva de paquetes transmitidos por segundo)

La Figura 144 muestra $E[N]$, $E[Q]$ y $E[X]$, así como los retardos normalizados con respecto al tiempo promedio de transmisión, $\mu E[T]$ y $\mu E[W]$. Las gráficas representan nuestra experiencia cotidiana haciendo fila en espera de algún servicio: Cuando hay baja demanda, el sistema tiende a estar desocupado y los tiempos de espera no son mucho mayores al propio tiempo de servicio; pero si la demanda es alta para la capacidad del servidor, el número de clientes en el sistema crece dramáticamente debido a que el denominador, $1-\rho$, que se acerca a cero a medida que ρ se acerca a uno, y los retardos experimentados crecen de la misma manera.

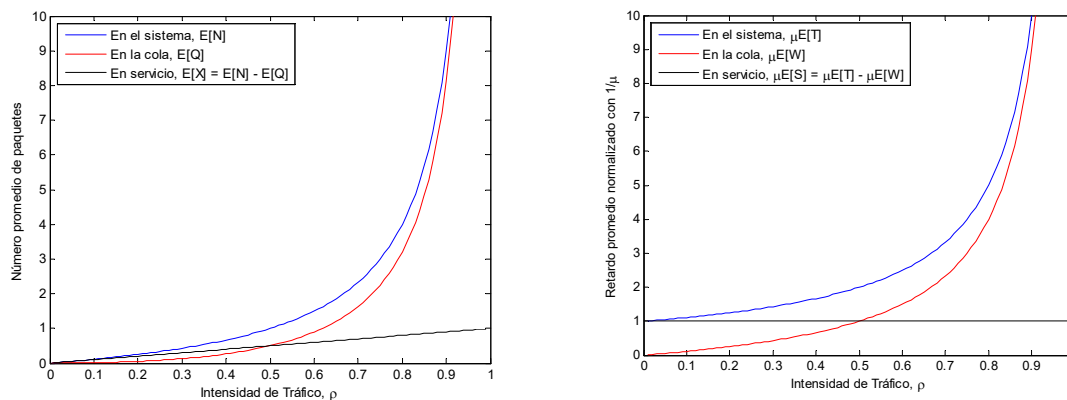


Figura 144. Medidas de desempeño de la cola M/M/1

137. Principio PASTA

En un sistema de colas en el que las llegadas forman un proceso de Poisson, la fracción de llegadas que encuentran el sistema en un estado particular es igual a la fracción de tiempo que el sistema permanece en ese estado. A este principio se le conoce como PASTA por las siglas de "Poisson Arrivals See Time Averages"

Ver el sistema en el instante de una llegada Poisson es estadísticamente equivalente a ver el sistema en un instante aleatorio, con lo cual la probabilidad de que haya n paquetes en el sistema es igual a la probabilidad de que un paquete que llega encuentre n paquetes en el sistema. En particular, el valor esperado de cualquier parámetro de la cola en el instante de una llegada es el promedio temporal de ese parámetro. Este resultado no es de ninguna manera intuitivo pues, en una cola D/D/1 con tasa de llegadas λ menor que la tasa de servicio, μ , cada paquete que llega encuentra el sistema desocupado, pero la probabilidad de que el sistema esté desocupado no es 1, sino $1 - \lambda/\mu$. De esta manera, en la cola D/D/1, $P[\text{hay un paquete en el sistema}] = \lambda/\mu$ pero $P[\text{un paquete que llega encuentra un paquete en el sistema}] = 0$: ¡Son muy distintos! El principio PASTA tampoco aplica si las llegadas forman un proceso de Poisson pero no son independientes de los tiempos de servicio. Por ejemplo, si el tiempo de servicio del paquete n es la mitad del tiempo entre las llegadas n y $n+1$, todos los paquetes encontrarán el sistema vacío, pero la probabilidad de que el sistema esté vacío es $1/2$.

El principio PASTA se puede ilustrar de una manera simple y convincente: Si llega un paquete en el intervalo $[0, t]$, ¿con qué probabilidad llegó en un intervalo de longitud $\tau < t$?

$$\begin{aligned} P[\text{una llegada en un intervalo de longitud } \tau | \text{Una llegada en } [0, t]] &= \dots \\ \frac{P[\text{una llegada en } [0, \tau]] P[\text{cero llegadas en } [0, t - \tau]]}{P[\text{una llegada en } [0, t]]} &= \dots \\ \dots = \frac{\lambda \tau e^{-\lambda \tau} \cdot e^{-\lambda(t-\tau)}}{\lambda t e^{-\lambda t}} &= \frac{\tau}{t} \end{aligned}$$

Esto indica que ver el sistema en el instante de una llegada equivale a ver el sistema en cualquier instante aleatorio.

Es interesante notar que este efecto de observar el sistema inmediatamente antes de una llegada, ocurre también al observar el sistema inmediatamente después de una salida.

Por cierto, hemos visto que el proceso Poisson posee algunas propiedades muy interesantes que facilitan el estudio de los sistemas de colas, pero esa no es una correcta justificación para su uso como modelo de tráfico. Debemos preguntarnos primero si las trazas de tráfico que ingresan a la red se pueden modelar adecuadamente mediante un proceso de Poisson. Lo cierto es que el comportamiento de un gran número de usuarios que actúan independientemente sí se puede aproximar a un proceso de Poisson si imaginamos que cada usuario envía paquetes al sistema como en un proceso de Bernoulli: En cada instante de tiempo el usuario lanza una moneda al aire para transmitir un paquete si cae sello o esperar si cae cara. Bajo esta suposición, la probabilidad de obtener n paquetes de N usuarios es una distribución binomial. Como vimos en la definición 43, cuando N tiende a infinito y

p tiende a cero, la distribución Binomial tiende a una distribución de Poisson. Así, la distribución de Poisson es adecuada cuando las llegadas pueden considerarse independientes unas de otras y provenientes de un gran número de usuarios donde cada usuario participa con una fracción muy pequeña del tráfico total. Esta es una suposición casi exacta para las llegadas de llamadas telefónicas en centrales de conmutación de circuitos con un alto número de abonados pero, desafortunadamente, esta técnica de conmutación está próxima a desaparecer y las llegadas de paquetes de diferentes tipos están lejos de satisfacer las suposiciones que conducen al modelo de Poisson. En definiciones posteriores veremos como adaptamos los modelos para incluir estos casos.

138. Ley de Little, $N = \lambda W$

Considere un sistema estable y conservador de trabajo, en el que los paquetes llegan a una tasa promedio efectiva λ . El promedio temporal del número de paquetes en el sistema, N , es igual a la tasa efectiva de llegadas por segundo, λ , multiplicada por el promedio temporal del tiempo que un paquete pasa en el sistema, W , esto es, $N = \lambda W$.

Esta es una fórmula muy general que aplica independientemente del proceso de llegadas, del proceso de servicios, de la disciplina de servicio e, incluso, de la estructura misma del sistema al que se refiere: Puede ser un segmento de cable, el buffer de un enrutador, un canal en una trama TDM, una red local, un sistema autónomo de Internet o hasta la Internet completa.

De cierta manera, la ley de Little es la extensión probabilística de la regla de tres simple directa, como ilustra el siguiente ejemplo: Imagine un cable de cobre de 2 km de largo por el que se transmite 1 millón de bits por segundo; si la velocidad de propagación de las señales electromagnéticas es de 200 000 km/s, ¿Cuántos bits permanecen en el cable? El tiempo que le toma a las señales atravesar el cable es $W = \text{longitud}/\text{velocidad} = 10 \mu\text{s}$. Entonces, mientras el primer bit va de un extremo a otro del cable, la fuente habrá inyectado $N=10$ bits. Si llamamos λ a la tasa con que entran los bits al cable, 10^6 bits/s, el resultado ilustra la ley de Little: $N = \lambda W$.

La ley de Little tiene importantes aspectos asociados con la relación entre las medias temporales y las medias probabilísticas. Supongamos, por ejemplo, que empezamos a observar el sistema en el instante $t_0=0$, en el que el sistema estaba vacío, y medimos algunas cantidades de interés:

$N(t)$ = Número de paquetes en el sistema en el instante t

$A(t)$ = Número de paquetes que han llegado al sistema hasta el instante t

W_i = Tiempo que tarda en el sistema el i -ésimo paquete en llegar

Desde un punto de vista puramente determinístico, diríamos que el número promedio de paquetes en el sistema hasta el instante T sería

$$N_T = \frac{1}{T} \int_0^T N(t) dt$$

que, aunque depende de T , supondremos que tiende a un valor de estado estable, N , a medida que T aumenta,

$$N = \lim_{T \rightarrow \infty} N_T$$

Llamaremos a N el "promedio temporal" del número de paquetes en el sistema. De igual manera, si han llegado $A(T)$ paquetes hasta el instante T , la tasa promedio de llegadas hasta ese instante sería

$$\lambda_T = \frac{A(T)}{T}$$

que, aunque depende de T , supondremos que tiende a un valor de estado estable, λ , a medida que T aumenta,

$$\lambda = \lim_{T \rightarrow \infty} \lambda_T$$

Llamaremos a λ el "promedio temporal" de la tasa de llegadas. Por último, si el i -ésimo paquete que llegó al sistema tardó en él W_i segundos desde su llegada hasta su salida, el tiempo promedio de permanencia en el sistema de cada paquete hasta el instante T sería

$$W_T = \frac{1}{A(T)} \sum_{i=1}^{A(T)} W_i$$

que, aunque depende de T , supondremos que tiende a un valor de estado estable, W , a medida que T aumenta,

$$W = \lim_{T \rightarrow \infty} W_T$$

Llamaremos a W el "promedio temporal" del tiempo de permanencia de un paquete en el sistema.

En este contexto determinístico, la ley de Little es casi una identidad que se puede demostrar gráficamente bajo algunas suposiciones particulares. Por ejemplo, si suponemos que el primer paquete en salir es el primero en llegar, la gráfica de las llegadas, $A(t)$ –arrivals–, y las salidas, $D(t)$ –departures–, podría tomar una forma como la mostrada en la Figura 145, que incluye también la diferencia $N(t)=A(t)-D(t)$, el número de paquetes en el sistema en el instante t . El i -ésimo paquete llega en el instante t_i y sale en el instante t'_i , de manera que su tiempo de permanencia en el sistema es $W_i = t'_i - t_i$. La figura muestra cómo el área debajo de la curva $N(t)$ se compone por rectángulos de altura 1 y anchura W_i . Si consideramos dos instantes 0 y T en los que el sistema esté vacío, los promedios anteriores toman la siguiente forma:

$$N_T = \frac{1}{T} \int_0^T N(t) dt = \frac{1}{T} \sum_{i=1}^{A(T)} W_i = \frac{A(T)}{T} \left(\frac{1}{A(T)} \sum_{i=1}^{A(T)} W_i \right) = \lambda_T W_T$$

Y, en el límite cuando $T \rightarrow \infty$, obtenemos la ley de Little, $N = \lambda W$.

Una representación gráfica semejante para una disciplina LIFO o cualquier otra conduce a una representación idéntica (repetimos, siempre que el sistema sea "conservador de trabajo").

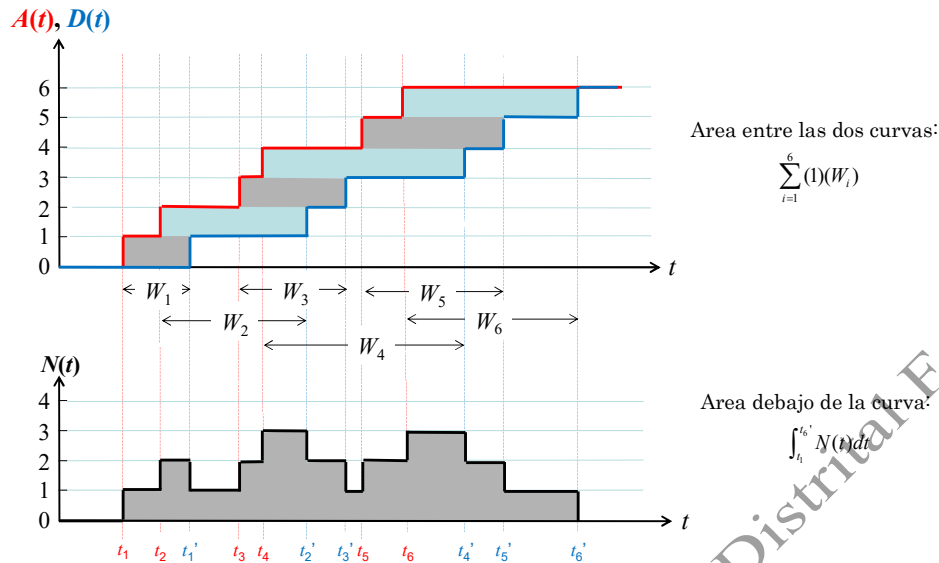


Figura 145. Verificación de la ley de Little en un esquema FIFO

Por supuesto, el anterior contexto puramente determinístico se puede aplicar en la interpretación probabilística que nos interesa, siempre que aplique el principio de ergodicidad, el cual permite reemplazar los promedios temporales por los promedio probabilísticos. Es decir, en vez de aplicar el anterior argumento a una traza muestral, podemos aplicarlo al ensamble entero:

$$E[N(t)] = \sum_{n=0}^{\infty} nP[N(t) = n]$$

Suponiendo que el sistema alcanza un estado estable en el que el número de paquetes en el sistema ya no depende del tiempo,

$$\lim_{t \rightarrow \infty} E[N(t)] = E[N] = \sum_{n=0}^{\infty} nP[N = n]$$

Bajo estas circunstancias, la ley de Little, $N = \lambda W$, sigue siendo válida cuando N, λ y W son los promedios probabilísticos.

Ya vimos un ejemplo simple pero revelador de la ley de Little cuando consideramos el número de bits "en vuelo" sobre un enlace: Si t_p es el tiempo de propagación de una señal electromagnética en el medio y C es la tasa de transmisión, el número de bits que permanecen en el cable es $b = Ct_p$. Este es el famoso "bandwidth-delay product" que tantos dolores de cabeza causa a los ingenieros de redes de comunicaciones, como se vio en la definición 46.

Las medidas de desempeño de la cola M/M/1 mostradas en la Tabla 1, nos revelan otros tres ejemplos de la ley de Little:

$$\begin{aligned} E[N] &= \lambda E[T] \\ E[Q] &= \lambda E[W] \\ E[X] &= \lambda(1/\mu) \end{aligned}$$

En el estudio de otras colas con llegadas Poisson y tiempos de servicio exponenciales (colas M/M), tendremos oportunidad de usar ampliamente la ley de Little para hallar diferentes medidas de desempeño. Sin embargo, enseguida mostraremos algunos ejemplos de aplicación de la ley de Little que demuestran su gran aplicabilidad en modelos muy generales.

139. Algunos ejemplos de aplicación de la Ley de Little

Consideremos una red de comunicaciones a la que llegan paquetes a través de n nodos de entrada, con tasas de llegada $\lambda_1, \lambda_2, \dots, \lambda_n$, respectivamente. Si en promedio cada paquete tarda $E[T]$ segundos en el sistema, sabemos que el número promedio de paquetes en la red es $E[N] = E[T] \sum_{i=1}^n \lambda_i$. Más aún, si los paquetes que ingresan a través del nodo i tardan en promedio $E[T_i]$ segundos en el sistema, el número promedio de paquetes en el sistema que entraron por el nodo i es $E[N_i] = E[T_i] \lambda_i$. Como $E[N] = \sum_{i=1}^n E[N_i]$, llegamos a la forma natural de calcular un promedio ponderado: $E[T] = (\sum_{i=1}^n E[T_i] \lambda_i) / (\sum_{i=1}^n \lambda_i)$. Nótese que estos resultados son válidos independientemente de las distribuciones de los tiempos entre llegadas, de los tiempos de servicio, de las disciplinas de servicio, de la administración de la memoria en los buffers, del número de nodos, de los procesos de enrutamiento, etc.

Consideremos ahora un solo nodo con un único enlace de salida al que llegan paquetes con tiempos determinísticos entre llegadas, a una tasa de λ paquetes por segundo. El tiempo de transmisión de cada paquete es determinístico, igual a una fracción $0 < \rho < 1$ de los tiempos entre llegadas, $1/\mu = \rho/\lambda$ constante para todos los paquetes. Hasta aquí se trata de una cola D/D/1 como la analizada en la definición 134. Sin embargo, ahora consideramos un tiempo fijo de propagación y procesamiento que sufre cada paquete, t_0 . Como los paquetes llegan con tiempo determinísticos entre llegadas y $\rho < 1$, no hay tiempo de espera en cola, por lo que el tiempo de permanencia de cada paquete en el sistema es $T = \rho/\lambda + t_0$. De acuerdo con la ley de Little, el número promedio de paquetes en el sistema es $N = \lambda T = \rho + \lambda t_0$. Por supuesto, éste no es el promedio de una variable aleatoria sino el nivel DC de una función determinística, $N(t)$.

Como tercer ejemplo, consideramos un protocolo de transporte que usa una ventana de hasta V paquetes para cada sesión, donde almacena aquellos paquetes que ya han sido transmitidos pero aún esperan ser reconocidos. El número de paquetes por sesión dentro del nivel de transporte siempre será menor o igual a V , por lo que la ley de Little afirma que $V \geq \lambda T$, donde T es el retardo promedio de cada paquete en el buffer de transporte y λ es la tasa de llegadas desde el nivel de aplicación. Si empieza a aparecer congestión, aumentará T hasta que la desigualdad se vuelva una igualdad. A partir de ese momento, λ debe decrementarse si T continúa incrementándose para mantener constante el producto λT , lográndose así el control de flujo deseado. Supongamos ahora que, en esas circunstancias, λ llegó al valor máximo que la red puede aceptar. ¿Tiene algún sentido aumentar la ventana para aliviar la congestión? No, porque, con λ constante, sólo se logrará aumentar el retardo.

Consideremos ahora un sistema G/G/K/M que se encuentra saturado (la tasa de llegadas es infinita o, mejor aún, se trata de un sistema cerrado en el que cada paquete que sale es inmediatamente

reemplazado por un nuevo paquete). Como el número de clientes en el sistema es M , según la ley de Little, el tiempo de permanencia en el sistema es, en promedio, $E[T] = M/\gamma$, donde γ es el caudal, esto es, la tasa efectiva de paquetes que se admiten por segundo. Si el tiempo promedio de servicio es $1/\mu$ y los K servidores permanecen ocupados, Little nos dice que $\gamma = K\mu$, por lo que el retardo normalizado es $\mu E[T] = M/K$.

Consideremos el mismo caso anterior sin saturación, esto es, llegan nuevos clientes a una tasa λ y pueden ser bloqueados al encontrar el sistema lleno, lo cual ocurrirá con una probabilidad P_B . En este caso, el número promedio de servidores ocupados puede ser menor a K , digamos $E[X]$ y, de acuerdo con Little, $E[X] = (1 - P_B)\rho$, con $\rho = \lambda/\mu$, de donde podemos encontrar $P_B = 1 - E[X]/\rho$. Como $E[X] \leq K$, hemos encontrado una cota inferior para la probabilidad de bloqueo: $P_B \geq 1 - K/\rho$.

Ahora miraremos un “sistema de sondeo”. Supongamos que un nodo con una sola línea de salida atiende a m usuarios (o m tipos de servicio diferentes) en ciclos periódicos –round robin–. En cada ciclo se atiende cierto número de paquetes del usuario 1, seguido por algunos paquetes del usuario 2 y así sucesivamente hasta finalmente atender algunos paquetes del usuario m . Hay un tiempo de sobrecarga de duración promedio A_i mientras se conmuta del usuario i al siguiente. La tasa promedio de llegadas de los paquetes del usuario i es λ_i y el tiempo promedio de servicio de un paquete del usuario i es $1/\mu_i$. La ley de Little nos dice que la fracción de tiempo que el servidor pasa atendiendo paquetes del usuario i es $\rho_i = \lambda_i/\mu_i$. El intervalo de sobrecarga para transitar del usuario i a $i+1$ se puede considerar como otro paquete con tiempo de transmisión A_i , cuya tasa de llegadas es $1/L$, siendo L la duración promedio de un ciclo. Little nos dice que la fracción de tiempo ocupado en “transmitir” estos paquetes es A/L , donde $A = \sum_{i=1..m} A_i$. Claramente, $1 = (A/L) + \sum_{i=1..m} \rho_i$, de donde podemos obtener la duración promedio de un ciclo,

$$L = \frac{A}{1 - \sum_{i=1}^m \rho_i}$$

140. Otras colas M/M

Cuando las llegadas forman un proceso de Poisson y los tiempos de servicio son exponenciales e independientes, la falta de memoria hace que el estado del sistema obedezca a una cadena de Markov en tiempo continuo, típicamente un proceso de nacimiento y muerte semejante al de la Figura 142. En estos casos, como con la cola M/M/1 de la definición 136, el análisis es muy metódico:

- Dibujar el diagrama de transición entre estados
- Plantear las ecuaciones de balance de flujo
- Normalizar para encontrar $p_n = P[N=n]$ (de aquí surgen las condiciones de estabilidad)
- Hallar las estadísticas de interés a partir de la distribución p_n (usando el principio PASTA, la ley de Little y otras consideraciones particulares de cada sistema)

a. M/M/1/K

En este sistema el cupo está limitado a K paquetes, de manera que si un paquete encuentra a su llegada que ya hay K paquetes en el sistema (uno en servicio y $K-1$ esperando en cola), deberá retirarse por falta de cupo. Al aplicar a este sistema los tres primeros pasos del proceso de análisis descrito antes, obtenemos los siguientes resultados:

Diagrama de transición entre estados	
Ecuaciones de balance de flujo	$\lambda p_0 = \mu p_1 \Rightarrow p_1 = \rho p_0, \quad \lambda p_1 = \mu p_2 \Rightarrow p_2 = \rho p_1 = \rho^2 p_0, \dots$ $\lambda p_{K-1} = \mu p_K \Rightarrow p_K = \rho p_{K-1} = \rho^K p_0 \Rightarrow p_n = \rho^n p_0, \quad 0 \leq n \leq K$
Normalización	$1 = \sum_{n=0}^K p_n = p_0 \sum_{n=0}^K \rho^n = \begin{cases} p_0(K+1) & \rho=1 \\ p_0 \frac{1-\rho^{K+1}}{1-\rho} & \rho \neq 1 \end{cases} \Rightarrow p_0 = \begin{cases} \frac{1}{K+1} & \rho=1 \\ \frac{1-\rho}{1-\rho^{K+1}} & \rho \neq 1 \end{cases} \Rightarrow p_n = \begin{cases} \frac{1}{K+1} & \rho=1 \\ \frac{\rho^n(1-\rho)}{1-\rho^{K+1}} & \rho \neq 1 \end{cases}, 0 \leq n \leq K$
Promedio	$E[N] = \sum_{n=0}^K n p_n = \frac{1-\rho}{1-\rho^{K+1}} \sum_{n=0}^K n \rho^n = \frac{\rho}{1-\rho} \left[1 - (K+1) \frac{\rho^K(1-\rho)}{1-\rho^{K+1}} \right] = \frac{\rho}{1-\rho} \cdot [1 - (K+1)p_K]$

La suma de la normalización se obtuvo así:

$$\sum_{n=0}^K \rho^n = 1 + \rho + \rho^2 + \dots + \rho^K = 1 + \rho(1 + \rho + \rho^2 + \dots + \rho^{K-1}) = 1 + \rho \left(\sum_{n=0}^{K-1} \rho^n - \rho^K \right) \Rightarrow (1-\rho) \sum_{n=0}^K \rho^n = 1 - \rho^{K+1} \Rightarrow \sum_{n=0}^K \rho^n = \frac{1-\rho^{K+1}}{1-\rho}$$

De manera que la convergencia se asegura para cualquier $\rho < \infty$. Es interesante notar que, si $\rho=1$, la distribución de probabilidad de estado es uniforme, $p_k = 1/(K+1)$, $k=0,1,\dots,K$; si $\rho < 1$, la probabilidad de estado disminuye con k y, si $\rho > 1$, la probabilidad de estado aumenta con k , como muestra la Figura 146. En los extremos, $\rho=0$ implica $p_0=1$ y $\rho=\infty$ implica $p_K=1$.

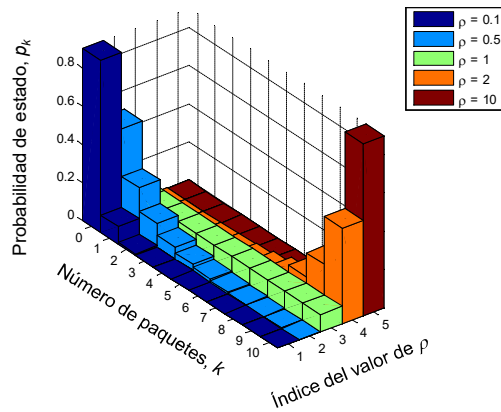


Figura 146. Distribución del número de paquetes en el sistema para una cola M/M/1/10

La suma del promedio se obtuvo así:

$$\sum_{n=0}^K n \rho^n = \rho \sum_{n=0}^K n \rho^{n-1} = \rho \sum_{n=0}^K \frac{d}{d\rho} \rho^n = \rho \frac{d}{d\rho} \sum_{n=0}^K \rho^n = \rho \frac{d}{d\rho} \left(\frac{1-\rho^{K+1}}{1-\rho} \right) = \rho \left[\frac{1-\rho^{K+1}}{(1-\rho)^2} - \frac{(K+1)\rho^K}{1-\rho} \right] = \frac{\rho}{1-\rho} \left[\frac{1-\rho^{K+1}}{1-\rho} - (K+1)\rho^K \right]$$

Aquí aparecen dos nuevas medidas de desempeño. Primero, como la fracción de tiempo que el sistema permanece completamente lleno es p_K , el principio PASTA nos permite saber que ésta es la misma probabilidad de que, cuando un paquete llegue, no encuentre cupo en el sistema. Le llamamos la Probabilidad de Bloqueo, P_B :

$$P_B = \begin{cases} \frac{1}{K+1} & \rho = 1 \\ \frac{\rho^K (1-\rho)}{1-\rho^{K+1}} & \rho \neq 1 \end{cases}$$

De donde surge una expresión más significativa para $E[N]$:

$$E[N] = \frac{\rho}{1-\rho} \cdot [1 - (K+1)P_B]$$

Segundo, si de los λ paquetes que llegan por segundo se pierde por falta de cupo una fracción P_B , la tasa efectiva de paquetes que se aceptan por segundo es

$$\Gamma = \lambda(1-P_B)$$

que se conoce como el caudal del sistema pues, efectivamente, es la misma tasa efectiva de paquetes que se transmiten por segundo:

$$\Gamma = \mu(1-p_0)$$

En efecto, si $\lambda = \mu$, la igualdad de estas dos tasas efectivas se reduce a $\lambda K / (K+1) = \mu K / (K+1)$. Pero si $\lambda \neq \mu$, obtenemos

$$\Gamma = \lambda \left[1 - \frac{\rho^K (1-\rho)}{1-\rho^{K+1}} \right] = \lambda \left[\frac{1-\rho^{K+1} - \rho^K (1-\rho)}{1-\rho^{K+1}} \right] = \lambda \left[\frac{1-\rho^K}{1-\rho^{K+1}} \right]$$

$$\Gamma = \mu \left[1 - \frac{1-\rho}{1-\rho^{K+1}} \right] = \mu \left[\frac{1-\rho^{K+1} - (1-\rho)}{1-\rho^{K+1}} \right] = \mu \left[\rho \frac{1-\rho^K}{1-\rho^{K+1}} \right] = \lambda \left[\frac{1-\rho^K}{1-\rho^{K+1}} \right]$$

Esto confirma la estabilidad del sistema para cualquier $\rho < \infty$: La tasa efectiva de entrada es la misma tasa efectiva de salida.

Para encontrar otras medidas de desempeño podemos usar la ley de Little. Al usar la ley de Little es importante notar que, cuando la expresamos como $N = \lambda W$, λ se refiere a la tasa efectiva de llegadas, pues los paquetes que no son aceptados no pueden afectar el retardo de los que sí son aceptados. En efecto, las gráficas de la Figura 145 muestran la cuenta de paquetes que ingresan al sistema. En la nomenclatura de las colas M/M que estamos viendo, la ley de Little debería escribirse $N = \Gamma T$, donde Γ es el caudal del sistema. Al utilizar la ley de little, el tiempo promedio en el sistema es $E[T] = E[N]/\Gamma$ y el número promedio de paquetes en servicio es $E[X] = \Gamma/\mu$; la diferencia nos da el número promedio de paquetes en cola, $E[Q] = E[N] - E[X]$, y, de nuevo, la ley de Little nos da el tiempo promedio de espera en la cola, $E[W] = E[Q]/\Gamma$. La Tabla 2 resume las medidas de desempeño de la cola M/M/1/K, las cuales se grafican en la Figura 147.

Tabla 2. Medidas de desempeño de la cola M/M/1/K

Medida	Significado
$P_B = \frac{\rho^K (1-\rho)}{1-\rho^{K+1}}$	Probabilidad de que un paquete sea rechazado
$E[N] = \frac{\rho}{1-\rho} \cdot [1 - (K+1)P_B]$	Número promedio de paquetes en el sistema
$\Gamma = \lambda(1-P_B)$	Caudal (tasa efectiva de paquetes transmitidos por segundo)
$E[T] = E[N]/\Gamma$	Tiempo promedio de permanencia en el sistema
$E[X] = \Gamma/\mu$	Número promedio de paquetes en servicio
$E[Q] = E[N] - E[X]$	Número promedio de paquetes en la cola
$E[W] = E[Q]/\Gamma$	Tiempo promedio de espera en la cola

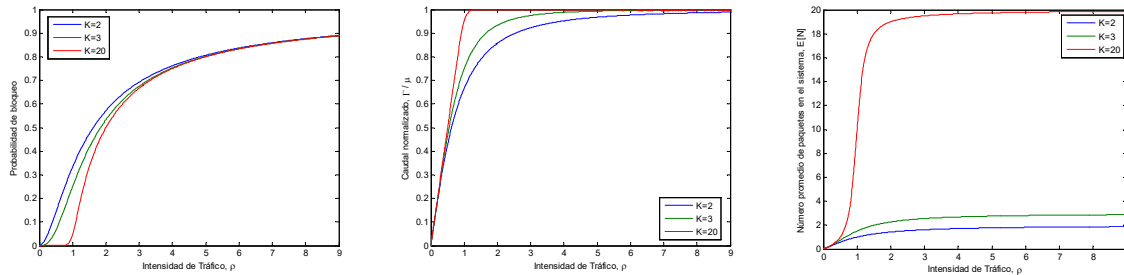


Figura 147. Algunas medidas de desempeño de la cola M/M/1/K, con $K=2,3,20$

La primera gráfica muestra la probabilidad de bloqueo, que tiende a uno a medida que ρ tiende a ∞ . Cuando $\rho=1$ se obtiene $P_B = 1/(K+1)$, de manera que, con K muy grande, la probabilidad de bloqueo permanece pequeña si $\rho < 1$ y, en consecuencia, el comportamiento de la cola M/M/1/K no es muy distinto del de la cola M/M/1 para valores pequeños de ρ , como se aprecia para el caso $K=20$ en la segunda y tercera gráfica de la Figura 147. Tan parecidos llegan a ser los comportamientos de estas dos colas con $\rho < 1$, que una buena ecuación de diseño para determinar el tamaño de la cola que se necesita para obtener una probabilidad de bloqueo dada (con $\rho < 1$) es usando la probabilidad de ocupación de M/M/1:

$$P_B = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}} \approx (1-\rho)\rho^K \Big|_{\rho^{K+1} \ll 1} \Rightarrow K \approx \left\lceil \frac{\log(P_B) - \log(1-\rho)}{\log(\rho)} \right\rceil$$

Por ejemplo, si con $\rho=0.8$ se desea $P_B \leq 0.001$, necesitaríamos $K = 24$ cupos en el sistema. Efectivamente, aplicando la ecuación de M/M/1/24 con $\rho=0.8$ obtenemos $P_B = 0.00095$.

b. M/M/K/K

Al igual que en la cola anterior, en esta el cupo está limitado a K clientes, aunque también hay K servidores homogéneos (todos tienen la misma capacidad y prestan el mismo servicio). Esto quiere decir que si un cliente es admitido en el sistema es porque hay un servidor disponible para atenderlo. Como los clientes que no encuentran un servidor disponible son rechazados, en este sistema no hay una cola de clientes en espera de servicio. Cuando hay n clientes en el sistema, cada uno está siendo atendido por un servidor de manera que la tasa total de servicio es $n\mu$ (tasa de servicio dependiente del estado). En el proceso de análisis usamos $A = \lambda/\mu$ como otra medida de intensidad de tráfico (tasa promedio de llegadas por tiempo promedio de servicio) en vez de $\rho = \lambda/K\mu$ (demanda sobre capacidad). El procedimiento estándar de análisis conduce a los siguientes resultados:

Diagrama de transición entre estados	
Ecuaciones de balance de flujo	$\lambda p_0 = \mu p_1 \Rightarrow p_1 = A p_0, \quad \lambda p_1 = 2\mu p_2 \Rightarrow p_2 = (A/2)p_1 = (A^2/2)p_0, \dots$ $\lambda p_{K-1} = K\mu p_K \Rightarrow p_K = (A/K)p_{K-1} = (A^K/K!)p_0 \Rightarrow p_n = (A^n/n!)p_0, \quad 0 \leq n \leq K$

Normalización	$1 = \sum_{n=0}^K p_n = p_0 \sum_{n=0}^K A^n / n! \Rightarrow p_0 = \frac{1}{\sum_{n=0}^K A^n / n!} \Rightarrow p_n = \frac{A^n / n!}{\sum_{m=0}^K A^m / m!}, 0 \leq n \leq K$
Promedio	$E[N] = \sum_{n=1}^K np_n = \frac{\sum_{n=1}^K nA^n / n!}{\sum_{m=0}^K A^m / m!} = \frac{A \sum_{n=1}^K A^{n-1} / (n-1)!}{\sum_{m=0}^K A^m / m!} = A \frac{\sum_{n=0}^{K-1} A^n / n!}{\sum_{m=0}^K A^m / m!} = A \frac{\left(\sum_{n=0}^K A^n / n!\right) - A^K / K!}{\sum_{m=0}^K A^m / m!} = A(1 - p_K)$

El término p_K de la expresión de $E[N]$, por el principio PASTA, es exactamente la probabilidad de bloqueo, que es la famosa **fórmula B de Erlang**:

$$P_B = \frac{A^K / K!}{\sum_{m=0}^K A^m / m!}$$

En efecto, este modelo M/M/K/K es una abstracción apropiada para representar un conmutador de circuitos con K canales y un número infinito de abonados. Como el tiempo promedio de permanencia en el sistema es el mismo tiempo promedio de servicio y el tiempo de espera en cola es cero (¡no hay cola!), las demás medidas resultan inmediatas usando la Ley de Little:

Tabla 3. Medidas de desempeño de la cola M/M/K/K

Medida	Significado
$P_B = \frac{A^K / K!}{\sum_{m=0}^K A^m / m!}$	Probabilidad de que un paquete sea rechazado
$E[N] = A \cdot (1 - P_B)$	Número promedio de paquetes en el sistema
$\Gamma = \lambda(1 - P_B) = \mu E[N]$	Caudal (tasa efectiva de paquetes transmitidos por segundo)
$E[T] = 1/\mu$	Tiempo promedio de permanencia en el sistema
$E[X] = \Gamma/\mu = E[N]$	Número pro medio de paquetes en servicio
$E[Q] = E[N] - E[X] = 0$	Número promedio de paquetes en la cola
$E[W] = E[Q]/\Gamma = 0$	Tiempo promedio de espera en la cola

La siguiente figura muestra la probabilidad de bloqueo en función de la intensidad de tráfico A , para diferentes valores del número de circuitos K .

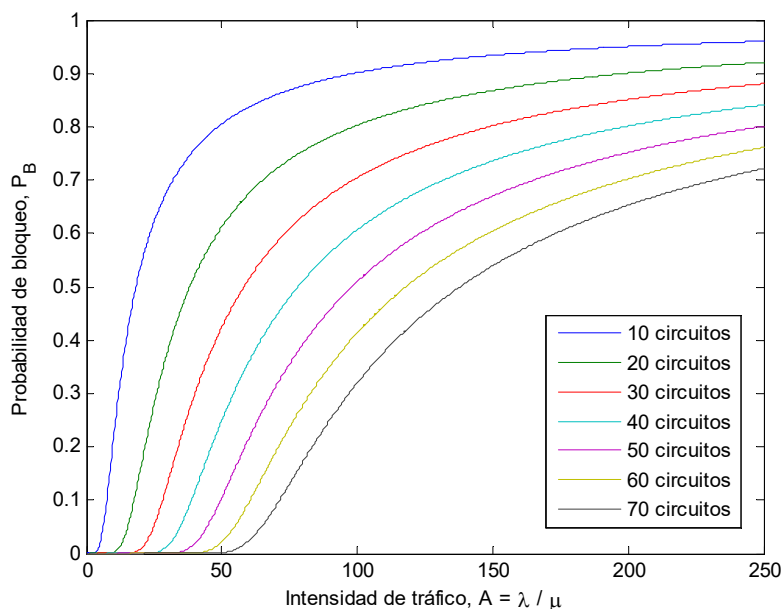


Figura 148. Probabilidad de bloqueo según la fórmula B de Erlang

Al evaluar la fórmula B de Erlang tal como se mostró anteriormente (calculando los factoriales y los exponenciales) se llega rápidamente a sobreflujos. Para elaborar la figura anterior multiplicamos por $K!$ y dividimos por A^K tanto en el numerador como en el denominador, luego hacemos el cambio de variable $n=K-m$ en el índice de la suma para obtener

$$P_B = \frac{1}{1 + \sum_{n=1}^K \frac{K!}{(K-n)! A^n}} = \frac{1}{1 + K/A + K(K-1)/A^2 + \dots + K!/A^K}$$

Esta expresión sugiere una implementación eficiente que evita los sobreflujos:

```

Producto = 1
Suma = 1
Para n desde 1 hasta K
    Producto = Producto*(K-n+1)/A
    Suma = Suma + Producto
Siguiendo n
PB = 1/Suma
    
```

c. M/M/K

En este sistema hay K servidores homogéneos (todos tienen la misma capacidad y prestan el mismo servicio), con un cupo infinito. El paquete que esté de primero en la cola será atendido por cualquiera de los enlaces que se desocupe primero.

Diagrama de transición entre estados	
Ecuaciones de balance de flujo	$\lambda p_{n-1} = \begin{cases} n\mu p_n & n < K \\ K\mu p_n & n \geq K \end{cases} \Rightarrow p_n = \begin{cases} p_0 \frac{(K\rho)^n}{n!} & n < K \\ p_0 \frac{K^K \rho^n}{K!} & n \geq K \end{cases}, \rho = \frac{\lambda}{K\mu} < 1$
Normalización	$\sum_{n=0}^{\infty} p_n = p_0 \left[\sum_{n=0}^{K-1} \frac{(K\rho)^n}{n!} + \sum_{n=K}^{\infty} \frac{K^K \rho^n}{K!} \right] = 1 \Rightarrow p_0 = \frac{1}{\frac{(K\rho)^K}{K!(1-\rho)} + \sum_{n=0}^{K-1} \frac{(K\rho)^n}{n!}}$
Promedio	$E[N] = \sum_{n=0}^{\infty} np_n = K\rho + \frac{\rho}{1-\rho} P_Q \text{ donde } P_Q = \sum_{n=K}^{\infty} p_n = p_0 \frac{(K\rho)^K}{K!(1-\rho)}$

En la normalización usamos el siguiente desarrollo:

$$\sum_{n=K}^{\infty} \frac{K^K \rho^n}{K!} = \frac{K^K}{K!} \sum_{n=K}^{\infty} \rho^n = \frac{(\rho K)^K}{K!} \sum_{n=0}^{\infty} \rho^n = \frac{(\rho K)^K}{K!(1-\rho)}$$

Esta expresión aparece al calcular la probabilidad de que todos los servidores estén ocupados que, de acuerdo con el principio PASTA, es igual a la probabilidad de que un paquete que llega deba esperar en cola:

$$P_Q = \sum_{n=K}^{\infty} p_n = p_0 \frac{(K\rho)^K}{K!} \sum_{n=0}^{\infty} \rho^n = p_0 \frac{(K\rho)^K}{K!(1-\rho)}$$

donde p_0 se obtuvo en el paso de normalización. Esta última expresión para P_Q se conoce como la **fórmula C de Erlang**. Las expresiones cerradas ya no lucen tan bonitas, pero se simplifican si las ponemos en términos de P_Q . Por ejemplo, el número promedio de paquetes en la cola se puede encontrar así:

$$E[Q] = \sum_{n=K}^{\infty} (n-K)p_n = \sum_{n=0}^{\infty} np_{n+K} = p_0 \frac{(K\rho)^K}{K!} \sum_{n=0}^{\infty} n\rho^n$$

Aquí es conveniente recordar la fórmula C de Erlang:

$$E[Q] = p_0 \frac{(K\rho)^K}{K!} \sum_{n=0}^{\infty} n\rho^n = P_Q (1-\rho) \sum_{n=0}^{\infty} n\rho^n$$

Y la derivación que hicimos al estudiar la cola M/M/1 (definición 136):

$$\sum_{n=0}^{\infty} n\rho^n = \rho \sum_{n=0}^{\infty} n\rho^{n-1} = \rho \sum_{n=0}^{\infty} \frac{d}{d\rho} \rho^n = \rho \frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^n = \rho \frac{d}{d\rho} \frac{1}{1-\rho} = \frac{\rho}{(1-\rho)^2}$$

Para llegar a

$$E[Q] = p_0 \frac{(K\rho)^K}{K!} \sum_{n=0}^{\infty} n\rho^n = P_Q \frac{\rho}{1-\rho}$$

Por el teorema de Little, el tiempo promedio que un paquete permanece en cola en espera de servicio es

$$E[W] = \frac{E[Q]}{\lambda} = P_0 \frac{\rho}{\lambda(1-\rho)} = P_0 \frac{1}{K\mu - \lambda}$$

Si añadimos el tiempo de servicio, obtenemos el tiempo promedio de permanencia en el sistema:

$$E[T] = \frac{1}{\mu} + E[W] = \frac{1}{\mu} + P_0 \frac{1}{K\mu - \lambda}$$

Podemos aplicar nuevamente la ley de Little para obtener

$$E[N] = \lambda E[T] = \frac{\lambda}{\mu} + P_0 \frac{\lambda}{K\mu - \lambda} = K\rho + P_0 \frac{\rho}{1-\rho}$$

En resumen,

Tabla 4. Medidas de desempeño de la cola M/M/K

Medida	Significado
$P_0 = \frac{1}{(K\rho)^K + \sum_{n=0}^{K-1} \frac{(K\rho)^n}{n!}}$	Probabilidad de que el sistema esté desocupado
$P_0 = p_0 \frac{(K\rho)^K}{K!(1-\rho)}$	Probabilidad de que un paquete deba esperar en cola
$E[N] = K\rho + P_0 \frac{\rho}{1-\rho}$	Número promedio de paquetes en el sistema
$E[Q] = P_0 \frac{\rho}{1-\rho}$	Número promedio de paquetes en la cola
$E[X] = K\rho$	Número promedio de paquetes en servicio
$E[T] = \frac{1}{\mu} + P_0 \frac{1}{K\mu - \lambda}$	Tiempo promedio de permanencia en el sistema
$E[W] = P_0 \frac{1}{K\mu - \lambda}$	Tiempo promedio de espera en la cola
$P_B = 0$	Probabilidad de que un paquete sea rechazado
$I = \lambda$	Caudal (tasa efectiva de paquetes transmitidos por segundo)

Con los resultados de las colas M/M/K y M/M/1 podemos hacernos una pregunta sencilla pero muy significativa: Al multiplexar paquetes sobre un enlace de alta velocidad ¿Es mejor usarlo como un canal veloz o como K canales lentos? Para responder, consideremos el caso K=2, donde cada uno de los canales tiene una capacidad de μ paquetes por segundo, y definamos $\rho = \lambda/2\mu$, pues ahora la capacidad total es 2μ paquetes por segundo. Los resultados de la **Tabla 4** muestran que

$$p_0 = \frac{1}{\frac{(2\rho)^2}{2(1-\rho)} + 1 + 2\rho} = \frac{1-\rho}{1+\rho} \quad \text{y} \quad P_Q = p_0 \frac{(2\rho)^2}{2(1-\rho)} = \frac{2\rho^2}{1+\rho}$$

De manera que el retardo promedio de cada paquete en el sistema es

$$E[T] = \frac{1}{\mu} + \frac{2\rho^2}{1+\rho} \frac{1}{2\mu-\lambda} = \frac{1}{\mu} + \frac{2\rho^2}{1+\rho} \frac{1/2\mu}{1-\rho} = \frac{1}{\mu} \left(1 + \frac{\rho^2}{1-\rho^2} \right) = \frac{1}{\mu} \frac{1}{1-\rho^2}$$

Queremos comparar este resultado con el retardo promedio en el sistema de la cola M/M/1, donde $E[T] = 1/(\mu-\lambda)$ si tenemos un solo servidor de capacidad μ , o $E[T] = 1/(2\mu-\lambda)$ si tenemos un solo servidor del doble de la capacidad. Al normalizar con respecto al tiempo de transmisión, los retardos normalizados de cada cola son:

$$\mu E[T]_{M/M/2} = \frac{1}{1-\rho^2} \Big|_{\rho=\lambda/2\mu} \quad \mu E[T]_{M/M/1(\mu)} = \frac{1}{1-2\rho} \Big|_{\rho=\lambda/2\mu} \quad \mu E[T]_{M/M/1(2\mu)} = \frac{1/2}{1-\rho} \Big|_{\rho=\lambda/2\mu}$$

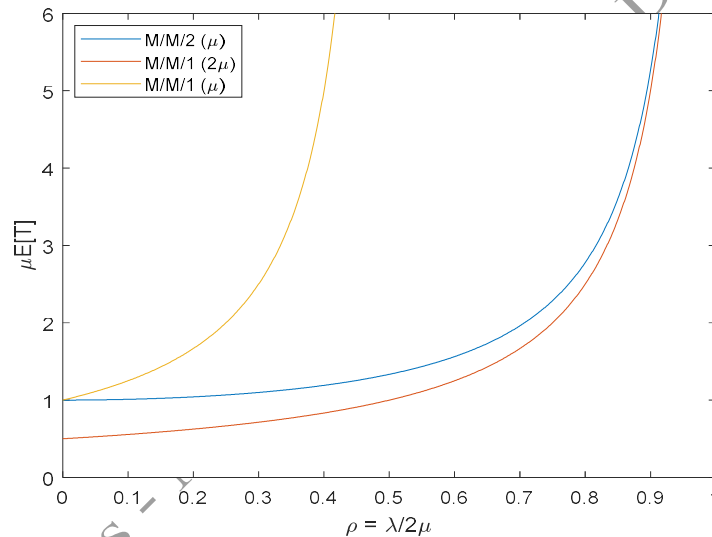


Figura 149. Comparación del retardo normalizado entre M/M/1 y M/M/2

Claramente, cuando añadimos un servidor a la cola M/M/1, mejoramos tanto el tiempo de espera como el caudal, pues ahora podemos atender el doble de la demanda con menores retardos. Sin embargo, si en vez de dos servidores usamos un servidor con el doble de la capacidad, obtenemos un mejor desempeño que con la cola M/M/2. En términos de desempeño es mejor aumentar la capacidad del canal que añadir más canales. Habría que tener en cuenta también criterios de robustez. Si uno de los servidores falla, con la cola M/M/2 todavía podremos prestar un servicio reducido.

d. M/M/∞

Este es el mismo caso anterior cuando K , el número de servidores, se hace infinitamente grande. En este caso no se necesita buffer, pues siempre habrá un servidor disponible para cualquier paquete que llegue al sistema. Aquí el concepto de demanda sobre capacidad carece de sentido, pues la capacidad es infinita. Sin embargo, definiremos el parámetro $\rho = \lambda/\mu$ para simplificar nuestras expresiones.

Diagrama de transición entre estados	
Ecuaciones de balance de flujo	$\lambda p_{n-1} = n\mu p_n \Rightarrow p_n = \frac{\lambda}{n\mu} p_{n-1} \Rightarrow p_n = \frac{\rho^n}{n!} p_0 \text{ donde } \rho = \frac{\lambda}{\mu}$
Normalización	$1 = \sum_{n=0}^{\infty} p_n = p_0 \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = p_0 e^{\rho} \Rightarrow p_0 = e^{-\rho} \Rightarrow p_n = \frac{\rho^n}{n!} e^{-\rho}$
Promedio	$E[N] = \sum_{n=0}^{\infty} np_n = \sum_{n=1}^{\infty} n \frac{\rho^n}{n!} e^{-\rho} = \rho \sum_{n=1}^{\infty} \frac{\rho^{n-1}}{(n-1)!} e^{-\rho} = \rho \sum_{m=0}^{\infty} \frac{\rho^m}{m!} e^{-\rho} = \rho$

El resultado es que la distribución de probabilidad de estado estable del número de paquetes en el sistema es Poisson con promedio ρ . Como ningún paquete pasa por el buffer, el tiempo de espera en cola es cero, el número promedio de paquetes en la cola es cero y el tiempo promedio de permanencia en el sistema es el tiempo de servicio, $E[T] = 1/\mu$. Este resultado, claro, se confirma mediante la regla de Little: $E[T] = E[N]/\lambda = 1/\mu$. No hay pérdidas y el caudal es la misma tasa de llegadas.

Tabla 5. Medidas de desempeño de la cola M/M/∞

Medida	Significado
$E[N] = \rho$	Número promedio de paquetes en el sistema
$E[Q] = 0$	Número promedio de paquetes en la cola
$E[X] = \rho$	Número promedio de paquetes en servicio
$E[T] = \frac{1}{\mu}$	Tiempo promedio de permanencia en el sistema
$E[W] = 0$	Tiempo promedio de espera en la cola
$P_B = 0$	Probabilidad de que un paquete sea rechazado
$I = \lambda$	Caudal (tasa efectiva de paquetes transmitidos por segundo)

Ciertamente parece muy irrealista pensar en un número infinito de servidores aunque podría corresponder a la idealización de un caso de abundancia de recursos. Sin embargo es interesante notar

que otros sistemas de colas pueden conducir a la misma distribución de estado estable pero con medidas de desempeño muy diferentes, como mostramos a continuación.

e. M/M/1 con control de flujo

En este caso consideramos que, cuando los clientes ven una larga fila, pueden preferir seguir derecho sin solicitar ningún servicio, esto es, el control de flujo es por desánimo de los clientes. Específicamente, cuando el número de paquetes en el sistema es n , la tasa de llegadas se reduce a $\lambda_n = \lambda/(n+1)$.

Diagrama de transición entre estados	
Ecuaciones de balance de flujo	$\frac{\lambda}{n} p_{n-1} = \mu p_n \Rightarrow p_n = \frac{\lambda}{n\mu} p_{n-1} \Rightarrow p_n = \frac{\rho^n}{n!} p_0 \text{ donde } \rho = \frac{\lambda}{\mu}$
Normalización	$1 = \sum_{n=0}^{\infty} p_n = p_0 \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = p_0 e^{\rho} \Rightarrow p_0 = e^{-\rho} \Rightarrow p_n = \frac{\rho^n}{n!} e^{-\rho}$
Promedio	$E[N] = \sum_{n=0}^{\infty} n p_n = \sum_{n=1}^{\infty} n \frac{\rho^n}{n!} e^{-\rho} = \rho \sum_{n=1}^{\infty} \frac{\rho^{n-1}}{(n-1)!} e^{-\rho} = \rho \sum_{m=0}^{\infty} \frac{\rho^m}{m!} e^{-\rho} = \rho$

Nótese que el proceso de análisis de esta cola, hasta la obtención de $E[N]$, resultó exactamente igual al análisis anterior de la cola $M/M/\infty$. Sin embargo, los dos sistemas son muy distintos en cuanto a las medidas de desempeño dada su diferencia dinámica. Primero, el caudal de este segundo sistema es $\gamma = \mu(1-p_0) = \mu(1-e^{-\rho})$. Segundo, ahora la cola si puede estar ocupada pues, de acuerdo con la ley de Little, $E[T] = E[N]/\gamma$. La diferencia entre el tiempo en el sistema y el tiempo de servicio es el tiempo promedio de espera en cola, $E[W] = E[T] - 1/\mu$. Y, nuevamente por Little, el número promedio de paquetes en la cola es $E[Q] = \gamma E[W]$. La siguiente tabla resume estos resultados.

Tabla 6. Medidas de desempeño de la cola M/M/1 con control de flujo

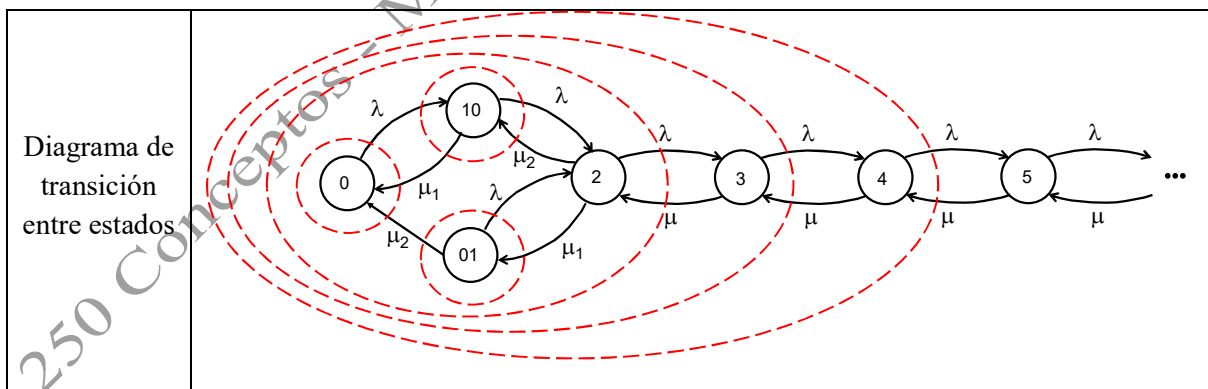
Medida	Significado
$E[N] = \rho$	Número promedio de paquetes en el sistema
$E[Q] = \rho + e^{-\rho} - 1$	Número promedio de paquetes en la cola
$E[X] = 1 - e^{-\rho}$	Número promedio de paquetes en servicio

$E[T] = \frac{1}{\mu} \frac{\rho}{1 - e^{-\rho}}$	Tiempo promedio de permanencia en el sistema
$E[W] = \frac{1}{\mu} \left(\frac{\rho}{1 - e^{-\rho}} - 1 \right)$	Tiempo promedio de espera en la cola
$P_B = 0$	Probabilidad de que un paquete sea rechazado
$\Gamma = \mu(1 - e^{-\rho})$	Caudal (tasa efectiva de paquetes transmitidos por segundo)

Insistamos en el interesante hecho de que las dos colas anteriores tienen exactamente la misma distribución de estado estable, pero sus medidas de desempeño son muy distintas.

f. M/M/2 con servidores no homogéneos

Como ejemplo de la cola M/M/K habíamos analizado el caso $K=2$ y pudimos comparar con dos casos M/M/1 semejantes. Ahora volveremos al caso M/M/2 pero supondremos que los servidores tienen tasas de servicio diferentes, μ_1 y μ_2 , con $\mu_1 > \mu_2$. Este caso es interesante porque el diagrama de transición entre estados debe considerar dos formas diferentes de tener un único paquete en el sistema, ya que dicho paquete puede estar siendo atendido por el enlace rápido o por el enlace lento, con un comportamiento diferente del sistema en cada caso. En efecto, suponemos que si un paquete llega y encuentra ambos servidores desocupados, ocupa el servidor rápido. Pero si el rápido está ocupado y el lento desocupado, se conforma con ocupar el servidor lento. Siendo $\mu = \mu_1 + \mu_2$ y $\rho = \lambda / \mu$, el análisis toma la siguiente forma:

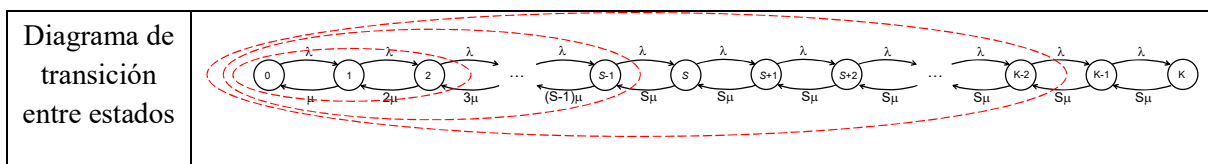


<p>Ecuaciones de balance de flujo</p>	$\begin{aligned} \lambda p_0 &= \mu_1 p_{10} + \mu_2 p_{01} \\ (\lambda + \mu_1) p_{10} &= \lambda p_0 + \mu_2 p_2 \\ (\lambda + \mu_2) p_{01} &= \mu_1 p_2 \\ \mu p_n &= \lambda p_{n-1}, \quad n = 2, 3, \dots \end{aligned} \Rightarrow \begin{aligned} p_{01} &= \frac{\lambda \rho}{\lambda + \mu_2 - \rho(\mu_1 - \mu_2)} p_0 \triangleq \alpha p_0 \\ p_{10} &= \frac{\lambda}{\mu_1} \frac{\lambda + \mu_2 - \rho \mu_1}{\lambda + \mu_2 - \rho(\mu_1 - \mu_2)} p_0 \triangleq \beta p_0 \\ p_2 &= \frac{\lambda + \mu_2}{\mu_1} p_{01} = \frac{\lambda + \mu_2}{\mu_1} \alpha p_0 \triangleq \gamma p_0 \\ p_n &= \rho^{n-2} p_2 = \gamma \rho^{n-2} p_0, \quad n = 3, 4, \dots \end{aligned}$
<p>Normalización</p>	$p_0 \left(1 + \alpha + \beta + \gamma \sum_{n=0}^{\infty} \rho^n \right) = 1 \quad \text{si } \rho < 1, \quad p_0 = \frac{1}{1 + \alpha + \beta + \frac{\gamma}{1 - \rho}}$
<p>Promedio</p>	$E[N] = p_{01} + p_{10} + p_2 \sum_{n=2}^{\infty} n \rho^{n-2} = p_0 \left(\alpha + \beta + \frac{\gamma}{\rho} \left(\sum_{n=0}^{\infty} n \rho^{n-1} - 1 \right) \right) = p_0 \left(\alpha + \beta + \frac{\gamma(2 - \rho)}{(1 - \rho)^2} \right)$ $E[N] = \frac{\gamma(2 - \rho) + (1 - \rho)^2(\alpha + \beta)}{(1 - \rho)(\gamma + (1 - \rho)(1 + \alpha + \beta))}$

Debemos confesar que las expresiones cerradas se hacen menos bonitas cuando el diagrama de transición entre estados ya no es un proceso de nacimiento y muerte como los de los casos anteriores. Más aún, el método de las ecuaciones de recurrencia obtenidas del diagrama de transición entre estados deja de ser útil cuando no se trata de procesos de nacimiento y muerte, por lo que pronto usaremos otros métodos basados en transformaciones o series geométricas matriciales. Por lo pronto, una simple verificación se consigue viendo el caso en que $\mu_1 = \mu_2$:

g. M/M/S/K

En este sistema hay S servidores homogéneos (todos tienen la misma capacidad y prestan el mismo servicio), con un cupo total en el sistema de K paquetes, $K \geq S$, de manera que pueden haber hasta $K - S$ paquetes esperando en cola su turno de servicio. Los paquetes que llegan cuando ya hay K paquetes en el sistema (S en servicio y $K - S$ en espera) serán rechazados por falta de cupo. El paquete que esté de primero en la cola será atendido por cualquiera de los enlaces que se desocupe primero.



<p>Ecuaciones de balance de flujo</p>	$\begin{aligned} \lambda p_{n-1} &= n\mu p_n, \quad n < S \\ \lambda p_{n-1} &= S\mu p_n, \quad S \leq n \leq K \end{aligned} \Rightarrow \begin{aligned} p_n &= \frac{\rho^n}{n!} p_0, \quad n < S \\ p_n &= \frac{\rho^n}{S^{n-S} S!} p_0, \quad S \leq n \leq K \end{aligned}$ <p style="text-align: center;">donde $\rho = \frac{\lambda}{\mu}$</p>
<p>Normalización</p>	$1 = \sum_{n=0}^K p_n = p_0 \left[\sum_{n=0}^{S-1} \frac{\rho^n}{n!} + \frac{\rho^S}{S!} \sum_{n=0}^{K-S} \left(\frac{\rho}{S}\right)^n \right] \Rightarrow p_0 = \frac{S!(S-\rho)}{S!(S-\rho) \sum_{n=0}^{S-1} \frac{\rho^n}{n!} + S\rho^S (1 - (\rho/S)^{K-S+1})}$ <p style="text-align: center;">Converge para $\rho < \infty$</p>
<p>Promedio</p>	$E[N] = \sum_{n=0}^K n p_n = \frac{\rho \sum_{n=0}^{S-2} \frac{\rho^n}{n!} + \frac{\rho^S}{S!} \sum_{n=0}^{K-S} (n+S) \left(\frac{\rho}{S}\right)^n}{\sum_{n=0}^{S-1} \frac{\rho^n}{n!} + \frac{\rho^S}{S!} \sum_{n=0}^{K-S} \left(\frac{\rho}{S}\right)^n}$

Se pueden buscar expresiones más concretas si hiciera falta, aunque ya las expresiones son fácilmente evaluables para cada par específico de valores de S y K y cada par de parámetros λ y μ . Lo importante es que ya debió quedar claro el procedimiento de análisis de desempeño de un sistema de cola M/M/... (llegadas Poisson y tiempos de servicio exponenciales e independientes).

141. La cola M/G/1

Esta cola tiene un único servidor y los clientes llegan según un proceso de Poisson con tasa λ paquetes/segundo, como en las anteriores colas. Sin embargo, los tiempos de servicio siguen una distribución general, con función de densidad de probabilidad $f_\tau(t)$, de la cual sólo conocemos su promedio ($1/\mu$ segundos por paquete) y su desviación estándar (σ_τ segundos por paquete). Suponemos, como antes, que la secuencia de tiempos de servicio forma un proceso IID –cada tiempo de servicio es una muestra independiente de todas la demás e idénticamente distribuida a cada una de las demás.

Considere el número de paquetes que deja en el sistema el j -ésimo paquete que sale, N_j . Si el paquete $j-1$ dejó el sistema vacío, el j -ésimo paquete encontró el enlace desocupado y, a su salida, sólo quedarán los paquetes que llegaron durante su servicio, v_j , esto es, $N_j = v_j$ si $N_{j-1} = 0$. Sin embargo, si el paquete anterior dejó $N_{j-1} > 0$ paquetes, el paquete j ya estaba en el sistema cuando salió el paquete $j-1$, de manera que $N_j = N_{j-1} - 1 + v_j$ si $N_{j-1} > 0$. Definiendo $u(n) = 1$ si $n > 0$ y $u(n) = 0$ si $n \leq 0$, podemos expresar ambas condiciones mediante la siguiente relación

$$N_j = N_{j-1} - u(N_{j-1}) + v_j \quad (8)$$

Cada uno de los términos de la ecuación anterior es una variable aleatoria a la que podemos aplicar el operador valor esperado para obtener

$$E[N_j] = E[N_{j-1}] - E[u(N_{j-1})] + E[v_j] \quad (9)$$

Si el sistema alcanza un estado estable, a la larga los valores esperados anteriores no dependerán en particular de los paquetes $j-1$ y j :

$$\lim_{j \rightarrow \infty} E[N_j] = E[N], \quad \lim_{j \rightarrow \infty} E[u(N_j)] = E[u(N)], \quad \lim_{j \rightarrow \infty} E[v_j] = E[v]$$

aplicando estos límites a la Ecuación (9) obtenemos $E[N] = E[N] - E[u(N)] + E[v]$ de manera que $E[u(N)] = E[v]$. Pero $E[u(N)]$ es la promedio de la ocupación del enlace, $1-p_0$, y $E[v]$ es el número promedio de llegadas durante el tiempo de servicio de un paquete, $\lambda/\mu = \rho$:

$$1 - p_0 = E[u(N)] = E[v] = \frac{\lambda}{\mu} \triangleq \rho \quad (10)$$

La Ecuación (10) es muy interesante, pero ya lo sabíamos y no es muy útil. En muchas ocasiones a un problema de modelamiento probabilístico se le da muchas vueltas hasta que inesperadamente aparece un paso simple que lo resuelve. Pollaczek y Khintchine, los dos matemáticos rusos que desarrollaron el modelo de la cola M/G/1, debieron sentir esa sensación cuando decidieron elevar al cuadrado cada lado de la Ecuación (8):

$$\begin{aligned} N_j^2 &= N_{j-1}^2 - 2N_{j-1}u(N_{j-1}) + 2N_{j-1}v_j + u^2(N_{j-1}) - 2u(N_{j-1})v_j + v_j^2 = \\ &= N_{j-1}^2 - 2N_{j-1} + 2N_{j-1}v_j + u(N_{j-1}) - 2u(N_{j-1})v_j + v_j^2 \end{aligned}$$

Para pasar de la primera a la segunda línea en la ecuación anterior simplemente notamos que, como $u(0)$ es 0 y $u(n)$ es 1 si $n > 0$, entonces $nu(n) = n$. Por otro lado, como los posibles valores de $u(n)$ son 0 ó 1, entonces $u^2(n) = u(n)$. Aplicando el operador de valor esperado y teniendo en cuenta la independencia de N_{j-1} y v_j ,

$$E[N_j^2] = E[N_{j-1}^2] - 2E[N_{j-1}] + 2E[N_{j-1}]E[v_j] + E[u(N_{j-1})] - 2E[u(N_{j-1})]E[v_j] + E[v_j^2]$$

Recurriendo nuevamente a la suposición de estabilidad, podemos aplicar valor esperado a la ecuación anterior para obtener

$$E[N^2] = E[N^2] - 2E[N] + 2E[N]E[v] + E[u(N)] - 2E[u(N)]E[v] + E[v^2]$$

Restando $E[N^2]$ a ambos lados, factorizando $E[N]$ y teniendo en cuenta la Ecuación (10)

$$2E[N](1 - \rho) = \rho - 2\rho^2 + E[v^2] = \rho - \rho^2 + (E[v^2] - E^2[v]) = \rho(1 - \rho) + \sigma_v^2$$

de donde podemos encontrar una expresión sencilla para $E[N]$:

$$E[N] = \frac{\rho}{2} + \frac{\sigma_v^2}{2(1 - \rho)} \quad (11)$$

Esta expresión está muy bien, excepto porque no conocemos la varianza del número de paquetes que llegan durante el tiempo de servicio de un paquete, σ_v^2 . Por definición,

$$\sigma_v^2 = \sum_{k=0}^{\infty} (k - \rho)^2 P[v = k]$$

La probabilidad $P[v=k]$ la podemos encontrar, como de costumbre, a partir del teorema de la probabilidad total (definición 19), condicionando en el valor que tome el tiempo de transmisión del paquete:

$$P[v = k] = \int_0^{\infty} P[v = k | \tau = t] f_{\tau}(t) dt$$

donde $P[v = k | \tau = t] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$. Entonces la varianza de v se puede expresar como $\lambda^2 \sigma_{\tau}^2 + \rho$

de acuerdo con la siguiente secuencia de pasos:

$$\begin{aligned} \sigma_v^2 &= \sum_{k=0}^{\infty} \left[(k - \rho)^2 \int_0^{\infty} P[v = k | \tau = t] f_{\tau}(t) dt \right] \\ &= \int_0^{\infty} f_{\tau}(t) \sum_{k=0}^{\infty} \left[(k - \rho)^2 P[v = k | \tau = t] \right] dt \\ \sigma_v^2 &= \int_0^{\infty} f_{\tau}(t) \sum_{k=0}^{\infty} \left[(k^2 - 2k\rho + \rho^2) \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right] dt \\ &= \int_0^{\infty} f_{\tau}(t) \left[\sum_{k=0}^{\infty} \left[k^2 \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right] - 2\rho \sum_{k=0}^{\infty} \left[k \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right] + \rho^2 \sum_{k=0}^{\infty} \left[\frac{(\lambda t)^k}{k!} e^{-\lambda t} \right] \right] dt \\ \sigma_v^2 &= \int_0^{\infty} f_{\tau}(t) \left[E[v^2 | \tau = t] - 2\rho E[v | \tau = t] + \rho^2 \right] dt \\ &= \int_0^{\infty} f_{\tau}(t) \left[\lambda^2 t^2 + \lambda t - 2\rho \lambda t + \rho^2 \right] dt \\ &= \lambda^2 E[\tau^2] + \lambda(1 - 2\rho) E[\tau] + \rho^2 \\ \sigma_v^2 &= \lambda^2 E[\tau^2] + \rho - \rho^2 = \lambda^2 (E[\tau^2] - E^2[\tau]) + \rho = \lambda^2 \sigma_{\tau}^2 + \rho \end{aligned}$$

Reemplazando $\sigma_v^2 = \lambda^2 \sigma_{\tau}^2 + \rho$ en la ecuación (11)

$$E[N] = \frac{\rho}{2} + \frac{\lambda^2 \sigma_{\tau}^2 + \rho}{2(1 - \rho)}$$

Haciendo las operaciones algebraicas necesarias para obtener la ecuación anterior con un factor $\rho/(1-\rho)$ que permita comparar con M/M/1, llegamos a

$$E[N] = \frac{\rho}{1 - \rho} \left(1 - \frac{\rho}{2} (1 - \mu^2 \sigma_{\tau}^2) \right) \quad (12)$$

donde $\mu^2 \sigma_{\tau}^2$ es el cuadrado del coeficiente de variación del tiempo de servicio de los paquetes. Esta es la fórmula de Pollaczek-Khintchine o fórmula P-K, escrita en una forma que permite comparar diferentes distribuciones del tiempo de servicio. Por ejemplo, supongamos que el tiempo de servicio es determinístico (todos los paquetes se demoran exactamente $1/\mu$ segundos en ser transmitidos).

$$M/D/1 : \sigma_{\tau}^2 = 0 \Rightarrow E_{M/D/1}[N] = \frac{\rho}{1-\rho} \left(1 - \frac{\rho}{2} \right)$$

Pero si los tiempos de servicio son exponenciales (el coeficiente de variación es 1) tenemos la fórmula ya conocida

$$M/M/1 : \sigma_{\tau}^2 = 1/\mu^2 \Rightarrow E_{M/M/1}[N] = \frac{\rho}{1-\rho}$$

Y si el tiempo de servicio está uniformemente distribuido entre 0 y $2/\mu$,

$$M/U/1 : \sigma_{\tau}^2 = 1/3\mu^2 \Rightarrow E_{M/U/1}[N] = \frac{\rho}{1-\rho} \left(1 - \frac{\rho}{3} \right)$$

Un caso particular es cuando los paquetes tienen una longitud de cola pesada, por ejemplo Pareto(a, b) –definición 44(d)–, cuya media y varianza son

$$\frac{1}{\mu} = \begin{cases} \frac{ab}{a-1} & a > 1 \\ \infty & a \leq 1 \end{cases}, \quad \sigma^2 = \begin{cases} \frac{ab^2}{(a-2)(a-1)^2} & a > 2 \\ \infty & a \leq 2 \end{cases}$$

Si escogemos como parámetros $a > 1$ y $b = (a-1)/a\mu$, el valor medio del tiempo de servicio será $1/\mu$, como esperamos. Con $1 < a \leq 2$, la varianza es infinita aunque la media sea $1/\mu$, así que $E[N]$ es infinito para cualquier $\rho > 0$. Pero si escogemos $a > 2$, el cuadrado del coeficiente de variación será $1/[a(a-2)]$:

$$M/P/1, a > 2 : \sigma_{\tau}^2 = 1/[a(a-2)\mu^2] \Rightarrow E_{M/P/1}[N] = \frac{\rho}{1-\rho} \left(1 - \frac{\rho}{2} \left(1 - \frac{1}{a(a-2)} \right) \right)$$

Se necesitaría $a = \infty$ para alcanzar el mismo promedio de M/D/1, pero basta con $a = 3$ para alcanzar el mismo promedio de M/U/1. Con $a = 1 + \sqrt{2}$ se alcanza el promedio de M/M/1, y a medida que a se acerca a 2 por la derecha, $E_{M/P/1}[N]$ se acerca a ∞ para cualquier $\rho > 0$. La Figura 150 muestra $E[N]$ en función de ρ para estas distribuciones.

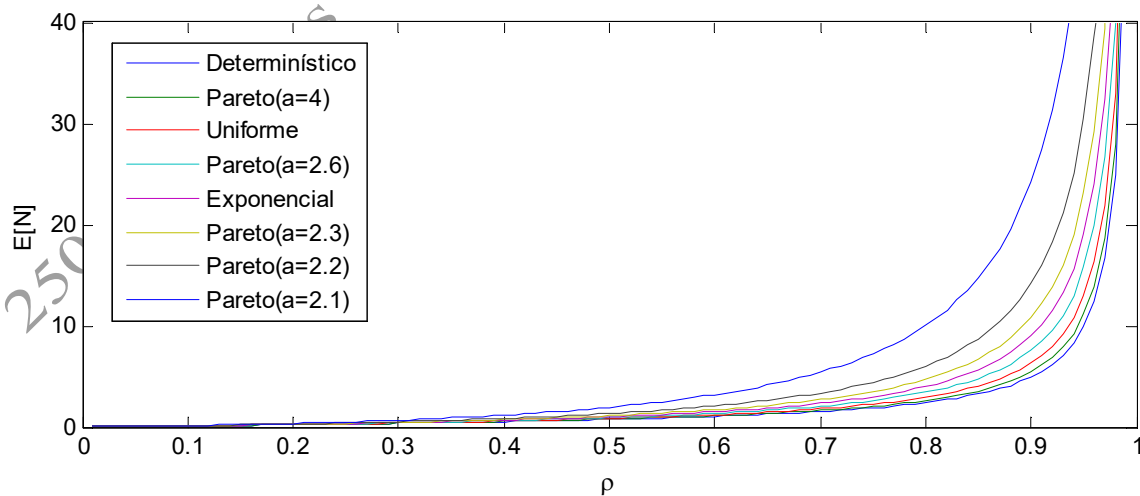


Figura 150. Número promedio de paquetes en algunos sistemas M/G/1

Veamos, por último, un ejemplo que muestra cómo se puede aprovechar la fórmula P-K este resultado M/G/1 en modelos de redes. Se trata del protocolo de retransmisión automática Go-Back-N, el cual ya se introdujo de manera simplificada en la definición 46(e). Mantendremos las ideas simplistas de que los paquetes se dañan con probabilidad p independiente de otros paquetes, que llegan según un proceso de Poisson con intensidad de λ paquetes por segundo y que toman un tiempo de transmisión constante de $1/\mu$ segundos. Como vimos, el número de transmisiones hasta que el paquete llegue correcto a su destino es k con probabilidad $p^{k-1}(1-p)$, para $k=1,2,3,\dots$. Sin embargo, en este caso cada error de transmisión implica tener que retransmitir nuevamente los n paquetes que se ubicaron en la ventana deslizante del transmisor mientras se esperaba el reconocimiento o el *time-out*, de manera que el tiempo de transmisión es $(1 + (k-1)n)/\mu$ con probabilidad $p^{k-1}(1-p)$ para $k=1,2,3,\dots$ o, equivalentemente, si contamos el número de retransmisiones,

$$P[T_{tx} = (1 + kn)/\mu] = p^k(1-p) \text{ para } k=0,1,2,3,\dots$$

El valor medio y el segundo momento de este tiempo de transmisión se encuentran fácilmente:

$$\mu E[T_{tx}] = \sum_{k=0}^{\infty} (1 + kn) p^k (1-p) = 1 + np(1-p) \sum_{k=0}^{\infty} kp^{k-1}$$

$$\mu^2 E[T_{tx}^2] = \sum_{k=0}^{\infty} (1 + kn)^2 p^k (1-p) = 1 + 2np(1-p) \sum_{k=0}^{\infty} kp^{k-1} + n^2 p^2 (1-p) \sum_{k=0}^{\infty} k^2 p^{k-2}$$

Donde

$$\sum_{k=0}^{\infty} kp^{k-1} = \sum_{k=0}^{\infty} \frac{d}{dp} p^k = \frac{d}{dp} \sum_{k=0}^{\infty} p^k = \frac{d}{dp} \frac{1}{1-p} = \frac{1}{(1-p)^2}$$

$$\sum_{k=0}^{\infty} k^2 p^{k-2} = \sum_{k=0}^{\infty} \left(\frac{d^2}{dp^2} p^k + kp^{k-2} \right) = \frac{d^2}{dp^2} \sum_{k=0}^{\infty} p^k + \frac{1}{p} \sum_{k=0}^{\infty} kp^{k-1} = \frac{d^2}{dp^2} \frac{1}{1-p} + \frac{1}{p(1-p)^2} = \frac{2}{(1-p)^3} + \frac{1}{p(1-p)^2} = \frac{1+p}{p(1-p)^3}$$

Esto es,

$$\mu E[T_{tx}] = 1 + \frac{np}{1-p}$$

$$\mu^2 E[T_{tx}^2] = 1 + \frac{2np}{1-p} + \frac{n^2 p(1+p)}{(1-p)^2}$$

La varianza del tiempo de transmisión es, entonces,

$$\mu^2 V[T_{tx}] = \frac{n^2 p}{(1-p)^2}$$

Por lo que el coeficiente de variación es $pn^2/(1+p(n-1))^2$. La fórmula P-K establece que



142. La cola M/G/1 con períodos de vacaciones

Un caso de mucho interés en diferentes aplicaciones es cuando, después de un período de ocupación, el servidor se toma un descanso durante un tiempo aleatorio. En este caso, un paquete que llegue y encuentre el sistema vacío podría tener que esperar el final del periodo de descanso antes de empezar

a ser transmitido. Si el servidor termina su período de descanso y el sistema sigue vacío, empezaría un nuevo período de descanso. Este esquema puede modelar muchos fenómenos típicos de redes de comunicaciones como la transmisión de información de control entre períodos de ocupación o la discretización del tiempo en el uso de diferentes canales FDM.

Los períodos sucesivos de vacaciones V_1, V_2, \dots son variables aleatorias idénticamente distribuidas independientes entre ellas e independientes de los tiempos entre llegadas y los tiempos de servicio de los paquetes. Las llegadas forman un proceso de Poisson y los tiempos de servicios también son idénticamente distribuidos e independientes con una distribución general. Este es un modelo extendido de la cola $M/G/1$, pero haremos el análisis de una manera diferente, considerando el tiempo residual para finalizar la transmisión actual o el tiempo de vacaciones actual.

(...)

143. Reservas y prioridades

(...)

144. La cola GI/GI/1

En esta cola sólo conocemos las medias y las varianzas de los tiempos entre llegadas ($1/\lambda, \sigma_a^2$) y de los tiempos de servicio ($1/\mu, \sigma_s^2$), pero no sus distribuciones.

Sea W_k el tiempo de espera en cola del k -ésimo paquete, S_k su tiempo de servicio y T_k el intervalo de tiempo que transcurre entre la llegada del k -ésimo paquete y la del $(k+1)$ -ésimo. Si $W_k + S_k < T_k$, el paquete $k+1$ encuentra el sistema vacío y su tiempo de espera es cero. En otro caso, el tiempo de espera del paquete $k+1$ es lo que le falta esperar desde que llega hasta que el paquete k salga del sistema, como muestra la Figura 151.

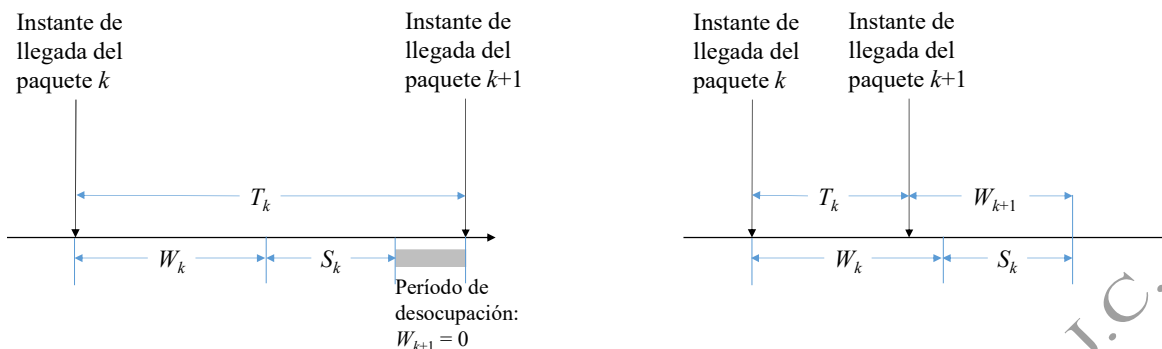


Figura 151. Tiempos de espera en cola de los paquetes k y $k+1$

Entonces,

$$W_{k+1} = \max(0, W_k + S_k - T_k)$$

El tiempo de desocupación del sistema entre los instantes de llegada de los paquetes k y $(k+1)$ es, entonces,

$$I_k = -\min(0, W_k + S_k - T_k)$$

(...)

145. Redes de Colas –El teorema de Burke

(...)

146. Redes de Colas –El teorema de Jackson

(...)

147. Integración de dos tipos de flujo

Hasta ahora hemos visto colas con un solo tipo de entrada, pero las redes de hoy deben integrar muchos tipos distintos de flujos con características y requerimientos muy diferentes entre ellos. Consideremos un flujo en tiempo real y un flujo elástico que comparten un enlace de salida. Los flujos en tiempo real, que pueden corresponder a tráfico de voz, audio o video, por ejemplo, requieren un

ancho de banda determinado ya que los paquetes se generan en la fuente a una tasa que debe conservarse durante su reproducción en el receptor. Tan importante es este requerimiento que no importa sacrificar las funciones de detección y corrección de errores con tal de que los paquetes lleguen a tiempo para su reproducción oportuna. Más aún, si un paquete del flujo en tiempo real llega después del momento en que debe reproducirse, será descartado a pesar de haber llegado a su destino sin errores. Los flujos elásticos, que puede corresponder a correo electrónico o transferencia de archivos, por ejemplo, no necesitan un ancho de banda específico, así que se pueden acomodar al que la red les ofrezca. En el flujo elástico los paquetes no se pueden perder y no pueden llegar con errores. Tan importante es este requerimiento que no importa sufrir retardos adicionales y variables debidos, por ejemplo, a retransmisiones de paquetes, con tal de que todos los paquetes lleguen a su destino sin errores, como se mostró en la Figura 64 y en la Figura 66. En contraste, la Figura 152 muestra cómo se procesa un flujo de tiempo real en el receptor. La primera línea muestra los instantes de transmisión de los paquetes, los cuales sufren retardos aleatorios en la red, como muestra la segunda línea. Para amortiguar esta variabilidad de los retardos, el receptor añade un retardo adicional al primer paquete con la esperanza de que aún los paquetes que sufran el máximo retardo en la red alcancen a llegar a tiempo para su reproducción oportuna. En la figura, los paquetes 0 a 3 se pueden reproducir a la tasa original pero, aún a pesar del amortiguador introducido por el receptor, el instante de reproducción del paquete 4 se presentó antes de que el paquete haya llegado. El receptor, entonces, prefiere descartar el paquete 4, a pesar de que llegó sin errores, antes que cambiar la tasa de reproducción para acomodar este paquete.

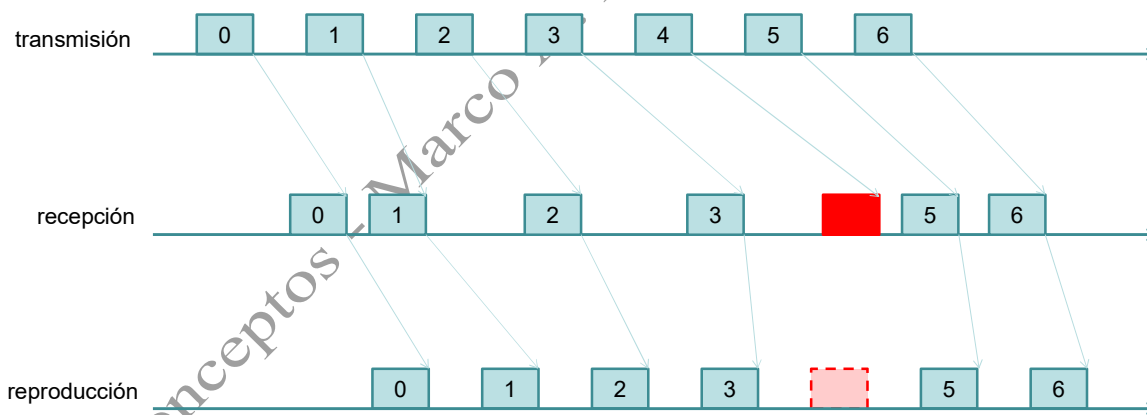


Figura 152. Recepción y reproducción de paquetes en tiempo real

Al comparar la Figura 66 y la Figura 152, queda claro que la red debe comportarse de manera diferente para cada tipo de tráfico. Por ejemplo, para el tráfico en tiempo real no vale la pena disponer de una cola de espera en caso de que los paquetes que no encuentran un servidor disponible, mientras que para tráfico elástico es importante almacenar los paquetes en una cola porque deben ser enviados sin errores sin importar cuánto tiempo tome la corrección de errores. En ese caso, para el tráfico en tiempo real lo ideal es acercarse a la conmutación de circuitos, en la que a cada flujo se le asigna una capacidad suficiente para atender el pico del flujo; de esta manera, una vez establecido el circuito, cada paquete sufre solamente el retardo de transmisión. En cambio, para el tráfico elástico lo ideal es

almacenar cada paquete en un buffer, no sólo para evitar que sea bloqueado sino también para tener la oportunidad de retransmitirlo si hace falta.

En términos de las colas M/M/... que hemos visto, diríamos que los flujos de tráfico en tiempo real desearían el servicio de un modelo M/M/K/K (no hay colas de espera) y el tráfico elástico desearía el servicio de un modelo M/M/K (no hay pérdidas). ¿Cómo ofrecer estos dos tipos de servicio en un único multiplexor? La Figura 153 muestra una alternativa de solución. El enlace de transmisión se divide en N canales (por ejemplo, TDM o FDM) que se asignan a los paquetes de cada tipo de tráfico según algunas políticas bien definidas. El tráfico tipo 1 es de tiempo real y genera un promedio de λ_1 paquetes por segundo según un proceso Poisson, donde cada paquete ocupa el canal que se le asigne durante un tiempo exponencialmente distribuido con promedio $1/\mu_1$. El tráfico tipo 2 es elástico y genera un promedio de λ_2 paquetes por segundo según un proceso Poisson, donde cada paquete ocupa el canal que se le asigne durante un tiempo exponencialmente distribuido con promedio $1/\mu_2$. Al tráfico tipo 1 le diremos "tráfico de voz" por ponerle un nombre, aunque los paquetes de voz no satisfacen los postulados de llegadas tipo Poisson y tiempos de servicio exponenciales; igualmente, al tráfico tipo 2 le diremos "tráfico de datos" por ponerle un nombre. Para satisfacer el comportamiento que cada flujo espera de la red, se reservan $N_1 < N$ canales para los paquetes de voz y $N_2 = N - N_1$ canales para los paquetes de datos. De esta manera el tráfico de voz ve un sistema M/M/ N_1 / N_1 y el tráfico de datos ve un sistema M/M/ N_2 , que es el tipo de servicio que cada uno desea. Como hay una frontera estática entre los N_1 canales de voz y los N_2 canales de datos, a este tipo de multiplexaje se le conoce como integración con **frontera fija**.

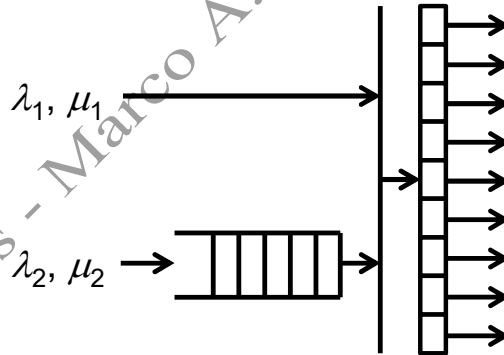


Figura 153. Integración de dos tipos de tráfico en un único multiplexor

El esquema de frontera fija tiene un inconveniente de ineficiencia. Si llega un paquete de voz cuando todos los N_1 servidores de voz están ocupados, el paquete será bloqueado a la entrada del multiplexor aunque haya canales de datos desocupados. Igualmente, si llega un paquete de datos cuando todos los N_2 servidores de datos están ocupados, el paquete deberá esperar en cola aunque haya canales de voz desocupados. Para ganar eficiencia, se puede usar un esquema de **frontera móvil** en el que un paquete de voz que encuentra a su llegada todos los N_1 servidores de voz ocupados y algunos servidores de datos desocupados, puede ocupar uno de los servidores de datos libres, siempre que se comprometa a entregarlo si un paquete de datos llega durante su tiempo de transmisión. Igualmente, un paquete de datos puede ocupar un servidor de voz, siempre que se comprometa a entregarlo si un paquete de voz lo requiere durante su tiempo de servicio.

Para intentar una solución analítica a este modelo matemático, como de costumbre, supondremos algunas simplificaciones que alejan el sistema de la realidad pero que hace tratable el sistema analíticamente y, de todas maneras, nos revela aspectos relevantes del sistema para darnos una mejor comprensión del mismo: Supondremos un único canal. Claro, aquí sólo es posible la frontera móvil (de otra manera el canal se asignaría a un solo tipo de tráfico y el otro nunca recibiría servicio), pero queda la pregunta de qué política usar para la asignación del canal. Primero supondremos un esquema FIFO: Cuando llega un nuevo paquete de voz, si encuentra el canal desocupado, lo ocupa durante un tiempo exponencialmente distribuido con promedio $1/\mu_1$; si no, el nuevo paquete de voz es bloqueado, independiente de si el paquete en servicio es de voz o de datos. De la misma manera, cuando llega un nuevo paquete de datos, si encuentra el canal desocupado, lo ocupa durante un tiempo exponencialmente distribuido con promedio $1/\mu_2$; si no, el nuevo paquete de datos esperará en la cola su turno de transmisión, independiente de si el paquete en servicio es de voz o de datos.

Aunque el esquema que acabamos de describir es estrictamente FIFO, representa de algún modo cierta prioridad para los datos porque los paquetes de voz sólo serán aceptados si la cola de datos está desocupada. De hecho, denotando el estado del sistema como la pareja (i,j) que indica la presencia de i paquetes de voz y j paquetes de datos, el diagrama de transición entre estados es como se muestra en la Figura 154. Si $\lambda_1=0$, los estados de la fila 1 serán transientes y, a la larga, sólo permanecerá la fila 0, que es el diagrama de una cola M/M/1, como esperamos. Igualmente, si $\lambda_2=0$, sólo serán recurrentes los dos estados la izquierda, que es el diagrama de la cola M/M/1/1. Pero si ambas tasas de llegadas son mayores a cero, los paquetes de datos podrán seguir llegando aun cuando haya un paquete de voz en servicio, en cuyo caso entran a la cola, pues no hay transiciones hacia la izquierda en la fila 1 ya que no hay paquetes de datos en servicio. Cuando salga el paquete de voz, el sistema puede quedar vacío si no habían paquetes de datos en cola, o el canal se asignará al primer paquete de la cola de datos, como muestran las transiciones verticales de la fila 1 a la fila 0. Nuevamente, aunque el esquema es FIFO, la ventaja para los paquetes de datos se evidencia en que sólo hay una transición etiquetada con λ_1 (los paquetes de voz sólo son aceptados en el estado $(0,0)$).

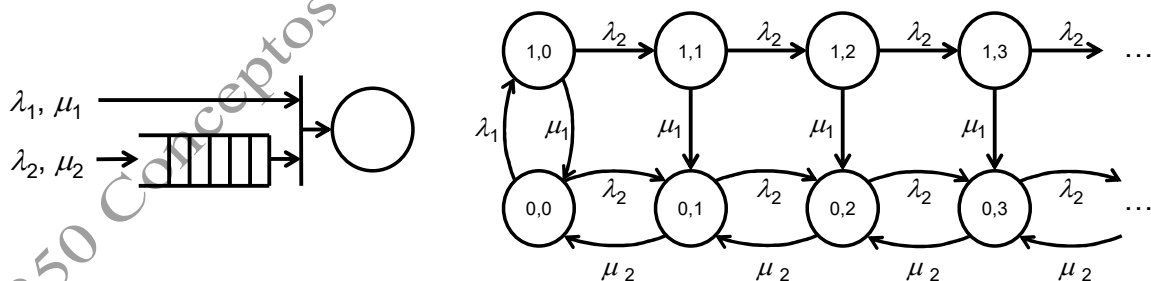


Figura 154. Un sólo canal FIFO y diagrama de transición entre estados

Como éste ya no es un proceso de nacimiento y muerte como las colas M/M/ que hemos visto hasta ahora, el proceso descrito en la definición **¡Error! No se encuentra el origen de la referencia.** ya no aplica directamente. Aunque existen diferentes alternativas de análisis para la cadena de Markov de la Figura 154, aquí usaremos el método de la función generadora de probabilidad de la definición 54. En general, se trata Y , para ilustrar el uso de la función generadora de probabilidad, pasaremos primero por el análisis de la cola M/M/1 desde este método.

a. Revisión de la cola M/M/1 mediante la función generadora de probabilidad

La idea de la aplicación de la función generadora de probabilidad es convertir una ecuación de recurrencia entre los elementos de una función de distribución de probabilidad en una ecuación algebraica, tal como se usa la transformada Z para convertir ecuaciones de diferencia en ecuaciones algebraicas, o como se usa la transformada de Laplace para convertir ecuaciones diferenciales en ecuaciones algebraicas. La función generadora de probabilidad y la transformada Z tienen una diferencia sutil de que la variable z de una de ellas es la variable $1/z$ de la otra pero, por lo demás, se usan casi de manera idéntica. Por ejemplo, para la cola M/M/1 habíamos encontrado la siguiente relación de recurrencia para las probabilidades de estado estable $p_n = P[N=n]$:

$$(\lambda + \mu) p_n = \lambda p_{n-1} + \mu p_{n+1}, n > 0$$

con $\lambda p_0 = \mu p_1$

que es una ecuación lineal de diferencias con coeficientes constantes. El método de la función generadora es realmente simple: Multipliquemos la ecuación de recurrencia por z^n y sumemos para el rango válido, $n > 0$.

$$(\lambda + \mu) \sum_{n=1}^{\infty} p_n z^n = \lambda \sum_{n=1}^{\infty} p_{n-1} z^n + \mu \sum_{n=1}^{\infty} p_{n+1} z^n \quad (13)$$

Si definimos la función generadora de probabilidad de la variable aleatoria N como

$$G_N(z) = \sum_{n=0}^{\infty} p_n z^n$$

podremos escribir cada una de las sumas de la Ecuación (13) en términos de $G_N(z)$ así:

$$\begin{aligned} \sum_{n=1}^{\infty} p_n z^n &= \sum_{n=0}^{\infty} p_n z^n - p_0 = G_N(z) - p_0 \\ \sum_{n=1}^{\infty} p_{n-1} z^n &= z \sum_{n=1}^{\infty} p_{n-1} z^{n-1} = z \sum_{n=0}^{\infty} p_n z^n = z G_N(z) \\ \sum_{n=1}^{\infty} p_{n+1} z^n &= \frac{1}{z} \sum_{n=1}^{\infty} p_{n+1} z^{n+1} = \frac{1}{z} \sum_{n=2}^{\infty} p_n z^n = \frac{1}{z} (G_N(z) - p_0 - p_1 z) \end{aligned}$$

Al reemplazar estas expresiones en la Ecuación (13) obtenemos

$$(\lambda + \mu)(G_N(z) - p_0) = \lambda z G_N(z) + \frac{\mu}{z} (G_N(z) - p_0) - \mu p_1$$

El término μp_1 es igual a λp_0 de acuerdo con la ecuación de recurrencia para $n=0$, por lo que el término $-\lambda p_0$ de la izquierda se puede eliminar con el término $-\mu p_1$ de la derecha

$$(\lambda + \mu)G_N(z) - \mu p_0 = \lambda z G_N(z) + \frac{\mu}{z}(G_N(z) - p_0)$$

Ahora podemos agrupar $G_N(z)$

$$[\mu(1 - 1/z) + \lambda(1 - z)]G_N(z) = p_0[\mu(1 - 1/z)]$$

y, al dividir por $\mu(1 - 1/z)$ a ambos lados,

$$\left[1 + \frac{\lambda(1 - z)}{\mu(1 - 1/z)}\right]G_N(z) = p_0 \Rightarrow G_N(z) = \frac{p_0}{1 - \rho z}$$

Para encontrar p_0 usamos una propiedad básica de la función generadora de probabilidad: $G_N(1) = 1$ (ver definición 54):

$$\frac{p_0}{1 - \rho} = 1 \Rightarrow p_0 = 1 - \rho$$

de manera que

$$G_N(z) = \frac{1 - \rho}{1 - \rho z} \quad (14)$$

que es la PGF de la distribución geométrica, como habíamos visto en la definición 54. Y, como

$\frac{d}{dz}G_N(z) = \frac{\rho(1 - \rho)}{(1 - \rho z)^2}$, entonces $E[N] = \frac{d}{dz}G_N(1) = \frac{\rho}{1 - \rho}$. Este es el resultado que obtuvimos en

la definición 136, pero trabajando directamente desde la ecuación de recurrencia (en el contexto de las señales, diríamos que allá usamos un método en el dominio del tiempo y aquí usamos un método en el dominio de la frecuencia, aunque no hay tiempo ni frecuencia en nuestras ecuaciones de recurrencia).

Ciertamente, la solución basada en la función generadora de probabilidad no es tan elegante, pero es simple y directa. Ahora veremos cómo usar este método en la solución de la cadena de Markov que representa el estado del multiplexor integrado de la Figura 154.

b. Integración en un solo canal con disciplina FIFO

La Figura 155 muestra el diagrama de transición de estado con algunas superficies cerradas dónde balancear el flujo.

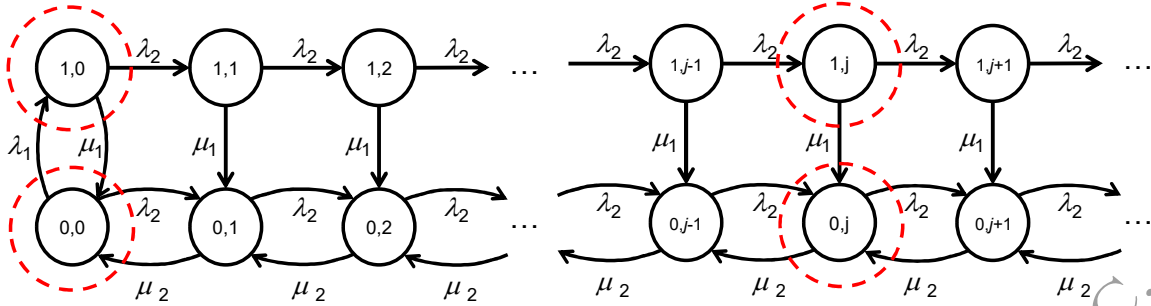


Figura 155. Superficies cerradas para establecimiento de ecuaciones de balance de flujo.

Veamos las ecuaciones de balance de flujo para cada fila:

$$(\lambda_1 + \lambda_2) p_{0,0} = \mu_1 p_{1,0} + \mu_2 p_{0,1} \quad (15)$$

Fila 0:

$$(\lambda_2 + \mu_2) p_{0,j} = \lambda_2 p_{0,j-1} + \mu_2 p_{0,j+1} + \mu_1 p_{1,j}, \quad j \geq 1 \quad (16)$$

$$(\lambda_2 + \mu_1) p_{1,0} = \lambda_1 p_{0,0} \quad (17)$$

Fila 1

$$(\lambda_2 + \mu_1) p_{1,j} = \lambda_2 p_{1,j-1}, \quad j \geq 1 \quad (18)$$

La fila 1 se puede iterar fácilmente:

De la Ecuación (17), $p_{1,0} = \frac{\lambda_1}{\lambda_2 + \mu_1} p_{0,0}$

De la Ecuación (18), con $j=1$, $p_{1,1} = \frac{\lambda_2}{\lambda_2 + \mu_1} p_{1,0} = \frac{\lambda_1}{\lambda_2 + \mu_1} \cdot \frac{\lambda_2}{\lambda_2 + \mu_1} p_{0,0}$

De la Ecuación (18), con $j=2$, $p_{1,2} = \frac{\lambda_2}{\lambda_2 + \mu_1} p_{1,1} = \frac{\lambda_1}{\lambda_2 + \mu_1} \cdot \left(\frac{\lambda_2}{\lambda_2 + \mu_1} \right)^2 p_{0,0}$

De la Ecuación (18), con $j=3$, $p_{1,3} = \frac{\lambda_2}{\lambda_2 + \mu_1} p_{1,2} = \frac{\lambda_1}{\lambda_2 + \mu_1} \cdot \left(\frac{\lambda_2}{\lambda_2 + \mu_1} \right)^3 p_{0,0}$

Así que, en general,

$$p_{1,j} = \frac{\lambda_1}{\lambda_2 + \mu_1} \cdot \left(\frac{\lambda_2}{\lambda_2 + \mu_1} \right)^j p_{0,0} \quad (19)$$

Una expresión semejante para $p_{0,j}$ no es tan fácil de encontrar porque la Ecuación (16) tiene un acople con la fila 1 a través del término $\mu_1 p_{1,j}$. Sin embargo podemos usar la técnica de la transformada Z definiendo una función auxiliar

$$G_0(z) = \sum_{j=0}^{\infty} p_{0,j} z^j \quad (20)$$

que no es exactamente una función generadora de probabilidad. En efecto, en vez de tener $G_0(1)=1$, tenemos que $G_0(1)$ es sólo la probabilidad de que no haya paquetes de voz en el sistema. Sin embargo podemos usar el mismo procedimiento, multiplicando cada término de la Ecuación (16) por z^j y sumando para todos los j válidos, $j \geq 1$:

$$(\lambda_2 + \mu_2) \sum_{j=1}^{\infty} p_{0,j} z^j = \lambda_2 \sum_{j=1}^{\infty} p_{0,j-1} z^j + \mu_2 \sum_{j=1}^{\infty} p_{0,j+1} z^j + \mu_1 \sum_{j=1}^{\infty} p_{1,j} z^j \quad (21)$$

Las tres primeras sumas de la Ecuación (21) se pueden expresar fácilmente en términos de $G_0(z)$:

$$\begin{aligned} \sum_{j=1}^{\infty} p_{0,j} z^j &= G_0(z) - p_{0,0} \\ \sum_{j=1}^{\infty} p_{0,j-1} z^j &= z \sum_{j=1}^{\infty} p_{0,j-1} z^{j-1} = z G_0(z) \\ \sum_{j=1}^{\infty} p_{0,j+1} z^j &= \frac{1}{z} \sum_{j=1}^{\infty} p_{0,j+1} z^{j+1} = \frac{1}{z} (G_0(z) - p_{0,0} - p_{0,1} z) \end{aligned}$$

La cuarta suma es igualmente fácil de calcular, pues es una serie geométrica:

$$\begin{aligned} \sum_{j=1}^{\infty} p_{1,j} z^j &= \frac{\lambda_1 p_{0,0}}{\lambda_2 + \mu_1} \sum_{j=1}^{\infty} \left(\frac{\lambda_2 z}{\lambda_2 + \mu_1} \right)^j = \frac{\lambda_1 \lambda_2 z p_{0,0}}{(\lambda_2 + \mu_1)^2} \sum_{j=0}^{\infty} \left(\frac{\lambda_2 z}{\lambda_2 + \mu_1} \right)^j = \frac{\lambda_1 \lambda_2 z p_{0,0}}{(\lambda_2 + \mu_1)^2} \frac{1}{1 - \lambda_2 z / (\lambda_2 + \mu_1)} \\ \sum_{j=1}^{\infty} p_{1,j} z^j &= \frac{\lambda_1 \lambda_2 z}{(\lambda_2 + \mu_1)(\mu_1 - \lambda_2(z-1))} p_{0,0} \end{aligned} \quad (22)$$

Remplazando estas cuatro sumas en la Ecuación (21)

$$(\lambda_2 + \mu_2)(G_0(z) - p_{0,0}) = \lambda_2 z G_0(z) + \frac{\mu_2}{z} (G_0(z) - p_{0,0} - p_{0,1} z) + \frac{\lambda_1 \lambda_2 \mu_1 z}{(\lambda_2 + \mu_1)(\mu_1 - \lambda_2(z-1))} p_{0,0}$$

Al expandir el segundo término de la derecha obtenemos el término $\mu_2 p_{0,1}$, el cual podemos poner en términos de $p_{0,0}$ mediante las ecuaciones (15) y (17):

$$\mu_2 p_{0,1} = (\lambda_1 + \lambda_2) p_{0,0} - \mu_1 p_{1,0} = (\lambda_1 + \lambda_2) p_{0,0} - \frac{\lambda_1 \mu_1}{\lambda_2 + \mu_1} p_{0,0} = \frac{\lambda_2 (\lambda_1 + \lambda_2 + \mu_1)}{\lambda_2 + \mu_1} p_{0,0}$$

obteniendo

$$(\lambda_2 + \mu_2)(G_0(z) - p_{0,0}) = \lambda_2 z G_0(z) + \frac{\mu_2}{z} (G_0(z) - p_{0,0}) - \frac{\lambda_2 (\lambda_1 + \lambda_2 + \mu_1)}{\lambda_2 + \mu_1} p_{0,0} + \frac{\lambda_1 \lambda_2 \mu_1 z}{(\lambda_2 + \mu_1)(\mu_1 - \lambda_2(z-1))} p_{0,0}$$

Agrupando a la izquierda los términos que contienen $G_0(z)$,

$$\left(\lambda_2 + \mu_2 - \lambda_2 z - \frac{\mu_2}{z} \right) G_0(z) = \left(-\frac{\mu_2}{z} - \frac{\lambda_2 (\lambda_1 + \lambda_2 + \mu_1)}{\lambda_2 + \mu_1} + \frac{\lambda_1 \lambda_2 \mu_1 z}{(\lambda_2 + \mu_1)(\mu_1 - \lambda_2(z-1))} + (\lambda_1 + \mu_2) \right) p_{0,0}$$

Factorizando a cada lado

$$\frac{1}{z}(z-1)(\mu_2 - \lambda_2 z)G_0(z) = \frac{(z-1)(\lambda_1 \lambda_2 z - \mu_2(\lambda_2(z-1) - \mu_1))}{z(\mu_1 - \lambda_2(z-1))} p_{0,0}$$

De manera que

$$G_0(z) = \frac{\lambda_1 \lambda_2 z + \mu_2(\mu_1 - \lambda_2(z-1))}{(\mu_2 - \lambda_2 z)(\mu_1 - \lambda_2(z-1))} p_{0,0}$$

Para poner la anterior expresión en parámetros más generales, dividamos arriba y abajo por $\mu_1 \mu_2$:

$$G_0(z) = \frac{\frac{\lambda_1}{\mu_1} \frac{\lambda_2}{\mu_2} z + \left(1 - \frac{\lambda_2}{\mu_1}(z-1)\right)}{\left(1 - \frac{\lambda_2}{\mu_2} z\right) \left(1 - \frac{\lambda_2}{\mu_1}(z-1)\right)} p_{0,0}$$

Y definamos $\rho_1 = \lambda_1/\mu_1$, $\rho_2 = \lambda_2/\mu_2$, $\alpha = \mu_2/\mu_1$ para obtener

$$G_0(z) = \frac{\rho_1 \rho_2 z + (1 - \alpha \rho_2)(z-1)}{(1 - \rho_2 z)(1 - \alpha \rho_2(z-1))} p_{0,0} \quad (23)$$

Los parámetros ρ_1 y ρ_2 son las intensidades de tráfico de voz y datos, respectivamente. El parámetro α es la relación entre el tiempo promedio de ocupación del canal por parte de los paquetes de voz y el tiempo promedio de ocupación del canal por parte de los paquetes de datos.

Para encontrar el valor de $p_{0,0}$, normalizamos usando el primer axioma de las probabilidades

$$\sum_{j=0}^{\infty} p_{0,j} + \sum_{j=0}^{\infty} p_{1,j} = 1$$

La primera suma, de acuerdo con la Ecuación (20), es simplemente $G_0(1)$. Evaluando la ecuación (23) en $z=1$,

$$\sum_{j=0}^{\infty} p_{0,j} = G_0(1) = \frac{1 + \rho_1 \rho_2}{1 - \rho_2} p_{0,0}$$

La segunda suma es la serie geométrica de los términos de la Ecuación (19) que, de acuerdo con las ecuaciones (17) y (22), corresponde a

$$\sum_{j=0}^{\infty} p_{1,j} = \left(\frac{\lambda_1}{\lambda_2 + \mu_1} + \frac{\lambda_1 \lambda_2}{\mu_1(\lambda_2 + \mu_1)} \right) p_{0,0} = \rho_1 p_{0,0}$$

De donde

$$\left(\frac{1 + \rho_1 \rho_2}{1 - \rho_2} + \rho_1 \right) p_{0,0} = \frac{1 + \rho_1}{1 - \rho_2} p_{0,0} = 1$$

$$p_{0,0} = \frac{1 - \rho_2}{1 + \rho_1} \quad (24)$$

Remplazando la Ecuación (24) en la Ecuación (23) obtenemos

$$G_0(z) = \frac{1 - \rho_2}{1 - \rho_2 z} \cdot \frac{\rho_1 \rho_2 z + (1 - \alpha \rho_2 (z - 1))}{(1 + \rho_1)(1 - \alpha \rho_2 (z - 1))} \quad (25)$$

Una verificación sencilla se obtiene evaluando $G_0(z)$ cuando $\rho_1=0$, que debe corresponder al caso M/M/1 para el tráfico de datos. En efecto,

$$G_0(z) \Big|_{\rho_1=0} = \frac{1 - \rho_2}{1 - \rho_2 z}$$

es la función generadora de probabilidad de la distribución geométrica $p_n = \rho_2^n (1 - \rho_2)$, como se mostró en la definición 54, que es la distribución del estado de la cola M/M/1, como se mostró en la definición 136.

De otro lado, la Ecuación (24) nos revela de una vez una importante medida de desempeño del multiplexor integrado: la probabilidad de bloqueo de los paquetes de voz, $P_B = 1 - p_{0,0} = 1 - (1 - \rho_2) / (1 + \rho_1)$

$$P_B = \frac{\rho_1}{1 + \rho_1} + \frac{\rho_2}{1 + \rho_1} \quad (26)$$

Otra verificación sencilla consiste en notar que, cuando $\rho_2=0$, la ecuación (26) conduce a la probabilidad de bloqueo de la cola M/M/1/1, tal como esperábamos. Pero, por tener que compartir el enlace, hay una degradación en el desempeño debida al tráfico de datos. Lo mismo debe ocurrirle al tráfico de datos, como veremos al calcular el número promedio de paquetes de datos en el sistema:

$$E[J] = \sum_{j=0}^{\infty} j p_{0,j} + \sum_{j=0}^{\infty} j p_{1,j} = \frac{d}{dz} G_0(1) + \sum_{j=0}^{\infty} j p_{1,j}$$

donde

$$\begin{aligned} \sum_{j=0}^{\infty} j p_{1,j} &= p_{0,0} \frac{\lambda_2}{\lambda_2 + \mu_1} \sum_{j=0}^{\infty} j \left(\frac{\lambda_2}{\lambda_2 + \mu_1} \right)^j = \frac{1 - \rho_2}{1 + \rho_1} \cdot \frac{\rho_1}{1 + \alpha \rho_2} \sum_{j=0}^{\infty} j \left(\frac{\alpha \rho_2}{1 + \alpha \rho_2} \right)^j \\ \sum_{j=0}^{\infty} j p_{1,j} &= \frac{1 - \rho_2}{1 + \rho_1} \cdot \frac{\rho_1}{1 + \alpha \rho_2} \cdot \frac{\alpha \rho_2}{1 + \alpha \rho_2} \sum_{j=0}^{\infty} j \theta^{j-1} \Big|_{\theta = \alpha \rho_2 / (1 + \alpha \rho_2)} \\ \sum_{j=0}^{\infty} j p_{1,j} &= \frac{1 - \rho_2}{1 + \rho_1} \cdot \frac{\rho_1}{1 + \alpha \rho_2} \cdot \frac{\alpha \rho_2}{1 + \alpha \rho_2} \cdot \left(\frac{1}{1 - \alpha \rho_2 / (1 + \alpha \rho_2)} \right)^2 = \frac{\alpha \rho_1 \rho_2 (1 - \rho_2)}{1 + \rho_1} \end{aligned}$$

Y, de la Ecuación (25),

$$\frac{d}{dz} G_0(z) = \frac{\rho_2 (1 - \rho_2)}{(1 - \alpha (1 - \rho_2))(1 + \rho_1)} \left[\frac{1 + \rho_1 - \alpha (1 - \rho_2)}{(1 - \rho_2 z)^2} - \frac{\alpha \rho_1 (1 + \alpha \rho_2)}{(1 - \alpha \rho_2 (z - 1))^2} \right]$$

de donde

$$\frac{d}{dz} G_0(1) = \frac{\rho_2(1-\rho_2)}{(1-\alpha(1-\rho_2))(1+\rho_1)} \left[\frac{1+\rho_1-\alpha(1-\rho_2)}{(1-\rho_2)^2} - \alpha\rho_1(1+\alpha\rho_2) \right] = \frac{\rho_2}{1-\rho_2} + \frac{\alpha\rho_1\rho_2^2}{1+\rho_1}$$

y

$$E[J] = \frac{\rho_2}{1-\rho_2} + \alpha \frac{\rho_1\rho_2}{1+\rho_1} \quad (27)$$

La última verificación consiste en notar que, cuando $\rho_1=0$, la ecuación (27) conduce al número promedio de paquetes de la cola M/M/1, tal como esperábamos. Pero, por tener que compartir el enlace, hay una degradación en el desempeño debida al tráfico de voz. Mediante la ley de Little podemos calcular el retardo promedio de un paquete de datos, el cual se presenta en la siguiente ecuación, normalizado con respecto al tiempo de transmisión de un paquete de datos:

$$\mu_2 E[T] = \frac{E[J]}{\rho_2} = \frac{1}{1-\rho_2} + \alpha \frac{\rho_1}{1+\rho_1} \quad (28)$$

Al comparar las ecuaciones (26) y (28), se puede notar que, tratándose de conmutación de paquetes de voz, es el tráfico de voz el que se ve más seriamente afectado. En efecto, típicamente un paquete de voz tiene algunos cientos de bytes mientras un paquete de datos suele tener más de 1000 bytes, por lo que el valor de α puede ser pequeño ($\sim 1/5$). Además, el término $\rho_1\rho_2/(1+\rho_1)$ suele ser, a su vez, significativamente menor que $\rho_2/(1-\rho_2)$. En consecuencia, la degradación en el desempeño experimentada por los paquetes de datos debido a que comparte el enlace con los paquetes de voz puede ser casi imperceptible. En cambio, la probabilidad de bloqueo de los paquetes de voz podría duplicarse fácilmente (si ρ_1 y ρ_2 son comparables). Claro, podríamos usar el mismo esquema para integrar llamadas de voz conmutadas por circuitos con paquetes de datos, en cuyo caso el valor de α puede ser de varios miles y así el tráfico de datos también se vería seriamente afectado. Pero, si hablamos de conmutación integrada de paquetes de voz y datos, parece razonable cambiar el esquema FIFO por un esquema de prioridad para el tráfico de voz.

c. Integración en un solo canal con prioridad para la voz

En el modelo anterior, cuando llegaba un paquete de voz y el canal estaba ocupado por cualquier tipo de paquete, el nuevo paquete de voz era rechazado a la entrada del multiplexor. Ahora el nuevo paquete de voz será rechazado sólo si el canal está ocupado por otro paquete de voz; si es un paquete de datos el que ocupa el canal, dicho paquete de datos será devuelto al buffer y el nuevo paquete de voz iniciará inmediatamente su transmisión. Esto es, la disciplina de servicio ya no es FIFO sino que se atiende con prioridad a los paquetes de voz.

La diferencia entre los diagramas de transición de estos dos modelos es que, ahora, los paquetes de voz serán aceptados desde cualquier estado de la fila cero, como muestra la Figura 156.

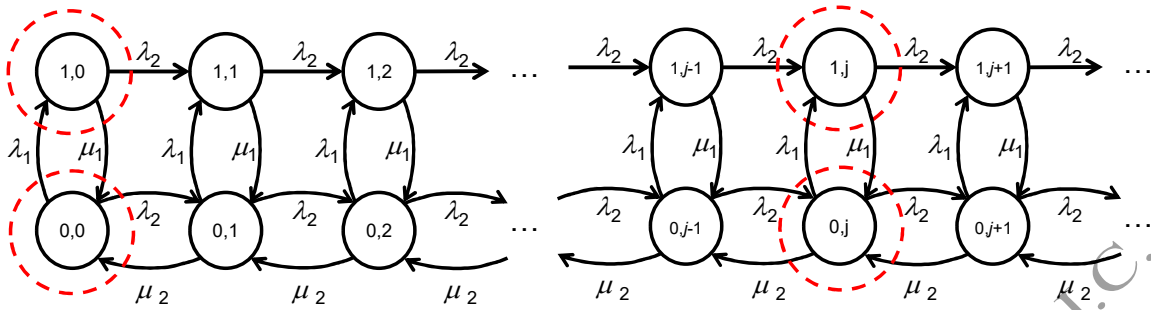


Figura 156. Diagrama de transición entre estados para el esquema con prioridad para la voz

Las ecuaciones de balance de flujo ahora son

$$(\lambda_1 + \lambda_2) p_{0,0} = \mu_1 p_{1,0} + \mu_2 p_{0,1}$$

Fila 0:

$$(\lambda_1 + \lambda_2 + \mu_2) p_{0,j} = \lambda_2 p_{0,j-1} + \mu_2 p_{0,j+1} + \mu_1 p_{1,j}, \quad j \geq 1$$

$$(\lambda_2 + \mu_1) p_{1,0} = \lambda_1 p_{0,0}$$

Fila 1

$$(\lambda_2 + \mu_1) p_{1,j} = \lambda_2 p_{1,j-1} + \lambda_1 p_{0,j}, \quad j \geq 1$$

Definiendo α , ρ_1 y ρ_2 de la misma manera que en el caso FIFO, podemos dividir las anteriores ecuaciones de balance de flujo por μ_1 para convertirlas en

$$(\rho_1 + \alpha \rho_2) p_{0,0} = p_{1,0} + \alpha p_{0,1} \tag{29}$$

Fila 0:

$$(\rho_1 + \lambda_2 + \mu_2) p_{0,j} = \alpha \rho_2 p_{0,j-1} + \alpha p_{0,j+1} + p_{1,j}, \quad j \geq 1 \tag{30}$$

$$(1 + \alpha \rho_2) p_{1,0} = \rho_1 p_{0,0} \tag{31}$$

Fila 1

$$(1 + \alpha \rho_2) p_{1,j} = \alpha \rho_2 p_{1,j-1} + \rho_1 p_{0,j}, \quad j \geq 1 \tag{32}$$

De las ecuaciones (29) y (31) obtenemos

$$p_{1,0} = \frac{\rho_1}{1 + \alpha \rho_2} p_{0,0}, \quad p_{0,1} = \rho_2 \left(1 + \frac{\rho_1}{1 + \alpha \rho_2} \right) p_{0,0} \tag{33}$$

Para procesar las ecuaciones (30) y (32), las multiplicamos por z^j y sumamos para $j \geq 1$, expresando las sumas obtenidas en términos de

$$G_0(z) = \sum_{j=0}^{\infty} p_{0,j} z^j \quad \text{y} \quad G_1(z) = \sum_{j=0}^{\infty} p_{1,j} z^j \quad (34)$$

Al hacerlo así surgen dos ecuaciones de las cuales podemos despejar $G_0(z)$ y $G_1(z)$ –los términos en $p_{0,1}$ y $p_{1,0}$ también los escribimos en términos de $p_{0,0}$ de acuerdo con la ecuación (33)–

$$G_0(z) = \frac{1 - \alpha \rho_2 (z - 1)}{1 + \rho_2 [\alpha - z(1 + \alpha(1 - \rho_2(z - 1)) + \rho_1)]} p_{0,0}$$

$$G_1(z) = \frac{\rho_1}{1 + \rho_2 [\alpha - z(1 + \alpha(1 - \rho_2(z - 1)) + \rho_1)]} p_{0,0}$$

Ninguna de estas dos funciones de z es una función generadora de probabilidad, pues $G_0(1) = P[\text{Hay 0 paquetes de voz en el sistema}]$ y $G_1(1) = P[\text{Hay 1 paquete de voz en el sistema}]$. Pero la suma

$$G(z) = G_0(z) + G_1(z) = \sum_{j=0}^{\infty} (p_{0,j} + p_{1,j}) z^j = \sum_{j=0}^{\infty} P[j \text{ paquetes de voz}] z^j$$

sí es la función generadora de la probabilidad del número de paquetes de datos en el sistema,

$$G(z) = \frac{1 + \rho_1 - \alpha \rho_2 (z - 1)}{1 + \rho_2 [\alpha - z(1 + \alpha(1 - \rho_2(z - 1)) + \rho_1)]} p_{0,0} \quad (35)$$

de manera que $p_{0,0}$ se puede encontrar por normalización, $G(1)=1$, para obtener

$$p_{0,0} = \frac{1}{1 + \rho_1} - \rho_2 \quad (36)$$

Al remplazar la Ecuación (36) en la ecuación (35) podemos hacer una verificación simple de que nuestros resultados hasta aquí parecen correctos: Al evaluar la expresión resultante de $G(z)$ cuando $\rho_1=0$ obtenemos

$$G(z) \Big|_{\rho_1=0} = \frac{1 - \alpha \rho_2 (z - 1)}{1 + \rho_2 [\alpha - z(1 + \alpha(1 - \rho_2(z - 1)))]} (1 - \rho_2) = \frac{1 - \rho_2}{1 - \rho_2 z}$$

que es exactamente la función generadora de probabilidad de la variable aleatoria geométrica, que corresponde a la distribución de la cola M/M/1 con intensidad de tráfico ρ_2 , como esperábamos.

La interpretación de $G_0(z)$ y $G_1(z)$ nos permitiría también despejar $p_{0,0}$ independientemente desde cada una de ellas, pues sabemos que, como el tráfico de datos es completamente transparente para el tráfico de voz, los paquetes de voz ven un sistema M/M/1/1 con intensidad de tráfico ρ_1 , de manera que $G_0(1)=P[0 \text{ paq. de voz}]=1/(1+\rho_1)$ y $G_1(1)=P[1 \text{ paq. de voz}]=\rho_1/(1+\rho_1)$. Esta última probabilidad es la probabilidad de que un paquete de voz sea bloqueado, por lo que la ecuación (36) se interpreta así: $p_{0,0} = (1 - P_B) - \rho_2$. Para que el sistema sea estable es necesario que esta probabilidad de que el sistema esté desocupado sea mayor que cero, lo cual nos lleva a una reveladora condición de estabilidad:

$$\rho_2 < 1 - P_B$$

En efecto, dada la prioridad que le otorgamos al tráfico de voz, el tráfico de datos sólo puede ocupar la fracción de capacidad que no utiliza el tráfico de voz: la capacidad disponible para el tráfico de datos son "las sobras" que deja el tráfico de voz.

Respecto al desempeño del tráfico de datos, podemos usar las ecuaciones (35) y (36) para evaluar

$$E[J] = \frac{d}{dz} G(z) \Big|_{z=1} = \frac{\rho_2}{1-\rho_2} + \rho_1 \frac{\rho_2 (1 + \rho_1 + \alpha(1-\rho_2))}{(1+\rho_1)(1-\rho_2)(1-\rho_2(1+\rho_1))}$$

La expresión anterior se buscó para hacer evidente dos aspectos interesantes: primero, que cuando $\rho_1=0$, $E[J]$ toma la ya conocida forma de la cola M/M/1, $E[J]=\rho_2/(1-\rho_2)$; segundo, el último factor del denominador, $(1-\rho_2(1+\rho_1))$, que muestra la necesidad de que ρ_2 sea menor a $1-P_B=1/(1+\rho_1)$ para que el tráfico de datos no exceda la capacidad que le deja el tráfico de voz. En términos del retardo promedio experimentado por un paquete de datos en el sistema, podemos aplicar la ley de Little y normalizar con respecto al tiempo de transmisión de un paquete de datos, $\mu_2 E[T] = E[J]/\rho_2$, para obtener

$$\mu_2 E[T] = \frac{1}{1-\rho_2} + \frac{\rho_1}{1+\rho_1} \cdot \frac{1+\rho_1+\alpha(1-\rho_2)}{(1-\rho_2)(1-\rho_2(1+\rho_1))}$$

Claro, el segundo término de la derecha es mayor que en la Ecuación (28), y su crecimiento hacia infinito se da a medida que ρ_2 tiende a $1-P_B$, no a medida que ρ_2 tiende a 1.

d. Integración en dos canales con frontera móvil para los datos

Si añadimos un canal adicional, podemos reservar uno para voz y otro para datos y así cada tipo de tráfico verá lo que desea: M/M/1/1 para voz y M/M/1 para datos. Sin embargo, como se mencionó anteriormente, se puede ganar en eficiencia si los períodos de desocupación de un canal se aprovechan para atender paquetes del otro tipo de tráfico.

Por ejemplo, si dejamos que el tráfico de voz siga en su mundo M/M/1/1 pero le permitimos al canal de voz atender paquetes de datos durante su tiempo libre, tendríamos el diagrama de transición entre estados de la Figura 157.

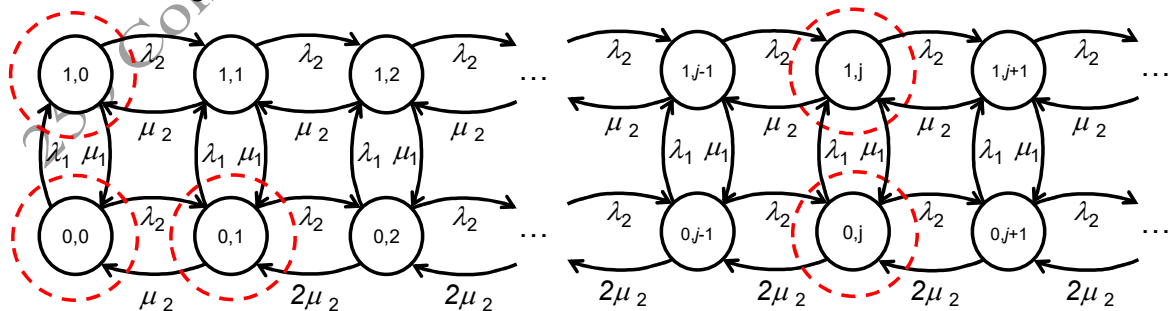


Figura 157. Un canal reservado para cada tipo de tráfico, con frontera móvil para los datos

En efecto, si no hay paquetes de voz y hay más de un paquete de datos, los dos servidores estarán atendiendo paquetes de datos, cada uno a una tasa μ_2 paquetes por segundo, para una tasa total de servicio de $2\mu_2$ paquetes por segundo. Pero si hay un solo paquete de datos, la tasa de servicio será la del canal que lo está atendiendo. Y si hay un paquete de voz, estará siendo atendido por su propio servidor de manera que los paquetes de datos sólo dispondrán de su propio servidor.

El procedimiento de análisis ya lo conocemos: Establecemos las ecuaciones de balance de flujo (que en este caso son cinco, como muestra la Figura 157), multiplicamos las dos ecuaciones genéricas por z^j y sumamos para los rangos de validez ($j \geq 1$ para la fila 1 y $j \geq 2$ para la fila 0), expresamos los resultados en términos de $G_0(z)$ y $G_1(z)$ como se definen en la Ecuación (34), y resolvemos las dos ecuaciones con las dos incógnitas para obtener expresiones cerradas para $G_0(z)$ y $G_1(z)$. De las tres superficies cerradas de la izquierda en la Figura 157,

$$(\lambda_1 + \lambda_2)p_{0,0} = \mu_1 p_{1,0} + \mu_2 p_{0,1}$$

$$(\mu_1 + \lambda_2)p_{1,0} = \lambda_1 p_{0,0} + \mu_2 p_{1,1}$$

$$(\lambda_1 + \lambda_2 + \mu_2)p_{0,1} = \lambda_2 p_{0,0} + \mu_1 p_{1,1} + 2\mu_2 p_{0,2}$$

Y de las dos superficies cerradas de la derecha en la misma figura

$$(\lambda_1 + \lambda_2 + 2\mu_2)p_{0,j} = \lambda_2 p_{0,j-1} + \mu_1 p_{1,j} + 2\mu_2 p_{0,j+1}, \quad j \geq 2$$

$$(\mu_1 + \mu_2 + \lambda_2)p_{1,j} = \lambda_2 p_{1,j-1} + \lambda_1 p_{0,j} + \mu_2 p_{1,j+1}, \quad j \geq 1$$

Multiplicamos las dos últimas ecuaciones por z^j y sumamos para el rango de validez de j :

$$(\lambda_1 + \lambda_2 + 2\mu_2) \sum_{j=2}^{\infty} p_{0,j} z^j = \lambda_2 \sum_{j=2}^{\infty} p_{0,j-1} z^j + \mu_1 \sum_{j=2}^{\infty} p_{1,j} z^j + 2\mu_2 \sum_{j=2}^{\infty} p_{0,j+1} z^j$$

$$(\mu_1 + \mu_2 + \lambda_2) \sum_{j=1}^{\infty} p_{1,j} z^j = \lambda_2 \sum_{j=1}^{\infty} p_{1,j-1} z^j + \lambda_1 \sum_{j=1}^{\infty} p_{0,j} z^j + \mu_2 \sum_{j=1}^{\infty} p_{1,j+1} z^j$$

Expresamos las sumas en términos de $G_0(z) = \sum_{j=0}^{\infty} p_{0,j} z^j$ y $G_1(z) = \sum_{j=0}^{\infty} p_{1,j} z^j$,

$$(\lambda_1 + \lambda_2 + 2\mu_2)(G_0(z) - p_{0,0} - p_{0,1}z) =$$

$$z\lambda_2(G_0(z) - p_{0,0}) + \mu_1(G_1(z) - p_{1,0} - p_{1,1}z) + 2\mu_2 \frac{1}{z}(G_0(z) - p_{0,0} - p_{0,1}z - p_{0,2}z^2)$$

$$(\mu_1 + \mu_2 + \lambda_2)(G_1(z) - p_{1,0}) = \lambda_2 z G_1(z) + \lambda_1(G_0(z) - p_{0,0}) + \mu_2 \frac{1}{z}(G_1(z) - p_{1,0} - p_{1,1}z)$$

Agrupando y reemplazando de las tres primeras ecuaciones

$$G_1(z)[z\mu_1 + (z-1)(\mu_2 - \lambda_2 z)] = \lambda_1 z G_0(z) + \mu_2(z-1)p_{1,0}$$

$$G_0(z)[z\lambda_1 + (z-1)(2\mu_2 - \lambda_2 z)] = \mu_1 z G_1(z) + \mu_2(z-1)(2p_{0,0} + zp_{0,1})$$

Dividiendo ambas ecuaciones a ambos lados por μ_1 , obtenemos las expresiones en términos de los parámetros α , ρ_1 y ρ_2 :

$$G_1(z)[z + \alpha(z-1)(1 - \rho_2 z)] = \rho_1 z G_0(z) + \alpha(z-1)p_{10}$$

$$G_0(z)[z\rho_1 + \alpha(z-1)(2 - \rho_2 z)] = zG_1(z) + \alpha(z-1)(2p_{00} + zp_{01})$$

Como una simple verificación, al evaluar en $z=1$ las dos ecuaciones anteriores obtenemos

$$G_1(1) = \rho_1 G_0(1) \text{ y, como } G_0(1) + G_1(1) = 1,$$

$$P_B = G_1(1) = \rho_1 / (1 + \rho_1)$$

$$1 - P_B = G_0(1) = 1 / (1 + \rho_1)$$

Que era lo que esperábamos porque el tráfico de voz experimenta el servicio de una cola M/M/1/1.

De otro lado, si derivamos respecto a z ,

$$[z + \alpha(z-1)(1 - \rho_2 z)] \frac{d}{dz} G_1(z) + G_1(z)[1 + \alpha(1 - \rho_2 z) - \rho_2 \alpha(z-1)] = \rho_1 z \frac{d}{dz} G_0(z) + \rho_1 G_0(z) + \alpha p_{10}$$

$$[z\rho_1 + \alpha(z-1)(2 - \rho_2 z)] \frac{d}{dz} G_0(z) + G_0(z)[\rho_1 + \alpha(2 - \rho_2 z) - \alpha(z-1)\rho_2] = G_1(z) + z \frac{d}{dz} G_1(z) + \alpha(z-1)p_{01} + \alpha(2p_{00} + zp_{01})$$

Evalutando en $z=1$,

$$\frac{d}{dz} G_1(z=1) + G_1(z=1)[1 + \alpha(1 - \rho_2)] = \rho_1 \frac{d}{dz} G_0(z=1) + \rho_1 G_0(z=1) + \alpha p_{10}$$

$$\rho_1 \frac{d}{dz} G_0(z=1) + G_0(z=1)[\rho_1 + \alpha(2 - \rho_2)] = G_1(z=1) + \frac{d}{dz} G_1(z=1) + \alpha(2p_{00} + p_{01})$$

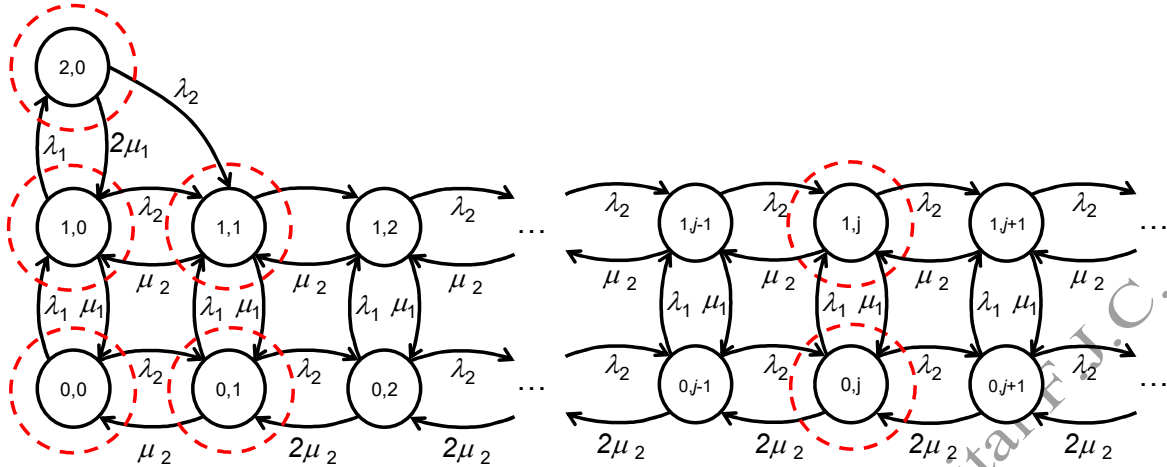
esto es,

$$\frac{d}{dz} G_1(1) + \frac{\rho_1}{1 + \rho_1} [1 + \alpha(1 - \rho_2)] = \rho_1 \frac{d}{dz} G_0(1) + \frac{\rho_1}{1 + \rho_1} + \alpha p_{10}$$

$$\rho_1 \frac{d}{dz} G_0(1) + \frac{1}{1 + \rho_1} [\rho_1 + \alpha(2 - \rho_2)] = \frac{\rho_1}{1 + \rho_1} + \frac{d}{dz} G_1(1) + \alpha(2p_{00} + p_{01})$$

(...)

e. Dos canales con frontera móvil para ambos tipos de tráfico



De las cinco superficies cerradas de la izquierda en la figura anterior obtenemos,

$$(\lambda_1 + \lambda_2)p_{0,0} = \mu_1 p_{1,0} + \mu_2 p_{0,1}$$

$$(\lambda_1 + \lambda_2 + \mu_2)p_{0,1} = \lambda_2 p_{0,0} + \mu_1 p_{1,1} + 2\mu_2 p_{0,2}$$

$$(\mu_1 + \lambda_1 + \lambda_2)p_{1,0} = \lambda_1 p_{0,0} + \mu_2 p_{1,1} + 2\mu_1 p_{2,0}$$

$$(\mu_1 + \mu_2 + \lambda_2)p_{1,1} = \lambda_2 p_{1,0} + \lambda_1 p_{0,1} + \lambda_2 p_{2,0} + \mu_2 p_{1,2}$$

$$(2\mu_2 + \lambda_2)p_{2,0} = \lambda_1 p_{1,0}$$

Estas ecuaciones se pueden agrupar así

$$\begin{bmatrix} \mu_1 & 0 & 0 & 0 & 0 \\ 0 & \mu_1 & 2\mu_2 & 0 & 0 \\ \lambda_1 + \lambda_2 + \mu_1 & -\mu_2 & 0 & -2\mu_1 & 0 \\ -\lambda_2 & \mu_1 + \mu_2 + \lambda_2 & 0 & -\lambda_2 & -\mu_2 \\ -\lambda_1 & 0 & 0 & 2\mu_2 + \lambda_2 & 0 \end{bmatrix} \begin{bmatrix} p_{10} \\ p_{11} \\ p_{02} \\ p_{20} \\ p_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 + \lambda_2 & -\mu_2 \\ -\lambda_2 & \lambda_1 + \lambda_2 + \mu_2 \\ \lambda_1 & 0 \\ 0 & \lambda_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_{00} \\ p_{01} \end{bmatrix}$$

Conviene dividir por μ_1 los elementos de las matrices a ambos lados para obtener

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2\alpha & 0 & 0 \\ \rho_1 + \alpha\rho_2 + 1 & -\alpha & 0 & -2 & 0 \\ -\alpha\rho_2 & 1 + \alpha + \alpha\rho_2 & 0 & -\alpha\rho_2 & -\alpha \\ -\rho_1 & 0 & 0 & 2\alpha + \alpha\rho_2 & 0 \end{bmatrix} \begin{bmatrix} p_{10} \\ p_{11} \\ p_{02} \\ p_{20} \\ p_{12} \end{bmatrix} = \begin{bmatrix} \rho_1 + \alpha\rho_2 & -\alpha \\ -\alpha\rho_2 & \rho_1 + \alpha\rho_2 + \alpha \\ \rho_1 & 0 \\ 0 & \rho_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_{00} \\ p_{01} \end{bmatrix}$$

para despejar las cinco probabilidades de la izquierda en términos de p_{00} y p_{01} :

$$\begin{bmatrix} p_{10} \\ p_{11} \\ p_{02} \\ p_{20} \\ p_{12} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2\alpha & 0 & 0 \\ \rho_1 + \alpha\rho_2 + 1 & -\alpha & 0 & -2 & 0 \\ -\alpha\rho_2 & 1 + \alpha + \alpha\rho_2 & 0 & -\alpha\rho_2 & -\alpha \\ -\rho_1 & 0 & 0 & 2\alpha + \alpha\rho_2 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 + \alpha\rho_2 & -\alpha \\ -\alpha\rho_2 & \rho_1 + \alpha\rho_2 + \alpha \\ \rho_1 & 0 \\ 0 & \rho_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_{00} \\ p_{01} \end{bmatrix}$$

De las dos superficies cerradas de la derecha en la misma figura obtenemos

$$(\lambda_1 + \lambda_2 + 2\mu_2)p_{0,j} = \lambda_2 p_{0,j-1} + \mu_1 p_{1,j} + 2\mu_2 p_{0,j+1}, \quad j \geq 2$$

$$(\mu_1 + \mu_2 + \lambda_2)p_{1,j} = \lambda_2 p_{1,j-1} + \lambda_1 p_{0,j} + \mu_2 p_{1,j+1}, \quad j \geq 2$$

Multiplicamos las dos últimas ecuaciones por z^j y sumamos para el rango de validez de j :

$$(\lambda_1 + \lambda_2 + 2\mu_2) \sum_{j=2}^{\infty} p_{0,j} z^j = \lambda_2 \sum_{j=2}^{\infty} p_{0,j-1} z^j + \mu_1 \sum_{j=2}^{\infty} p_{1,j} z^j + 2\mu_2 \sum_{j=2}^{\infty} p_{0,j+1} z^j$$

$$(\mu_1 + \mu_2 + \lambda_2) \sum_{j=2}^{\infty} p_{1,j} z^j = \lambda_2 \sum_{j=2}^{\infty} p_{1,j-1} z^j + \lambda_1 \sum_{j=2}^{\infty} p_{0,j} z^j + \mu_2 \sum_{j=2}^{\infty} p_{1,j+1} z^j$$

Expresamos las sumas de las dos ecuaciones anteriores en términos de $G_0(z) = \sum_{j=0}^{\infty} p_{0,j} z^j$ y

$$G_1(z) = \sum_{j=0}^{\infty} p_{1,j} z^j,$$

$$(\lambda_1 + \lambda_2 + 2\mu_2)(G_0(z) - p_{0,0} - p_{0,1}z) = \lambda_2 z(G_0(z) - p_{0,0}) + \mu_1(G_1(z) - p_{1,0} - p_{1,1}z) + 2\frac{\mu_2}{z}(G_0(z) - p_{0,0} - p_{0,1}z - p_{0,2}z^2)$$

$$(\mu_1 + \mu_2 + \lambda_2)(G_1(z) - p_{1,0} - p_{1,1}z) = \lambda_2 z(G_1(z) - p_{1,0}) + \lambda_1(G_0(z) - p_{0,0} - p_{0,1}z) + \frac{\mu_2}{z}(G_1(z) - p_{1,0} - p_{1,1}z - p_{1,2}z^2)$$

Nuevamente dividimos por μ_1 ambas ecuaciones a ambos lados para obtener

$$(\rho_1 + \alpha\rho_2 + 2\alpha)(G_0(z) - p_{0,0} - p_{0,1}z) = \alpha\rho_2 z(G_0(z) - p_{0,0}) + (G_1(z) - p_{1,0} - p_{1,1}z) + 2\alpha(G_0(z) - p_{0,0} - p_{0,1}z - p_{0,2}z^2) / z$$

$$(1 + \alpha + \alpha\rho_2)(G_1(z) - p_{1,0} - p_{1,1}z) = \alpha\rho_2 z(G_1(z) - p_{1,0}) + \rho_1(G_0(z) - p_{0,0} - p_{0,1}z) + \alpha(G_1(z) - p_{1,0} - p_{1,1}z - p_{1,2}z^2) / z$$

Estas dos ecuaciones se pueden agrupar de la siguiente manera:

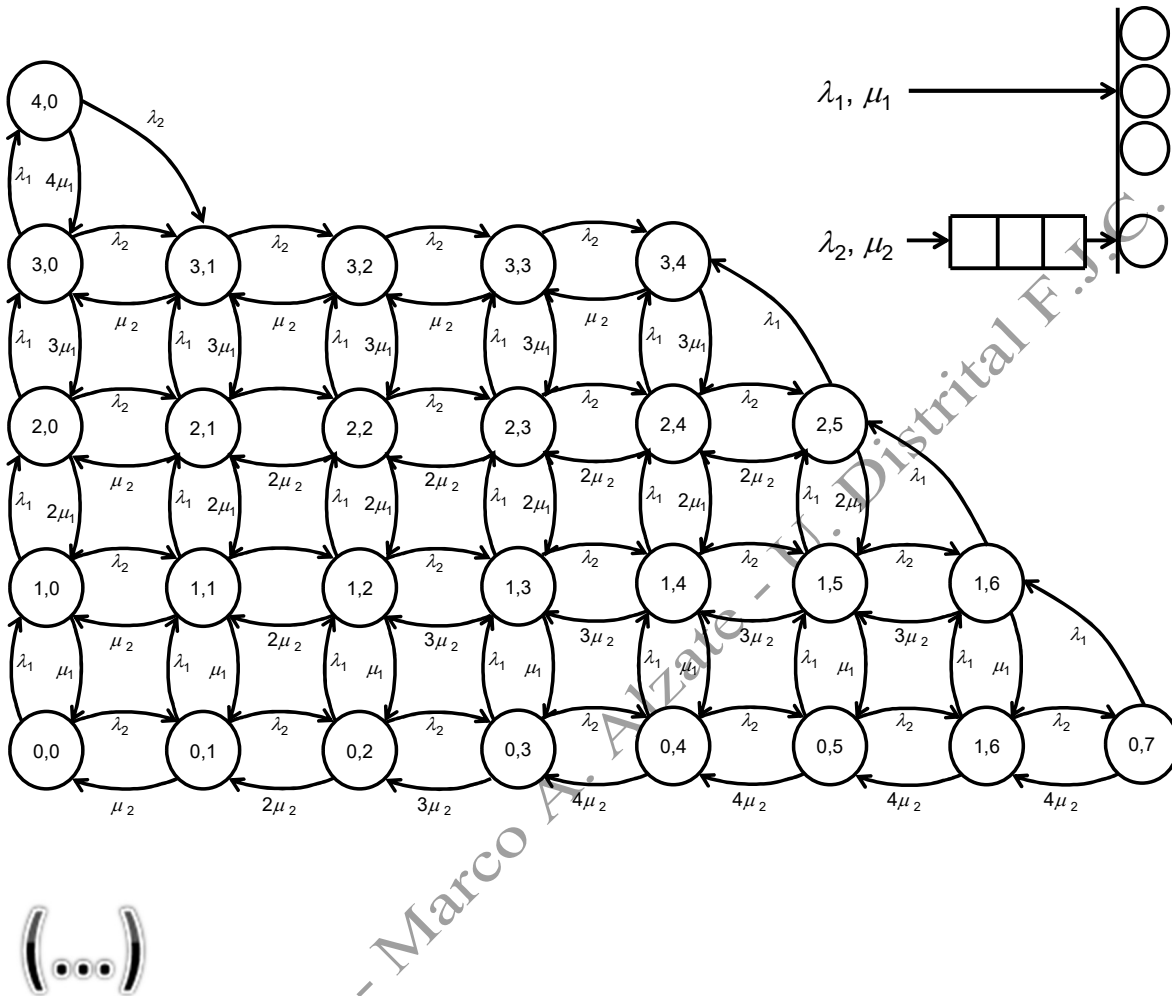
$$\begin{bmatrix} \alpha(z-1)(z\rho_2-2)/z-\rho_1 & 1 \\ \rho_1 & -1-\alpha(1-1/z-\rho_2(z-1)) \end{bmatrix} \begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \begin{bmatrix} \alpha(z-1)(z\rho_2-2)/z-\rho_1 & 2\alpha-z(2\alpha+\rho_1+\alpha\rho_2) \\ \rho_1 & z\rho_1 \end{bmatrix} \begin{bmatrix} p_{00} \\ p_{01} \end{bmatrix} + \dots$$

$$\begin{bmatrix} 1 & z & 2\alpha z & 0 \\ -1-\alpha(1-1/z-\rho_2(z-1)) & \alpha-z(1+\alpha+\alpha\rho_2) & 0 & \alpha z \end{bmatrix} \begin{bmatrix} p_{10} \\ p_{11} \\ p_{02} \\ p_{12} \end{bmatrix}$$

Combinando las matrices, obtenemos expresiones para $G_0(z)$ y $G_1(z)$ en términos de los parámetros ρ_1 , ρ_2 y α , y de las probabilidades p_{00} y p_{01} . Como $G_0(z) + G_1(z) = 1$, podemos dejarlo todo en términos de p_{00} .



f. N canales con frontera móvil



El siguiente programa simula un multiplexor integrado de dos flujos de Poisson, uno que no usa cola porque no tolera el retardo y otro que usa cola porque no tolera pérdidas (les llamaremos voz y datos, respectivamente, por darles algún nombre). El primer flujo se caracteriza por una tasa de llegadas λ_1 y un tiempo promedio de servicio $1/\mu_1$, mientras el segundo flujo se caracteriza por una tasa de llegadas λ_2 y un tiempo promedio de servicio $1/\mu_2$. Hay N_1 canales dedicados al tráfico de voz y N_2 canales dedicados al tráfico de datos, aunque cada tipo de tráfico puede usar los canales del otro para aumentar la eficiencia en el uso de los recursos. El programa sólo muestra la creación de una traza muestral del número de paquetes de cada tipo en el sistema y el número de paquetes de datos en la cola, pues los detalles de la adquisición de estadísticas y el manejo de la cola se pueden añadir fácilmente a partir de la discusión de la definición 135.

```

1 function traza = DosFlujos(lambda1, lambda2, mu1, mu2, N1, N2, tiempoSimulacion)
2     N = N1+N2; % Número total de canales
3     eventosFuturos = inf*ones(N+2,1); % 1,2: llegadas.
4     eventosFuturos(1) = random('exp',1/lambda1); % i+2: salida por el canal i, i=1:N
                    % Programa las próximas llegadas

```

```
5 eventosFuturos(2) = random('exp',1/lambda2);
6 paquetesSistema = [0 0 0]; % Número de paquetes de voz, de datos, de
% datos en cola
7 lambda = lambda1 + lambda2; % Tasa total
8 traza = zeros(ceil(2*tiempoSimulacion*lambda),4); % El evento i ocurrió en el instante
% traza(i,1) y dejó traza(i,2)
% paquetes de voz, traza(i,3) paquetes de
% datos en el sistema, traza(i,4) paquetes
% de datos en la cola
9 numeroEvento = 0; % Reloj de la simulación
10 reloj = 0; % Inicia el ciclo de simulación
11 while reloj<tiempoSimulacion % Un evento más
12 numeroEvento = numeroEvento+1; % Registra (tiempo,estado)
13 traza(numeroEvento,:) = [reloj paquetesSistema]; % Actualiza el reloj y determina el
14 [reloj,evento] = min(abs(eventosFuturos)); % próximo evento
15 switch evento % Llega un paquete de voz
16 case 1 % Programa la próxima
17 eventosFuturos(1) = reloj + random('exp',1/lambda1); % llegada de voz
% Busca un canal de voz desocupado
18 I = find(eventosFuturos(3:2+N1)==inf);
19 if ~isempty(I)
20 eventosFuturos(I(1)+2) = reloj + random('exp',1/mu1); % Programa la salida
21 paquetesSistema = paquetesSistema + [1 0 0]; % Paquete de voz admitido
22 else
23 I = find(eventosFuturos(3:2+N1)<0); % Busca un canal de voz ocupado por un
24 if ~isempty(I) % paquete de datos
25 eventosFuturos(I(1)+2) = reloj + random('exp',1/mu1); % Programa la
% salida de este paquete
26 paquetesSistema = paquetesSistema + [1 0 1]; % Paquete de voz admitido. El
27 else % paquete de datos vuelve a la cola
28 I = find(eventosFuturos(3+N1:2+N1)==inf); % Busca un canal de datos libre
29 if ~isempty(I)
30 eventosFuturos(N1+I(1)+2) = -(reloj + random('exp',1/mu1)); % Programa
% la salida de este paquete
31 paquetesSistema = paquetesSistema + [1 0 0]; % Paquete de voz admitido
32 end
33 end
34 end
35 case 2 % Llega un paquete de datos
36 eventosFuturos(2) = reloj + random('exp',1/lambda2); % Programa la próxima
% llegada de datos
37 paquetesSistema = paquetesSistema + [0 1 0]; % Siempre será aceptado
38 I = find(eventosFuturos(N1+3:N+2)==inf); % Busca un canal de datos desocupado
39 if ~isempty(I)
40 eventosFuturos(N1+I(1)+2) = reloj + random('exp',1/mu2); % Programa la
41 else % salida de este paquete
42 I = find(eventosFuturos(N1+3:N+2)<0); % Busca un canal de datos ocupado por
43 if ~isempty(I) % un paquete de voz
44 eventosFuturos(N1+I(1)+2) = reloj + random('exp',1/mu2); % Programa la
% salida de este paquete
45 paquetesSistema = paquetesSistema + [-1 0 0]; % el paquete de voz que
46 else % ocupaba esta canal es descartado
47 I = find(eventosFuturos(3:N1+2)==inf); % Busca un canal de voz desocupado
48 if ~isempty(I)
49 eventosFuturos(I(1)+2) = -(reloj + random('exp',1/mu2)); % Programa la
50 else % salida de este paquete
51 paquetesSistema = paquetesSistema + [0 0 1]; % Admitido en la cola
52 end
53 end
54 end
55 otherwise % Sale un paquete
56 if evento > N1+2
57 if eventosFuturos(evento) < 0
58 paquetesSistema = paquetesSistema + [-1 0 0]; % Era un paquete de voz en
59 else % un canal de datos
```

```

60         paquetesSistema = paquetesSistema + [0 -1 0]; % Era un paquete de datos
61     end % en un canal de datos
62     signo = 1;
63 else
64     if eventosFuturos(evento) < 0
65         paquetesSistema = paquetesSistema + [0 -1 0]; % Era un paquete de datos
66     else % en un canal de voz
67         paquetesSistema = paquetesSistema + [-1 0 0]; % Era un paquete de voz en
68     end % un canal de voz
69     signo = -1;
70 end
71 if paquetesSistema(3) > 0
72     paquetesSistema = paquetesSistema + [0 0 -1];
73     eventosFuturos(evento) = signo*(reloj + random('exp',1/mu2)); % Programa la
74     % salida de este paquete
75     eventosFuturos(evento) = inf; % este enlace queda desocupado
76 end
77 end
78 end
79 traza = traza(1:numeroEvento,:); % Grafica la ocupación del sistema
80 stairs(traza(:,1),traza(:,2:4)); % (número de paquetes de cada tipo en el sistema)
81 legend('Paquetes de voz','Paquetes de datos','Paquetes en cola')
82 end

```

Listado 10. Simulador de un multiplexor integrado de dos flujos Poisson

La primera línea muestra los parámetros de entrada y salida del programa. Además de las tasas λ_1 , μ_1 , λ_2 y μ_2 y del número de canales N_1 y N_2 , se debe especificar el tiempo de simulación. Retorna un arreglo de 4 columnas, *traza*, que trae una línea por cada evento procesado durante la simulación con los siguientes campos: El tiempo de ocurrencia del evento ocupa la primera columna; el número de paquetes de voz en el sistema en ese instante ocupa la segunda columna; el número de paquetes de datos en el sistema en ese instante ocupa la tercera columna; la cuarta columna contiene el número de paquetes de datos en cola. El número de paquetes de datos en servicio es la diferencia entre la tercera y la cuarta columnas.

En las líneas 2 y 3 se dimensiona la lista de eventos futuros. La salida por cada canal es un tipo de evento distinto y, como hay llegadas de paquetes de voz y llegadas de paquetes de datos, en total hay $N_1 + N_2 + 2$ tipos de eventos. Los dos primeros tipos de eventos son los dos tipos de llegadas, que se inicializan en las líneas 4 y 5 con instantes exponencialmente distribuidos, mientras que los demás eventos son las salidas que se inicializan en infinito, indicando que todos los canales empiezan desocupados. La línea 6 inicializa el número de paquetes en el sistema, que es un arreglo de tres variables, indicando el número de paquetes de voz en el sistema, el número de paquetes de datos en el sistema y el número de paquetes de datos en cola, respectivamente. Las líneas 7 y 8 inicializan en ceros el arreglo *traza* y las líneas 9 y 10 inicializan el contador de eventos y el reloj de simulación. La línea 11 da inicio a la simulación, que durará mientras no se cumpla el tiempo máximo de simulación.

Las líneas 12 y 13 actualizan la *traza* con el evento que haya acabado de transcurrir, mientras la línea 14 selecciona el próximo evento y actualiza el reloj al instante de ocurrencia del mismo. La línea 15 selecciona el proceso a seguir de acuerdo con el tipo de evento. Si es la llegada de un paquete de voz (línea 16), se programa la próxima llegada (línea 17) y se busca un canal de voz desocupado, esto es, con tiempo de salida infinito (línea 18). Si se encuentra, se programa la próxima salida por ese canal y se incrementa el número de paquetes de voz en el sistema (líneas 19 a 21). En otro caso, en la línea

23 se busca un canal de voz ocupado por un paquete de datos. Para esto se usa un truco muy simple: se hacen negativos los tiempos de salida de los paquetes de un tipo que ocupen un canal reservado para el otro tipo (por eso en la línea 14 se usó la función valor absoluto). Así, si en los primeros N_1 canales hay una futura salida en un tiempo negativo, es porque ese canal lo ocupa un paquete de datos y el nuevo paquete de voz tiene derecho a reclamar el uso de su canal. Si lo encuentra, reprograma la salida por este canal interrumpiendo el servicio del paquete de datos (línea 25) y enviándolo de nuevo a la cola al tiempo que cuenta la admisión de este paquete de voz en el sistema (línea 26). Si hasta aquí no ha encontrado un canal que lo atienda es porque todos los canales de voz están ocupados atendiendo paquetes de voz, así que ahora debe buscar un paquete de datos libre (línea 28). Si lo encuentra, el paquete es admitido y el próximo tiempo de salida del canal seleccionado se hace negativo indicando que, aunque es un canal de datos, lo ocupa un paquete de voz (línea 30). Si ninguna de las búsquedas fue exitosa, el paquete de voz es rechazado. El proceso de llegada de un paquete de datos (líneas 35 a 54) es casi idéntico, excepto que el paquete siempre es aceptado (línea 37), si escoge un canal de datos ocupado por un canal de voz, el paquete seleccionado será rechazado (línea 45) y, si no encuentra ningún canal dispuesto a atenderlo, el nuevo paquete no es bloqueado sino que entra a la cola (línea 51).

Las líneas 55 a 77 procesan la salida de un paquete. Si el canal que se desocupa era de datos (línea 56) y el tiempo del evento era negativo (línea 57), el paquete que sale era de voz (línea 58) o, si no, era de datos (línea 60). En cualquier caso, si hemos de atender un paquete de datos en la cola, el signo del tiempo de salida debe ser positivo, pues se trata de un canal de datos (línea 62). Las líneas 63 a 70 hacen el mismo proceso para un canal de voz. Si quedan paquetes de datos en la cola (línea 71), se atiende el primero de ellos en el canal recién desocupado (líneas 72 y 73). Si no, el canal queda libre (línea 75). Este proceso de eventos se repite mientras dure la simulación y, al final de la simulación, se trunca el arreglo traza para que incluya sólo los eventos que alcanzaron a ocurrir y se grafica la ocupación del sistema en función del tiempo (líneas 79 a 81).

148. Modelo MMDP de un abonado VoIP

Un abonado telefónico transita entre estados de actividad e inactividad según una cadena de Markov de dos estados como muestra la siguiente figura. α es la tasa de activación dado que está en el estado inactivo ($1/\alpha$ es el promedio del tiempo que el abonado dura en el estado inactivo, el cual está exponencialmente distribuido). β es la tasa de desactivación dado que se está en el estado activo ($1/\beta$ es el promedio del tiempo que el abonado dura en el estado activo, el cual está exponencialmente distribuido). La probabilidad del estado activo es $\pi_1 = \alpha/(\alpha+\beta)$ y la probabilidad del estado inactivo es $\pi_0 = \beta/(\alpha+\beta)$. Durante el estado activo se generan V paquetes por segundo según una tasa constante y durante el estado inactivo no se generan paquetes. En promedio, el abonado genera $\lambda = \alpha V/(\alpha+\beta)$. Como en cada estado se tienen llegadas determinísticas, este modelo se conoce como MMDP –Markov Modulated Deterministic Process–.

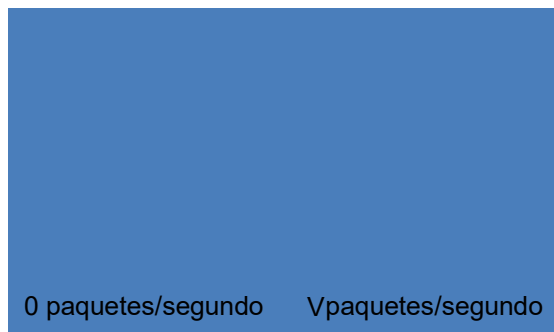


Figura 158. Modelo CTMC de dos estados para un abonado de voz

La transición entre estados de sonido y de silencio se aprecia en la siguiente figura, donde se muestra la actividad de un abonado B durante 8,5 segundos de una conversación telefónica con otro abonado A. Hay silencios largos mientras se escucha al abonado A y silencios cortos correspondientes a las pausas durante el monólogo debidas a fenómenos prosódicos asociados a signos de puntuación (coma, punto seguido, puntos suspensivos, etc.). Otros silencios aún más breves corresponden a la producción de fonemas explosivos como /k/, /t/ o /p/, en los que el tracto vocal se cierra completamente durante una fracción de segundo (entre la parte posterior de la lengua y el velo del paladar para /k/, entre la punta de la lengua y el paladar duro para /t/ o entre los labios para /p/). Con este cierre del tracto vocal se acumula energía que después se libera abruptamente. A tasas entre 16 y 64 kbps, típicas para conversaciones telefónicas, aún durante estos pequeños silencios se pueden transmitir varios miles de bits.

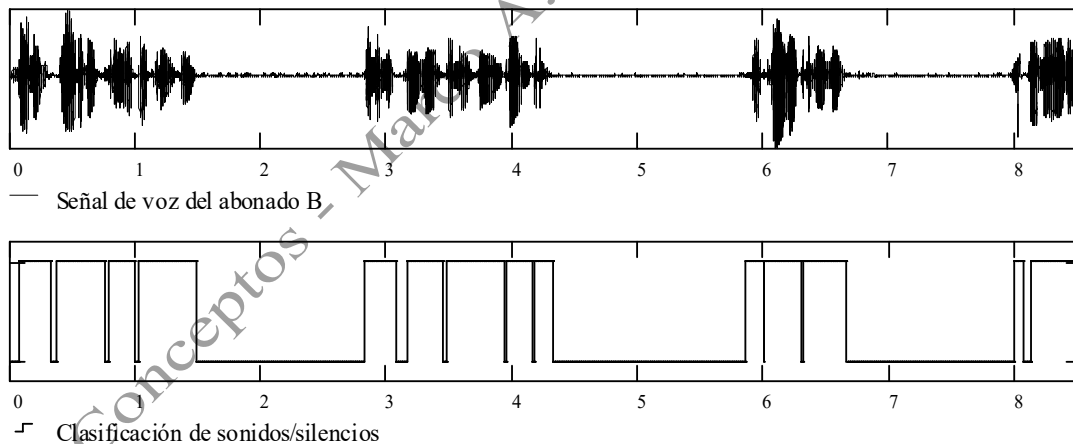


Figura 159. Intervalos de sonido y silencio de un abonado en una conversación telefónica

Los períodos de sonido y silencio intercalados permiten a las señales de voz compartir dinámicamente los enlaces en una red de comunicación. La idea principal es la de producir paquetes de voz solamente durante los períodos de sonido, de manera que los canales se puedan utilizar para otra información durante los períodos de silencio, como muestra la siguiente figura.

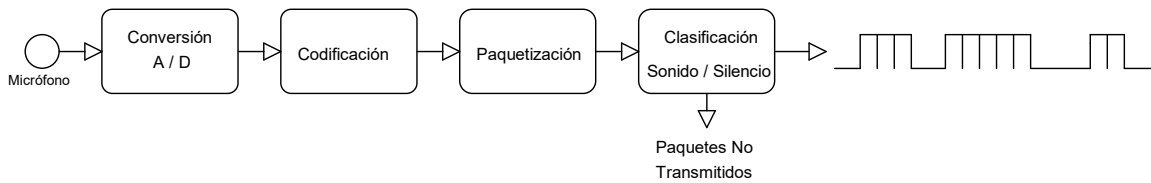


Figura 160. Transmisión de paquetes de voz sólo durante los períodos de sonido

Durante un período de sonido el abonado produce V paquetes/segundo, donde el valor específico de V depende del codec utilizado (por ejemplo, si se usa codificación lineal predictiva a 16 kbps, V podría corresponder a 20 paquetes por segundo, de 100 bytes cada paquete). Si el tiempo promedio de un período de silencio es $1/\alpha$ y el tiempo promedio de un período de sonido es $1/\beta$, la fracción de tiempo que un abonado permanece activo es, en promedio, $\alpha/(\alpha+\beta)$, de manera que el abonado genera en promedio sólo $\lambda = V\alpha/(\alpha+\beta)$ paquetes por segundo. Surge entonces la pregunta ¿cómo se ve afectado el desempeño ofrecido por un multiplexor de paquetes de voz cuando se quiere aprovechar la ventaja de la multiplexión? Para responder, debemos construir un modelo de llegada de paquetes de voz que represente las características mencionadas de la actividad de los abonados.

Como de costumbre, el modelo debe obedecer a una simplificación que represente el compromiso entre la suposición de independencia (lo ideal desde el punto de vista de la facilidad del tratamiento analítico) y la inclusión de la compleja estructura de correlación que existe en el proceso de llegadas a múltiples escalas de tiempo (lo ideal desde el punto de vista del realismo de los modelos). Un compromiso intermedio es suponer un modelo de Markov, donde se acepta una dependencia entre el pasado y el futuro, pero se supone que ellos son condicionalmente independientes dado el presente. Esto es, supondremos que los promedios $1/\alpha$ y $1/\beta$ de los períodos de silencio y de sonido, respectivamente, son los parámetros de dos distribuciones exponenciales independientes, lo que genera una cadena de Markov de dos estados como se mostró en la **Figura 158**.

Como muestra la siguiente figura, la suposición de tiempos exponenciales de sonido es bastante precisa, pero no así la suposición de tiempos exponenciales de silencio, donde parece ajustarse mejor una distribución con cola pesada. La suposición de independencia es inexacta para ambos estados del abonado. Sin embargo, este modelo simplificado ha demostrado ser suficientemente útil en el diseño de multiplexores de paquetes de voz, donde el tamaño de los buffers y la capacidad de los enlaces se diseñan para una tasa de pérdida o un retardo promedio dado.

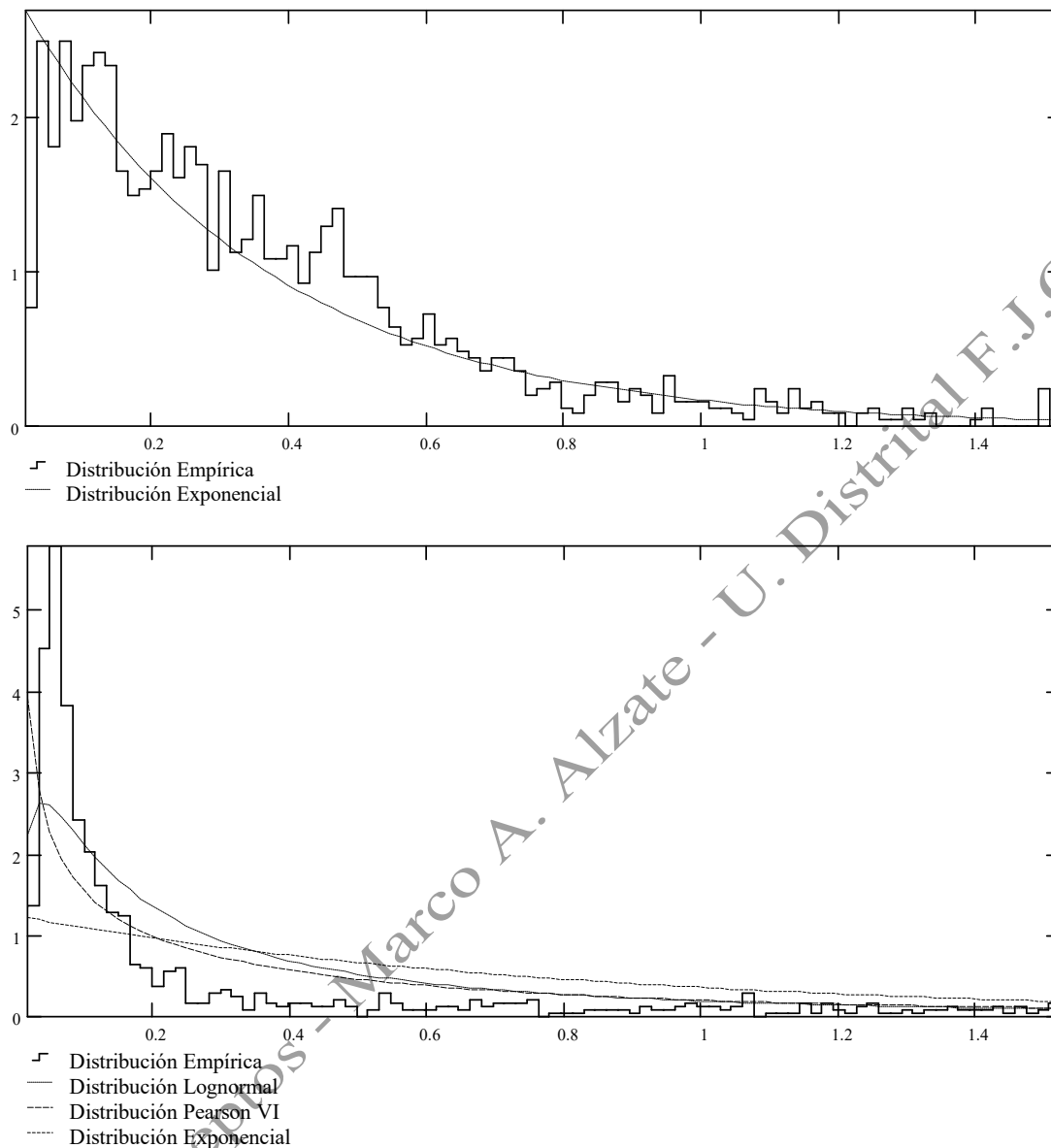


Figura 161. Los periodos de sonido son, efectivamente, exponenciales. Los periodos de silencio no.

La siguiente figura muestra los instantes de llegada de paquetes correspondientes a un flujo de voz y un flujo concentrado de múltiples sesiones de datos interactivos.

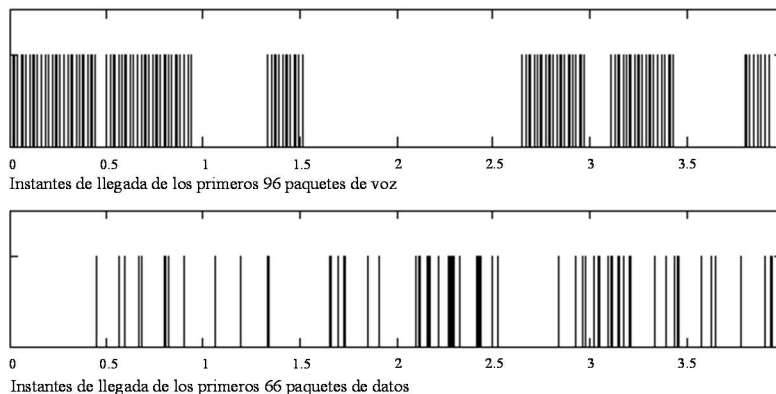


Figura 162. Instantes de llegadas de paquetes de voz y datos

149. Modelo MMDP para N abonados de voz

Al considerar el flujo mezclado de N abonados de voz se tiene otro proceso determinístico markovianamente modulado, donde la tasa de llegadas es jV cuando hay j abonados activos. En ese estado cada uno de los j abonados activos se desactiva a una tasa β , de manera que la tasa total de transición al estado $j-1$ es $j\beta$. Igualmente, en ese estado existen $N-j$ abonados inactivos, cada uno activándose a una tasa α , de manera que la tasa total de transición al estado $j+1$ es $(N-j)\alpha$. El diagrama de transición entre estados correspondiente es el proceso de nacimiento y muerte de la siguiente figura.

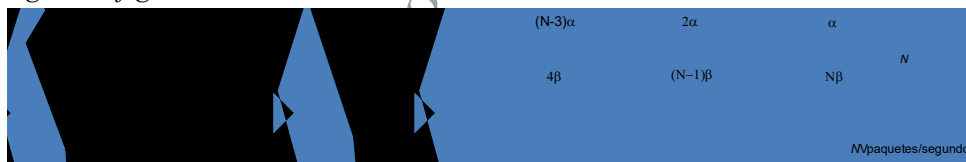


Figura 163. Modelo CTMC para los estados de actividad de N abonados de voz

La probabilidad de que haya j abonados activos es

$$P[j \text{ activos}] \triangleq \pi_j = \binom{N}{j} \left(\frac{\alpha}{\alpha + \beta} \right)^j \left(\frac{\beta}{\alpha + \beta} \right)^{N-j}, \quad j = 0, 1, \dots, N,$$

con promedio $N\alpha/(\alpha + \beta)$, de manera que, en promedio, se generan $\lambda = N\alpha/(\alpha + \beta)$ paquetes por segundo.

En efecto, la matriz generadora de esta cadena de Markov es

$$Q = \begin{bmatrix} -N\alpha & N\alpha & 0 & 0 & \dots & 0 & 0 \\ \beta & -(N-1)\alpha - \beta & (N-1)\alpha & 0 & \dots & 0 & 0 \\ 0 & 2\beta & -(N-2)\alpha - 2\beta & (N-2)\alpha & \dots & 0 & 0 \\ 0 & 0 & 3\beta & -(N-3)\alpha - 3\beta & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\alpha - (N-1)\beta & \alpha \\ 0 & 0 & 0 & 0 & \dots & N\beta & -N\beta \end{bmatrix}$$

Las ecuaciones de balance de flujo, $\bar{\pi}Q = \bar{0}$, establecen que

$$j\beta\pi_j = (N-j+1)\alpha\pi_{j-1}, \quad j=1\dots N$$

de donde

$$\pi_1 = \frac{N\alpha}{\beta} \pi_0, \quad \pi_2 = \frac{(N-1)\alpha}{2\beta} \pi_1 = \frac{N(N-1)}{1 \cdot 2} \left(\frac{\alpha}{\beta}\right)^2 \pi_0, \quad \pi_3 = \frac{(N-2)\alpha}{3\beta} \pi_2 = \frac{N(N-1)(N-2)}{1 \cdot 2 \cdot 3} \left(\frac{\alpha}{\beta}\right)^3 \pi_0$$

y, en general,

$$\pi_j = \binom{N}{j} \left(\frac{\alpha}{\beta}\right)^j \pi_0$$

Aplicando la condición de normalización,

$$\sum_{j=0}^N \pi_j = 1 = \pi_0 \sum_{j=0}^N \binom{N}{j} \left(\frac{\alpha}{\beta}\right)^j = \pi_0 \left(1 + \frac{\alpha}{\beta}\right)^N \Rightarrow \pi_0 = \left(\frac{\beta}{\alpha + \beta}\right)^N$$

Remplazando en la expresión anterior para π_j ,

$$\pi_j = \binom{N}{j} \left(\frac{\alpha}{\beta}\right)^j \left(\frac{\beta}{\alpha + \beta}\right)^N = \binom{N}{j} \left(\frac{\alpha}{\alpha + \beta}\right)^j \left(\frac{\beta}{\alpha + \beta}\right)^{N-j}$$

que es, por supuesto, la distribución binomial que resulta de sumar N variables de Bernoulli independientes (el estado de actividad de cada abonado), como vimos en la definición 43(c). En esa misma definición encontramos que el número promedio de abonados activos es $N\alpha/(\alpha+\beta)$ y, como cada abonado activo genera exactamente V paquetes por segundo, la tasa promedio de generación de paquetes de los N abonados es $\lambda = NV\alpha/(\alpha+\beta)$.

El estado de actividad de N abonados de voz es fácil de simular, pues sabemos que el tiempo de permanencia en cada estado es una variable aleatoria distribuida exponencialmente con promedio igual al inverso de la tasa de salida de este estado, por lo que, una vez alcanzamos un estado particular, podremos programar el instante futuro en que saldremos de él. Por ejemplo, del estado j se sale con una tasa total $r_j = (N-j)\alpha + j\beta$, por lo que el tiempo de permanencia en el estado j es una variable aleatoria exponencialmente distribuida con promedio $1/r_j$. ¿Y hacia dónde debemos dirigirnos? La fracción de transiciones por segundo a cada estado vecino sobre el número total de transiciones por segundo nos da la probabilidad de escoger cada destino. Por ejemplo, una vez salimos del estado j , una fracción $j\beta/r_j$ de las veces nos dirigimos al estado $j-1$ y una fracción $(N-j)\alpha/r_j$ de las veces nos dirigimos al estado $j+1$. Mientras permanece en el estado j , la cadena produce jV paquetes por segundo (los distribuimos uniformemente para que se genere exactamente un paquete cada $1/jV$ segundos). El programa del Listado 11 realiza dicha simulación.

```

1 function MMDP(alpha, beta, N, V, tiempoSimulacion)
2   reloj = 0; % Reloj de simulación
3   Estado = 0; % Número de abonados activos
4   numeroCambioDeEstado = 0; % Contador de cambios de estado
5   numeroDeLlegadas = 0; % Contador de llegadas de paquetes
6   traza = zeros(ceil(ceil(3*N*(alpha+beta)*tiempoSimulacion)),2);
7   % Numero de abonados activos en función del tiempo
8   InstanteDeLlegada = zeros(ceil(N*V*tiempoSimulacion),1);
9   % Instantes de llegada de los paquetes
10  TN = zeros(N+1,1); % Acumula el tiempo que permanece en cada estado
11  tue = 0; % Instante del último cambio de estado
12  tasa = Estado*beta + (N-Estado)*alpha; % Tasa de cambio de estado (en este estado)
13  ProximoCambioDeEstado = random('exp',1/tasa); % Duración en el estado actual
14  ProximaLlegadaDePaquete = 1/(Estado*V); % Cada abonado activo produce V paquetes/segundo
15  while reloj<tiempoSimulacion % Mientras dure la simulación
16    [reloj, evento] = min([ProximoCambioDeEstado ProximaLlegadaDePaquete]);
17    switch evento
18      case 1 % Actualiza el reloj y selecciona el próximo evento
19        % Cambia el estado de actividad de los abonados
20        numeroCambioDeEstado = numeroCambioDeEstado+1; % Un cambio de estado más
21        traza(numeroCambioDeEstado,:) = [tue Estado]; % Registra el evento
22        TN(Estado+1)=TN(Estado+1) + reloj - tue; % Acumula el tiempo en el estado actual
23        tue = reloj; % Recuerda este instante como el último cambio de estado
24        Pmas_uno = (N-Estado)*alpha/tasa; % Probabilidad de pasar al siguiente estado
25        if rand() <= Pmas_uno
26          Estado = Estado + 1; % Un abonado inactivo se activa
27        else
28          Estado = Estado - 1; % Un abonado activo de inactiva
29        end
30        tasa = Estado*beta + (N-Estado)*alpha; % Nueva tasa de cambio de estado
31        ProximoCambioDeEstado = reloj + random('exp',1/tasa); % Programa el próximo camb
32        if ((ProximaLlegadaDePaquete == inf) && (Estado>0)) % Si no habian paquetes por llegar
33          ProximaLlegadaDePaquete = reloj; % Este abonado se activa generando un nuevo paquete
34        end
35      case 2 % Llega un paquete de voz
36        numeroDeLlegadas = numeroDeLlegadas + 1; % Un paquete más que llega
37        InstanteDeLlegada(numeroDeLlegadas) = reloj; % Registra la llegada
38        ProximaLlegadaDePaquete = reloj + 1/(Estado*V); % Programa la próxima llegada
39    end
40  end
41  traza = traza(1:numberCambioDeEstado,:); % Tamaño efectivo de este arreglo
42  InstanteDeLlegada=InstanteDeLlegada(1:numberDeLlegadas); % Tamaño efectivo de este arreglo
43  subplot(131); stairs(traza(:,1),traza(:,2)) % Grafica el número de abonados activos
44  title('Abonados activos'); xlabel('Tiempo en segundos'); axis([0 reloj 0 max(traza(:,2))+1])
45  TN = TN/sum(TN); % Fracción de tiempo en que permanecieron n abonados activos
46  subplot(132); stem(0:N,TN)
47  p = binopdf(0:N,N,alpha/(alpha+beta)); % Probabilidad de que haya n abonados activos
48  hold on; stem(0:N,p,'rx') % Compara los resultados experimentales con los teóricos
49  axis([-1 N+1 -0.02 1.5*max(p)])
50  title('Distribución de probabilidad del número de abonados activos')
51  legend('Simulación', 'Teoría'); xlabel('Número de abonados activos'); ylabel('Probabilidad')
52  subplot(133); stem(InstanteDeLlegada,ones(size(InstanteDeLlegada)),'.');
53  axis([0 reloj -0.1 1.1])
54  title('Instantes de llegadas de los paquetes de voz'); xlabel('Tiempo en segundos')

```

Listado 11. Simulador de las llegadas de paquetes de voz generadas por N abonados

Los parámetros de entrada del programa son la tasa de activación α , la tasa de desactivación β , el número de abonados N , la tasa de paquetes por segundo generada por un abonado activo, v , y el tiempo máximo de simulación, tiempoSimulacion , (línea 1). El reloj de simulación y el estado del sistema (número de abonados activos) inician en cero (líneas 2 y 3). El programa genera una traza de instantes de llegadas de paquetes (InstanteDeLlegada) y una traza de estado de actividad de los abonados (traza)—líneas 4 a 7—. En el arreglo TN mantenemos un registro del tiempo que el sistema ha permanecido en cada uno de los $N+1$ estados (líneas 8 y 9). La lista de eventos futuros está

compuesta por dos variables, `ProximoCambioDeEstado` y `ProximaLlegadaDePaquete`. En las líneas 10 a 11 se programan los primeros eventos (en el Estado 0, donde inicia, el tiempo del próximo cambio de estado está exponencialmente distribuido con promedio $N\alpha$ y la próxima llegada se espera que ocurra en un tiempo infinito). Mientras dure la simulación (línea 13), se selecciona el evento futuro más próximo y se actualiza el reloj al instante de ocurrencia de ese evento inminente (línea 14). Si es un cambio de estado (líneas 15 y 16), se actualiza la traza de número de abonados activos (líneas 17 y 18), se acumula el tiempo que los abonados permanecieron en el estado actual (líneas 19 y 20), y se selecciona el próximo estado (líneas 21 a 26). Después se programa el próximo cambio de estado (líneas 27 y 28) y, si no había ninguna llegada programada y se acaba de activar un abonado, se programa la llegada del primer paquete de este nuevo período con abonados activos (líneas 29 a 31). Si el evento es la llegada de un paquete (línea 32), se registra el instante de esta llegada (líneas 33 y 34) y se programa la próxima llegada –que puede ser dentro de infinitos segundos si estamos en el estado 0, pues en el estado j se genera un paquete cada $1/jV$ segundos– (línea 35). Al final de la simulación se dibuja la traza del estado de actividad de los abonados (líneas 40 y 41), se calcula y se grafica la fracción de tiempo que los abonados estuvieron en cada estado (líneas 42 y 43), comparando esta última fracción con la probabilidad teórica de cada estado (líneas 44 a 48). Finalmente se grafican los instantes de llegada de cada paquete (líneas 49 a 51). Esta última traza sería la que se use para alimentar el simulador de una cola, como veremos en la siguiente definición.

La Figura 164 muestra el resultado de una simulación de 10 abonados durante 100 segundos con $\alpha=0.56 \text{ s}^{-1}$, $\beta=0.83 \text{ s}^{-1}$ y $V=10 \text{ p/s}$. Con una sola simulación durante tan breve tiempo, se aprecia la tendencia de las fracciones de tiempo de permanencia en cada estado hacia la verdadera distribución de probabilidad, lo cual es una validación inicial del programa simulador. De otro lado, la Figura 165 muestra cómo la tasa de llegadas de paquetes cambia con el estado de actividad de los abonados.

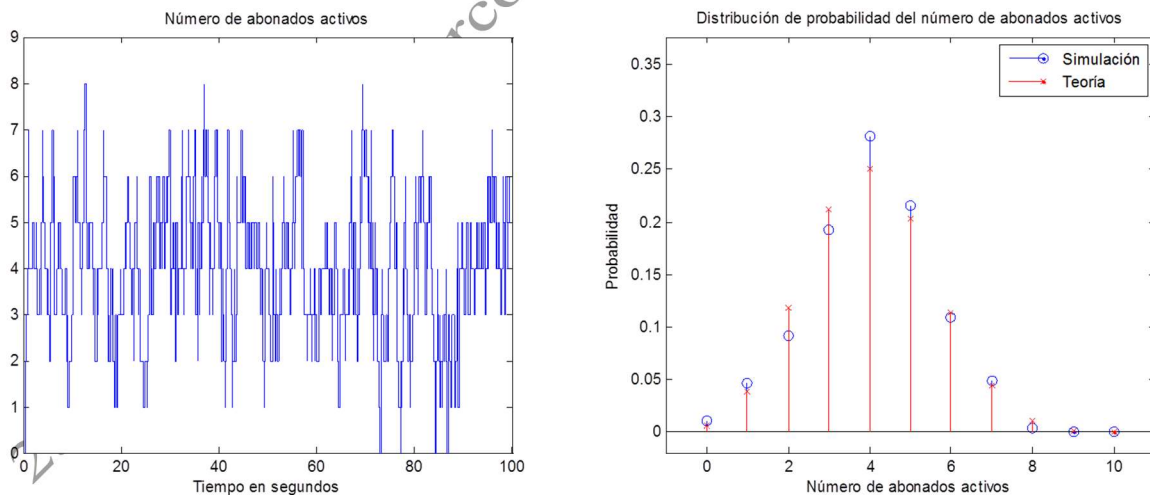


Figura 164. Simulación del estado de actividad de 10 abonados durante 100 segundos

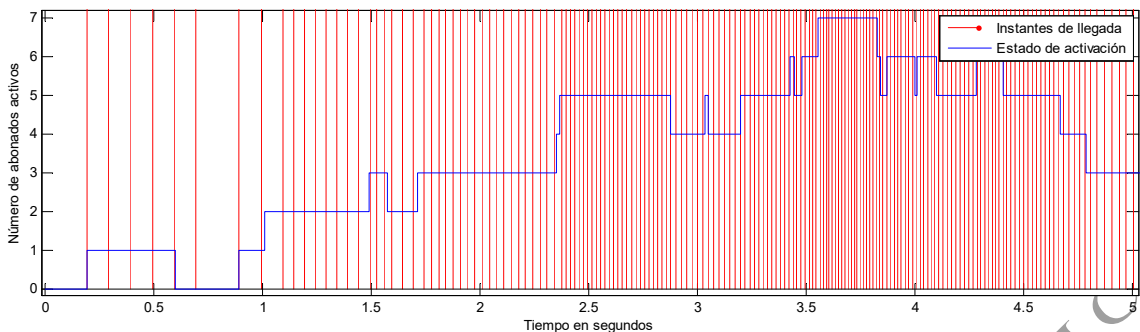


Figura 165. Estado de actividad de los abonados e instantes de llegadas de paquetes

150. Modelo de flujo continuo para un multiplexor de paquetes de voz MMDP/D/1

Consideramos un multiplexor que dirige los paquetes de N abonados de voz a través de un único enlace, como muestra la siguiente figura. Cada abonado se comporta como en la definición **¡Error! No se encuentra el origen de la referencia.**, de manera que la entrada a la cola forma un proceso MMDP con $N+1$ estados, como el de la definición **¡Error! No se encuentra el origen de la referencia.** Todos los paquetes tienen la misma longitud, de manera que el tiempo de servicio es determinístico con promedio $1/\mu$. Este es un modelo de cola MMDP/D/1.

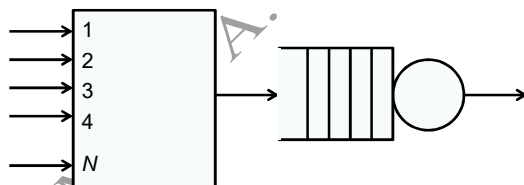


Figura 166. Multiplexor estadístico para los paquetes generados por N abonados de voz

La suposición del flujo continuo se refiere a que la granularidad de los paquetes es infinitesimal comparada con la tasa de llegadas durante un período de sonido. En particular, si el número de fuentes N y la capacidad del canal μ son suficientemente grandes, se puede despreciar la naturaleza discreta del buffer y suponer que existe un flujo continuo de paquetes. Entonces el número de paquetes en el sistema en el instante t , $X(t)$, corresponde a una variable aleatoria continua no negativa, $X(t) \in \mathbb{R}$. Suponiendo que el sistema se estabiliza, a la larga la función de distribución acumulativa de $X(t)$ se hace independiente del tiempo y toma la siguiente forma:

$$F_X(x) = 1 + \sum_{k: \text{Re}(\gamma_k) < 0} a_k \left(\sum_{j=0}^N \phi_{k,j} \right) e^{\gamma_k x}$$

donde γ_k y ϕ_k , $k=0..N$, son los valores propios y los vectores propios de la matriz $M=QD^{-1}$, siendo Q la matriz generadora de la cadena de Markov que representa el número de abonados activos, y D^{-1} la inversa de la matriz diagonal $D = \text{diag}(jV-\mu)$, $j=0..N$. Los

coeficientes a_k se obtienen del siguiente sistema de $N - \lfloor \mu/V \rfloor$ ecuaciones con $N - \lfloor \mu/V \rfloor$ incógnitas:

$$0 = \pi_j + \sum_{k: \text{Re}(\gamma_k) < 0} a_k \phi_{k,j}, \quad j > \lfloor \mu/V \rfloor$$



Figura 167. Modelos de flujo discreto y de flujo continuo

La manera como obtendremos los resultados de esta definición constituye una metodología típica de análisis de sistemas de colas con llegadas tipo MMDP.

Como la tasa promedio de llegada de paquetes por segundo es $\lambda = NV\alpha/(\alpha + \beta)$, la condición de estabilidad que necesitamos es $\mu > \lambda$. Claro, si $\mu \geq NV$, hay suficiente capacidad para darle a cada abonado el ancho de banda pico, V , con lo que estamos desaprovechando la ventaja del multiplexaje de paquetes. Por lo tanto, buscamos poder asignar una capacidad en el rango

$$NV \frac{\alpha}{\alpha + \beta} < \mu < NV$$

que satisfaga algún requerimientos de retardo. Suponiendo que $\mu/V \notin \mathbb{Z}$, sin perder generalidad, notamos que para los estados $j < \mu/V$ la cola está decreciendo y para los estados $j > \mu/V$ la cola está creciendo. En general, mientras permanezcan j abonados activos, la tasa de crecimiento de la cola es de $jV - \mu$ paquetes por segundo, donde nuestra suposición de flujo continuo implica que, en un intervalo infinitesimal Δt durante el cual permanezcan j abonados activos, el incremento de la cola es $\Delta x = (jV - \mu)\Delta t$ paquetes. Esta relación básica entre el estado de actividad de los abonados y la longitud de la

cola la podemos capturar si consideramos la distribución conjunta de dos procesos relacionados: $A(t)$ =Número de abonados activos en el instante t , $X(t)$ = Número de paquetes en el sistema en el instante t . Para cualquier instante t , $A(t) \in \{0, 1, 2, \dots, N\}$, pero nuestra suposición de flujo continuo implica que, en cualquier instante t , $X(t) \in \mathbb{R}^+$. La distribución conjunta de estos dos procesos la representaremos mediante la función

$$F_j(t, x) = P[A(t) = j, X(t) \leq x], j \in \{0, 1, \dots, N\}, x \in \mathbb{R}^+$$

que indica que en el instante t hay j abonados activos y x paquetes o menos en el sistema. Nos preguntaremos por $F_j(t+\Delta t, x)$, lo que pueda ocurrir dentro de Δt segundos con esta distribución. Si durante esos Δt segundos permanecieron j abonados activos, la cola creció en una cantidad $\Delta x = (jV - \mu)\Delta t$, de manera que en el instante t debió suceder el evento $X(t) \leq x - \Delta x$ para que en $t+\Delta t$ ocurra el evento $X(t+\Delta t) \leq x$. Respecto a la activación de los abonados, si en $t+\Delta t$ hay $A(t+\Delta t)=j$ abonados activos, lo más probable es que en t hubiera $A(t)=j$ activos y no haya habido ninguna transición, lo cual ocurriría con probabilidad $1 - [(N-j)\alpha + j\beta]\Delta t + o(\Delta t)$, pues $(N-j)\alpha + j\beta$ es la tasa con que la cadena sale del estado j (recuerde que $o(\Delta t)$ es cualquier función que tienda a cero más rápidamente que Δt , $(\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0)$). Pero pudo ocurrir también que hubiera $j-1$ abonados activos y se haya activado otro, lo cual ocurre con probabilidad $(N-j+1)\alpha\Delta t + o(\Delta t)$. Si había $j+1$ abonados activos, se debió desactivar uno de ellos, lo cual ocurre con probabilidad $(j+1)\beta\Delta t + o(\Delta t)$. O pudo haber cualquier número de abonados activos y se dio un conjunto de activaciones y desactivaciones que condujeron a j abonados activos, lo cual ocurre con probabilidad $o(\Delta t)$. En consecuencia, particionando el espacio muestral en el número de abonados activos que había en t , el teorema de la probabilidad total toma la siguiente forma:

$$\begin{aligned} F_j(t + \Delta t, x) = & o(\Delta t) \quad \dots \\ & + F_{j-1}(t, x - \Delta x)(N - j + 1)\alpha\Delta t \quad \dots \\ & + F_{j+1}(t, x - \Delta x)(j + 1)\beta\Delta t \quad \dots \\ & + F_j(t, x - \Delta x)[1 - ((N - j)\alpha + j\beta)\Delta t] \end{aligned} \quad , \quad j=0, 1, 2, \dots, N$$

Restemos a ambos lados $F_j(t, x - \Delta x)$ para obtener

$$\begin{aligned} F_j(t + \Delta t, x) - F_j(t, x - \Delta x) = & o(\Delta t) \quad \dots \\ & + F_{j-1}(t, x - \Delta x)(N - j + 1)\alpha\Delta t \quad \dots \\ & + F_{j+1}(t, x - \Delta x)(j + 1)\beta\Delta t \quad \dots \\ & - F_j(t, x - \Delta x)((N - j)\alpha + j\beta)\Delta t \end{aligned} \quad , \quad j=0, 1, 2, \dots, N$$

Ahora restemos y sumemos $F_j(t, x)$ a la izquierda

$$\begin{aligned} (F_j(t + \Delta t, x) - F_j(t, x)) + (F_j(t, x) - F_j(t, x - \Delta x)) = & o(\Delta t) \quad \dots \\ & + F_{j-1}(t, x - \Delta x)(N - j + 1)\alpha\Delta t \quad \dots \\ & + F_{j+1}(t, x - \Delta x)(j + 1)\beta\Delta t \quad \dots \\ & - F_j(t, x - \Delta x)((N - j)\alpha + j\beta)\Delta t \end{aligned}$$

Dividiendo a ambos lados por $\Delta t = \Delta x / (jV - \mu)$ obtenemos

$$\begin{aligned} \frac{F_j(t + \Delta t, x) - F_j(t, x)}{\Delta t} + (jV - \mu) \frac{F_j(t, x) - F_j(t, x - \Delta x)}{\Delta x} &= o(\Delta t) / \Delta t \quad \dots \\ &+ F_{j-1}(t, x - \Delta x)(N - j + 1)\alpha \quad \dots \\ &+ F_{j+1}(t, x - \Delta x)(j + 1)\beta \quad \dots \\ &- F_j(t, x - \Delta x)((N - j)\alpha + j\beta) \end{aligned}$$

Sólo resta tomar el límite cuando Δt tiende a cero para obtener el sistema de ecuaciones diferenciales parciales que describe el comportamiento dinámico de la distribución de probabilidad conjunta de los procesos $A(t)$ y $X(t)$:

$$\begin{aligned} \frac{\partial}{\partial t} F_j(t, x) + (jV - \mu) \frac{\partial}{\partial x} F_j(t, x) &= F_{j-1}(t, x)(N - j + 1)\alpha \quad \dots \\ &+ F_{j+1}(t, x)(j + 1)\beta \quad \dots, \quad j=0, 1, 2, \dots, N \\ &- F_j(t, x)((N - j)\alpha + j\beta) \end{aligned}$$

Esta ecuación es válida bajo cualquier condición de los parámetros pero, si suponemos que $\mu > \lambda = NV\alpha / (\alpha + \beta)$, la solución transiente se desvanecerá y sólo quedará la solución estable en la que $\lim_{t \rightarrow \infty} F_j(t, x) = F_j(x)$, de manera que su derivada parcial con respecto al tiempo será cero. Siendo así, una vez alcanzado el estado estable, la distribución conjunta de las variables aleatorias A y X (ya no son procesos estocásticos) obedecerá el siguiente sistema de $N+1$ ecuaciones diferenciales ordinarias:

$$(jV - \mu) \frac{d}{dx} F_j(x) = F_{j-1}(x)(N - j + 1)\alpha + F_{j+1}(x)(j + 1)\beta - F_j(x)((N - j)\alpha + j\beta), \quad j=0 \dots N$$

Nótese que la suposición $\mu/V \notin \mathbb{Z}$ no implica ninguna pérdida de generalidad: Si no fuera así, al evaluar la anterior ecuación en $j = \mu/V$ obtendríamos

$$F_j(x) = \frac{(N - j + 1)\alpha}{(N - j)\alpha + j\beta} F_{j-1}(x) + \frac{(j + 1)\beta}{(N - j)\alpha + j\beta} F_{j+1}(x), \quad j = \mu/V$$

con lo que podemos eliminar la ecuación diferencial correspondiente a $F_j(x)$ del sistema reemplazándola por la anterior combinación; el sistema se reduce de $N+1$ ecuaciones diferenciales ordinarias a sólo N ecuaciones.

Volviendo a la suposición $\mu/V \notin \mathbb{Z}$, si agrupamos las diferentes distribuciones en un vector fila,

$$\vec{F}(x) = [F_0(x) \quad F_1(x) \quad \dots \quad F_N(x)]$$

el sistema de ecuaciones toma la forma

$$\frac{d}{dx} \vec{F}(x) D = \vec{F}(x) Q$$

donde Q es la matriz generadora de la cadena de Markov que representa el número de abonados activos y D es la matriz diagonal $\text{diag}(jV - \mu), j=0 \dots N$. Definiendo $M = QD^{-1}$, el sistema toma la forma

$$\frac{d}{dx} \vec{F}(x) = \vec{F}(x) M$$

Donde $M = QD^{-1}$, esto es,

$$M = \begin{bmatrix} \frac{N\alpha}{\mu} & \frac{N\alpha}{V-\mu} & 0 & 0 & \dots & 0 & 0 \\ \frac{\beta}{\mu} & \frac{-(N-1)\alpha-\beta}{V-\mu} & \frac{(N-1)\alpha}{2V-\mu} & 0 & \dots & 0 & 0 \\ 0 & \frac{2\beta}{V-\mu} & \frac{-(N-2)\alpha-2\beta}{2V-\mu} & \frac{(N-2)\alpha}{3V-\mu} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \frac{3\beta}{2V-\mu} & \frac{-(N-3)\alpha-3\beta}{3V-\mu} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{-\alpha-(N-1)\beta}{(N-1)V-\mu} & \frac{\alpha}{NV-\mu} \\ 0 & 0 & 0 & 0 & \dots & \frac{N\beta}{(N-1)V-\mu} & \frac{N\beta}{NV-\mu} \end{bmatrix}$$

Esta matriz tiene $N+1$ vectores propios a la izquierda con sus respectivos valores propios,

$$\vec{\phi}_k M = \gamma_k \vec{\phi}_k, \quad k = 0, 1, \dots, N$$

de manera que la solución de la ecuación diferencial es

$$\vec{F}(x) = \sum_{k=0}^N a_k \vec{\phi}_k e^{\gamma_k x}$$

donde los coeficientes $\{a_k, k=0 \dots N\}$ se obtienen de un conjunto adecuado de "condiciones iniciales". Por ejemplo, como el buffer es infinito, por estabilidad los coeficientes correspondientes a valores propios con parte real mayor que cero deben ser cero: $a_k=0$ si $\text{Re}(\gamma_k) > 0$, y la forma de la solución se reduce a

$$\vec{F}(x) = \sum_{k: \text{Re}(\gamma_k) \leq 0} a_k \vec{\phi}_k e^{\gamma_k x}$$

Pero uno de los valores propios es necesariamente cero, pues $M=QD^{-1}$ y la matriz generadora Q satisface la propiedad

$$\vec{\pi} Q = \mathbf{0}, \quad \pi_j = \binom{N}{j} \left(\frac{\alpha}{\alpha + \beta} \right)^j \left(\frac{\beta}{\alpha + \beta} \right)^{N-j}$$

siendo $\vec{\pi}$ la distribución de estado estable del número de abonados activos. Como los vectores propios se pueden especificar hasta una constante (si $\vec{\phi}_k$ es un vector propio también lo es $\rho \vec{\phi}_k \forall \rho \in \mathbb{R}$), debemos escoger para el valor propio $\gamma_k=0$ el vector propio $a_k \vec{\phi}_k = \vec{\pi}$ pues, en efecto,

$$P[A=j] = P[A=j, X \leq \infty] = F_j(\infty) = \sum_{k: \text{Re}(\gamma_k) \leq 0} a_k \phi_{k,j} e^{\gamma_k x} = a_{k_0} \phi_{k_0,j} \Big|_{\gamma_{k_0}=0} = \pi_j \Rightarrow \vec{\pi} = a_{k_0} \vec{\phi}_{k_0}$$

Entonces la solución toma la forma

$$\vec{F}(x) = \vec{\pi} + \sum_{k: \text{Re}(\gamma_k) < 0} a_k \vec{\phi}_k e^{\gamma_k x}$$

Pero se puede demostrar que, para esta estructura de la matriz M , el número de valores propios con parte real negativa es exactamente $N - \lfloor \mu/V \rfloor$, que es el número de estados que hacen que la cola esté creciendo, $j > \lfloor \mu/V \rfloor$. En esos estados la probabilidad de que en el sistema haya 0 paquetes debe

ser nula, por lo que tenemos $N - \lfloor \mu / V \rfloor$ condiciones iniciales de donde obtener los restantes coeficientes a_k :

$$0 = \pi_j + \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \phi_{k,j}, \quad j > \lfloor \mu / V \rfloor$$

Una vez determinadas las distribuciones conjuntas $F_j(x) = P[A=j, X \leq x]$, podemos encontrar la función de distribución acumulativa del número de paquetes en el sistema:

$$F_X(x) = P[X \leq x] = \sum_{j=0}^N P[A=j, X \leq x] = \sum_{j=0}^N F_j(x) = 1 + \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \left(\sum_{j=0}^N \phi_{k,j} \right) e^{\gamma_k x}$$

Conociendo la CDF de la longitud de la cola podemos conocer todas las estadísticas que queramos. Por ejemplo, la densidad de probabilidad de la longitud de la cola es

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \gamma_k \left(\sum_{j=0}^N \phi_{k,j} \right) e^{\gamma_k x}$$

y, por consiguiente, el valor esperado de la longitud de la cola es

$$E[X] = \int_0^{\infty} x dF_X(x) = \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \gamma_k \left(\sum_{j=0}^N \phi_{k,j} \right) \int_0^{\infty} x e^{\gamma_k x} dx$$

$$E[X] = \sum_{k: \operatorname{Re}(\gamma_k) < 0} \frac{a_k}{\gamma_k} \left(\sum_{j=0}^N \phi_{k,j} \right)$$

de manera que, por la ley de Little, podemos conocer el retardo promedio de un paquete de voz en el sistema,

$$E[T] = \frac{1}{\lambda} E[X] = \frac{\alpha + \beta}{\alpha N V} \sum_{k: \operatorname{Re}(\gamma_k) < 0} \frac{a_k}{\gamma_k} \left(\sum_{j=0}^N \phi_{k,j} \right)$$

El modelo completo se puede evaluar fácilmente mediante un programa de computador:

1. Introduzca los parámetros del modelo α, β, N, V, μ
2. Construya la matriz $M_{(N+1) \times (N+1)}$
3. Determine los valores propios y los vectores propios DE LA IZQUIERDA, γ_k y $\vec{\phi}_k$ ($\vec{\phi}_k M = \gamma_k \vec{\phi}_k, \quad k=0,1,\dots,N$)
4. Encuentre los coeficientes $\{a_k, k: \operatorname{Re}(\gamma_k) < 0\}$ resolviendo el sistema de ecuaciones lineales

$$0 = \pi_j + \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \phi_{k,j}, \quad j > \lfloor \mu / V \rfloor$$

5. La función de distribución acumulativa de la ocupación del buffer es

$$F_X(x) = P[X \leq x] = 1 + \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \left(\sum_{j=0}^N \phi_{k,j} \right) e^{\gamma_k x}, \text{ y su función de densidad de probabilidad es}$$

$$f_X(x) = \sum_{k: \operatorname{Re}(\gamma_k) < 0} a_k \gamma_k \left(\sum_{j=0}^N \phi_{k,j} \right) e^{\gamma_k x}. \text{ Con ellas se puede calcular todas las estadísticas de}$$

interés.

6. Por ejemplo, el retardo promedio de cada paquete en el sistema es

$$E[T] = \frac{\alpha + \beta}{\alpha N V} \sum_{k: \text{Re}(\gamma_k) < 0} \frac{a_k}{\gamma_k} \left(\sum_{j=0}^N \phi_{k,j} \right)$$

Veamos un ejemplo sencillo con una única fuente de voz, $N=1$. La matriz de coeficientes del sistema de dos ecuaciones diferenciales es

$$M = \begin{bmatrix} \frac{\alpha}{\mu} & \frac{\alpha}{V-\mu} \\ -\frac{\beta}{\mu} & -\frac{\beta}{V-\mu} \end{bmatrix} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \begin{bmatrix} -\mu & 0 \\ 0 & V-\mu \end{bmatrix}^{-1}$$

La ecuación característica $|M^T - \gamma I| = \gamma[\gamma - (\alpha/\mu - \beta/(V-\mu))] = 0$ conduce a los eigenvalores $\gamma_0=0$ y $\gamma_1 = \alpha/\mu - \beta/(V-\mu)$, cuyos respectivos eigenvectores son $\phi_0 = [\beta/\alpha, 1]$ y $\phi_1 = [(V-\mu)/\mu, 1]$. En efecto, si escogemos $a_0 = \alpha/(\alpha+\beta)$ obtenemos $a_0 \phi_0 = \pi = [\beta, \alpha]/(\alpha+\beta)$, la distribución de probabilidad del número de abonados activos. Para que el sistema sea estable, es necesario que el segundo eigenvalor tenga parte real negativa: $\alpha/\mu - \beta/(V-\mu) < 0$ o, lo que es lo mismo, $V\alpha/(\alpha+\beta) = \lambda < \mu$ (la capacidad debe ser mayor a la demanda). En este caso, el coeficiente de ϕ_1 se encuentra de: $\pi_1 + a_1 \phi_{1,1} = 0$, de donde $a_1 = -a_0 = -\alpha/(\alpha+\beta)$. La distribución conjunta del estado del abonado y la longitud de la cola es

$$\bar{F}(x) = \frac{1}{\alpha + \beta} [\beta \quad \alpha] - \frac{\alpha}{\alpha + \beta} \begin{bmatrix} V-\mu & 1 \\ \mu & 1 \end{bmatrix} \exp\left(\begin{bmatrix} \alpha & \beta \\ \mu & V-\mu \end{bmatrix} x \right)$$

De manera que la CDF de la longitud de la cola es

$$F_x(x) = 1 - \frac{\alpha V}{\mu(\alpha + \beta)} \exp\left(\begin{bmatrix} \alpha & \beta \\ \mu & V-\mu \end{bmatrix} x \right)$$

Si consideramos la intensidad de tráfico

$$\rho = \frac{\lambda}{\mu} = \frac{\alpha V}{\mu(\alpha + \beta)} = \frac{\alpha}{C(\alpha + \beta)}$$

donde $C = \mu/V$ es una capacidad normalizada, podemos notar que la CDF de la longitud de la cola es la combinación convexa de dos distribuciones:

$$F_x(x) = (1 - \rho)u(x) + \rho(1 - \exp(-\theta x))$$

donde $u(x)$ es el escalón unitario (la CDF de la variable que vale cero con probabilidad 1) y $1 - \exp(-\theta x)$ es la CDF de una variable aleatoria exponencial con media $\frac{1}{\theta} = \mu \frac{1-C}{C(\alpha + \beta) - \alpha}$. La función

de densidad de probabilidad es

$$f_x(x) = (1 - \rho)\delta(x) + \rho\theta \exp(-\theta x)$$

de manera que $P[X=0] = 1 - \rho$ (ya lo sabíamos, $\rho = \lambda/\mu$ es la utilización del enlace en estos sistemas con un único servidor) y el valor esperado del número de paquetes en el sistema es

$$E[X] = (1 - \rho) \int_0^\infty x \delta(x) dx + \rho \theta \int_0^\infty x \exp(-\theta x) dx = \frac{\rho}{\theta} = \frac{\alpha V (V - \mu)}{(\alpha + \beta)(\beta \mu - \alpha (V - \mu))}$$

Estos resultados teóricos para un abonado de voz pueden compararse con el correspondiente resultado de simulación que considera la naturaleza discreta de los paquetes. Para esto construimos el programa de simulación del multiplexor de paquetes de voz mostrados en el Listado 12. Existen tres posibles eventos en la lista de eventos futuros: Un cambio en el número de abonados activos, la llegada un paquete y la salida de un paquete. En el primer evento se decide a qué estado transitar, se programa