PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU

Estadística Inferencial

para

Economía

2017

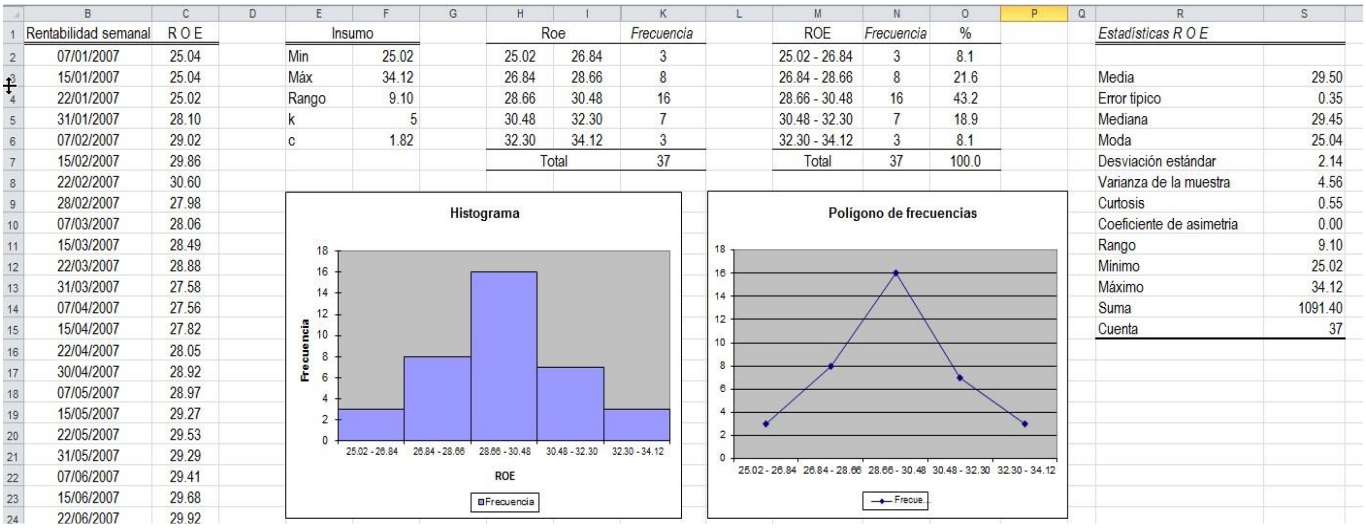
Capítulo 1

Probabilidad y Variable aleatoria

Introducción

En Economía, Finanzas, Gestión, etc., el analista trata con cifras que miden o representan resultados de procesos donde interactúan diversos agentes en condiciones de incertidumbre parcial, ve tendencias pero éstas no son exactas sino “patrones” que presentan cierta variabilidad.

Por ejemplo, la rentabilidad financiera (ROE=Beneficio neto/Fondos propios) de un banco, medida semana a semana no es constante, pero tampoco es caótica, examinando cifras (como las de abajo) se encuentran tendencias, valores que oscilan entre extremos bien definidos. Estas tendencias, cuantificadas adecuadamente, nos dicen “el estado” de una población o grupo, indicando qué es *lo que predomina*, qué es lo más frecuente y también *cuánta variabilidad* (diferencia arriba o debajo de lo predominante) *existe*.



Rentabilidad semanal	R O E
07/01/2007	25.04
15/01/2007	25.04
22/01/2007	25.02
31/01/2007	28.10
07/02/2007	29.02
15/02/2007	29.86
22/02/2007	30.60
28/02/2007	27.98
07/03/2007	28.06
15/03/2007	28.49
22/03/2007	28.88
31/03/2007	27.58
07/04/2007	27.56
15/04/2007	27.82
22/04/2007	28.05
30/04/2007	28.92
07/05/2007	28.97
15/05/2007	29.27
22/05/2007	29.53
31/05/2007	29.29
07/06/2007	29.41
15/06/2007	29.68
22/06/2007	29.92

Insumo	
Min	25.02
Máx	34.12
Rango	9.10
k	5
c	1.82

Roe		*Frecuencia*
25.02	26.84	3
26.84	28.66	8
28.66	30.48	16
30.48	32.30	7
32.30	34.12	3
Total		37

ROE	*Frecuencia*	%
25.02 - 26.84	3	8.1
26.84 - 28.66	8	21.6
28.66 - 30.48	16	43.2
30.48 - 32.30	7	18.9
32.30 - 34.12	3	8.1
Total	37	100.0

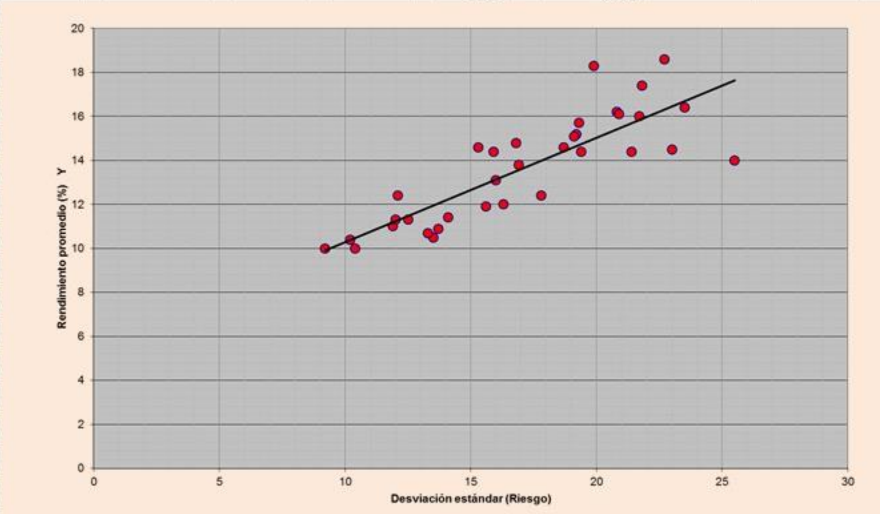
Estadísticas R O E	
Media	29.50
Error típico	0.35
Mediana	29.45
Moda	25.04
Desviación estándar	2.14
Varianza de la muestra	4.56
Curtosis	0.55
Coeficiente de asimetría	0.00
Rango	9.10
Mínimo	25.02
Máximo	34.12
Suma	1091.40
Cuenta	37

Saber más de un proceso debe ponerlo a uno en condición de indicar “*el estado de la población*” y también de “*explicar ese estado*” o sea de *decir el porqué del estado*. Para ello uno suele apoyarse en algunas características (o variables) que hayan mostrado estar asociadas a la variable de interés que mide el estado de la población o grupo. Si la explicación es buena, eso nos pone incluso en capacidad de *predecir*.

Por ejemplo, si tenemos datos sobre la rentabilidad promedio de diversos fondos mutuos y de la variación en rentabilidad de esos mismos fondos y juntamos las cifras para analizarlas, podemos distinguir un patrón de asociación que es muy conocido en el campo de las finanzas, la relación entre “rentabilidad y riesgo”. En este caso, el riesgo podemos medirlo como el promedio de variaciones tanto hacia arriba como debajo de la rentabilidad media de cada fondo durante el año. Como se trata de las oscilaciones arriba o debajo de lo esperado, es claro que miden de alguna manera el “riesgo”, riesgo entendido como posibilidad de recibir bastante menos de lo esperado con una inversión.

Las cifras de abajo muestran rentabilidades promedio y riesgos respectivos para un grupo de fondos mutuos. Graficando Riesgo versus Rentabilidad media se ve una clara tendencia lineal creciente, representable por una recta de la forma y = ax+b, que incluso puede ser estimada o cuantificada:

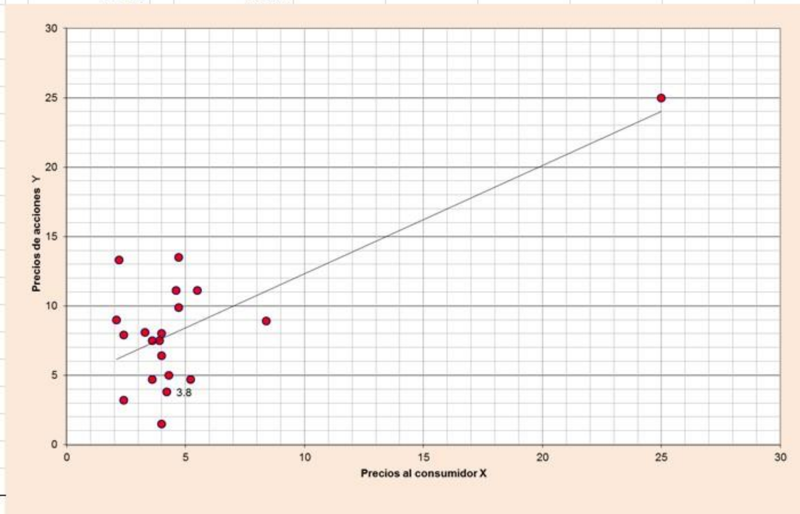
Fondo mutuo	Desviación estándar Riesgo X	Rendimiento anual promedio (%) Y	r_{XY}	y = ax + b	a	b	X	Y estimado
1	15.3	14.60	0.84		0.47	5.54	15.3	12.8
2	9.2	10.00					9.2	9.9
3	13.5	10.50					13.5	11.9
4	16.3	12.00					16.3	13.3
5	15.6	11.90						
6	12.1	12.40						
7	16.8	14.80						
8	19.3	15.70						
9	13.7	10.90						
10	21.4	14.40						
11	15.9	14.40						
12	11.9	11.00						
13	19.2	15.20						
14	18.7	14.60						
15	23.5	16.40						
16	23.0	14.50						
17	21.7	16.00						
18	19.1	15.10						
19	14.1	11.40						
20	25.5	14.00						



Pero no siempre las cosas lucen claras, incluso pueden ser engañosas, como lo muestran los siguientes datos relativos a las variaciones del Indice de precios al consumidor y las del Indice de bolsa de valores en una serie de países en los años 70:

Variación anual de acciones y precios al consumidor

	Tasa de variación anual (%)			
País	Precio de acciones Y	Precios al consumidor X	r_{XY} con Chile	r_{XY} sin Chile
Australia	5	4.3	0.755	0.096
Austria	11.1	4.6		
Belgica	3.2	2.4		
Canada	7.9	2.4		
Dinamarca	3.8	4.2		
Finlandia	11.1	5.5		
Francia	9.9	4.7		
Alemania	13.3	2.2		
India	1.5	4		
Irlanda	6.4	4		
Israel	8.9	8.4		
Italia	8.1	3.3		
Japón	13.5	4.7		
México	4.7	5.2		
Holanda	7.5	3.6		
Nueva Zelandia	4.7	3.6		
Suecia	8	4		
Reino Unido	7.5	3.9		
USA	9	2.1		
Chile	25	25		



La situación es más complicada: si procediéramos a estimar ***Y*** con una recta, usando ***X*** como variable explicativa o predictor, el error sería serio, como lo muestra el gráfico *XY*; en realidad no hay ninguna tendencia, pero el caso de Chile genera una tendencia artificial. ¿Qué ha ocurrido?

El problema subyacente es que nunca tendremos a mano información completa sino sólo muestras o partes de la información y esto induce variabilidad "azarosa", no sistemática, que debemos separar de las diferencias que sí son sistemáticas, que tienen fuente conocida. Es decir necesitamos herramientas analíticas para trabajar confiablemente con muestras. Eso lleva a la *Probabilidad* primero y a la *Estadística* después.

- **Estadística ⇔** Análisis racional de datos provenientes de muestras con fines de investigación.

- **“Racional”** quiere decir con fundamento no subjetivo, con fundamento científico ⇒ Fundamento matemático: En la Teoría de probabilidad, que no es Estadística pero sí le da su base lógica.

- Con resultados de una muestra siempre hay posibilidad que no sean “representativos” de aquello que realmente sucede en la población: que lo encontrado haya sido ocasionado básicamente por el azar, por la suerte.

- Para distinguir cuándo y en qué medida los resultados se deben al azar, se necesita aplicar herramientas conceptuales especiales: la Teoría de Probabilidad, que estudia “las leyes del azar”.

- “Las leyes del azar” sirven para tomar decisiones en **condiciones de incertidumbre**, como ocurre en el análisis de muestras, tomadas **“con fines de investigación” de algún proceso de interés**

- Investigar un proceso y explicarlo: Es poder asociar ciertos Resultados con determinadas Condiciones. Por ejemplo:
  Papiloma virus y cáncer uterino
  Días de la semana o días festivos y Alza en Precio de pasajes
  Maltrato infantil y autoestima baja.

- Estrategia de investigación:
  (1) Observar el proceso bajo diversas condiciones o ir variándolas y tomar nota de los cambios en los resultados.
  (2) Establecer o identificar algún patrón o “ley” asociando condiciones con determinado tipo de resultado. **(1) y (2) forman un “experimento”.**

- En lo que sigue desarrollaremos con más amplitud las ideas esbozadas líneas arriba.

1. Probabilidad: Enfoque Axiomático

En general, investigar o estudiar un proceso puede ser visto como establecer relaciones o conexiones entre determinados hechos o sucesos, que llamamos "condiciones" y otros sucesos de interés, que denominamos "resultados"

El camino usual es probar con distintas condiciones y ver los cambios en los resultados, hasta distinguir algún "patrón" o "ley", como por ejemplo, la Ley de Hooke de la Física o la Ley de Oferta y Demanda en Economía. Todo el proceso anterior es lo que constituye un experimento (o sea Condiciones y sus Resultados asociados).

Cuando dadas las condiciones no hay resultado único, sino un conjunto de resultados que ocurren siguiendo una distribución de frecuencias estable, el experimento se llama Experimento aleatorio y el estudio de estos experimentos es la teoría de probabilidad

En Economía y Ciencias Sociales abundan procesos como los descritos líneas arriba. Se caracterizan porque existe incertidumbre sobre los resultados, proveniente de nuestra falta de control total de las condiciones que los rigen. Esta incertidumbre dificulta el análisis y la toma de decisiones. Si estudiamos los precios de un valor bursátil, por ejemplo, encontraremos fuertes fluctuaciones. Quizá haya una tendencia (a la baja o al alza) pero además, alrededor de esa tendencia, veremos variaciones que impiden hacer un pronóstico exacto. Y es que en la determinación del precio hay dos componentes: una económica, que se origina en la interacción de los agentes y el estado de la economía, y otra que ya no puede explicarse económicamente, que tiene que ver con la subjetividad de las personas en el mercado, con sus miedos, gustos e intuiciones. Esta segunda componente resume la falta de conocimiento y control que tenemos sobre el proceso de oferta y demanda. Un análisis fino implica separar la primera componente de la segunda. Para ello podemos asumir (y esto parece razonable) que existen múltiples factores no controlables y fortuitos, que son responsables de las variaciones observadas. Variaciones que matizan un poco el "efecto" de la componente económica y que se distribuyen a lo largo del tiempo de manera relativamente estable, repartiendo por igual la buena y mala fortuna. Cuando hacemos esto, estamos aceptando la noción de "azar".

Hecho lo anterior, tenemos a continuación la tarea aislar el efecto del azar y ver qué tan fuerte es el efecto de la componente económica. Para hacer esta tarea de limpieza, necesitamos estudiar sistemática y racionalmente el azar, delimitarlo y encontrar las leyes que lo rigen. La manera más eficiente de tratar con este incómodo ente es analizarlo formalmente primero y luego, respetando sus propiedades, usar éstas para retirarlo de escena.

Una manera confiable de construir una teoría racional del azar, es usar un método que permita deducir sus propiedades a partir de un conjunto mínimo de premisas, de modo que dichas propiedades sean puramente lógicas, libres de error de inferencia. La Teoría Axiomática de Probabilidades hace lo anterior.

1.1 Conceptos Primitivos

Suceso: Es cualquier hecho que cuya ocurrencia o presentación nos interesa. Podemos clasificar a los sucesos en una de dos categorías a saber, Condiciones y Resultados.

Condición: Es un suceso cuya ocurrencia podemos controlar, es decir podemos hacer que suceda.

Resultado: Es un suceso cuya ocurrencia depende de un conjunto de condiciones que lo rigen.

Observación:
Como ya dijimos, una manera racional de investigar un proceso y explicarlo, es determinar las Condiciones en las que discurre, y luego ir variándolas, tomando nota de los cambios que se presentan en

los Resultados. Si llegamos a establecer algún patrón, estamos ante el germen de una "ley", e incluso, si nuestras observaciones las codificamos numéricamente, podemos enunciar la ley en la forma de alguna ecuación. Este sistema de análisis, consistente en realizar experiencias y ver cambios en los resultados asociados a ciertas condiciones es lo que llamaremos un "experimento".

Experimento: Es un conjunto de Condiciones (que definen el experimento) asociado a un conjunto de Resultados.

Para determinar la relación entre las condiciones y los resultados, necesitamos estar en la posibilidad de repetir o replicar el experimento, de modo que las sucesivas réplicas permitan la identificación y el estudio de la relación, incluyendo la posibilidad de hacer predicciones contrastables. En este contexto debemos distinguir dos tipos de experimentos o ensayos: los Determinísticos y los No Determinísticos.

Experimentos Determinísticos: Son aquellos experimentos en los cuales las Condiciones determinan unívocamente el Resultado.

En estos experimentos, sucesivas réplicas permiten establecer el resultado asociado a condiciones específicas, y la variación de las condiciones y el registro de los cambios en los resultados, conducen a una ley que se puede verificar empíricamente y con certeza.

Por ejemplo, si dejamos caer un dado desde una cierta altura y registramos el tiempo que tarde en llegar al piso, podemos medir el tiempo con exactitud razonable si tenemos los instrumentos adecuados. Variando la altura, encontraremos que el tiempo cambia y al final podemos establecer una "fórmula" que liga tiempo con altura. La verificación de la validez de la fórmula así deducida se puede hacer pronosticando tiempos y confrontándoles con los resultados de nuevos experimentos.

Experimentos No Determinísticos: Aquellos en donde las Condiciones no fijan el Resultado de manera unívoca.

En estos experimentos no hay un Resultado sino un Conjunto de Resultados, por tanto es más difícil establecer una correspondencia entre cambios en las Condiciones y cambios en los Resultados, pues al variar sistemáticamente las condiciones y registrar los cambios en los resultados, enfrentamos el problema de que los conjuntos de resultados no necesariamente son excluyentes. Sin embargo, en ciertos casos se encuentran regularidades aprovechables: al repetir el experimento, los resultados se tienden a presentarse manteniendo un patrón en la frecuencia con que ocurren.

Por ejemplo, en el caso del Consumo e Ingreso disponible de las familias en una economía, si analizamos datos de alguna encuesta grande, encontraremos que para cada nivel de ingreso habrá familias con distintos niveles de consumo. Sin embargo, trabajando sobre todos los ingresos y consumos registrados, podremos distinguir una "tendencia" creciente: a mayor ingreso mayor consumo. La relación no es exacta, determinista, pero existe y es directa. Y si calculamos en cuánto crece el consumo por cada unidad adicional de ingreso, probablemente encontraremos que esta tasa es casi constante, que hay una cierta regularidad. Esta regularidad, permite un tipo de pronóstico "relativo", así como una "ley" no exacta, que si bien no elimina la incertidumbre, la disminuye y la administra. Los experimentos donde se presenta este tipo de estabilidad estadística de resultados, reciben un nombre especial. Se llaman Experimentos Aleatorios.

Experimento Aleatorio: Denotado $\boldsymbol{\varepsilon}$, es un experimento donde las Condiciones no determinan de manera unívoca un resultado pero sí permiten establecer un Conjunto de posibles Resultados, de modo que en sucesivas réplicas del experimento, los resultados o grupos de resultados, se presentan con una frecuencia relativa (o porcentual) estable.

Ejemplo 1

(a) Soltar una tiza desde 1.5 metros y observar en cuántos trozos se parte

(b) Contar la cantidad de establecimientos que visita un consumidor hasta que se decide a comprar un bien específico.

(c) Escoger una muestra al azar de manzanas de una ciudad, entrevistar a los hogares de cada manzana y registrar el número de miembros de cada hogar que estaban sin trabajo la semana anterior a la entrevista.

Observación:
Aunque hay un conjunto de resultados posibles, en cada réplica del experimento, **sólo se presenta uno de ellos**, pudiendo variar el resultado de réplica en réplica.

En un experimento aleatorio no es posible saber con certeza el resultado del experimento, pero sí es factible establecer el conjunto de posibles resultados y se puede determinar (al menos conceptualmente) la frecuencia relativa (o porcentual) conque se presentan diferentes grupos de resultados.

Espacio Muestral: Por construcción, en todo experimento aleatorio ε estamos en condiciones de determinar el conjunto de posibles resultados. Este conjunto debidamente representado, se conoce como Espacio Muestral y lo denotaremos con la letra S.

Evento: Un evento es un subconjunto del espacio muestral. Los eventos se suelen denotar con letras mayúsculas: A, B, etc.

Por ejemplo, en el experimento (b), podemos representar S mediante el conjunto S = {1, 2, 3, 4,...}.

Ocurrencia de un evento: Diremos que un evento A "ocurre" si el resultado del experimento aleatorio es elemento de A.

Sobre un espacio S podemos definir o distinguir muchos subconjuntos y éstos pueden tener elementos comunes, es perfectamente posible que dos o más eventos ocurran simultáneamente. Por otra parte, el que un evento haya ocurrido quiere decir que uno de sus elementos fue el resultado del experimento. Obviamente, esto no significa que todos sus resultados se han presentado.

Evento Seguro: Es el espacio muestral. Por definición, ocurre siempre pues tiene todos los resultados que se pueden presentar.

Evento Imposible: Es el vacío $\varnothing$ y como no tiene elementos, nunca ocurre.
Eventos Mutuamente Excluyentes: Dos eventos A y B se dicen mutuamente excluyentes si carecen de elementos comunes. A y B no se presentan a la vez, nunca ocurren juntos. Formalmente se tiene que $A \cap B = \varnothing$

Observaciones:

(1) Para reducir la incertidumbre podemos usar la estabilidad estadística de la frecuencia relativa de aparición u ocurrencia de los distintos eventos de un espacio muestral, para "medir" en ellos su propensión a ocurrir y tener una herramienta que permita hacer pronósticos relativos.

(2) Como S ocurre siempre, nuestra medida de la propensión a ocurrir de los eventos, debe darle al espacio muestral S, un valor máximo. Análogamente, como $\varnothing$ nunca ocurre, debe recibir la medida más pequeña, acorde con su nula propensión a presentarse. Finalmente, los eventos de real interés, aquellos intermedios entre el vacío $\varnothing$ y el espacio total S, deben recibir una medida intermedia que los jerarquice desde menos propensos a ocurrir a más propensos a presentarse. Esta medida existe y

se llama “Probabilidad”. Por conveniencia la probabilidad se define de modo que esté entre 0 y 1, correspondiendo el 0 al vacío $\varnothing$ y el 1 al espacio S.

1.2 σ-álgebra de eventos

Dado un espacio muestral S, no siempre estaremos interesados en medir la opción de ocurrir de cualquier subconjunto de S, sino sólo de algunos básicos y otros adicionales que podamos obtener combinando los primeros. Es decir, *nos interesa una determinada familia de subconjuntos* de S. Esta familia será llamada σ-álgebra de eventos, la denotaremos con $\mathscr{A}$ y asumiremos que tiene las propiedades siguientes:

(**A$_1$**) Si $A \in \mathbf{A} \;\Rightarrow\; A^C \in \mathscr{A}$
(**A$_2$**) $\varnothing \in \mathscr{A}$
(**A$_3$**) Si $A_1, A_2, \cdots, A_k, \cdots$ es cualquier sucesión numerable de eventos en **A,** entonces $\bigcup_{k=1}^{\infty} A_k \in \mathscr{A}$

Propiedad 1
S $\in \mathscr{A}$**,** pues S **=** $\varnothing^C$, así que (**A$_1$**) implica la propiedad

Propiedad 2
Si $A_1, A_2, \cdots, A_N$ es una sucesión finita de eventos en $\mathscr{A}$**,** entonces $\bigcup_{k=1}^{N} A_k \in \mathscr{A}$ **,** pues basta definir los eventos $A_{N+1} = A_{N+2} = \cdots = \varnothing$ y entonces $\bigcup_{k=1}^{N} A_k = \bigcup_{k=1}^{\infty} A_k \in \mathscr{A}$ debido a las propiedades (**A$_2$**) y (**A$_3$**)

Propiedad 3
Si $A_1, A_2, \cdots, A_k, \cdots$ es cualquier sucesión numerable de eventos en $\mathscr{A}$**,** entonces $\bigcap_{k=1}^{\infty} A_k \in \mathscr{A}$ **,** pues por (**A$_1$**), $\forall\; A_k \in \mathscr{A} \;\Rightarrow\; A_k^C \in \mathscr{A}$**.** Además sabemos que $\left(\bigcap_{k=1}^{\infty} A_k\right)^C = \bigcup_{k=1}^{\infty} A_k^C$
y por (**A$_3$**), esta última unión también pertenece a la familia $\mathscr{A}$

Propiedad 4
Si $A_1, A_2, \cdots, A_N$ es una sucesión finita de eventos en $\mathscr{A}$**,** entonces $\bigcap_{k=1}^{N} A_k \in \mathscr{A}$ **,** ya que basta definir los eventos $A_{N+1} = A_{N+2} = \cdots = S$ y entonces $\bigcap_{k=1}^{N} A_k = \bigcap_{k=1}^{\infty} A_k \in \mathscr{A}$ debido a la propiedad 3.
Las propiedades anteriores muestran que una familia definida según **(A$_1$)** a **(A$_3$)**, contiene a todos los eventos que podamos construir por unión e intersección de conjuntos, o sea es “cerrada” bajo estas operaciones.

Ejemplo 2
Lanzamos un dado y observamos el número que muestra en su cara superior. En este caso $S = \{1,2,3,4,5,6\}$. Sea $A = \{2,4,6\}$.

(a) Si definimos $\mathscr{A} = \{S, \varnothing, A, A^C\}$**,** entonces $\mathscr{A}$ es una σ-álgebra
(b) Si definimos $\mathscr{A} = \{S, \varnothing\}$**,** entonces $\mathscr{A}$ es también una σ-álgebra
(c) Si definimos $\mathscr{A} = 2^S$ (el “conjunto de partes” o “conjunto potencia” de S), luego $\mathscr{A}$**,**así definida, es otra σ-álgebra

Observaciones:

(1) Nótese que sobre un mismo espacio S hemos definido varias σ-álgebras. Por otra parte puede ser curioso el caso (a), pero no es difícil imaginar un juego de azar cuyas reglas impliquen que sólo interese si ocurre un número par o no, más que estar pendientes de resultados individuales.

(2) También vale la pena notar que aunque (**A₃**) alude a una sucesión numerable e infinita de subconjuntos de S, este axioma sí es aplicable a las tres familias definidas en el ejemplo, pues basta "completar" cualquier sucesión finita definiendo más eventos, todos de la forma $A_k = \varnothing$.

(3) Si $\mathscr{C}$ es cualquier colección de eventos que no es σ-álgebra, siempre podemos "completarla" de modo que se obtenga una σ-álgebra, añadiendo subconjuntos de S convenientemente. Por ejemplo completando hasta llegar a 2^S **,** aunque esta extensión puede ser excesiva, dando una familia demasiado "grande". La alternativa más económica es definir $\mathscr{A}$ como intersección de todas las σ-álgebras que contienen a $\mathscr{C}$. Este caso especial se denota $\sigma(\mathscr{C})$ y es $\sigma(\mathscr{C}) = \bigcap_{\mathscr{C}\in\mathscr{A}_i} \mathscr{A}_i$. **Un caso importante ocurre cuando $\mathscr{C}$ es la familia de todos los intervalos del eje real, en este contexto $\sigma(\mathscr{C})$ es llamada "σ-álgebra de Borel".**

(4) De aquí en adelante reservamos la palabra evento para los subconjuntos de S que además son elementos de una σ-álgebra. La razón es que cuando S es conjunto no numerable (como el intervalo [0,1]) puede encontrase subconjuntos de S a los cuales no se les puede asignar ninguna probabilidad sin generar contradicciones lógicas, cosa que no ocurre cuando se trata de eventos que pertenecen a una σ-álgebra.

1.3 Axiomas y propiedades de la Probabilidad

La medición de la incertidumbre mediante la probabilidad se ha intentado varias veces en la historia de la Matemática, así tenemos:

1.3.1 Definición Clásica de Probabilidad (o de La Place)

Si un espacio muestral S tienen n(S) elementos todos con similar opción de presentarse, n(A) de los cuales también son elementos de un evento A, la probabilidad de A, denotada P(A), se define como

$$P(A) = \frac{n(A)}{n(S)}$$

Observaciones:

(1) Esta definición es la de los juegos de azar; la que se aplica en los juegos de cartas, por ejemplo.

(2) El defecto de la definición es que no siempre es aplicable, pues hay espacios con infinitos elementos (el ejemplo (b) es de ese tipo) o siendo finitos, sus elementos no son equiprobables. Por ejemplo, si en un dado borramos el número 6 y escribimos un 1, es claro que este último caso tiene el doble de opción de ocurrir que otros, sin embargo al ser cinco los elementos de S, la definición clásica le asigna una probabilidad de 1/5 y no de 2/6 como debiera ser.

Ejemplo 3

Si una persona contesta al azar una pregunta de opción múltiple con cinco opciones de respuesta, digamos a, b, hasta e:

$S = \{a, b, c, d, e\}$ y si **a** es la opción correcta, el evento A="La persona acierta" = {a}, y $P(A) = \frac{n(A)}{n(S)} = \frac{1}{5} = 0.20$; Si B es el evento B = "No acierta", entonces B={b,c,d,e} y $P(B) = \frac{n(B)}{n(S)} = \frac{4}{5} = 0.80$

Ejemplo 4

Si una persona contesta al azar dos preguntas de opción múltiple con cinco opciones de respuesta cada una, digamos a_1, b_1,…,e_1 para la primera pregunta y a_2,b_2,…,e_2 para la segunda pregunta, entonces

S = {(X,Y)| X= a_1, b_1,…,e_1; Y= a_2,b_2,…,e_2 }

y si a_1 es la opción correcta en la primera pregunta y c_2 es la opción correcta en la segunda:
Sea el evento A="La persona acierta en todo" = {(a_1,c_2)}, entonces:
n(S) = 5×5=25 y n(A) = 1, luego $P(A)=\frac{n(A)}{n(S)}=\frac{1}{25}=0.04$

1.3.2 Definición Frecuencial de Probabilidad (o de Von Mises)

Si un experimento aleatorio **ε** se repite N veces y el evento A ocurre en N_A de esas veces, la probabilidad de A, denotada P(A) es

$$P(A)=\lim_{N\to\infty}\frac{N_A}{N}$$

Observaciones:

(1) Esta definición es estadística y según ella, la probabilidad es el límite de una frecuencia relativa. En este caso, el punto de vista de la probabilidad es actuarial.

(2) Es difícil de aplicar, implica repetir el experimento aleatorio un número grande de veces para poder medir la probabilidad. Basada en la regularidad estadística de los resultados asociados al experimento, es útil para interpretar la probabilidad pero no para investigar sus propiedades.

Ejemplo 5

Si la distribución del número de trabajadores ("tamaño de la empresa) en las 80 empresas de metalmecánica de un distrito es:

Tamaño	f	%
0 - 4	30	38
4 - 8	20	25
8 - 12	15	19
12 - 16	10	13
16 - 20	5	6
Total	80	100

La probabilidad de A="La empresa tiene entre 4 y 8 trabajadores" es
$P(A)=\frac{N_A}{N}=\frac{20}{80}=0.25$

Nota: En ambas definiciones ocurre que $0\le P(A)\le 1$, $P(S)=1$ y $P(\phi)=0$

1.3.3 Definición Axiomática (o de Kolmogorov)

Sea S un espacio muestral asociado a un experimento aleatorio **ε y sea** $\mathscr{A}$ una σ-álgebra de eventos definida sobre S **.** Una Probabilidad P definida sobre los eventos de $\mathscr{A}$ es una función $P:\mathscr{A}\to R$ que a cada evento A le asigna un número real, denotado P(A) y llamado Probabilidad de A, de modo que se satisfacen los axiomas:

(1) $0\le P(A)$

(2) $P(S)=1$

(3) Si $A_1, A_2,...A_n,...$ es una sucesión de eventos mutuamente excluyentes, esto es $A_i\cap A_k=\varnothing \quad \forall i\ne k$,
entonces $P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$

Observaciones:

(1) Esta definición, a diferencia de las anteriores, es "no constructiva", o sea, no dice cómo calcular la probabilidad sino que sólo indica los requisitos que debe satisfacer una asignación de probabilidades a eventos para ser considerada "correcta", en el sentido de estar libre de contradicciones lógicas.

(2) Según esta definición, **es posible hacer diferentes asignaciones de probabilidades sobre un mismo conjunto de eventos y si se cumplen los tres axiomas, todas las asignaciones son formalmente correctas**. Esta posición a primera vista parece un contrasentido, pero en realidad convierte a la probabilidad en una poderosa herramienta de investigación de procesos no determinísticos: Dado un proceso, es posible elaborar diferentes teorías explicativas para el mismo y cada una inducirá una asignación de probabilidades. Al efectuar el experimento aleatorio, el resultado que se presente será contradictorio con algunas asignaciones y confirmatorio de otras. Como las asignaciones están libres de contradicciones lógicas, el no ajuste del resultado del experimento no se debe a la matemática usada, sino que tiene base real. Por tanto, aquellas asignaciones no ratificadas por los datos empíricos pueden ser descartadas y con ellas, las teorías que les sirvieron de base.

1.3.4 Propiedades de la Probabilidad

La teoría de la probabilidad es resultado del estudio de las “leyes del azar”, esto es, se deduce como consecuencia de un proceso de derivaciones lógicas, de un análisis formal o racional, a partir del cual se llega a descubrir propiedades que luego serán herramientas del análisis. Es así que tenemos las siguientes proposiciones:

Proposición 1: $P(\varnothing) = 0$

Demostración

Definamos la sucesión de eventos $\{A_i, i = 1,2,3\ldots\}$ donde $A_i = \varnothing \quad \forall i$. Entonces es claro que $\bigcup_{i=1}^{\infty} A_i = \varnothing$ y como la Probabilidad es una función, se cumple aquello de "a igualdad de puntos de partida, igualdad de puntos de llegada"; esto es, se puede aplicar la probabilidad a la identidad anterior y ésta se preserva. Por tanto tenemos:

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\varnothing)$$

y aplicando el axioma (3):

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = P(\varnothing)$$

o equivalentemente:

$$P(\varnothing) + P(\varnothing) + ... + P(\varnothing) = P(\varnothing)$$

Obviamente el único número real que satisface esta ecuación es 0, esto es:

$$P(\varnothing) = 0$$

Proposición 2: Si $A_1, A_2, ...A_N$ es una sucesión de N eventos mutuamente excluyentes, entonces

$$P(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$$

Demostración

Completemos la sucesión finita de modo que sea equivalente a una sucesión infinita y apliquemos el tercer axioma y la propiedad 1. Entonces, sean los eventos $A_{N+1} = \varnothing$; $A_{N+2} = \varnothing$; $A_{N+3} = \varnothing$; ...; se cumple que:

$$\bigcup_{j=1}^{N} A_j = \bigcup_{j=1}^{\infty} A_j$$

y por tanto:

$$P(\bigcup_{j=1}^{N} A_j) = P(\bigcup_{j=1}^{\infty} A_i) = \sum_{j=1}^{\infty} P(A_j) = \sum_{j=1}^{N} P(A_j) + \sum_{j=N+1}^{\infty} P(A_j)$$

Pero por construcción, $A_j = \varnothing$ si $j \geq N+1$. Aplicando la proposición anterior tenemos que $P(A_j) = 0$ si $j \geq N+1$, lo que anula la segunda sumatoria y así se llega a:

$$P(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$$

Proposición 3: $P(A) + P(A^C) = 1$

Demostración
Como $A \cup A^C = S$, siendo A y A^C mutuamente excluyentes, aplicando la proposición anterior y el axioma 3 tenemos:

$$P(A \cup A^C) = P(A) + P(A^C) = P(S) = 1$$

O equivalentemente:

$$P(A) + P(A^C) = 1$$

Proposición 4: Si A y B son eventos arbitrarios, entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demostración
En general $B = B \cap S$ y $S = A \cup A^C$. Luego, podemos escribir:

$$B = (B \cap A) \cup (\mathrm{B} \cap \mathrm{A}^C) \text{ y } A \cup B = A \cup (\mathrm{B} \cap \mathrm{A}^C)$$

Aplicando el axioma 2:

$$P(B) = P(B \cap A) + P(\mathrm{B} \cap \mathrm{A}^C)$$

o equivalentemente: $P(\mathrm{B} \cap \mathrm{A}^C) = P(B) - P(B \cap A)$.

También: $P(A \cup B) = P(A) + P(\mathrm{B} \cap \mathrm{A}^C)$ y reemplazando $P(\mathrm{B} \cap \mathrm{A}^C)$ por $P(B) - P(B \cap A)$, llegamos a:

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

Proposición 5: Si A y B son eventos tales que $A \subseteq B$, entonces:

$$P(A) \leq P(\mathrm{B})$$

Demostración
De la proposición anterior sabemos que:

$$P(\mathrm{B} \cap \mathrm{A}^C) = P(B) - P(\mathrm{B} \cap \mathrm{A})$$

Y como $A \subseteq B$, es claro que $(B \cap A) = A$. Luego:

$$P(\mathrm{B} \cap \mathrm{A}^C) = P(B) - P(A)$$

Pero toda probabilidad, según el axioma 1 es no negativa, en particular:

$$P(B \cap A^C) = P(B) - P(A) \geq 0$$

y se deduce que $P(B) \geq P(A)$ o equivalentemente:

$$P(A) \leq P(B)$$

Corolario: En el contexto de la proposición anterior, se cumple:

$$0 \leq P(A) \leq P(B) \leq 1$$

Demostración

Es consecuencia directa de la propiedad: $\varnothing \subseteq A \subseteq B \subseteq S$ y la aplicación sucesiva de la Proposición 5.

Proposición 6: Si $A_1, A_2, ...$ son eventos tales que $A_1 \subseteq A_2 \subseteq ...$ y $A = \bigcup_{n=1}^{\infty} A_n$ entonces

$$P(A) = \lim_{n \to \infty} P(A_n)$$

Demostración

Definamos los eventos excluyentes $B_1, B_2, ..., B_i, ...$ mediante

$B_1 = A_1, B_2 = A_2 - A_1, B_3 = A_3 - A_2, .., B_i = A_i - A_{i-1}, ...,$, donde recuérdese $A_i - A_{i-1} = A_i \cap A_{i-1}^C$

En este contexto $A = \bigcup_{n=1}^{\infty} A_n = \bigcup_{i=1}^{\infty} B_i$ y $B_1, B_2, ..., B_i, ...$ son obviamente disjuntos. Luego por el axioma (3)

$P(A) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \to \infty} \sum_{i=1}^{n} P(B_i)$, pero por construcción tenemos $A_n = \bigcup_{i=1}^{n} B_i$ y por la Proposición 2, $P(A_n) = P(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} P(B_i)$, así que llegamos a $P(A) = \lim_{n \to \infty} \sum_{i=1}^{n} P(B_i) = \lim_{n \to \infty} P(A_n)$

Comentario: La probabilidad como medida racional de incertidumbre

La proposición anterior permite asegurar que en relación a la propiedad de inclusión de eventos, la probabilidad mide la "propensión a ocurrir" al menos a nivel ordinal, donde el vacío $\varnothing$ ocupa el menor puesto y el espacio S tiene la mayor jerarquía, correspondiéndole a otros eventos los puestos intermedios**. Sin embargo, debemos notar que existen casos donde hay eventos que reciben probabilidad 0 y son distintos del vacío.**

Por ejemplo, lancemos un dardo al azar sobre un blanco circular y midamos la probabilidad de que el dardo caiga en una región, como el cociente del área de la región sobre el área del círculo. Muchos eventos tendrán probabilidad obvia y razonable con esta medida. Así por ejemplo, al evento A definido como: "El dardo cae en el semicírculo inferior", la asignación de probabilidades definida antes, le dará a una probabilidad de $1/2$ o 0.5, i.e. P(A) = 0.5, algo que está de acuerdo con nuestra intuición.

Sin embargo, si definimos el evento B como: "El dardo cae exactamente en el centro del blanco", resulta que como el área de un punto es 0, la asignación de probabilidades le dará a este evento una probabilidad también 0 o sea P(B)= 0. Así tenemos el caso curioso de un evento perfectamente factible pero de probabilidad nula.

La explicación intuitiva es que la probabilidad mide la opción de ocurrencia entendida como nuestro grado de incertidumbre con respecto a los eventos, y esta incertidumbre está asociada al conocimiento que poseamos de los mismos. Ahora bien, en relación al evento B, materialmente es imposible que sepamos exactamente si ocurrió o no el evento, pues nuestros medios físicos de verificación tienen un límite en su precisión. Nunca podremos estar seguros de si el dardo cayó en el centro o si cayó a una millonésima de milímetro del centro. La probabilidad cero refleja este estado de información.

En verdad, la aparente paradoja es irrelevante, porque si bien el blanco circular es un objeto material, de existencia cierta a nuestros sentidos, su "centro" es una abstracción, una región ideal. Por tanto no debemos preocuparnos y sólo nos queda tomar nota de esta sutil diferencia entre lo "improbable" y lo "imposible", y confortarnos con el hecho de tener una herramienta de propiedades conocidas para trabajar en contextos de incertidumbre.

Ejemplo 6

Un consumidor encuentra dos productores w_1 y w_2, que le ofrecen el mismo bien al mismo precio. El consumidor puede comprar a w_1 con probabilidad "p" o a w_2 con probabilidad "q". ¿Cuáles de los siguientes valores de p y q son formalmente correctos?

a) $p = q = 1/2$
b) $p = 2/3$ y $q = 1/3$
c) $p = 2/8$ y $q = 4/5$

Solución:

Sea S el espacio muestral dado por $S = \{w_1, w_2\}$. Entonces $S = \{w_1\} \cup \{w_2\}$ y naturalmente $1 = P(S) = P(\{\text{w}_1, \text{w}_2\}) = P(\{\text{w}_1\} \cup \{\text{w}_2\}) = P(\{\text{w}_1\}) + P(\{\text{w}_2\}) = p + q$ es una igualdad que debe cumplirse para que la asignación de probabilidades sea correcta (en el sentido de Kolmogorov). Por tanto, verificando cada caso:

En a) $p + q = 1/2 + 1/2 = 1 \Rightarrow$ Asignación correcta
En b) $p + q = 2/3 + 1/3 = 1 \Rightarrow$ Asignación correcta
En c) $p + q = 2/8 + 4/5 = 21/20 > 1$, lo que es una contradicción. Esta asignación **no** es correcta.

Nótese que tanto las asignaciones (a) como (b) son matemáticamente correctas y sin embargo la intuición dice que la primera es más compatible con la realidad. Esta última idea sólo es verificable con datos, con "evidencia empírica". Dicho sea de paso, el investigador no debe escatimar esfuerzos para obtener toda la información relevante sobre el proceso que pretende explicar, y debe asignar probabilidades de acuerdo a esa información. Los cambios pueden ser notables. Por ejemplo, si supiéramos que w_1 gasta en propaganda el doble que w_2 y asumimos que esto afecta las preferencias del consumidor de modo que la probabilidad es directamente proporcional a la propaganda, entonces $p = \alpha G(w_1)$ y $q = \alpha G(w_2)$, donde $G(w_i)$ es el gasto del productor i, siendo α la constante de proporcionalidad. Como $G(w_1) = 2G(w_2)$, es fácil ver que, en este caso, $p = 2/3$ y $q = 1/3$ sería la asignación correcta.

Ejemplo 7.1

El precio p de un bien agrícola puede ser de 1, 2, 3 o 4 unidades monetarias y la cantidad demandada q de este bien responde al precio de modo que $q = 5 - p$. Un cambio climático origina una caída en la producción del bien de modo que la probabilidad de que el precio tome un valor p resulta directamente proporcional a p. En este contexto, explicite el espacio muestral S asociado a observar las parejas de precios y cantidades posibles en el mercado de este bien, halle la distribución de probabilidades en S y diga si la cantidad demandada tomará su mínimo valor posible.

Solución:

$S = \{(p,q) | q = 5 - p, p = 1,2,3,4\}$ y por dato $P(p = k) = \alpha k,\ k = 1,2,3,4$ donde k es la constante de proporcionalidad. Como $\sum_{k=1}^{4} P(p = k) = 1 \Rightarrow \alpha + 2\alpha + 3\alpha + 4\alpha = 1 \Rightarrow 10\alpha = 1 \Rightarrow \alpha = \frac{1}{10}$ y la distribución de probabilidades en S es:

(p,q)	(1,4)	(2,3)	(3,2)	(4,1)
$P\{(p,q)\}$	1/10	2/10	3/10	4/10

Finalmente, $P(\text{La cantidad demandada toma su valor mínimo}) = P\{(4,1)\} = \frac{4}{10} = 0.4$

Ejemplo 7.2
Un dado está "cargado", de modo que al lanzarlo y observar la cara que muestra en su lado superior, cada número tiene una opción de presentarse directamente proporcional a dicho número.

a) Halle la distribución de probabilidades compatible con esta información
b) Halle la probabilidad de que al lanzar el dado ocurra un número par
c) Halle la probabilidad de que al lanzar el dado ocurra un número primo

Solución:
(a) $S=\{1,2,3,4,5,6\}$ y de los datos $P(\{k\})=\alpha k$ para $k=1,2,...,6$ donde $\alpha>0$ es la constante de proporcionalidad

Además como $S=\{1\}\cup\{2\}\cup\{3\}\cup\{4\}\cup\{5\}\cup\{6\}$, aplicando el axioma 2 y sucesivas veces el axioma 3 se tiene:
$P(S)=P(\{1\}\cup\{2\}\cup\{3\}\cup\{4\}\cup\{5\}\cup\{6\})=1 \Leftrightarrow P(\{1\})+P(\{2\})+P(\{3\})+P(\{4\})+P(\{5\})+P(\{6\})=1$
$\Leftrightarrow \alpha+2\alpha+3\alpha+...+6\alpha=\sum_{j=1}^{6}k\alpha=1 \Leftrightarrow 1=21\alpha$. Despejando se obtiene $\alpha=1/21$ y la asignación o distribución de probabilidades es:

$$P(\{k\})=k/21 \text{ para } k=1,2,...,6$$

(b) Si A denota el evento "El número es par", entonces $A=\{2,4,6\}$ y se tiene
$P(A)=P(\{2,4,6\})=P(\{2\})+P(\{4\})+P(\{6\})=\frac{2}{21}+\frac{4}{21}+\frac{6}{21}=\frac{12}{21}$

(c) Análogamente a (b), sea B es el evento pedido, entonces $B=\{1,2,3,5\})$ y se tiene
$P(B)=\frac{1}{21}+\frac{2}{21}+\frac{3}{21}+\frac{5}{21}=\frac{11}{21}$

Observación:
Nótese que escribimos $P(\{k\})$ y no $P(k)$ porque la probabilidad está definida sobre eventos o conjuntos (se dice que es una 'función de conjuntos') y no sobre elementos. Por eso, siempre debiéramos escribir $P(\{w\})$ para denotar la probabilidad de un elemento w de S; sin embargo, para no recargar la notación, podemos tomarnos la licencia de usar $P(w)$ siempre y cuando esto no produzca confusión.

Ejemplo 8
Un bien puede costar 1 o 2 unidades monetarias y un consumidor puede comprar 1,2 o 3 unidades del bien. Sabiendo que todas las parejas de precios y cantidades (p,q) son posibles y que en general la probabilidad $P\{(p,q)\}$ es <u>directamente</u> proporcional a la razón (q/p): $P\{(p,q)\}=\alpha(q/p)$

a) Describa el espacio muestral S asociado a este experimento y halle α.
b) Identifique el evento A = "El gasto del consumidor es de 2 u.m." y calcule su probabilidad.
c) Identifique el evento B = "El consumidor adquiere 2 unidades del bien" y calcule su probabilidad.
d) Halle la probabilidad de A∪B.
e) El consumidor tiene un ingreso de 6 u.m. ¿Pronosticaría Ud. que gastará todo en el bien?

Solución:
a) El conjunto de todas las parejas (p,q) posibles es S, y podemos escribir

$S=\{(p,q)\mid p=1,2;q=1,2,3\}=\bigcup_{p=1}^{2}\bigcup_{q=1}^{3}\{(p,q)\}$. Como $P(S)=1$ y también $P(S)=P(\bigcup_{p=1}^{2}\bigcup_{q=1}^{3}\{(p,q)\})$

$=\sum_{p=1}^{2}\sum_{q=1}^{3}P(p,q)=\sum_{p=1}^{2}\sum_{q=1}^{3}\alpha(q/p)=\alpha\sum_{p=1}^{2}\sum_{q=1}^{3}(q/p)$, tenemos

$1=\alpha\sum_{p=1}^{2}\sum_{q=1}^{3}(q/p)=\alpha\sum_{p=1}^{2}(1+2+3)/p=\alpha 6(1+\frac{1}{2}))=9\alpha$, esto es, $\alpha=\frac{1}{9}$

b) $A=\{(p,q)\in S\mid pq=2\}=\{(1,2),(2,1)\}$ y $P(A)=P(1,2)+P(2,1)=\frac{1}{9}\times\frac{2}{1}+\frac{1}{9}\times\frac{1}{2}=\frac{5}{18}$

c) $B=\{(p,q)\in S\mid q=2\}=\{(1,2),(2,2)\}$ y $P(B)=P(1,2)+P(2,2)=\frac{1}{9}\times\frac{2}{1}+\frac{1}{9}\times\frac{2}{2}=\frac{6}{18}$

d) $P(A\bigcup B)=P(A)+P(B)-P(A\bigcap B)=\frac{5}{18}+\frac{6}{18}-\frac{4}{18}=\frac{7}{18}$

e) Con seis u.m. de ingreso, sólo hay un caso en que se gasta todo, que es cuando el consumidor se sitúa en el punto (2,3). La probabilidad de que esto ocurra es $P(2,3)=\frac{1}{9}\times\frac{3}{2}=\frac{1}{6}=0.17$ o 17%, que está más cerca de 0 que de 1, esto es, no pronosticaríamos la ocurrencia de este evento.

1.4 Casos Especiales de asignación de Probabilidades

Examinaremos algunas formas de asignación de probabilidades que siendo compatibles con el sistema de Kolmogorov, serán de utilidad en el futuro.

1.4.1 Probabilidad Geométrica

Sea un experimento aleatorio ε consistente en tomar un punto al azar de un conjunto geométrico S que tiene una medida $m(S)$ y sea A un evento del espacio muestral resultante. Si $m(A)$ denota la medida de este evento, entonces la probabilidad de A es:

$$P(A)=\frac{m(A)}{m(S)}$$

Observaciones:

(1) Si S es un intervalo, la “medida” m es la longitud; si S es una región de R^2 donde está definida un área, la “medida” es el área.

(2) Se trata de una extensión de la definición clásica, que aparece en situaciones especiales, como lanzar un dardo sobre un blanco o tomar un punto al azar de un segmento.

Ejemplo 9.1

Dos proveedores se han presentado a un concurso de precios. Del proveedor A se sabe que puede ofrecer el bien a un precio que estará indistintamente entre 1 y 10 dólares; del proveedor B se sabe que su precio podría estar indistintamente entre 1 y 5 dólares. Sin más información y asumiendo un experimento aleatorio:

a) Describa el espacio muestral S asociado al experimento
b) Calcule la probabilidad de que A resulte ganador
c) Calcule la probabilidad de que el precio ganador no pase de US$ 3
d) ¿Esperaría Ud. que A superara a B en dos o más dólares?

Solución:

a) El experimento consiste en observar los precios ofrecidos por A y B. De este modo, si X = Precio ofrecido por A e Y = Precio ofrecido por B, entonces $S=\{(x,y)\mid 1\le x\le 10, 1\le y\le 5\}$.

Geométricamente S es **un rectángulo en el plano XY, esto es, S tiene área. Por tanto, la probabilidad de un evento E se puede definir como cociente de áreas y así tenemos:**

$P(E)=\dfrac{Area(E)}{Area(S)}$ $\forall$ E **evento de** S

b) Sea A = "A resulta ganador" $\Rightarrow A$ = {(X,Y)∈ S| X < Y }, pues como se trata de proveedores, gana quien ofrece menor precio. Graficando A, **resulta ser la parte de S que está arriba de la recta identidad Y=X.**

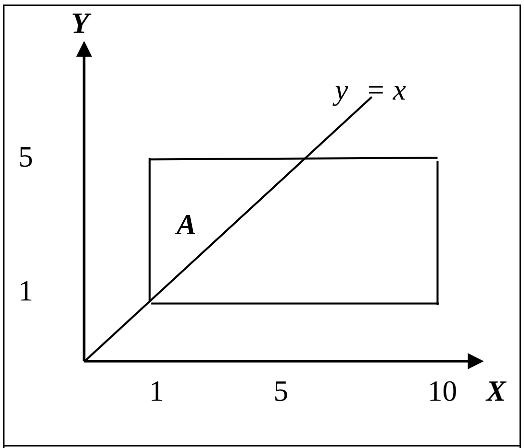


Fig. 1.1 Evento ***A*** en espacio *S*

***S* tiene** $Area = base \times altura = (10-1)\times(5-1)=36;$ ***A* es un triángulo y tiene área** $Area = base \times altura/2=(5-1)\times(5-1)/2=8$

La probabilidad de A es entonces: $P(A)=\dfrac{Area(A)}{Area(S)}=\dfrac{(4\times 4/2)}{9\times 4}=\dfrac{8}{36}=0.22$

c) Si C ="El precio ganador no pasa de US$ 3" $\Rightarrow$ $C=\{(X,Y)\in S \mid Mín\{X,Y\}\le 3\}$ y $P(C)=1-P(C^C)$.
Como $C^C=\{(X,Y)\in S \mid Mín\{X,Y\}>3\}=\{(X,Y)\in S \mid (X>3)\cap(Y>3)\}$

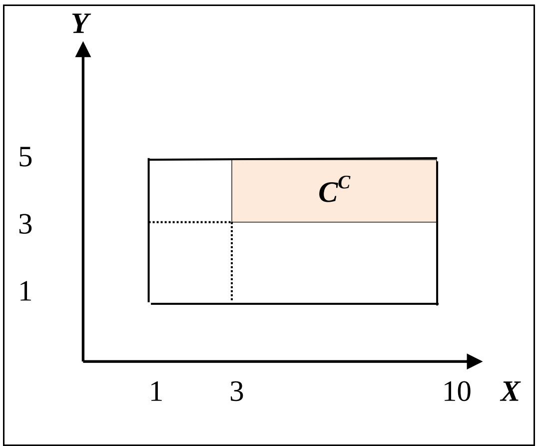


Fig. 1.2 Evento $\boldsymbol{C^C}$ en espacio *S*

Tenemos:
$Area \quad de \quad C^C = base\times Altura=(10-3)\times(5-3)=14; P(C^C)=14/36=0.39$ y $P(C)=0.61$

d) Si D ="A supera a B en dos o más dólares" $\Rightarrow$ $D=\{(X,Y)\in S \mid Y+2\le X\}$, pues en el caso de la igualdad exacta (A supera a B en dos dólares) esta equivale a decir que a Y tendríamos que sumarle 2 para que alcance a X, esto es $Y+2=X$, y en la desigualdad X excede a Y en más de 2. Graficando tenemos que D es el trapecio dibujado abajo (figura 1.3). Descomponiendo D en un triángulo y un rectángulo y calculando áreas tenemos:

Area del triángulo $=(7-3)\times 4/2=8$; Area del rectángulo $=(10-7)\times 4=12$; Area total $=12$ y $P(D)=20/36=0.56>0.5$, **luego, podemos esperar que el evento D sí ocurrirá**

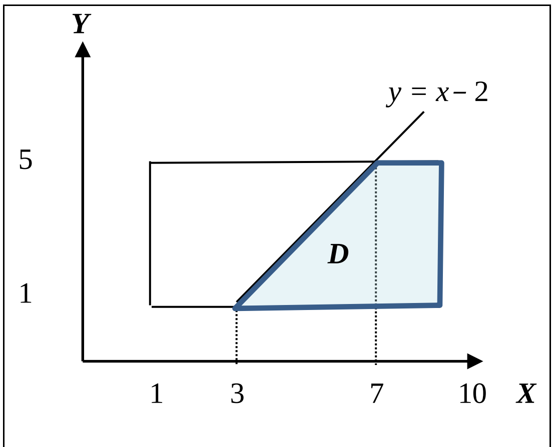
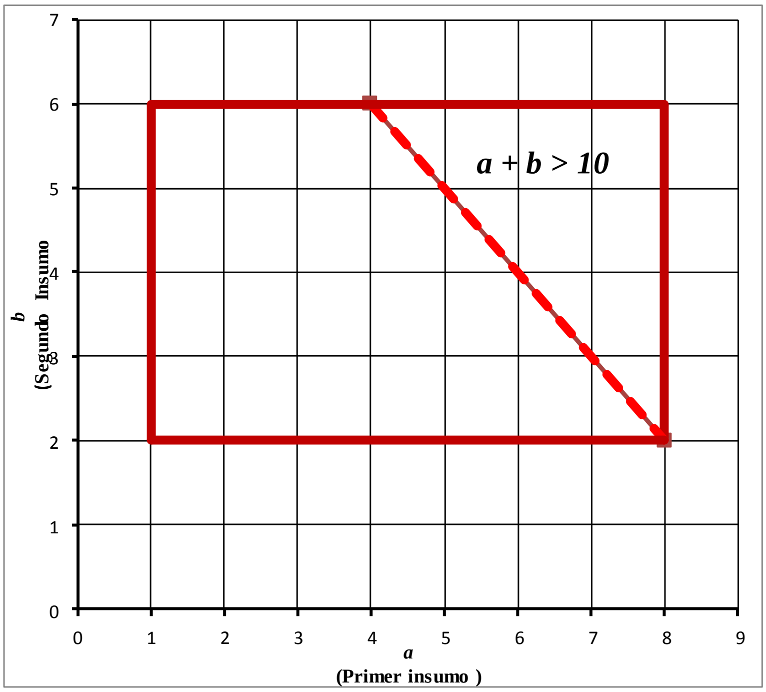


Fig. 1.3 Evento ***D*** en espacio *S*

Ejemplo 9.2
Para realizar un trabajo se necesitan dos insumos, que entran en cantidades variables según el trabajo, de modo que el costo por el primer insumo puede tomar aleatoriamente cualquier valor entre 1 y 8 unidades monetarias, mientras que el costo por el segundo insumo puede tomar aleatoriamente cualquier valor entre 2 y 6 unidades monetarias. Sin más información, explicite el espacio muestral *S* asociado a costo por insumos de un trabajo y diga si el costo por insumos pasará de las 10 unidades monetarias.

Solución:



$S=\{(a,b)|1<a<8;2<b<6\}$, donde
a = costo del primer insumo y
b = costo del segundo insumo.
Para responder la pregunta necesitamos la probabilidad del evento
$D=$ "$Costo\ por\ insumos\ pasa\ de$ 10".
Aplicando probabilidad geométrica:
$P(D\}=P(\{(a,b)\in S|a+b>10\})=\frac{Area(D)}{Area(S)}=\frac{(4\times 4)/2}{7\times 4}=\frac{2}{7}=0.29<0.5,$

Pronosticamos que el costo **no** pasará de 10 unidades.

1.4.2 Probabilidad en Espacios Numerables

Sea $S=\{w_1,w_2,...,w_k,...\}$ un espacio muestral infinito numerable. Una asignación de probabilidades compatible con los axiomas es una asignación de la forma:

$$P(\{w_k\}) = p_k \quad k = 1,2,...$$

donde $p_k \geq 0$ y $\sum_k p_k = 1$

Ejemplo 10

Si en la definición anterior $p_k = r^k$ establezca condiciones sobre r para que la asignación de probabilidades sea correcta en el sentido de Kolmogorov

Solución:

Como $p_k = r^k \geq 0 \Rightarrow 0 \leq r$. Por otra parte $\sum_k p_k = 1 \Rightarrow \sum_{k=1}^{\infty} r^k = 1$ y nuestro problema es calcular la suma infinita $\sum_{k=1}^{\infty} r^k$. Podemos escribir $\sum_{k=1}^{\infty} r^k = \lim_{N\to\infty} \sum_{k=1}^{N} r^k$ y el cálculo de la suma finita $\sum_{k=1}^{N} r^k$ es algo estándar:

Sea $S_N := \sum_{k=1}^{N} r^k$, entonces podemos escribir $S_N := r + r^2 + r^3 + \ldots + r^{N-1} + r^N$ y multiplicando por la razón r obtenemos $rS_N := r^2 + r^3 + r^4 + \ldots + r^N + r^{N+1}$. Restando llegamos a

$S_N - rS_N = (r + r^2 + r^3 + \ldots + r^{N-1} + r^N) - (r^2 + r^3 + r^4 + \ldots + r^N + r^{N+1}) = r - r^{N+1}$ y entonces

$S_N = (r - r^{N+1})/(1-r)$ (lo que de paso nos da una nueva restricción: $r \neq 1$ y por tanto $0 < r < 1$).

Aplicando la condición $\sum_{k=1}^{\infty} r^k = 1 \Rightarrow \lim_{N\to\infty} \sum_{k=1}^{N} r^k = 1 \Rightarrow \lim_{N\to\infty} S_N = 1 \Rightarrow \lim_{N\to\infty} (r - r^{N+1})/(1-r) = 1$ pero $\lim_{N\to\infty} (r - r^{N+1})/(1-r) = r/(1-r)$ (pues $\lim_{N\to\infty} r^{N+1} = 0$). Llegamos así a $r/(1-r) = 1$ y despejando, nos queda finalmente $r = (1-r) \Rightarrow r = 1/2$ es el valor pedido.

1.4.3 Probabilidad Clásica y Técnicas de Conteo

La definición clásica requiere que podamos contar la cantidad de elementos que tienen tanto S como el evento A cuya probabilidad queremos calcular. Para hacer esto, la enumeración o conteo directo es por lo general un sistema ineficiente y por ello pasaremos revista a ciertas técnicas de conteo rápido.

A. Dos Principios Básicos

- **Principio de la Multiplicación**

Si una 'operación' A puede realizarse u ocurrir de **a** maneras diferentes y otra 'operación' B puede realizarse de **b** maneras diferentes, entonces la operación compuesta AXB consistente en realizar A primero y luego realizar B, se puede realizar de (**axb**) maneras distintas.

- **Principio de la Adición**

Si una 'operación' A puede realizarse u ocurrir de **a** maneras diferentes y otra 'operación' B puede realizarse de **b** maneras diferentes, siendo ambas operaciones excluyentes, entonces la operación compuesta A ó B consistente en realizar A o realizar B pero no ambas, se puede realizar de (**a+b**) maneras distintas.

B. Permutaciones y Combinaciones

Sea un conjunto L con n elementos y sea r un entero fijo, conocido y no mayor que n.

Definición
Una Permutación de tamaño r, formada a partir de los n elementos de L, es una 'arreglo' de r elementos de L donde se distingue o impone un orden entre ellos, sin repetición de elementos

Definición
Una Combinación de tamaño r, formada a partir de los n elementos de L, es un subconjunto de r elementos de L.

Ejemplo 11

- Si L={A,B,C,D} y tomamos r=3 entonces algunas 'permutaciones de tamaño r=3' son: (A,B,C); (A,C,B); (A,C,D); (D,C,B). Nótese que hay más permutaciones; sólo hemos escrito cuatro de ellas.
- Si L={A,B,C,D} y tomamos r=3 entonces algunas 'combinaciones' de tamaño r=3' son: {A,B,C}; {A,C,D} . Nótese que hay más combinaciones y sólo hemos escrito dos de ellas. Observe además que de la combinación {A,B,C} podemos formar seis permutaciones del mismo tamaño, con la misma composición pero con diferentes órdenes. A saber: (A,B,C); (A,C,B); (C,A,B); (C,B,A); (B,A,C) y (B,C,A)

Observaciones:

(1) Informalmente, una permutación es una "cola" de objetos y una combinación es una colección de objetos
(2) En las permutaciones importa el orden; en la combinaciones no importa el orden

Factorial de un entero.
Si N es un entero no negativo, el "Factorial de N" denotado $N!$ se define mediante $N! = N\times(N-1)\times(N-2)\times\ldots\times3\times2\times1$. Adicionalmente definimos $0! = 1!$

Proposición
En el contexto de las definiciones anteriores se cumple que el número total de Permutaciones de tamaño r es: $n!/(n-r)!$

Demostración

Aplicando el principio de multiplicación y considerando cualquier permutación como una 'cola' compuesta de r elementos, podemos desagregar la operación de formar una permutación en r sub-operaciones consistentes en asignar o "llenar" el primer lugar de la cola, luego el segundo, el tercero, etc.

Para el primer lugar tenemos n maneras o posibilidades, y llenado éste, para el segundo lugar tenemos (n-1) maneras o posibilidades y luego, para el tercer lugar hay (n-2) maneras y así sucesivamente.

Aplicando el principio de multiplicación, la operación conjunta de formar la cola se puede hacer de:

$$n(n-1)(n-2)\ldots(n-[r-1]) = \frac{n(n-1)(n-2)\ldots(n-[r-1])x(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

Nota: El total de permutaciones de tamaño r lo denotaremos P_r^n **o** $n\mathrm{Pr}$

Proposición

En el contexto de las definiciones anteriores se cumple que el número total de Combinaciones de tamaño r es: $n!/r!(n-r)!$

Demostración

Primero notemos que si tomamos **una** combinación cualquiera de tamaño r, ésta genera $r!$ permutaciones distintas de tamaño r. Esto implica que hay una proporcionalidad entre el número total de permutaciones de tamaño r y el número total de combinaciones de tamaño r

En segundo lugar el conjunto total de Combinaciones de tamaño r genera el conjunto total de Permutaciones de tamaño r.

Entonces, si denotamos mediante C_r^n el número total de combinaciones podemos hacer una regla de tres simple:

$$\begin{array}{c} 1 \text{ Combinación } \rightarrow r! \text{ Permutaciones} \\ C_r^n \text{ Combinaciones} \rightarrow P_r^n \text{ Permutaciones} \end{array}$$

Por tanto:

$$C_r^n = P_r^n / r! = n!/(n-r)!\,r!$$

Nota: El número C_r^n también se escribe $\binom{n}{r}$ y se llama 'número combinatorio'. Tiene algunas propiedades como:

- $C_r^n = C_{n-r}^n$
- $C_r^n = C_{r-1}^{n-1} + C_r^{n-1}$
- $(a+b)^n = \sum_{j=0}^{n} C_j^n a^j b^{n-j}$ (El Binomio de Newton)

Ejemplo 12
En una manzana hay 20 hogares, de los cuales 10 son de clase popular, 6 de clase media y 4 de clase acomodada. En una encuesta se tomó al azar una muestra de 5 hogares de la manzana. Halle la probabilidad de que:

a) En la muestra haya hogares de clase media.
b) En la muestra la mayoría de hogares resulte de clase popular.
c) En la muestra haya dos hogares de clase popular, dos de clase media y uno de clase acomodada.

Solución:
Se trata de un experimento en donde se toma un subconjunto (una muestra) de cinco hogares del conjunto mayor L={H1,H2,...,H20} de veinte hogares de la manzana. El espacio S es el conjunto de todas las muestras posibles de tamaño 5 (o sea el conjunto de combinaciones de tamaño r=5), pues sólo importa identificar los hogares que debemos entrevistar y no el orden en que formemos la muestra, no hay un orden especial. Además, como no tenemos razones para pensar que algunas muestras de hogares tienen mayor opción de presentarse, podemos usar la definición clásica de probabilidad. Esto es:

$S = \{\{x_1, x_2, ..., x_5\} \mid x_i \in L \wedge x_i \neq x_j \; si \quad i \neq j\}$ y $n(S) = C_5^{20}$

a) Si A ="En la muestra hay hogares de clase media" entonces:
A^C ="En la muestra no hay hogares de clase media" y $P(A) = 1 - P(A^C)$. Luego $n(A^C) = C_5^{14}$ pues hay 10+4=14 hogares que no son de clase media y de ellos debo tomar la muestra de 5 hogares si deseo que ocurra A^C. Finalmente tenemos $P(A) = 1 - C_5^{14} / C_5^{20}$

b) Si $B =$ "La mayoría de hogares en la muestra son de clase popular", eso equivale a que en la muestra hay al menos tres hogares de ese estrato; por tanto $n(B) = C_3^{10} \times C_2^{10} + C_4^{10} \times C_1^{10} + C_5^{10} \times C_0^{10}$, donde:
C_3^{10}=maneras de tomar 3 hogares de estrato popular,
C_2^{10}=maneras de tomar otros 2 hogares del resto de estratos, luego $C_3^{10} \times C_2^{10}$ es el número de maneras en que podemos tomar 3 hogares de clase popular y el resto de la muestra en los otros estratos.
Análogamente sucede con los siguientes sumandos y $P(B) = \dfrac{C_3^{10} \times C_2^{10} + C_4^{10} \times C_1^{10} + C_5^{10} \times C_0^{10}}{C_5^{20}}$

b) Sea C ="En la muestra hay dos hogares de clase popular, dos de clase media y uno de clase acomodada", entonces:

Hay C_2^{10} maneras de tomar 2 hogares de clase popular;
Hay C_2^{6} maneras de tomar 2 hogares de clase media;
Hay C_1^{4} maneras de tomar 2 hogares de clase popular.

Por tanto, $n(C) = C_2^{10} \times C_2^{6} \times C_1^{4}$ y $P(C) = C_2^{10} \times C_2^{6} \times C_1^{4} / C_5^{20}$

Ejemplo 13
Un sistema de seguridad posee un código de entrada de 8 casillas con las 2 primeras blancas, las 3 siguientes negras y las 3 últimas rojas. Asuma que Ud. sabe la cantidad de colores pero desconoce el código e intenta adivinarlo: ¿Cree que logrará hacerlo en un intento? Use probabilidades para responder

Solución:
S es el conjunto de todos los códigos posibles conformados por dos letras B, tres N y tres R. Formar un código equivale a seleccionar lugares para las letras, dentro de los 8 que forman el código:

Hay C_2^8 maneras de seleccionar dos de los ocho lugares para colocar las 2 B

Hay C_3^6 maneras de seleccionar tres de los seis lugares restantes para colocar las 3 N

Hay C_3^3 maneras de seleccionar tres de los tres lugares restantes para colocar las 3 R.

Finalmente $n(S) = C_2^8 C_3^6 C_3^3$ y $P(Acertar) = \frac{1}{C_2^8 C_3^6 C_3^3} = \frac{1}{560} = 0.0018$ que es casi cero: No se logrará adivinar la clave en un intento.

Ejemplo 14
En el mercado de un bien con 6 productores se sabe que hay al menos dos coaliciones y un organismo de regulación se interesa por el estado del mercado. Halle la probabilidad de que el mercado esté formado por dos coaliciones, cada una con tres empresas.

Solución:
En este contexto, y si no hay otra información, el experimento aleatorio ε consiste en observar el estado del mercado, que puede lograrse de las siguientes maneras:

a) Dos coaliciones de 2 empresas cada una y las otras empresas libres, o
b) Dos coaliciones, una de 2 empresas y la otra de 3, con la empresa restante libre, o
c) Dos coaliciones, una de 2 empresas y la otra de 4 o
d) Dos coaliciones, cada una con tres empresas, o finalmente,
e) Tres coaliciones, cada una con dos empresas.

El tamaño de S es $n(S) = C_2^6 C_2^4 + C_2^6 C_3^4 + C_2^6 C_4^4 + C_3^6 C_3^3 + C_2^6 C_2^4 C_2^2 = 275$ estados posibles y nos interesa que se haya dado la situación d). Denotando D a este evento, entonces

$$P(D) = \frac{C_3^6 C_3^3}{275} = \frac{20}{275} = 0.07$$

1.5 Probabilidad Condicional e Independencia

La probabilidad, como ya vimos, mide la propensión o tendencia a ocurrir que le asignamos a un evento. Esta asignación inicialmente se hace partiendo de la información que proporciona el experimento aleatorio que sirve de base. Pero pueden presentarse circunstancias que pongan en nuestras manos información adicional y que aconsejen un reasignación de probabilidades. Por ejemplo, si lanzamos un dado cerrando los ojos, antes de abrirlos y ver el número que muestra la cara superior es natural pensar que la probabilidad de que ocurra el # 5 es 1/6 ; pero si mientras estamos pensando, nos dicen que el número que ha salido es par, entonces tendríamos que reasignar probabilidades y en este caso, lo natural sería decir que la probabilidad del #5 es 0, dada la nueva información. El proceso que está debajo es simple: Inicialmente el espacio muestral es S={1,2,3,4,5,6} y si B denota el evento "Ocurre el #5 " y A denota el evento "El número es par", entonces B={5} y A={2,4,6}. Así P(B)=1/6, pues B tiene un sólo elemento. Pero si sabemos que A se ha presentado, entonces los únicos resultados posibles son 2 o 4 o 6: {2,4,6} y dentro de este **nuevo** espacio muestral, B no tiene ningún elemento, por tanto, su nueva probabilidad debe ser cero.

1.5.1 Probabilidad Condicional

Dados dos eventos A y B, tales que $P(A) > 0$, definimos la Probabilidad Condicional de B dado A, denotada $P(B \mid A)$, mediante:

$$P(B\mid A)=\frac{P(B\cap A)}{P(A)}$$

Observaciones:

(1) Si usamos la noción clásica de probabilidad, en donde ésta se mide como el cociente del número de elementos de un evento sobre el número de elementos en el espacio muestral, tenemos:

$$P(B\mid A)=\frac{P(B\cap A)}{P(A)}=\frac{n(B\cap A)/n(S)}{n(A)/n(S)}=\frac{n(B\cap A)}{n(A)}$$

Como se ve, P(B|A) es la probabilidad de B cuando S se reduce al evento A. Esto puede usarse para distinguirla de la probabilidad incondicional:

Al hacer el cálculo, uno debe preguntarse sobre cuál conjunto se esta trabajando; si el cálculo o porcentaje se está haciendo sin ninguna restricción (o sea sobre S) la probabilidad es **incondicional**. En cambio, si el universo se ha restringido, se trata de una probabilidad **condicional**. Por lo anterior, a veces nos referiremos al evento A como evento "condicionante".

(2) Observando el numerador del cociente en (1), se encuentra la razón por la cual a veces se confunde $P(B\mid A)$ con $P(B\cap A)$: En ambos casos el numerador es el mismo. De nuevo, la mejor manera de evitar la confusión es preguntarse por el universo sobre el cual se efectúa el cálculo, **si no es S sino otro evento (o sea A), se trata de una probabilidad condicional.**

Ejemplo 15

En el ejemplo de la encuesta de hogares, calcule la probabilidad de que en la muestra esté presente la clase acomodada, dado que tres hogares de la muestra resultaron de clase popular.

Solución:

Sean E = "En la muestra está presente la clase acomodada" y D = "Tres hogares de la muestra resultaron de clase popular", nos piden $P(E\mid D)$.

Como $P(E\mid D)$=1- $P(E^C\mid D)$ y tenemos $n(D)=C_3^{10}xC_2^{10}$; $n(E^C\cap D)=C_3^{10}xC_2^{6}$, entonces

$P(E^C\mid D)=\dfrac{C_3^{10}C_2^{6}}{C_3^{10}C_2^{10}}=\dfrac{C_2^{6}}{C_2^{10}}$ y la probabilidad pedida es $P(E\mid D)=1-\dfrac{C_2^{6}}{C_2^{10}}$

Regla del Producto

Se trata de un simple despeje en la definición, que permite reconstruir la probabilidad de A∩B a partir de la probabilidad condicional y de la probabilidad del evento condicionante. A saber, la Regla del Producto establece que:

$$P(A\cap B)=P(B\mid A)P(A)$$

Nótese que si $P(B)>0$, $P(A\mid B)$ también está definida y se puede escribir:

$$P(A\cap B)=P(A\mid B)P(B)$$

Observación:

La propiedad anterior se generaliza inductivamente a tres o más eventos. Por ejemplo, si las probabilidades condicionales están definidas, entonces

$$P(A \cap B \cap C) = P(C \mid A \cap B)P(B \mid A)P(A)$$

Ejemplo 16
Tres personas A, B y C se ponen en cola y toman una tras otra una carta de una baraja normal con 52 cartas. Gana quien tiene el as de espadas. Antes de hacer juego ¿Algún lugar en la cola es más conveniente?

Solución:
Definamos los eventos A = "Gana A", B = "Gana B" y C = "Gana C" y para simplificar supongamos que extraen sus cartas en ese orden: Primero A y si no gana, entonces extrae B y si no gana, extrae C y ahí acaba el juego. En este contexto y aplicando la regla del producto:

$P(A)$=1/52
$P(B)=P(A^C \cap B)= P(B|A^C)\times P(A^C)$=(1/51)×(51/52)=1/52
$P(C)=P(A^C \cap B^C \cap C)$=(1/50)×(50/51)×(51/52)= 1/52

Las tres probabilidades son iguales: Antes de empezar, todos los lugares son equiprobables para ganar.

1.5.2 Independencia Probabilística

Más de una vez, sucede que tenemos elementos para concluir que un evento A no es "condicionante" de otro evento B, es decir, que la ocurrencia de A no afecta las posibilidades de B, y por tanto la probabilidad de éste no se altera. Formalmente $P(B \mid A) = P(B)$. Y si reemplazamos en la Regla del Producto tenemos:

$$P(A \cap B) = P(B \mid A)P(A) = P(A)P(B)$$

Por razones de conveniencia matemática, la definición del concepto de independencia probabilística usa la última igualdad.

Independencia

Dos eventos A y B se dicen Independientes si y sólo se cumple la igualdad:

$$P(A \cap B) = P(A)P(B)$$

Observaciones:

(1) Esta definición se extiende inductivamente al caso de tres o más eventos.
(2) Se demuestra que si A y B son independientes, también lo son sus complementos y en general todos los eventos de la serie A, B, A^C, B^C.
(3) Si A y B no son independientes, diremos que están "asociados". En este contexto, el cociente $[P(A \cap B) - P(A)P(B)]/P(A)P(B)$ servirá de base para medir el "grado de asociación"

Ejemplo 17
La probabilidad del evento A = 'El lunes habrá baja de precios en el mercado de minerales' es 0.7 y la probabilidad del evento B = 'El martes habrá baja de precios en el mercado de minerales' es 0.8

a) Si el 65% de las veces hay dos días consecutivos de baja ¿Cuál es la probabilidad de que haya baja el martes, si el lunes sí la hubo? ¿Son independientes los eventos?

b) Si los eventos anteriores fueran independientes ¿Con qué probabilidad habrá baja de precios en el mercado en alguno de estos días?

Solución:
a) Sabemos que $P(A)=0.7$, $P(B)=0.8$, $P(A\cap B)=0.65$ y nos piden $P(B\,|\,A)$. Aplicando la definición de probabilidad condicional obtenemos $P(B\,|\,A)=\dfrac{P(B\cap A)}{P(A)}=\dfrac{0.65}{0.7}=0.93$;
En cuanto a la independencia $0.65=P(A\cap B)\neq P(A)P(B)=0.7\times 0.8=0.56$, es decir, los eventos no son independientes. Lo anterior también se deduce al ser $P(B\,|\,A)$ distinta de $P(B)$
b) Nos piden $P(A\cup B)=P(A)+P(B)-P(A\cap B)$. En esta nueva situación, dada la independencia, $P(A\cap B)=P(A)P(B)=0.56$ y por tanto $P(A\cup B)=0.7+0.8-0.56=0.94$

Ejemplo 18
Si A y B son independientes, pruebe que A^C y B^C son independientes

Solución:
Tenemos que demostrar $P(A^C\cap B^C)=P(A^C)P(B^C)$ sabiendo que $P(A\cap B)=P(A)P(B)$. Como
$A^C\cap B^C=(A\cup B)^C \Rightarrow P(A^C\cap B^C)=P((A\cup B)^C)=1-P(A\cup B)$ o equivalentemente

$$P(A^C\cap B^C)=1-P(A)-P(B)+P(A\cap B)=1-P(A)-P(B)+P(A)P(B)=P(A^C)-P(B)[1-P(A)]$$
$$=P(A^C)-P(B)P(A^C)=P(A^C)[1-P(B)]=P(A^C)P(B^C)\Rightarrow P(A^C\cap B^C)=P(A^C)P(B^C)$$

Ejemplo 19
Hay dos vías que van de la ciudad A a la ciudad B y dos vías que van de B a la ciudad C. Si cada una de las 4 vías tiene probabilidad 1/4 de ser interrumpida por huaycos, independientemente de las otras. ¿Con qué probabilidad se podrá pasar de A a C?

Solución:
Sean los eventos *G* =“Se puede pasar de A a C”; *E*=“Hay vía libre A a B” y *F*=“Hay vía libre de B a C”. Entonces:

$G=E\cap F$ y E,F son independientes. Luego $P(G)=P(E\cap F)=P(E)\times P(F)$ y como:

$P(E)=1-P(E^C)$; y $P(E^C)=P($ Cae huayco en las dos vías de A a B $)=\dfrac{1}{4}\times\dfrac{1}{4}=\dfrac{1}{16}$

entonces $P(E)=1-\dfrac{1}{16}=\dfrac{15}{16}$. Análogamente $P(F)=\dfrac{15}{16}$ y $P(G)=P(E\cap F)=\left(\dfrac{15}{16}\right)^2$

Ejemplo 20
En un mercado hay 3 productores del bien A y 3 productores del bien B y un consumidor puede elegir indistintamente a cualquiera de los productores de A, pero la probabilidad de que elija al productor j de B es directamente proporcional al gasto γ_j que dicho productor haga en propaganda, habiendo independencia entre las elecciones en A y en B. Sea S el espacio muestral dado por S={w_{ij}| w_{ij}=El consumidor elige al productor i de A y al productor j de B}. Asuma precio único para A y también para B.

a) Halle la asignación o distribución de probabilidades en S si $\gamma_j = j$

b) Un inversionista adquiere las empresas de los productores 1 y 2 de A y la del productor 3 de B. Calcule la probabilidad de que el consorcio formado por el inversionista no logre realizar ventas.

c) Si el inversionista de b) sí ha logrado hacer ventas: ¿Cuál sería la probabilidad de que esto haya ocurrido sólo en B?

Solución:

a) $S = \{w_{ij} \mid w_{ij} = el\ \ consumidor\ \ elige\ \ al\ \ productor\ \ i\ \ de\ \ A\ \ y\ \ al\ \ j\ \ de\ \ B\}$ (ver figura 1.4), por dato tenemos que $P(i) = 1/3 \quad i = 1, 2, 3$ y $P(j) = \alpha j \quad j = 1, 2, 3$ y como hay independencia $P(w_{ij}) = P(i)P(j) = \alpha j/3$ *i=1,2,3; j=1,2,3.*

S tiene 9 puntos (1,1),(1,2),(1,3),...,(3,3) y como $P(S) = \sum_{i=1}^{3}\sum_{j=1}^{3} P(w_{ij}) = 1$, evaluando la probabilidad $P(i)P(j) = \alpha j/3$ en cada pareja de posibles elecciones del consumidor y sumando, obtenemos la ecuación:

$\frac{1}{3}(\alpha + 2\alpha + 3\alpha) + \frac{1}{3}(\alpha + 2\alpha + 3\alpha) + \frac{1}{3}(\alpha + 2\alpha + 3\alpha) = 1 \Rightarrow 6\alpha = 1 \Rightarrow \alpha = 1/6$ y la distribución de probabilidades en S es $P(w_{ij}) = j/18$ *i=1,2,3; j=1,2,3.*

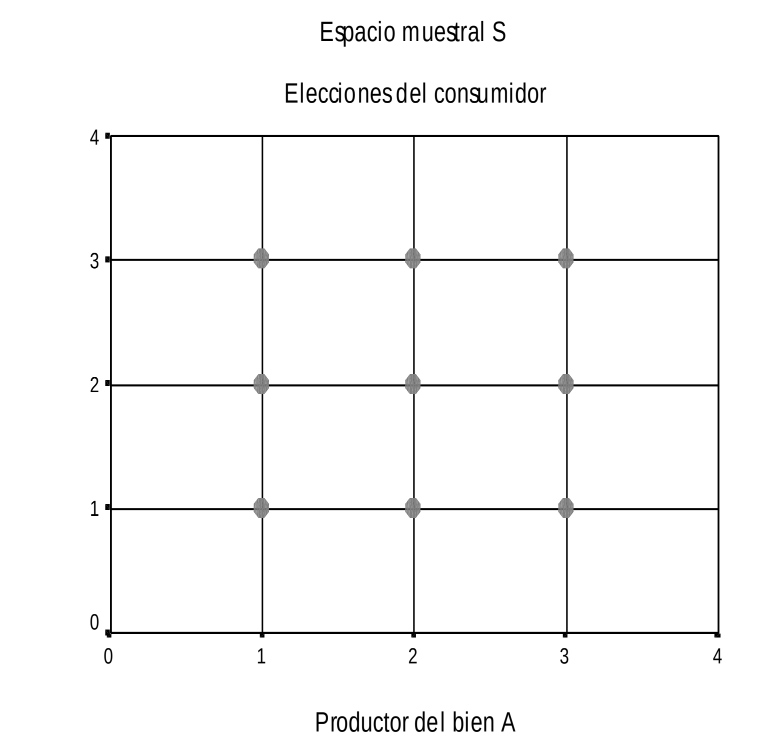


Figura 1.4

b) Sea C = "El consorcio del inversionista no realiza ventas", entonces $C = \{w_{31}, w_{32}\}$ pues "no realiza ventas" equivale a "no vende A ni B" que a su vez equivale a que el consumidor elige comprar A y B fuera del consorcio (ver figura 1.5)**.** Evaluando:

$P(C) = P(\{w_{31}, w_{32}\}) = P(\{w_{31}\}) + P(\{w_{32}\}) = (1/18) + (2/18) = 3/18$

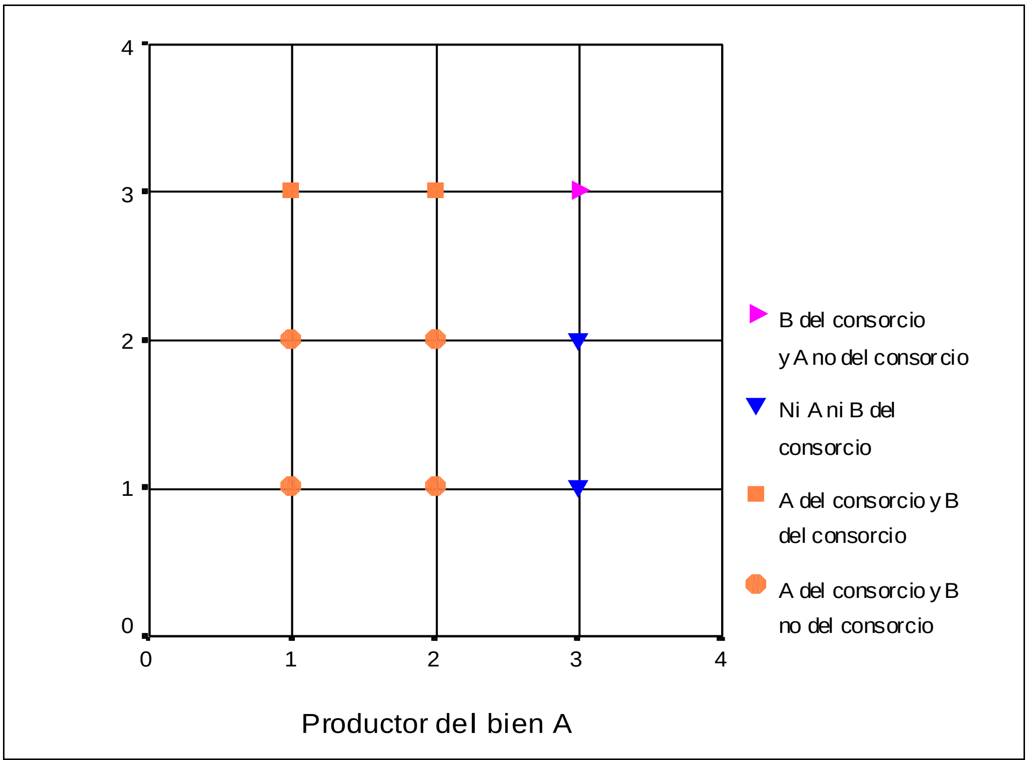


Figura 1.5

d) Nos piden $P(\{w_{33}\} \mid C^C) = \dfrac{P(\{w_{33}\} \cap C^C)}{P(C^C)} = \dfrac{P(\{w_{33}\})}{P(C^C)} = \dfrac{3/18}{1-(3/18)} = \dfrac{3}{15}$ (ver figura 1.6)

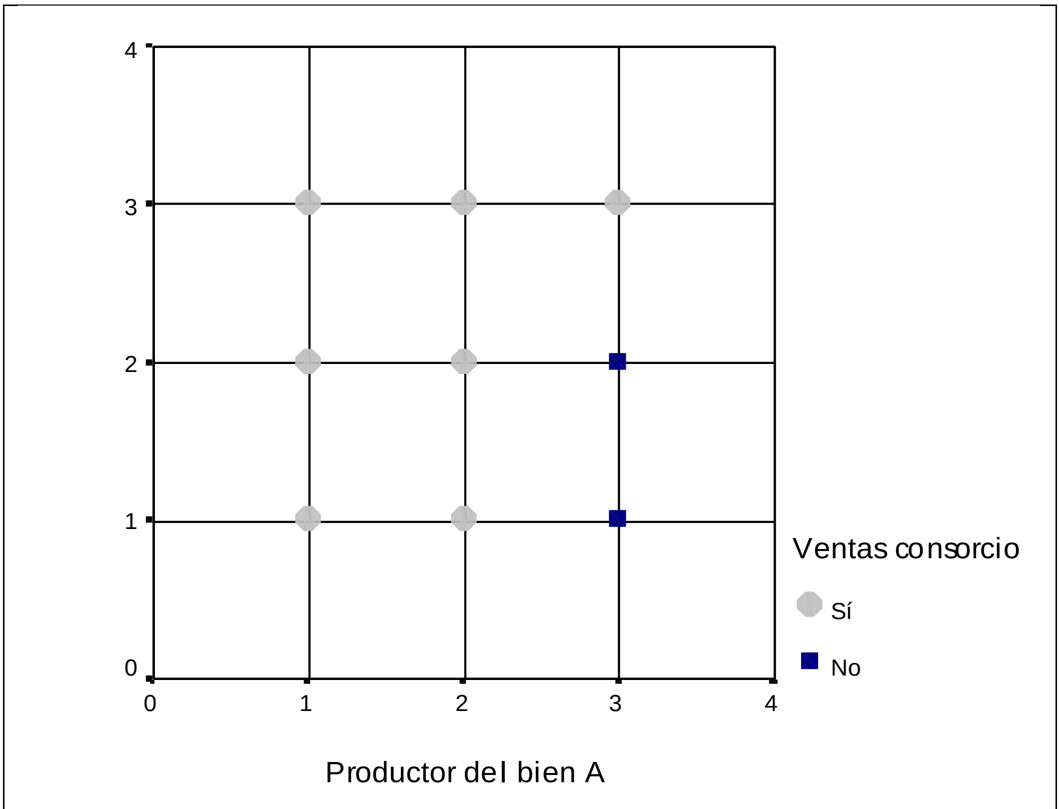


Figura 1.6

1.6 Probabilidad Total y Teorema de Bayes

Una aplicación de la probabilidad condicional es permitir la reconstrucción de probabilidades previas o incondicionales, a partir de las condicionales. Lo anterior se hace gracias a dos teoremas simples pero importantes.

1.6.1 Teorema de Probabilidad Total

Sean $\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_N$ eventos mutuamente excluyentes, todos con probabilidad positiva y tales que $\bigcup_j A_j = \mathrm{S}$. Sea B otro evento de S. Entonces se cumple:

$$P(B) = \sum_j (B \mid A_j) P(A_j)$$

Demostración:

$$B = B \cap S = B \cap \left[\bigcup_j A_j \right] = \bigcup_j \left[B \cap A_j \right]$$

Aplicando la Regla del Producto a $P(B \cap A_j)$ obtenemos:

$$P(B \cap A_j) = P(B|A_j)P(A_j) \quad \forall j$$

luego:

$$P(B) = P\left(\bigcup_j \left[B \cap A_j \right] \right) = \sum_j P(B \cap A_j) = \sum_j P(B|A_j)P(A_j)$$

1.6.2 Teorema de Bayes

En el contexto del Teorema de Probabilidad Total, si además P(B)>0, entonces se cumple:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_j P(B|A_j)P(A_j)} \quad \forall\, k$$

Demostración:

Usando la definición de Probabilidad Condicional y luego la Probabilidad Total:

$$P(A_k \mid B) = \frac{P(B \cap A_k)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{\sum_j P(B \mid A_j)P(A_j)} \quad \forall k$$

Ejemplo 21

Ante una pregunta de opción múltiple con 5 opciones, el examinado puede saber la respuesta, no saberla en absoluto o tener dudas. Si no sabe marca al azar y si tiene dudas, después de un análisis de opciones puede reducir las mismas a las 3 más plausibles, una de las cuales es la correcta. Datos previos indican que el 55% sabe la respuesta y el 15% no la sabe en absoluto.

a) ¿Qué proporción de aciertos se espera en esta pregunta?
b) Un examinado acertó en la pregunta ¿Sabrá verdaderamente la respuesta?

Solución:

Sean los eventos:

A_1 = "El examinado sabe la respuesta"

A_2 = "El examinado no sabe la respuesta en absoluto"

A_3 = "El examinado tiene dudas"
B = "El examinado acierta en la respuesta"

Tenemos los datos iniciales: $P(A_1) = 0.55$; $P(A_2) = 0.15$ y podemos completar entonces: $P(A_3) = 1 - 0.55 - 0.15 = 0.30$;

Por otra parte

$P(B \mid A_1) = 1$, pues si sabe la respuesta, obviamente marca lo correcto;

$P(B \mid A_2) = 1/5 = 0.2$, pues si no sabe, contesta al azar sobre el total de 5 preguntas, y

$P(B \mid A_3) = 1/3 = 0.33$, pues si tiene dudas, siempre puede reducir el conjunto de casos posibles a 3, uno de los cuales es el correcto.

Por tanto:

a) En este caso nos piden $P(B)$ y aplicando probabilidad total:

$$P(B) = \sum_{j=1}^{3} P(B \mid A_j)P(A_j) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3) =$$
$$(1 \times 0.55) + (0.2 \times 0.15) + (0.33 \times 0.30) = 0.678$$

b) Aquí debemos calcular $P(A_1 \mid B)$. Aplicando el Teorema de Bayes:

$$P(A_1 \mid B) = \frac{P(B \cap A_1)}{P(B)} = \frac{P(B \cap A_1)P(A_1)}{\sum_{j=1} P(B \mid A_j)P(A_j)} = \frac{(1x0.55)}{0.679} = 0.81$$

que es lo bastante alta como para asegurar que lo más probable es que, si acertó en la respuesta es porque de verdad la sabía.

Nótese que también se puede calcular $P(A_2 \mid B) = 0.146$ y $P(A_3 \mid B) = 0.044$ y con estas cifras a la mano, podemos decir que dentro del grupo de personas que aciertan el ítem o pregunta, aquellos que no saben nada de él, son la minoría absoluta, el 4.4% aproximadamente.

Ejemplo 22

E un mercado laboral, el 20% de trabajadores tiene 5 años de escolaridad, el 70% tiene 10 años y el resto tiene 15 años. La probabilidad ***p*** de que un trabajador sea estable, está condicionada por sus años de escolaridad ***E*,** a través de $p = 1/(1 + e^{-0.04E})$.
Dado que Ud. entrevista a un trabajador de este mercado y resulta que tiene empleo estable ¿Qué nivel de escolaridad sería el más razonable imputarle? ¿Por qué?

Solución:

Sabemos que la persona tiene empleo estable y nos preguntamos por su escolaridad, que puede ser de cinco, diez o quince años. Aunque no podemos decir con seguridad cuál es la escolaridad, sí podemos identificar la más probable, que sería entonces nuestra mejor conjetura.

Como dato, tenemos una fórmula general que condiciona la estabilidad del empleo a la escolaridad, vía $p = 1/(1 + e^{-0.04E})$, donde E puede valer 5, 10 o 15.

Sean entonces los eventos A_1 = `Escolaridad de cinco años´; $A_2$ = `Escolaridad de diez años´; A_3 = `Escolaridad de quince años´ y B = 'Empleo estable'. Necesitamos calcular $P(A_k \mid B)$ para k = 1, 2, 3 y determinar cuál es mayor. Según el Teorema de Bayes

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_j P(B|A_j)P(A_j)} \quad \forall k$$

Ahora bien, de acuerdo a los datos y aplicando la fórmula general de las probabilidades condicionales del empleo estable:

Si ocurre $A_1 \Rightarrow E = 5 \Rightarrow P(B \mid A_1) = 1/(1+e^{-0.04\times 5}) = 0.55$;
Si ocurre $A_2 \Rightarrow E = 10 \Rightarrow P(B \mid A_2) = 1/(1+e^{-0.04\times 10}) = 0.60$ y finalmente,
Si ocurre $A_3 \Rightarrow E = 15 \Rightarrow P(B \mid A_3) = 1/(1+e^{-0.04\times 15}) = 0.65$

Por otra parte $P(A_1) = 0.20$; $P(A_2) = 0.70$ y $P(A_3) = 0.10$

Evaluando tenemos:

$$P(B) = \sum_j P(B|A_j)P(A_j) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) =$$

$= 0.2\times 0.55 + 0.7\times 0.6 + 0.1\times 0.65 = 0.595$, luego:

$$P(A_1 \mid B) = \frac{0.2\times 55}{0.595} = 0.18;\ P(A_2 \mid B) = \frac{0.7\times 0.60}{0.595} = 0.71;\ P(A_3 \mid B) = \frac{0.1\times 0.65}{0.595} = 0.11$$

Comparando probabilidades, concluimos que la escolaridad más razonable para esta persona es 10 años.

Ejemplo 23

Tres personas se dividen un trabajo de Estadística, de modo que cada una trabaja su parte independientemente y se ponen de acuerdo para integrar sus trabajos en uno solo. Las probabilidades de fallar en sus respuestas son 0.2, 0.15 y 0.25 respectivamente: Halle la probabilidad de que presenten bien resuelto el trabajo; De que el trabajo tenga algún error; De que el más capaz de los tres haya fallado, si el trabajo entregado tenía error.

Solución:

Sean A= "A falla en su parte"; B= "B falla en su parte" y C= "C falla en su parte", por dato, $P(A)$ = 02; $P(B)$=0.15 y $P(C)$=0.25, siendo independientes los eventos. En este contexto:

- Si D = "Trabajo bien resuelto" = "Ningún error o nadie falla"=$A^C \cap B^C \cap C^C$, luego: P(Trabajo bien resuelto)=P(Nadie falla)=$P(A^C \cap B^C \cap C^C)$= $P(A^C)\times P(B^C)\times P(C^C)$=0.8×0.85×0.75 =0.51
- P(Trabajo con algún error)=$P(D^C)$=1 – P(Ningún error) =1 – $P(D)$=1-0.51=0.49
- "El más capaz" es el que tiene menor probabilidad de fallar, o sea B. En este contexto se pide:

$$P(B \mid D^C) = \frac{P(B\cap D^C)}{P(D^C)} = \frac{P(D^C \mid B)P(B)}{P(D^C)} = \frac{1\times 0.15}{0.49} = \frac{0.15}{0.49} = 0.31$$

Note que <u>si</u> B falla en su parte, entonces necesariamente ya hay error en el trabajo, o sea ocurre D^C y por eso $P(D^C \mid B) = 1$

Ejemplo 24

Un economista que trabaja en una agencia de publicidad sabe que con una probabilidad de 80% las amas de casa de un sector social ven telenovelas románticas y que el 70% de amas de casa ve 'talk shows'. También se sabe que el 10% de amas de casa no ve ninguno de los dos tipos de programa. ¿Cuál es la probabilidad de que un ama de casa entrevistada al azar vea ambos tipos de programa? ¿Cuál es la probabilidad de que un ama de casa entrevistada al azar vea telenovelas románticas si ya declaró no ver 'talk shows'?

Solución:

Sean $\boldsymbol{A}$ = "El ama de casa ve telenovelas románticas" y B = "El ama de casa ve talkshows". Tenemos como datos $P(A)=0.80$; $P(B)=0.70$; $P(A^C \cap B^C)=0.10$ y nos piden $P(A\cap B)$ y $P(A|B^C)$. Mejor usamos una tabla de probabilidades o de contingencia:

Tabla 5 Datos iniciales y faltantes en la tabla de contingencia de A y B

$\boldsymbol{S}$	$\boldsymbol{A}$	$\boldsymbol{A^C}$	**Total**
$\boldsymbol{B}$	$P(A\cap B)$	$P(A^C \cap B)$	**0.7**
$\boldsymbol{B^C}$	$P(A\cap B^C)$	**0.10**	$P(B^C)$
Total	**0.8**	$P(A^C)$	1

Tabla 6 Tabla de contingencia de A y B

$\boldsymbol{S}$	$\boldsymbol{A}$	$\boldsymbol{A^C}$	**Total**
$\boldsymbol{B}$	0.6	0.1	**0.7**
$\boldsymbol{B^C}$	0.2	**0.1**	0.3
Total	**0.8**	0.2	1

Explicación del llenado de la tabla:

Datos: $\boldsymbol{P(A)=0.8}$; $\boldsymbol{P(B)=0.7}$ y $\boldsymbol{P(A^C\cap B^C)=0.10}$

Con los datos pasamos a calcular y completar la tabla:

$\boldsymbol{P(A^C)}$=1-$P(A)$=1-0.8=**0.2** y como $A^C=(A^C\cap B)\cup(A^C\cap B^C)$ $\Rightarrow P(A^C)=P(A^C\cap B)+P(A^C\cap B^C) \Rightarrow 0.2= P(A^C\cap B)+0.1 \Rightarrow$ $\boldsymbol{P(A^C\cap B)=0.1}$; También $\boldsymbol{P(B^C)}$=1-$P(B)$=**0.3** y como $B^C= (B^C\cap A) \cup(B^C\cap A^C) \Rightarrow P(B^C)=P(B^C\cap A)+ P(B^C\cap A^C)\Rightarrow 0.3= P(B^C\cap A)+0.1\Rightarrow$ $\boldsymbol{P(B^C\cap A)=0.2}$

Finalmente $P(A)=P(A\cap B)+ P(A\cap B^C)\Rightarrow 0.8= P(A\cap B)+0.2\Rightarrow$ $\boldsymbol{P(A\cap B)=0.6}$

- "Ama de casa ve ambos tipos de programa"=$A\cap B$ y leyendo en la tabla se ve $\boldsymbol{P(A\cap B)=0.6}$
- P(Ama de casa ve telenovelas románticas si ya declaró no ver 'talk shows')=$\boldsymbol{P(A|B^C)}= \frac{P(A\cap B^C)}{P(B^C)} = \frac{0.2}{0.3} = 0.67$

Ejemplo 25

Un economista genera un ranking de las tres empresas que considera con mejores perspectivas para el año actual, a partir de la lista de las diez empresas que tuvieron mayores utilidades el año pasado, digamos las empresas A1,A2,…,A10.

Las empresas A1, A7 y A5 forman parte de la corporación ABEP. En este contexto, usando probabilidad clásica y principios básicos de conteo, calcule la probabilidad de que:

a) ABEP cope el ranking confeccionado por el economista.
b) ABEP figure en el ranking.
c) ABEP ocupe los puestos primero y tercero del ranking

Solución:

$S = \{(x_1,x_2,x_3)|\ x_i \in\{A1,A2,\ldots,A10\},\ x_1 \neq x_2 \neq x_3\}$. La restricción $x_1 \neq x_2 \neq x_3$ es natural pues una misma empresa no puede estar en dos puestos del ranking a la vez.

Aplicando el Principio de multiplicación:

$n(S) = 10 \times 9 \times 8$: pues hay diez maneras de asignar una empresa al primer lugar (x_1) y nueve maneras de asignar una empresa al segundo lugar (x_2) y ocho maneras de asignar una empresa al tercer lugar (x_3)

a) Sea A el evento A = "ABEP copa el ranking", entonces para que eso ocurra, el economista debe

haber seleccionado para los tres puestos sólo empresas de la corporación ABEP, o sea, debe haber seleccionado sólo entre A1, A7 y A5.
Aplicando el Principio de multiplicación a este caso:
$n(A) = 3 \times 2 \times 1$: pues hay tres maneras de asignar una empresa de ABEP al primer lugar (x_1) y luego hay dos maneras de asignar una empresa al segundo lugar (x_2) y hecho esto, ya sólo queda una manera de asignar una empresa de ABEP al tercer lugar (x_3). Entonces, la probabilidad pedida es:
$P(A) = \frac{n(A)}{n(S)} = \frac{3 \times 2 \times 1}{10 \times 9 \times 8} = \frac{1}{120}$

b) Si B = "ABEP figura en el ranking", en este caso es mejor calcular $P(B) = 1 - P(B^C)$ y como
$P(B^C) = \frac{n(B^C)}{n(S)} = \frac{7 \times 6 \times 5}{10 \times 9 \times 8} = \frac{210}{720} = \frac{21}{72} \Rightarrow P(B) = 1 - \frac{21}{72} = \frac{51}{72}$

c) Si definimos D = "ABEP ocupa los puestos primero y tercero del ranking" entonces
$P(D) = \frac{n(D)}{n(S)} = \frac{3 \times 7 \times 2}{10 \times 9 \times 8} = \frac{7}{120}$

Ejemplo 26
Una acción valuada inicialmente en S/.10 , se cotiza en dos periodos. Durante cada periodo la acción puede bajar, permanecer igual o subir en S/. 1. Las probabilidades de estos eventos en el primer periodo son 0.2, 0.5 y 0.3 respectivamente. En el segundo la acción puede volver a subir con probabilidad 0.3 y bajar después de subir en el primero, con probabilidad 0.1; Ahora, si la acción se mantiene igual en el periodo inicial, puede subir en el segundo con probabilidad 0.4, y puede mantenerse igual con probabilidad 0.35; Finalmente, si la acción baja en el primer periodo, lo seguirá haciendo en el segundo con probabilidad 0.5, y subirá con 0.1 de probabilidad.

a) Halle la probabilidad de que el valor final de la acción sea S/. 12
b) Halle la probabilidad de que el valor final de la acción sea S/. 11
c) Si no cambia el valor final de la acción, en relación al valor inicial ¿Cuál sería la probabilidad de que la acción haya bajado?

Solución:

Sean los eventos:
$A1$ = Acción sube en el primer periodo, $A2$ = Acción permanece igual en el primer periodo, $A3$ = Acción baja en el primer periodo.
$B1$ = Acción sube en el segundo periodo, $B2$ = Acción permanece igual en el segundo periodo, $B3$ = Acción baja en el segundo periodo

a) Si C = Valor final de la acción es S/.12, entonces $C = A1 \cap B1$ y aplicando la regla del producto:
$P(C) = P(A1 \cap B1) = P(B1|A1)P(A1) = 0.3 \times 0.3 = 0.09$

b) Con los mismos eventos de a), sea ahora D = Valor final de la acción es S/.11, entonces
$D = (A1 \cap B2) \cup (A2 \cap B1)$ y $P(D) = P(A1 \cap B2) + P(A2 \cap B1) = 0.6 \times 0.3 + 0.4 \times 0.5 = 0.38$

c) Con los mismos eventos de a) sea ahora E = Valor final es igual al valor inicial, entonces
$E = (A1 \cap B3) \cup (A2 \cap B2) \cup (A3 \cap B1)$ y $P(E) = P(A1 \cap B3) + P(A2 \cap B2) + P(A3 \cap B1)$, etc.

Pero se pregunta por $P\big((A1 \cap B3) \cup (A3 \cap B1)|E\big)$. Aplicando la definición de probabilidad condicional:
$P\big((A1 \cap B3) \cup (A3 \cap B1)|E\big) = P\big((A1 \cap B3)|E\big) + P\big((A3 \cap B1)|E\big) = \frac{P(A1 \cap B3 \cap E)}{P(E)} + \frac{P(A3 \cap B1 \cap E)}{P(E)} =$
$\frac{P(E|A1 \cap B3)P(A1 \cap B3)}{P(E)} + \frac{P(E|A3 \cap B1)P(A3 \cap B1)}{P(E)}$ y donde $P(E|A1 \cap B3) = 1;\ P(E|A3 \cap B1) = 1$, etc. El resto se deja al lector.

2. Variable Aleatoria

La construcción de "Leyes del Azar" que estamos intentando, se complica innecesariamente si insistimos en trabajar con espacios muestrales generales, y la razón es simple: tal como S es definido, puede estar constituido de objetos bastante variados: letras, símbolos, números; lo que hace difícil elaborar enunciados teóricos generales que se refieran a los elementos de S y peor aún, también impide aplicar la mayor parte de herramientas matemáticas conocidas para análisis teóricos (como la diferenciación, integración, etc.), obstaculizando el uso de "fórmulas" simplificadoras.
Por otra parte, muchas veces el interés está centrado en determinados aspectos del espacio muestral y no en todo el conjunto. Entonces, podemos ganar mucho si, por ejemplo, "codificamos" el espacio muestral original S en un conjunto de números R, transfiriendo a su vez las probabilidades, de modo que en R podamos aplicar las operaciones de suma, multiplicación, etc., y usarlas para facilitar el cálculo de probabilidades.

Ejemplo 27
En el mercado de un bien con 6 productores, se sabe que al menos hay dos coaliciones y un organismo de regulación se interesa por el tamaño X de la mayor coalición.

En este contexto, y si no hay otra información, el experimento aleatorio ε consiste en observar el estado del mercado, que puede lograrse de las siguientes maneras:
a) Dos coaliciones de 2 empresas cada una y las otras empresas libres, o
b) Dos coaliciones, una de 2 empresas y la otra de 3, con la empresa restante libre, o
c) Dos coaliciones, una de 2 empresas y la otra de 4 o
d) Dos coaliciones, cada una con tres empresas, o finalmente,
e) Tres coaliciones, cada una con dos empresas.

El tamaño de S es $n(S) = C_2^6 C_2^4 + C_2^6 C_3^4 + C_2^6 C_4^4 + C_3^6 C_3^3 + C_2^6 C_2^4 C_2^2 = 275$ estados posibles; y si sólo interesa el tamaño de la coalición más grande, denotemos con X este tamaño. Entonces:
(i) X puede tomar los valores 2, 3 o 4;
(ii) Las probabilidades de que ocurran estos tamaños de coalición son respectivamente:

$$P(X=2) = \frac{C_2^6 C_2^4 + C_2^6 C_2^4 C_2^2}{275} = \frac{180}{275};\ P(X=3) = \frac{C_2^6 C_3^4 + C_3^6 C_3^3}{275} = \frac{80}{275} \text{ y } P(X=4) = \frac{C_2^6 C_4^4}{275} = \frac{15}{275}$$

Como se ve, el sistema de codificación usado, en realidad es una función X que convierte cada punto de S en un número y simplifica la presentación de probabilidades al darle a S el formato de un conjunto numérico R. Ciertamente toda la información presente en S no se trasladó a R, pero sí aquella que era de nuestro particular interés: el tamaño de la mayor coalición y sus respectivas probabilidades. Y podemos resumir todo en una tabla de distribución de probabilidades:

x	$P(X=x)$
2	$180/275$
3	$80/275$
4	$15/275$
Total	1

2.1 Definición y clasificación

2.1.1 Definición

Si S es un espacio muestral, sobre el cual se ha construido una σ-álgebra de eventos $\mathscr{A}$ **,** una variable aleatoria X definida sobre S, es una función cuyo dominio es S y cuyo rango es un conjunto de números reales que denotaremos R_X, que además satisface que para todo $t \in \mathrm{R}$ se cumple $\{w \in S | X(w) \leq t\} \in \mathscr{A}$ **.**

La última condición permite "probabilizar" todo el eje real y es necesaria para evitar patologías cuando se trata con espacios S continuos, como los que aparecen en los modelos probabilísticos que describen procesos en finanzas. No verificaremos la condición de aquí en adelante y asumiremos que se cumple en condiciones bastante generales.

Ejemplo 28

Si una persona contesta al azar un pregunta de opción múltiple con cinco opciones, una de las cuales es verdadera, un espacio muestral apropiado sería $S = \{C, I\}$, donde C indica el resultado "La persona acierta" e I denota el resultado "La persona no acierta". Una asignación o distribución de probabilidades natural es, en este caso:

$$P(\{C\}) = \frac{1}{5} \text{ y } P(\{I\}) = \frac{4}{5}$$

Sea la variable aleatoria (v.a.) X = Número de errores. Aplicando la regla que define X, tenemos $X(C) = 0$ y $X(I) = 1$, de modo que el rango (de valores posibles) de X es $R_X = \{0,1\}$ y una distribución de probabilidades en R_X es:

$$P(X = 0) = \frac{1}{5} \text{ , pues (} X = 0 \text{) ocurre si y sólo si ocurre \{C\}}$$

$$P(X = 1) = \frac{4}{5} \text{ , pues (} X = 1 \text{) ocurre si y sólo si ocurre \{I\}}$$

Nótese que también podemos escribir, de manera compacta:

$$P(X = k) = \frac{4^k}{5} \quad k = 0,1$$

que proporciona una fórmula para distribuir probabilidades en R_X

Ejemplo 29

Si una persona contesta al azar 2 preguntas tipo V o F y se define X como el número de errores, hallar R_X y la distribución de probabilidades en R_X.

Solución:

Si denotamos mediante pares ordenados los resultados posibles, tenemos que S se puede escribir:

$S = \{(C,C),(C,I),(I,C),(I,I)\}$ y aplicando la v.a. X

X =# de Errores

$R_X = \{\ 0,\quad 1,\quad 2\ \}$

Como en S hay 4 resultados que podemos considerar equiprobables, las probabilidades se pueden "transferir" del espacio muestral original S, al espacio transformado o "codificado" R_X, de modo que tenemos:

$$P(X=0)=P[\{(C,C)\}]=\frac{1}{4};\ P(X=1)=P[\{(C,I),(I,C)\}]=\frac{2}{4} \text{ y } P(X=2)=P[\{(I,I)\}]=\frac{1}{4}$$

Ordenando en una Tabla los valores de X y sus probabilidades:

x	0	1	2
$P(X=x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

o también, y de manera compacta:

$$P(x)=\frac{C_x^2}{4} \quad donde \quad x=0,1,2.$$

Nota:
Si las preguntas no tienen dos opciones sino cinco, aunque R_X se mantiene, el cálculo de la distribución de probabilidades $P(x)$ se complica algo y se tiene $P(x)=\frac{C_x^2 4^x}{25} \quad donde \quad x=0,1,2.$
Análogamente, si fueran 20 preguntas y no 2, todas de cinco opciones, tendríamos
$P(x)=\frac{C_x^{20} 4^x}{5^{20}} \quad donde \quad x=0,1,2,...20$

Ejemplo 30
En una privatización hay dos postores de los cuales sólo se sabe que pueden ofrecer precios entre 0 y 1 millón de unidades monetarias por la la empresa que desean comprar. Sea X el precio de venta. Determine la transformación que define a X así como el rango de valores posibles de X (R_X) y una fórmula para $P(X\le x)$, donde $x \in R_X$.

Solución:
Podemos representar adecuadamente los resultados posibles de este experimento aleatorio mediante:
$S=\{(a,b)|0<a\le 1, \quad 0<b\le 1\}$ donde a y b son los precios ofrecidos por A y B respectivamente.

Como se trata de una privatización, el mayor precio será el ganador, esto es $X=Máx\{a,b\}$ define la transformación que va de S a $\mathbb{R}$ y por tanto $R_X=]0,1]$.

Un punto más delicado es hallar una fórmula general para $P(X\le x)$, donde x es un valor genérico pero dado de R_X:

El evento $[X\le x]$ (definido en R_X) ocurre si y sólo si 0 < a ≤ x y 0 < b ≤ x
Sea el evento $D=\{(a,b)\in S|0<a\le x, \quad 0<b\le x\}$, entonces podemos decir que $[X\le x]$ y D son equivalentes. Por tanto:

$$P(X\le x)=P(D)$$

Ahora, aplicando probabilidad geométrica sobre los eventos de S tenemos:

$$P(X \leq x) = P(D) = \frac{Area(D)}{Area(S)} = \frac{x^2}{1} = x^2 \text{ para } 0 \leq x \leq 1$$

y así obtenemos una fórmula que distribuye probabilidades sobre R_X .

Observaciones:
A partir de los ejemplos anteriores, podemos concluir que:

(1) En general, R_X puede verse como el resultado de una "codificación" de S , pues $\begin{array}{c} X : S \to R \\ w \mapsto X(w) \end{array}$ convierte cada punto w de S en un número $X(w)$ que está en R_X y donde las probabilidades definidas sobre eventos de S se transfieren a los eventos del nuevo espacio muestral R_X . Ahora bien, la función X que usamos para codificar, la definimos de modo que capte la información de particular interés para nosotros, y si deseamos información relativa a otro aspecto de S , podemos definir un sistema adicional de codificación Y , y así sucesivamente. Por ejemplo, en el caso de la licitación, otra variable de interés puede ser Y = Diferencia entre las ofertas $= |a-b|$, cuyo rango es $R_Y = [0,1]$.

(2) La función X , como tal, no tiene nada aleatorio, pues como toda función que se respete, debe ser una regla de correspondencia bien definida. Pero al tomar como "insumo" a los elementos w de S , que sí son aleatorios, su "producto" $X(w)$ deviene en azaroso, no podemos pronosticar cuál valor $X(w)$ ocurrirá, de ahí el nombre de "variable aleatoria" para X : Sus valores cambian con w y lo hacen al azar. Por lo mismo, una manera laxa, pero útil, de ver a una v.a. X es como *"una variable que toma sus valores al azar"*.

(3) La utilidad de trabajar con un espacio muestral transformado como lo es R_X , reside en que al ser sus elementos números, es factible aprovechar las propiedades de éstos para simplificar los resultados, construyendo, por ejemplo, fórmulas que proporcionen distribuciones de probabilidades. Esto último, no se puede hacer con espacios muestrales generales.

(4) Aunque ya lo mencionamos en la definición, se exige que el conjunto $\{w \in S | X(w) \leq t\}$ sea un evento de S para cualquier número real *t*, de modo que ese conjunto tenga probabilidad bien definida. El objetivo de esta condición es "probabilizar" **todo** R . Es posible construir transformaciones de S a R que no cumplen esta condición, pero son poco útiles y no aparecen con frecuencia en situaciones de aplicación.

2.1.2 Clasificación de las variables aleatorias

De acuerdo al rango R_X de una variable aleatoria, clasificaremos a ésta como:

Variable aleatoria Continua: Si R_X es un intervalo
Variable aleatoria Discreta: Si R_X es un conjunto finito o numerable

Ejemplos 31
La variable X = Número de errores en el ejemplo de las respuestas al azar en dos preguntas de opción múltiple, es discreta.
La variable X = Precio de venta en el ejemplo de la licitación es continua.

Existen variables mixtas, que se comportan como continuas en ciertos tramos de su rango y como discretas en otros. No son tan frecuentes, pero tienen su importancia, aunque no las trataremos aquí.

Observaciones:

(1) Cuando X es discreta la probabilidad se "concentra" en determinados puntos del eje real, aquellos que constituyen R_X; en cambio si X es continua, la probabilidad se distribuye sobre intervalos contenidos en R_X

(2) Como notaciones usaremos las siguientes:

$$(X = k) := \{w \in S | X(w) = k\}$$
$$(X \leq t) := \{w \in S | X(w) \leq t\}$$

2.2 Variable discreta y Función de Probabilidad

Definición

Si X es v.a. discreta, la función de probabilidad de X, denotada $P_X(x)$, o también $f_X(x)$, se define mediante $P_X(x) = P(X = x)$

Observaciones:

(1) $P_X(x) = 0$ si $x \notin R_X$

(2) $P_X(x)$ es llamada también "función de distribución de probabilidades" pues indica cómo se redistribuye la probabilidad total de S (que es 1), entre los valores alternativos (y discretos) que puede asumir X.

Ejemplo 32

Si la probabilidad de que una perforación petrolera resulte en un pozo rentable es p y se define X = # de pozos perforados hasta que se descubre el primer pozo rentable, hallar $P_X(x)$

Solución:

Es claro que $R_X = \{1,2,3,...\}$. Sea x un valor genérico pero dado de X, i.e. $x \in R_X$, tenemos que:

$[X = x]$ ocurre si y sólo si 'Las (x-1) primeras perforaciones son no rentables y la x-ésima perforación es rentable'. Entonces podemos escribir:

$P[X = x] = \overbrace{(1-p)(1-p)(1-p)...(1-p)}^{(x-1)\text{veces}}p = p(1-p)^{x-1}$ y por tanto la función de probabilidad de X es

$$P_X(x) = P(X = x) = \begin{cases} p(1-p)^{x-1} & \text{si} \quad x = 1,2,3,... \\ 0 & \text{en otro caso} \end{cases}$$

Ejemplo 33

Un etólogo construye un modelo para estudiar la capacidad de memoria de las ratas albinas. Para ello, diseña el experimento de colocar una rata de laboratorio en un laberinto con cinco salidas, de las cuales sólo una conduce al exterior y las otras, después de un recorrido, retornan a la rata al centro del laberinto; luego estimula al animal para que intente salir. Sea X la v.a. definida como el número de intentos de escape hasta que la rata logra salir. Halle R_X y $P_X(x)$ si:

a) La rata no tiene memoria alguna

b) La rata tiene memoria perfecta

Solución:

(a) **Si no hay memoria,** en cada intento se repite la misma situación, la rata escoge "al azar" una salida, pudiendo elegir inclusive la misma salida seleccionada en anteriores intentos. Es claro que con este modelo la rata puede estar tratando de salir "ad infinitum". Luego $R_X = \{1,2,3,...\}$ y si x es un valor dado y genérico de R_X, necesitamos calcular $P(X = x)$:

Sean los eventos $A_i =$ "La rata escapa en el intento #i" , i = 1,2,3,... entonces tenemos:

$$P(X = x) = P(\overbrace{A_1^C \cap A_2^C \cap A_3^C ... \cap A_{x-1}^C}^{(x-1)veces} \cap A_X) = \overbrace{(4/5)(4/5)(4/5)...(4/5)}^{(x-1)veces}(1/5) = (4/5)^{x-1}(1/5)$$

Lo anterior sale aplicando la regla del producto repetidas veces. Entonces llegamos a

$P_X(x) = (4/5)^{x-1}(1/5) \qquad x = 1,2,3,...$

(b) **Si la memoria es perfecta**, la rata también saldrá por azar, pero podrá ir "controlándolo" de intento en intento, eliminando de su conjunto de salidas u opciones, aquellas que mostraron ser falsas en intentos anteriores. De acuerdo con esto, el número máximo de intentos es 5 y en cada intento la probabilidad (condicional) de escapar va creciendo. Así $R_X = \{1,2,3,4,5\}$ como en (a), definamos $A_i =$ "La rata escapa en el intento # i", i=1,2,3, 4,5. Procediendo inductivamente $P(X = 1) = P(A_1) = 1/5$, pues a la primera todas las salidas son válidas como opciones y sólo una de ellas es la "verdadera".

$P(X = 2) = P(A_1^C \cap A_2) = P(A_2 \mid A_1^C)P(A_1^C) = (1/4)(4/5) = 1/5$, pues $(X = 2)$ equivale a decir que la rata falla en el primer intento y acierta en el segundo; al aplicar la Regla de Producto, la probabilidad de acertar en el segundo intento es $1/4$ pues la salida usada en el primero ya no es retomada (debido al supuesto de "memoria perfecta" en la rata) y la probabilidad de que falle en el primer intento es $1/4$.

Análogamente, para el tercer intento:

$$P(X = 3) = P(A_1^C \cap A_2^C \cap A_3) = \overbrace{P(A_3 \mid A_1^C \cap A_2^C)P(A_1^C \cap A_2^C)}^{\text{R.delProducto}} = \overbrace{P(A_3 \mid A_1^C \cap A_2^C)P(A_2^C \mid A_1^C)P(A_1^C)}^{R.deProducto}$$

$= (1/3)(3/4)(4/5) = 1/5$ pues $P(A_3 \mid A_1^C \cap A_2^C) = 1/3$, ya que para el tercer intento, sólo quedan 3 salidas alternativas, una de las cuales es la correcta; y también $P(A_2^C \mid A_1^C) = (3/4)$ ya que si la rata falló en el primer intento, para el segundo "descuenta" una de las salidas falsas y quedan 4 en total, de las cuales una es verdadera y las otras tres son falsas.

Extendiendo el razonamiento a los otros dos valores posibles de X, $P(X = 4)$
$= P(A_1^C \cap A_2^C \cap A_3^C \cap A_4) = (1/2)(2/3)(3/4)(4/5) = 1/5$ y $P(X = 5) = P(A_1^C \cap A_2^C \cap A_3^C \cap A_4^C \cap A_5)$
$= (1)(1/2)(2/3)(3/4)(4/5) = 1/5$

Así pues, tenemos $P_X(x) = (1/5) \qquad x = 1,2,3,4,5.$

Obsérvese que para un mismo experimento hemos formulado dos "modelos" distintos que intentan explicar los posibles resultados (o mejor dicho, las frecuencias de los resultados) obtenibles, pero bajo teoría explicativas distintas: sin memoria y con memoria. Sólo repeticiones reales del experimento dirán cuál modelo se ajusta mejor a la realidad, e inclusive, esta evidencia puede recusar a los dos modelos.

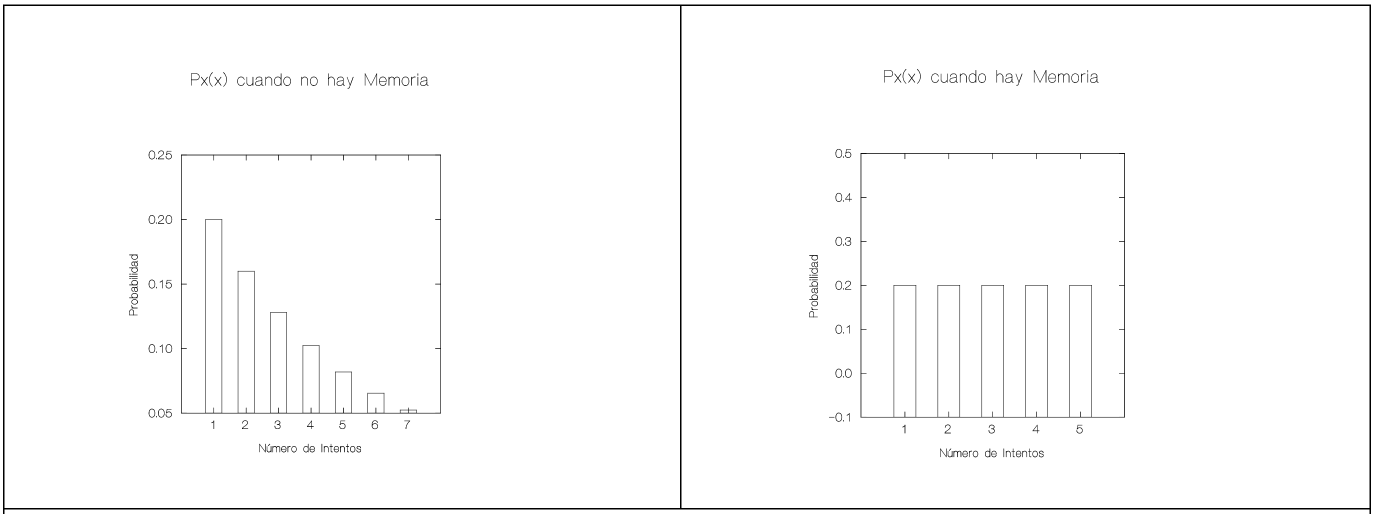


Fig. 1.7

Nota:

(1) En general, $P_X(x)$ también es llamada "función de distribución de probabilidades" porque, en efecto, distribuye la probabilidad total de S (que es 1) sobre el espacio codificado R_X, asignando a cada elemento de R_X una probabilidad no negativa, que mide la "propensión a ocurrir" de dicho valor o elemento.

(2) Como X es un sistema de codificación que captura cierta información de interés, entonces podemos considerar que $P_X(x)$ "modela" el comportamiento del azar, en lo que concierne a nuestro objeto de análisis, cuando éste es de naturaleza numérica discreta.

Propiedades de $P_X(x)$.

Una función de probabilidad $P_X(x)$ tiene las siguientes propiedades:

(a) $0 \le P_X(x) \le 1 \quad \forall x$

(b) $\sum_{x \in R_X} P_X(x) = 1$ donde $x \in R_X$ indica que la suma se hace sobre todos los x que pertenecen a R_X

(c) $P(X \in A) = \sum_{x \in A} P_X(x) \quad \forall A \subseteq R_X$

Todo lo anterior se debe a que $P_X(x)$, evaluada en un punto x, proporciona una probabilidad, de modo que necesariamente debe estar entre 0 y 1. Si sumamos todas las probabilidades de los elementos de R_X, en realidad estamos hallando la probabilidad total del espacio muestral, y ésta vale 1.

Ejemplo 34

Sea X v.a. discreta (v.a.d) tal que $R_X = \{1,2,3,...,N\}$ y $P(X = x) = \alpha x$, con $\alpha > 0$ una constante. Hallar α y P_X

Solución:

Aplicando la propiedad (b):

$$\sum_{x=1}^{n} P_x(x) = 1 \Leftrightarrow \sum_{x=1}^{n} \alpha x = 1 \Leftrightarrow \alpha N(N+1)/2 = 1 \Leftrightarrow \alpha = 2/N(N+1)$$ y por tanto:

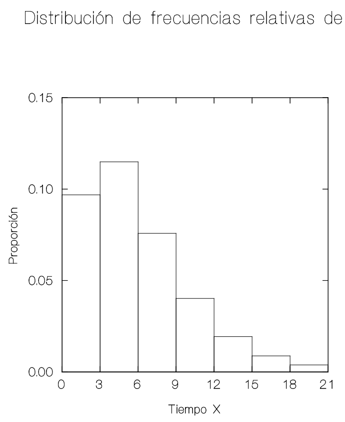
$$P_X(x) = \begin{cases} 2x/\mathrm{N(N+1)} & \text{si} \quad \mathrm{x = 1,2,...,N} \\ 0 & \text{en otro caso} \end{cases}$$

2.3 Variable Continua y Función de Densidad

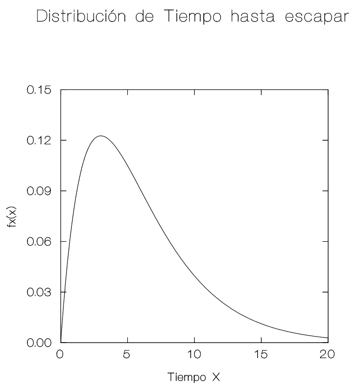
Cuando la variable es discreta, podemos distribuir probabilidades sobre sus valores asignando a cada uno, una probabilidad, por pequeña que ella sea. Pero cuando el rango de posibles valores es un intervalo, esto es, cuando X es continua, hay que adaptar el sistema de distribución de probabilidades sobre R_X. Al respecto, un matemático ruso explicaba el punto así: "Si la abuelita tiene 4 papas para distribuir entre 4 nietos, la abuelita da a cada nieto una papa; si son 8 los nietos, parte cada papa en dos y da a cada nieto media papa; pero si son 11 nietos, entonces … la abuelita hace puré y reparte la masa entre todos".

Notemos que si R_X es un intervalo, físicamente es imposible registrar la ocurrencia exacta de un valor específico de X; por ejemplo, el tiempo X que demora una rata sin memoria en escapar del laberinto del ejemplo (2) puede ser 18 minutos o 18 minutos y 01 segundo o 17 minutos y 58 segundos, etc. de modo que nunca podremos estar seguros de haber registrado el valor real. Lo más sano es considerar que el valor registrado representa a un entorno de valores cercanos a él. Es decir, con variables continuas, más que estar interesados en la ocurrencia de valores, debemos pensar en la *ocurrencia de intervalos de valores*. Por tanto, *necesitamos una función que distribuya probabilidades sobre los intervalos contenidos en R_X, no sobre los valores de X. Esta función debe distribuir la probabilidad total que es 1, de modo mas bien continuo*, haciendo "más densos en probabilidad" a algunos intervalos, y "menos densos" a otros. El modo natural de hacer esta distribución de densidades de probabilidad es mediante la gráfica (continua) de la función.

En realidad, este proceso ya se conoce desde la estadística descriptiva, cuando los datos se agrupan en intervalos y se registra la frecuencia relativa de cada uno, y cuando luego se dibuja el histograma o el polígono de frecuencias. Recordemos que, tomando convenientemente el ancho de cada intervalo, el área total del histograma vale 1 y que lo mismo pasa con respecto al área debajo del polígono. Ahora bien, si se toma un número muy grande de intervalos, el polígono se "suaviza" y deviene en una curva continua. Esta curva, llamémosla $f_X(x)$, es candidato natural para distribuir probabilidades mediante las áreas que corresponden a cada intervalo contenido en R_X



Distribución Empírica de X



Distribución Teórica de X

Figura 1.8

Ejemplo 35

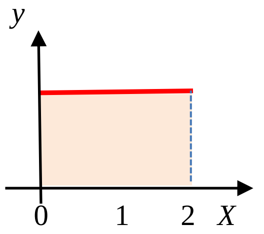
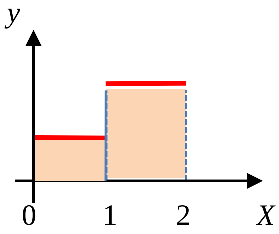
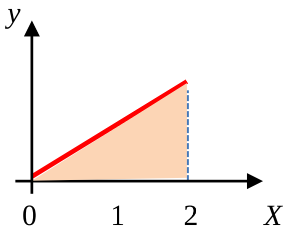
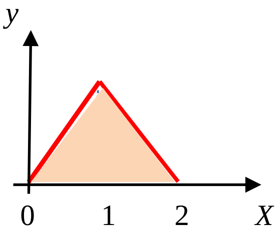
Suponga que para la rentabilidad anual X de un fondo mutuo sabemos que X puede estar entre 0 y 2% de modo que el rango de X es R_X=[0,2]. Para la distribución de probabilidades de X tenemos cuatro alternativas plausibles según la experiencia de cuatro expertos, que nos inducen a postular cuatro "modelos" para X:

Modelo 1: Si pensamos que puede ocurrir cualquier valor con la misma verosimilitud, entonces podemos usar probabilidad geométrica y distribuir la probabilidad sobre R_X de manera que cualquier intervalo de la misma longitud tenga similar área (o sea probabilidad)

Modelo 2: Si pensamos que valores de X entre 1 y 2 son el doble de probables que valores entre 0 y 1, entonces podemos distribuir la probabilidad sobre R_X de manera que el intervalo [0,1[reciba la mitad del área que el intervalo [1,2]

Modelo 3: Si más bien pensamos que la probabilidad de una rentabilidad es proporcional a ésta, de modo que valores de X cercanos a 2 tienen más probabilidad, podemos asignar áreas de modo que ésta "crezca" proporcionalmente a X

Modelo 4: La probabilidad crece hasta que se llega a 1 y luego decrece a la misma tasa hasta llegar a 2

Modelo 1	**Modelo 2**	**Modelo 3**	**Modelo 4**
			
$y=f_X(x)=c$ $0\le x\le 2$	$y=f_X(x)=\begin{cases}a & 0\le x<1\\ b & 1\le x\le 2\end{cases}$	$y=f_X(x)$ $=cx \quad 0\le x\le 2$	$y=f_X(x)=\begin{cases}a+bx & 0\le x<1\\ c+dx & 1\le x\le 2\end{cases}$
$y=f_X(x)=\dfrac{1}{2}$ $0\le x\le 2$	$y=f_X(x)$ $=\begin{cases}\dfrac{1}{3} & 0\le x<1\\ \dfrac{2}{3} & 1\le x\le 2\end{cases}$	$y=f_X(x)$ $=\dfrac{1}{2}x \quad 0\le x\le 2$	$y=f_X(x)=\begin{cases}0+x & 0\le x<1\\ 2-x & 1\le x\le 2\end{cases}$

Siendo el rango el mismo R_X=[0,2], es claro que la asignación de áreas la determina la "parte superior del área total" que es una línea o conjunto de líneas y que puede ser descrita muy bien mediante una función $y=f_X(x)$, que se llama función de densidad de X

Definición

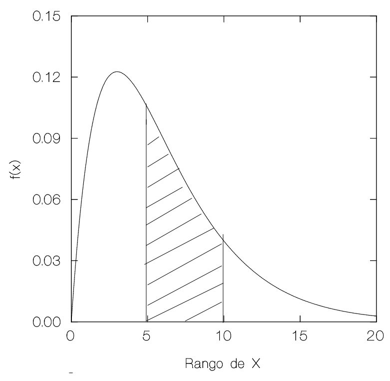
Si X es una v.a. continua, una función de densidad de X, denotada $f_X(x)$, es una función no negativa y continua, tal que para todo intervalo]a,b] $\subseteq$ R_X se cumple:

$$P(a<X\le b)=\int_a^b f_X(x)dx$$

Observaciones:

(1) $f_X(x)$ no proporciona ninguna probabilidad directamente; sólo al ser integrada sobre un intervalo determina la probabilidad del mismo. Por eso no es problema que $f_X(x)$ sea mayor que 1. También llamaremos a $f_X(x)$ función de distribución de probabilidades de X.

(2) Como sabemos, $P(a<X\le b)$=Área debajo de la gráfica de $f_X(x)$ entre a y b:



P(5 < X ≤ 10) = Área debajo de $f_X(x)$ entre 5 y 10

Figura 1.9

(3) Como $(X = a) = (a < X \leq a)$ y el área debajo de $f_X(x)$ sobre este intervalo es *cero*, se concluye que con variables continuas, los puntos tomados aisladamente tienen probabilidad *cero*. Este valor no quiere decir que no pueden ocurrir, sino que no tenemos manera de verificar exactamente la ocurrencia de un valor específico de X

(4) Se desprende que $P[a < X \leq b] = P[a \leq X \leq b] = P[a \leq X < b] = P[a < X < b]$ pues en todos los casos, se añade o se quita un punto de probabilidad cero, lo que no altera la probabilidad del intervalo.

Propiedades.
Una función de densidad $f_X(x)$ tiene las siguientes propiedades:

(a) $0 \leq f_X(x) \quad \forall x$

(b) $\int_{-\infty}^{+\infty} f_X(x)dx = \int_{R_X} f_X(x)dx = 1$

(c) $P(X \in A) = \int_A f_X(x)dx$

Ejemplo 36
En el ejemplo de la licitación, halle una función de densidad para la v.a. X definida como X = Precio de venta.

Solución:
Ya vimos que $P[X \leq t] = t^2$ para un valor *t* en R_X. Luego si $f_X(x)$ es una función de densidad para X, se debe cumplir:

$$P(0 < X \leq t) = \int_0^t f_X(x)dx = t^2$$

Y si derivamos con respecto a t, obtenemos: $f_X(t) = 2t$, luego podemos definir como una función de densidad de X a $f_X(x) = 2x$ para $0 \leq x \leq 1$. Trabajando con más generalidad:

$$f_X(x) = \begin{cases} 2x & si \quad 0 \leq x \leq 1 \\ 0 & en \quad otro \quad caso \end{cases}$$

Ejemplo 37
Asuma que el tiempo que el tiempo X (en años) que demora una empresa nueva en consolidarse en el mercado, es una v.a.c. con función de densidad

$$f_X(x) = \begin{cases} ax & si \quad 0 \le x \le 0.5 \\ b & si \quad 0.5 < x \le 1 \end{cases} \qquad \text{(a y b constantes positivas)}$$

Halle *a* y *b*, si se sabe que con 60% de probabilidad, la empresa estará consolidada antes de medio año.

Solución:
La gráfica, de forma genérica, es:

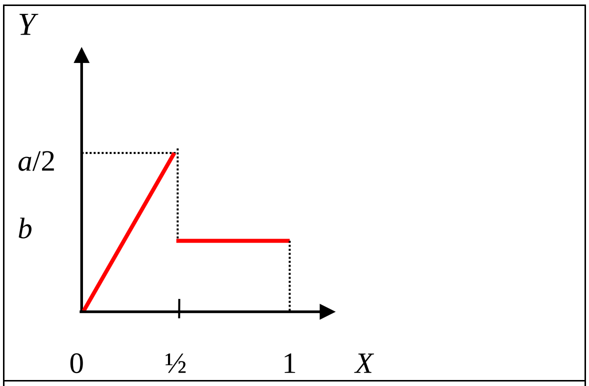


Figura 2

Como $P[0 \le X \le 0.5] = 0.6$ entonces $0.25a/2 = 0.6$ y despejando tenemos $a = 4.8$. Por complemento $P[0.5 \le X \le 1] = 0.4$ y por tanto $(1-0.5)b = 0.4$ así que $b = 0.4/0.5 = 0.8$ y finalmente

$$f_X(x) = \begin{cases} 4.8x & si \quad 0 \le x \le 0.5 \\ 0.8 & si \quad 0.5 < x \le 1 \end{cases}$$

2.4 Función de Distribución Acumulativa

Si X es una v.a., se define la Función de Distribución Acumulativa de X, denotada F_X, mediante la regla de correspondencia:

$$F_X(x) = P(X \le x) \quad \forall x \text{ real}$$

Observación
También se escribe $F_X(t) = P(X \le t)$ al definir F_X cuando se quiere resaltar que la función se evalúa o está definida, no sólo sobre el rango de X sino en general sobre todo el eje real.

F_X tiene como principal función simplificar el cálculo de probabilidades, en particular con variables continuas, las principales distribuciones vienen en tablas donde figuran las diversas probabilidades acumuladas $F_X(x)$ en sucesivos valores de x

Ejemplo 38
Para X discreta con función de probabilidad $P_X(x) = 1/5$ x = 1,2,3,4,5, es fácil verificar que $F_X(x) = P(X \le x) = x/5$ cuando x = 1,2,3,4,5**.**

En cambio, para X del ejemplo (1) de variables continuas, tenemos $F_X(x) = P(X \leq x) = x^2$ cuando $0 < x \leq 1$

Propiedades de $F_X(x)$

(1) $0 \leq F_X(t) \leq 1$ para todo t real.

(2) $a < b \Rightarrow F_X(a) \leq F_X(b)$

(3) F_X es continua a la derecha ('diestro continua'), ie. $F_X(t) = \lim\limits_{h \to 0^+} F_x(t+h)$ para todo t real

(4) $\lim\limits_{t \to -\infty} F_x(t) = 0$ y $\lim\limits_{t \to +\infty} F_x(t) = 1$

(5) $P[a < X \leq b] = F_X(b) - F_X(a)$

(6) $P[X = b] = F_X(b) - \lim\limits_{h \to 0^+} F_x(b+h)$

Demostración:

(1) Es consecuencia directa de la definición

(2) Se deduce de $(X \leq a) \subseteq (X \leq b)$ y la propiedad de monotonía de la probabilidad

(3) La demostración no la haremos pues requiere usar la continuidad de la probabilidad, tópico que no hemos tratado extensamente en el curso

(4) Obviamente $(X \leq -\infty) = \varnothing$ y $(X \leq +\infty) = S$ de donde se deduce la propiedad

(5) Usemos conjuntos y propiedades de la probabilidad:
$(X \leq b) = (X \leq a) \cup (a < X \leq b) \Rightarrow P(X \leq b) = P(X \leq a) + P(a < X \leq b) \Rightarrow F_X(b) =$
$F_X(a) + P(a < X \leq b) \Rightarrow P[a < X \leq b] = F_X(b) - F_X(a)$

(6) No haremos la demostración que requiere usar la continuidad de la probabilidad, tópico que no hemos tratado extensamente en el curso

Propiedades adicionales:

(1) Si X es v.a. discreta con función de probabilidad $P_X(x)$. Entonces $F_X(b) = \sum\limits_{x_j | x_j \leq b} P_X(x)$

(2) Si X es v.a. continua con función de densidad $f_X(x)$. Entonces $\int_{-\infty}^{t} f_X(x)dx = F_X(t)$ para todo t real

Demostración:

(1) Si X es discreta $\Rightarrow [X \leq b] = \{x_j \mid x_j \leq b\} = \bigcup\limits_{x_j | x_j \leq b} \{x_j\}$. Por tanto:

$$F_X(b) = P(X \leq b) = P(\bigcup_{x_j | x_j \leq b} \{X_j\}) = \sum_{x_j | x_j \leq b} P(X = x_j) = \sum_{x_j | x_j \leq b} P_X(x_j)$$

(2) Si X es continua $\Rightarrow F_X(t) = P(X \leq t) = P(-\infty < X \leq t) = \int_{-\infty}^{t} f_X(x)dx$.

Observación:

- Como consecuencia de (1) tenemos que
  Si X es discreta con $R_X = \{x_1, x_2, ..., x_{N-1}, x_N, ...\}$ donde $x_1 < x_2 < ... < x_{N-1} < x_N, ...$ entonces
  $F_X(x_N) = \sum\limits_{j=1}^{N} P_X(x_j)$ y también $P_X(x_N) = F_X(x_N) - F_X(x_{N-1})$
- Análogamente, de (2) se deduce que si X es continua, entonces $f_X(x) = F_X^{'}(x)$

Ejemplo 39

Se toman *k* números al azar y con reposición del conjunto {1,2,…,*n*}, 1< *k* < *n*. Halle la función de probabilidad de $X =$ Máximo de los k números

Solución

Es claro que $R_X = \{1,2,...,n\}$. Sea $x \in R_X$ un valor dado, entonces por combinatoria

- Hay n posibilidades en cada una de las k extracciones así que el número total de posibilidades es $\underbrace{n \times n \times ... \times n}_{k \quad veces} = n^k$
- $X =$Máximo de los k números menor o igual que x implica que todos los k números extraídos son menores o iguales que x, lo que puede ocurrir de $\underbrace{x \times x \times ... \times x}_{k \quad veces} = x^k$ maneras.

Tenemos $F_X(x) = P(X \leq x) = \frac{x^k}{n^k} = \left(\frac{x}{n}\right)^k$ y así $P_X(x) = F_X(x) - F_X(x-1) = \left(\frac{x}{n}\right)^k - \left(\frac{x-1}{n}\right)^k$

Ejemplo 40

Una v.a.c. X positiva satisface $P(t < X \leq t + dt \mid X > t) = \alpha dt$ para dt suficientemente pequeño. Hallar la f. de densidad de X

Solución

$P(t < X \leq t + dt \mid X > t) = \alpha dt \Leftrightarrow \frac{F_X(t+dt) - F_X(t)}{1 - F_X(t)} = \alpha dt \Leftrightarrow \frac{F_X(t+dt) - F_X(t)}{dt} = \alpha(1 - F_X(t))$ y con dt en el límite a cero tenemos:

$$\frac{dF_X(t)}{dt} = \alpha(1 - F_X(t)) = \alpha - \alpha F_X(t).$$

Para simplificar escribamos $y = F_X(t)$ y entonces tenemos la ecuación diferencial $y' + \alpha y = \alpha$ cuya solución es $y(t) = 1 - e^{-\alpha t}$ y así llegamos a $F_X(t) = 1 - e^{-\alpha t}$ y por tanto la función de densidad de X es $f_X(x) = e^{-\alpha x} \quad x > 0$

Observación:

(a) Si una v.a. X es discreta, la correspondiente función de distribución acumulativa $F_X(x)$ es "función escalera" con una cantidad finita o numerable de saltos; en cambio si X es continua, $F_X(x)$ es absolutamente continua.

(b) En tratamientos más avanzados del concepto de variable aleatoria, es preferible definir primero la función de distribución acumulativa $F_X(x)$ y luego, según las propiedades de esta función, definir a la correspondiente variable aleatoria como "discreta", "continua" o "mixta".

2.5 Valor Esperado o Esperanza Matemática

Si bien en la función de Probabilidad o de Densidad, está toda la información acerca del comportamiento de una variable aleatoria, a veces es menester representar toda la distribución mediante unos pocos indicadores, representar todo el rango de valores posibles y aleatorios de X, mediante una constante que los represente. Esto se hace recurriendo al concepto de Valor Esperado.

2.5.1 Definición

Sea X variable aleatoria y $H(X)$ una función de X, se define el Valor Esperado de $H(X)$, denotado $E[H(X)]$, mediante:

$$E[H(X)] = \begin{cases} \sum_{x \in Rx} H(x) P_X(x) & si \quad X \quad es \quad Discreta \\ \int_{-\infty}^{+\infty} H(x) f_X(x) dx & si \quad X \quad es \quad Continua \end{cases}$$

Observaciones:

(1) Aunque no lo hemos mencionado en la definición, se requiere que haya **convergencia absoluta, i.e.** $\sum_{x \in Rx} |H(x)| P_X(x) < \infty$ o $\int_{-\infty}^{+\infty} |H(x)| f_X(x) dx < \infty$. Esto es así para que no haya ambigüedad en el número obtenido, pues de no haber convergencia absoluta, el valor de la serie puede depender del orden en la suma, o el valor de la integral puede depender del orden al tomar límites al infinito. Como consecuencia, hay casos en donde $E[H(X)]$ no existe.

(2) Como X es aleatoria, $H(X)$ también lo es, pudiendo tomar distintos valores, según lo que el azar determine para X. En este sentido, es necesario calcular un "valor resumen", que **represente a** $H(X)$ y permita tomar decisiones. El Valor esperado (o la Esperanza matemática) $E[H(X)]$ es ese "representante" o "valor resumen". Nótese que se trata de un número real que ya no depende de X y que va sus mismas unidades.

(3) Operacionalmente**,** $E[H(X)]$ **es un promedio ponderado de los valores de** $H(X)$, donde el factor de ponderación (el "peso") está asociado a la probabilidad de X vía $P_X(x)$ o $f_X(x)$ según el caso.

Ejemplo 41

Se lanza un dado según la apuesta: Si sale el 1 se gana U$S 2, si sale el 6 se gana US$ 10, en otro caso se pierde US$ 6. Sea X el número que muestra el dado y sea $H(X)$ la utilidad, para el jugador, en esta apuesta. Halle $E[H(X)]$. ¿Le conviene este juego al apostador?

Solución:

La función de probabilidad de X es:

x	1	2	3	4	5	6
$P_X(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Y $H(X)$ responde a $H(X) = \begin{cases} 2 & si \quad x = 1 \\ -6 & si \quad x = 2,3,4,5 \\ 10 & si \quad x = 6 \end{cases}$

Escribiendo en una tabla, para ordenar datos antes del cálculo:

x	1	2	3	4	5	6
$P_X(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$H(x)$	2	-6	-6	-6	-6	10

Entonces, aplicando la definición del valor esperado:

$$E[H(X)] = \sum_{x\in Rx} H(x)P_X(x) = (2)\times\frac{1}{6} + (-6)\times\frac{1}{6} + (-6)\times\frac{1}{6}(-6)\times\frac{1}{6}(-6)\times\frac{1}{6} + (10)\times\frac{1}{6} = \frac{-12}{6} = -2$$

Interpretando $E[H(X)]$:

- Si usamos la noción frecuencial de probabilidad, esperaríamos que de seis lanzamientos, en uno de ellos ocurriera el 1 (por tanto se gana US$ 2), en otro se presentaría el 6 (ganándose entonces US$ 10) y en el resto de casos se perdería (a razón de de US$ 6 por cada vez). O sea que al cabo de 6 lanzamientos, esperamos retirarnos de la mesa de juego con 12-24 =-12 US$. *Es decir, en esta apuesta, la tendencia es a perder a razón de US$12 por cada 6 jugadas.*
- Si lanzáramos el dado 12 veces, la pérdida sería de US$ 24; Si jugamos 18 veces, perderíamos US$ 36; en 36 jugadas, perderíamos US$ 72, etc. En general, si hacemos N lanzamientos, perderemos (N/6)x12 = N(12/6) = Nx2 dólares en total.
- Es decir, podemos calcular un índice que indica la pérdida esperada por cada lanzamiento y permite prever la pérdida en una cantidad general de lanzamientos. Este índice es precisamente $E[H(X)] = -2$ *El signo negativo muestra que la tendencia es a la pérdida, y el valor 2 indica el monto de ésta "por jugada", para poder calcular la pérdida global en general.*
- $E[H(X)] = -2$ *no es un valor "real" , sino sólo un índice que representa la tendencia de los valores de* $H(X)$*, un índice útil para cálculos posteriores.*

Ejemplo 42

En el ejemplo de X = Precio pagado en una privatización, calculemos el esperado de la misma v.a. (i.e., trabajamos con $H(X) = X$)

Recordemos que la f. de densidad de X es:

$$f_X(x) = \begin{cases} 2x & si \quad 0 \le x \le 1 \\ 0 & en \quad otro \quad caso \end{cases}$$

Luego, y siendo rigurosos en la presentación formal:

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx = \int_{-\infty}^{0} xf_X(x)dx + \int_{0}^{1} xf_X(x)dx + \int_{1}^{+\infty} xf_X(x)dx = \int_{0}^{1} x2xdx = 2x^3]_0^1 = 2/3 = 0.67 \text{ millones}$$

Podemos decir que, en promedio, esperamos que la empresa se venda en 670 mil unidades monetarias.

2.5.2 Casos especiales de Valor Esperado

La Media Poblacional

Se denota μ o μ_X y se define como el valor esperado de la misma v.a. X, o sea $\mu_X = E(X)$. Se dice que μ_X representa a X, mejor dicho, representa a la mayoría de los valores de X. En ese sentido es el "valor típico de X".

La Varianza Poblacional

Se denota σ^2 o σ_X^2 o $V(X)$ y se define como el valor esperado de la diferencia al cuadrado entre X y su representante μ_X. Es decir, $\sigma_X^2 = V(X) := E[(X-\mu_X)^2]$

σ_X^2 es la distancia al cuadrado y promedio entre un valor cualquiera de X y el representante de X, μ_X. Mide la variabilidad presente en los valores de X.

La Desviación Estándar

Se denota σ o σ_X y se define como la raíz cuadrada de la varianza. Esto es, $\sigma_X = \sqrt{\sigma_X^2}$.

Proposición (Desigualdad de Tchebychev)

Si X es v.a. con media μ_X y desviación estándar σ_X. Sea k una constante positiva dada, entonces:

$$P[|X-\mu_X| < k\sigma_X] \geq 1 - \frac{1}{k^2}$$

Observaciones:

- Esta desigualdad es general y muy informativa. Por ejemplo, si tomamos $k=3$, entonces $P[|X-\mu_X| < 3\sigma_X] \geq 1 - \frac{1}{3^2} = \frac{8}{9} = 0.89$ o equivalentemente el intervalo $[\mu_X - 3\sigma_X, \mu_X + 3\sigma_X]$ contiene por lo menos al 89% de los valores de X.

- Aunque no es propiamente lo mismo***, se interpreta a σ_X como la "distancia promedio" entre X y su representante μ_X. Interpretada así, como una distancia promedio entre X y μ_X, σ_X mide el "margen de error" de μ_X como representante de X.***

2.5.3 Propiedades del Valor Esperado

El trabajo con el valor esperado, para ser simple, necesita del estudio de algunas propiedades. Veremos algunas, desarrollando la demostración para el caso discreto. En el caso continuo basta cambiar la sumatoria por la integral, de modo que las propiedades podemos considerarlas generales. Las más importantes son:

Propiedad 1

Sea X v.a. discreta, con función de probabilidad $P_X(x)$ y sea c una constante. Entonces se cumple que $E(c) = c$.

En efecto, esto se verifica aplicando la definición de valor esperado:

$E(c) = \sum_{x\in R_X} cP_X(x) = c\sum_{x\in R_X} P_X(x) = c$ pues $\sum_{x\in R_X} P_X(x) = 1$

Propiedad 2

Sea X v.a. discreta, con función de probabilidad $P_X(x)$. Sea a constante o variable no aleatoria y sea $G(X)$ una función de X.

Entonces se cumple que $E[aG(X)]= aE[G(X)]$.

La verificación es directa, aplicando la definición y las propiedades de la sumatoria. Sobreentendiendo que la suma se hace sobre $x \in R_X$ y recordando que a no depende de x tenemos:

$$E[aG(X)] = \sum_x aG(x)P_X(x) = a\sum_x G(x)P_X(x) = aE[G(X)]$$

Propiedad 3
Sea X v.a. discreta, con función de probabilidad $P_X(x)$. Sean a y b constantes dadas y sean $H_1(X)$ y $H_2(X)$ funciones de X.
Entonces se cumple que $E[aH_1(X)+bH_2(X)]= aE[H_1(X)]+bE[H_2(X)]$.

Otra vez, la demostración se apoya en la definición y las propiedades de la sumatoria. Sobreentendiendo que la suma se hace sobre R_X, tenemos:

$$E[aH_1(X)+bH_2(X)] = \sum[aH_1(X)+bH_2(X)]P_X(x) = \sum[aH_1(X)P_X(x)+bH_2(X)P_X(x)] =$$

$$\sum aH_1(X)P_X(x) + \sum bH_2(X)P_X(x) = a\sum H_1(X)P_X(x) + b\sum H_2(X)P_X(x) = aE[H_1(X)] + bE[H_2(X)]$$

Corolarios
Si X es v.a. discreta con función de probabilidad $P_X(x)$, se cumple que

(1) $V(X) = E(X^2) - \mu_X^2$.
(2) Si $Y = a + bX \quad \forall X$, entonces $E(Y) = a + bE(X)$ y $V(Y) = b^2V(X)$.

La demostración de estos resultados es consecuencia directa de las propiedades 1 y 2 anteriormente desarrolladas.

Note que en (2) la linealidad debe cumplirse exactamente, en caso contrario la propiedad no se cumple. Por ejemplo si tenemos $Y = \begin{cases} 2 & x \le 0 \\ 3x & x > 0 \end{cases}$ entonces $E(Y) \ne 2 + 3E(X)$

Observación:
Si consideramos el símbolo E[•] como un 'operador' , resulta que tiene las propiedades básicas de un 'operador lineal', compartidas con otros operadores análogos y ya conocidos, como la derivada por ejemplo. **Pero debe recordarse que en general** $E[H(X)] \ne H(E(X))$

Ejemplo 43.1
En el caso de X = Precio de venta en una privatización:

a) Halle el rango de valores más probables para X : $\mu_X \pm \sigma_X$.
b) Si la privatización implica el pago de 5 mil unidades monetarias a una empresa tasadora y el pago de un 4% de la venta a una empresa encargada de la privatización ¿Cuál es el Ingreso esperado para el Estado?
c) Verifique que $E[X^3] \ne (E[X])^3$

Solución:

a) Sólo faltaría hallar σ_X^2. Apliquemos la propiedad $V(X) = E(X^2) - \mu_X^2$; Como $E(X^2) = \int_0^1 x^2 2xdx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1}{2} \Rightarrow \sigma_X^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18} \Rightarrow \sigma_X = \sqrt{1/18} = 0.24$. Luego, el intervalo de 'valores más probables' para X es 0.67 ± 0.24 o sea $[0.43, 0.91]$, i.e. **esperamos un precio de venta entre 430 mil y 910 mil unidades monetarias.**

b) Sea $Y = H(X)$=Ingreso del Estado $\Rightarrow Y = X - (0.05 + 0.04X) = 0.96X - 0.05 \Rightarrow$ $E(Y) = E(0.96X - 0.05) = 0.96E(X) - 0.05 = 0.5932$, es decir e**l estado espera recibir 593,200 unidades monetarias por la empresa**

c) $E[X^3] = \int_0^1 x^3 2xdx = \int_0^1 2x^4 dx = 2\left[\frac{x^5}{5}\right]_0^1 = \frac{2}{5}$ y $(E[X])^3 = \left(\frac{2}{3}\right)^3 = \frac{8}{27}$, esto es $E[X^3] \neq (E[X])^3$

Ejemplo 43.2

Si en el ejemplo de la privatización, se paga un impuesto "por tramos": Si el precio es inferior a 0.5 millones se paga 1% del ingreso; si es de 0.5 o más, se paga 4%.
¿Cuál sería la recaudación por impuestos esperada para privatizaciones de este tipo?

Solución:

La recaudación por impuesto no es función lineal sino lineal por tramos, es decir:

$$Y = H(X) = \text{Recaudación} = \begin{cases} 0.01X & si\ 0 < X < 0.5 \\ 0.04X & si\ 0.5 \leq X \leq 1 \end{cases}$$

No podemos usar la propiedad de linealidad $E(Y) = a + bE(X)$, sino calcular el valor esperado aplicando la definición misma:

$$\text{Recaudación esperada} = E[H(X)] = \int_0^1 H(x) f_X(x)dx = \int_0^{0.5} H(x) f_X(x)dx + \int_{0.5}^1 H(x) f_X(x)dx =$$

$$\int_0^{0.5} 0.01x2xdx + \int_{0.5}^1 0.04x2xdx = 0.02\left[\frac{x^3}{3}\right]_0^{0.5} + 0.08\left[\frac{x^3}{3}\right]_{0.5}^1 = 0.0242$$

El valor esperado $E[\bullet]$ es una herramienta que se puede usar no sólo para medir o identificar la "tendencia principal" de una función $H(X)$ de una v.a. X, sino también para "dirigir" $H(X)$, que dependiendo del azar vía X, en principio toma valores fuera de nuestro control. El método es introducir en el problema alguna variable no aleatoria (variable matemática) S en $H(X)$ y pasar a algo como $H(X,S)$, luego tomar el valor esperado con respecto a X: $E(H(X,S))$ que resultará en una función de S, digamos $\varphi(S) := E(H(X,S))$, que pone la "tendencia principal" de $H(X,S)$ como dependiente de S, que es una variable que sí podemos controlar para "optimizar" (en promedio) a $H(X,S)$.

Ejemplo 44

El distribuidor de un solvente industrial tiene la política de comprar al inicio de la temporada de ventas una existencia ('stock') de S unidades de volumen a 4 unidades monetarias. Durante la temporada vende el producto a 7 unidades monetarias por unidad de volumen; al final de temporada remata el sobrante a 3 unidades monetarias por unidad de volumen.
Si la demanda (cantidad demandada) de solvente al distribuidor es una v.a.c. X con función de densidad $f_X(x) = \frac{1}{100}\ \ 0 < x < 100$

a) Escriba la función de utilidad $U = U(X,S)$ del distribuidor.
b) Determine el valor óptimo de S.

Solución:

a) Si U es la utilidad del distribuidor, entonces U depende de S y de X, pues:

- Si $X \leq S$, durante la temporada vende X unidades y al final de la temporada remata las $(S - X)$ unidades sobrantes, de modo que $U = 7X + 3(S - X) - 4S = 4X - S$
- Si $X > S$, durante la temporada vende todo su stock S y nada más. En este caso $U = 3S$, aunque quede una demanda insatisfecha por $(X - S)$ unidades de volumen.

  En resumen $U = U(X, S) = \begin{cases} 4X - S & si\ X \leq S \\ 3S & si\ X > S \end{cases}$

b) U tiene una componente aleatoria X y otra no aleatoria S y "lo óptimo" sería maximizar la utilidad U; pero cómo ésta es aleatoria, maximicemos el valor esperado de la utilidad $E(U)$:
$E(U) = E[U(X, S)] =: \varphi(S)$ que es una función de S. Calculando S de modo que se maximice la utilidad esperada $\varphi(S)$, determinamos el sotck "óptimo" S que induce una 'tendencia óptima' para la utilidad U. Veamos: $\varphi(S) \coloneqq E[U(X, S)] = \int_{-\infty}^{+\infty} U(x, S) f_X(x)dx = \int_0^{100} U(x, S) f_X(x)dx =$
$\int_0^S U(x, S) f_X(x)dx + \int_S^{100} U(x, S) f_X(x)dx = \int_0^S (4x - S) f_X(x)dx + \int_S^{100} 3S f_X(x)dx =$
$\int_0^S 4x f_X(x)dx - \int_0^S S f_X(x)dx + 3S \int_S^{100} f_X(x)dx = 4\int_0^S x f_X(x)dx - SF_X(S) + 3S\big(1 - F_X(S)\big) =$
$4\int_0^S x f_X(x)dx - 4SF_X(S) + 3S$.
$\varphi(S) \coloneqq E[U(X, S)] = 4\int_0^S x f_X(x)dx - 4SF_X(S) + 3S$ es explícitamente una función diferenciable de S que podemos maximizar mediante derivación:

$\varphi'(S) = \frac{d\varphi(S)}{ds} = 4Sf_X(S) - 4F_X(S) - 4SF_X'(S) + 3 = 4Sf_X(S) - 4F_X(S) - 4Sf_X(S) + 3 =$
$-4F_X(S) + 3$ y $\varphi'(S) = \frac{d\varphi(S)}{ds} = 0$ equivale a $F_X(S) = \frac{3}{4}$ y como $\varphi''(S) = -4f_X(S) > 0$, se trata de un máximo.
Finalmente, como $f_X(x) = \frac{1}{100}$ $0 < x < 100 \Rightarrow F_X(S) = \int_0^S \frac{1}{100} dx = \frac{S}{100} = \frac{3}{4} \Rightarrow S = 75$ es el valor "óptimo" del stock S (u "stock óptimo")

Ejemplo 45
Una empresa contratada para la purificación de un lote de mineral, tiene un proceso automatizado en el cual el operario debe fijar el número *k* de horas que el proceso debe de trabajar y una vez empezado ya no se detiene hasta cumplir ese plazo. Si el mineral no alcanza el nivel de pureza suficiente hay que aplicar un segundo proceso manual, más costoso, hasta lograr la purificación. El primer proceso le cuesta a la empresa 500 soles por hora y el segundo 800 soles por hora. Además activar el segundo proceso cuesta 1,000 soles. Por otro lado la empresa cobra el precio de mercado por la purificación del mineral, que es 2,000 soles por hora de trabajo.

a) Si el tiempo X (en horas) que se necesita para lograr la purificación de un lote, es una v.a. continua cuya función de densidad tiene la gráfica que figura abajo, halle la fórmula de la función de densidad y la probabilidad de que el tiempo de purificación no pase de 2 Horas

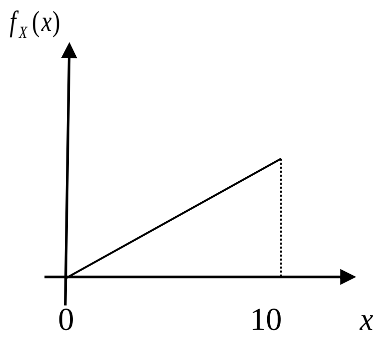


Figura 3

b) En el problema anterior halle el valor óptimo de *k*

Solución
a) La gráfica corresponde a una recta de la forma $f_X(x) = bx \quad 0 < x \le 10$ donde b es la constante normalizadora que hace que el área total valga 1. Aplicando esta condición y resolviendo se obtiene $b = 1/50$ y así tenemos que $f_X(x) = x/50 \quad 0 < x \le 10$

b) En cuanto a la utilidad, para un número prefijado *k* de horas de trabajo con el sistema automático la empresa asume un costo de 500*k* soles, y si por mala suerte debe emplear el segundo proceso, hay un costo de 800 soles por cada hora adicional más los 1000 soles que cuesta arrancar este otro proceso. El ingreso en cualquier caso es $2000X$ soles, donde X es el tiempo total hasta purificar el mineral. En este contexto, la utilidad U de todo el trabajo es:

$$U = U(X,k) = \begin{cases} 2000X - 500k & si \quad X \le k \\ 2000X - 500k - 800(X-k) - 1000 & si \quad X > k \end{cases} \Leftrightarrow$$

$$U = U(X,k) = \begin{cases} 2000X - 500k & si \quad X \le k \\ 1200X + 300k - 1000 & si \quad X > k \end{cases}$$

Tomando valor esperado:

$$\varphi(k) := E(U(X,k)) = \int_0^{10} U(x,k) f_X(x)dx = \int_0^{k} U(x,k) f_X(x)dx + \int_k^{10} U(x,k) f_X(x)dx =$$

$$\int_0^k (2000x - 500k) f_X(x)dx + \int_k^{10} (1200X + 300k - 1000) f_X(x)dx =$$
$$2000\int_0^k xf_X(x)dx - 500k\int_0^k f_X(x)dx + 1200\int_k^{10} xf_X(x)dx + (300k - 1000)\int_k^{10} f_X(x)dx =$$
$$2000\int_0^k xf_X(x)dx - 500kF_X(k) + 1200\int_k^{10} xf_X(x)dx + (300k - 1000)[1 - F_X(k)]$$

Derivando $\varphi(k)$ para maximizar:

$$\varphi'(k) = 2000kf_X(k) - 500F_X(k) - 500kf_X(k) - 1200kf_X(k)$$
$$+300 - 300F_X(k) - 300kf_X(k) + 1000 f_X$$
$$= -800F_X(k) + 1000 f_X(k) + 300$$

Igualando a cero la derivada de $\varphi(k)$: $\varphi'(k) = 0 \Leftrightarrow -800F_X(k) + 1000 f_X(k) + 300 = 0$ $\Leftrightarrow -8F_X(k) + 10 f_X(k) + 3 = 0$ y como $f_X(x) = x/50 \quad 0 < x \le 10$ y $F_X(x) = x^2/100 \quad 0 < x \le 10$, reemplazando en la ecuación obtenemos $-\frac{8k^2}{100} + \frac{10k}{50} + 3 = 0 \Leftrightarrow 8k^2 - 20k - 300 = 0$

$$\Rightarrow k = \frac{20 \pm \sqrt{10000}}{16} = \begin{cases} 7.5 \\ -5 \end{cases}$$

La solución negativa es absurda, así que $k = 7.5$ es la respuesta.

Ejemplo 46

Un distribuidor de un solvente industrial tiene la política de comprar al inicio de la temporada de ventas una existencia (‘stock’) de S unidades de volumen a C unidades monetarias. Durante la temporada vende el producto a V unidades monetarias; al final de temporada remata el sobrante a R unidades. Sabemos que la demanda de solvente al distribuidor es una v.a.c. X con función de densidad $f_X(x)$ y que $R < C < V$

a) Escriba la función de utilidad $U = U(X,S)$ del distribuidor
b) Determine el valor óptimo de S

Solución:

a) Sea U la utilidad, entonces U depende de S y de X, pues:

- Si $X \le S \Rightarrow$ durante la temporada vende X unidades y al final de la temporada remata $(S - X)$ unidades, de modo que $U = VX + R(S - X) - CS = (V - R)X + (R - C)S$
- Si $X > S \Rightarrow$ durante la temporada vende todo su stock S, quedando demanda insatisfecha por $(X - S)$ unidades de volumen. En este caso $U = (V - C)S$

De lo anterior podemos escribir:

$$U = U(X,S) = \begin{cases} (V - R)X + (R - C)S & si \quad X \le S \\ (V - C)S & si \quad X > S \end{cases}$$

b) De a) vemos que U tiene una componente aleatoria y otra no aleatoria. Desde el punto de vista económico lo racional es maximizar utilidades, pero la componente aleatoria X impide una maximización clásica, pues por definición X no es controlable. Tenemos que imponer un control ‘indirecto’. Aquí es donde el Valor Esperado muestra sus ventajas: $E[U] = E[U(X,S)]$ es una función (digamos, $\varphi(S) = E[U(X,S)]$) que depende sólo de S y no de X, pues precisamente integramos con respecto a X al calcular el esperado. Por tanto, si luego calculamos S de modo que maximice $\varphi(S) = E[U(X,S)]$ lo que estamos haciendo es determinar una ‘tendencia óptima’ para la utilidad U.

Veamos:

$$\varphi(S)=E[U(X,S)]=\int_{-\infty}^{+\infty}U(x;S)f_X(x)dx=\int_{-\infty}^{0}U(x;S)f_X(x)dx+\int_{0}^{S}U(x;S)f_X(x)dx+\int_{S}^{+\infty}U(x;S)f_X(x)dx$$
$$=\int_0^S(V-R)xf_X(x)dx+\int_0^S(R-C)Sf_X(x)dx+\int_S^{+\infty}(V-C)Sf_X(x)dx$$
$$=(V-R)\int_0^S xf_X(x)dx+(R-C)S\int_0^S f_X(x)dx+(V-C)S\int_S^{+\infty}f_X(x)dx$$
$$=(V-R)\int_0^S xf_X(x)dx+(R-C)SF_X(S)+(V-C)S[1-F_X(S)]$$
$$=(V-R)\int_0^S xf_X(x)dx+[(R-C)-(V-C)]SF_X(S)+(V-C)S$$
$$=(V-R)\int_0^S xf_X(x)dx+(R-V)SF_X(S)+(V-C)S$$

Por tanto $\varphi(S)=E[U(X,S)]=(V-R)\int_0^S xf_X(x)dx+(R-V)SF_X(S)+(V-C)S$ que es explícitamente una función de S

Examinando $\varphi(S)=(V-R)\int_0^S xf_X(x)dx+(R-V)SF_X(S)+(V-C)S$ se ve que es diferenciable y para maximizarla podemos aplicar derivación:

La condición de primer orden $\varphi'(S)=0$ equivale a

$$\varphi'(S)=(V-R)Sf_X(S)+(R-V)F_X(S)+(R-V)Sf_X(S)+(V-C)=(R-V)F_X(S)+(V-C)=0$$

$\Rightarrow F_X(S)=\dfrac{(C-V)}{(R-V)}\Rightarrow S=F_X^{-1}\left[\dfrac{(V-C)}{(V-R)}\right]$ y como $\varphi''(S)=(R-V)f_X(S)<0$, se trata de un máximo.

En definitiva, el stock óptimo S debe ser $S=F_X^{-1}\left[\dfrac{(V-C)}{(V-R)}\right]$

Ejemplo 47

Un mayorista compra un bien a 3 u.m. la unidad y lo vende a 2 u.m. la unidad. La cantidad de ese bien que le pueden demandar al mayorista es una variable aleatoria discreta X con rango $R_X=\{1,2,\ldots,N\}$ y función de probabilidad $P(x)$. El mayorista tiene la política de comprar, al inicio de cada campaña de ventas, una cantidad predeterminada ("stock") S de unidades del bien y nada más, así quede demanda insatisfecha. Por otra parte, si al terminar la campaña, quedan sobrantes, éstos se pierden y asume el costo de ello.

a) Escriba la utilidad U del comerciante como función de la demanda aleatoria X y el stock S.
b) Verifique que la utilidad esperada del comerciante $E(U)$ queda como función general del stock S, de la forma $E(U)=3\sum_{x=1}^{S}xP(x)-3SF(S)+S$, donde $P(x)$ es la función de probabilidad de X y $F(S)$ es la función de probabilidad acumulativa de X evaluada en S.
c) Si denotamos mediante $\varphi(S)$ a $E(U)$, para resaltar su dependencia de S, demuestre que la variación de la utilidad esperada, cuando se pasa de un stock de S unidades a un stock de (S+1) unidades, es $\varphi(S+1)-\varphi(S)=1-3F(S)$
d) Suponga que la función de probabilidad de X es la que figura más abajo, tabule la función de distribución acumulativa y úsela para identificar el stock S más conveniente para el comerciante, esto es, aquél que maximiza su utilidad esperada

x	1	2	3	4	5	6	7	8
$P(x)$	0.05	0.08	0.09	0.1	0.15	0.18	0.2	0.15

Solución:

a)

$$E(U)=\sum_{x=1}^{S}U(x,S)P(x)=\sum_{x=1}^{S}U(x,S)P(x)+\sum_{x=S+1}^{N}U(x,S)P(x)=\sum_{x=1}^{S}(3x-2S)P(x)+\sum_{x=S+1}^{N}SP(x)=$$

$$=3\sum_{x=1}^{S}xP(x)-2S\underbrace{\sum_{x=1}^{S}P(x)}_{F(S)}+S\underbrace{\sum_{x=S+1}^{N}P(x)}_{1-F(S)}=3\sum_{x=1}^{S}xP(x)-2SF(S)+S\big(1-F(S)\big)=$$

$$=3\sum_{x=1}^{S}xP(x)-3SF(S)+S\Rightarrow \boldsymbol{E(U)=3\sum_{x=1}^{S}xP(x)-3SF(S)+S}$$

b)

$$U(X,S)=\begin{cases}3X-2S & 1\le X\le S\\ S & S+1\le X\le N\end{cases}$$

c)

Sea $\varphi(S)=E(U)=3\sum_{x=1}^{S}xP(x)-3SF(S)+S$, entonces

$$\varphi(S+1)=3\sum_{x=1}^{S+1}xP(x)-3(S+1)F(S+1)+(S+1)=$$

$$=3\left(\sum_{x=1}^{S}xP(x)+(S+1)P(S+1)\right)-3(S+1)[F(S)+P(S+1)]+(S+1)=$$

$$=\left(3\sum_{x=1}^{S}xP(x)+3(S+1)P(S+1)\right)-3(S+1)F(S)-3(S+1)P(S+1)+(S+1)=$$

$$=3\sum_{x=1}^{S}xP(x)+\boldsymbol{3SP(S+1)}+\boldsymbol{3P(S+1)}-3SF(S)-3F(S)\boldsymbol{-3SP(S+1)}\boldsymbol{-3P(S+1)}+S+1=$$

$$\Rightarrow\varphi(S+1)=3\sum_{x=1}^{S}xP(x)-3SF(S)-3F(S)+S+1\Rightarrow$$

$$\varphi(S+1)-\varphi(S)=\left(3\sum_{x=1}^{S}xP(x)-3SF(S)-3F(S)+S+1\right)-\left(3\sum_{x=1}^{S}xP(x)-3SF(S)+S\right)=$$

$$=\boldsymbol{3\sum_{x=1}^{S}xP(x)}\boldsymbol{-3SF(S)}-3F(S)\boldsymbol{+S}+1\boldsymbol{-3\sum_{x=1}^{S}xP(x)}\boldsymbol{+3SF(S)}\boldsymbol{-S}=-3F(S)+1\Rightarrow$$

$$\boldsymbol{\varphi(S+1)-\varphi(S)=1-3F(S)}$$

d) La distribución acumulativa $F(x)$ se obtiene sumando probabilidades:

x	1	2	3	**4**	5	6	7	8
$P(x)$	0.05	0.08	0.09	**0.1**	0.15	0.18	0.2	0.15
$F(x)$	0.05	0.13	0.22	**0.32**	0.47	0.65	0.85	1

La variación en la utilidad, conforme aumentamos el stock S, es $\varphi(S+1)-\varphi(S)=1-3F(S)$ y mientras esta variación sea positiva estamos bien, o sea mientras $1-3F(S)>0\ \Rightarrow F(S)<\frac{1}{3}=0.33;$ luego la condición para detenernos es que es $F(S)<\frac{1}{3}=0.33$ o sea, mientras no pasemos de una probabilidad acumulada de 0.33 podemos seguir aumentando S. Examinando la tabla de probabilidades acumulada, $F(x)\ pasa\ de\ 0.33$ cuando $x=5$, o sea no debemos llegar a 5 **y por tanto el stock "optimo" que maximiza la utilidad esperada $\boldsymbol{\varphi(S)=E(U)}$ es $\boldsymbol{S}$=4.**

Nota:

Como se tiene una fórmula $\varphi(S)=E(U)=3\sum_{x=1}^{S}xP(x)-3SF(S)+S$ que explicita la dependencia de la utilidad esperada con respecto al valor del stock S, otra alternativa de solución es la "computacional": Para localizar el stock optimo S, se calcula $\varphi(S)$ para los distintos valores posibles del S, desde 1 hasta 8 (no tiene sentido pasar de 8 porque la demanda posible no lo hace). Trabajar así, a mano o con calculadora no es práctico, pero con ayuda de una hoja de cálculo y una computadora, es sencillo realizar esta operación:

x	1	2	3	4	5	6	7	8
$P(x)$	0.05	0.08	0.09	0.1	0.15	0.18	0.2	0.15
$F(x)$	0.05	0.13	0.22	0.32	0.47	0.65	0.85	1
$xP(x)$	0.05	0.16	0.27	0.4	0.75	1.08	1.4	1.2
$\sum_{x=1}^{S} xP(x)$	0.05	0.21	0.48	0.83	1.42	2.23	3.23	3.68
$3\sum_{x=1}^{S} xP(x)$	0.15	0.63	1.44	2.49	4.26	6.69	9.69	11.04
$3SF(S)$	0.15	0.78	1.98	3.84	7.05	11.7	17.85	24
S	1	2	3	4	5	6	7	8
$\varphi(S) = E(U)$	1.00	1.85	2.46	**2.65**	2.21	0.99	-1.16	-4.96

En la última línea están las diversas utilidades esperadas y por inspección se encuentra que con *S*=4, la utilidad esperada es máxima. Este método computacional a veces es el único que se puede aplicar cuando no hay fórmula explícita ("fórmula cerrada") para resolver el problema.

Ejemplo 48

Un empresario enfrenta el problema de introducir un nuevo producto en el mercado, para lo cual dispone de cuatro procesos de producción alternativos y excluyentes: a_1 , a_2, a_3 y a_4. La utilidad obtenible con cada proceso depende de estado del mercado, clasificado según los niveles de demanda que haya: Baja, Media o Alta, que pueden ocurrir con probabilidades 0.1, 0.5 y 0.4 respectivamente. Las correspondientes utilidades por tipo de proceso según nivel de demanda son (en miles de unidades monetarias):

Proceso	Utilidad según Nivel de demanda		
	D. Baja	D. Media	D. Alta
a_1	70	120	200
a_2	80	120	180
a_3	100	125	160
a_4	100	120	150

Considerando aleatoria la utilidad asociada a cada proceso, a largo plazo ¿Cuál sería el proceso más conveniente?

Solución:

Primero evaluemos cada proceso para ver si alguno puede descartarse de antemano: un proceso es descartable (no admisible) si es superado en todo por alguno de los otros procesos.
a_1 no es superado por ningún otro proceso si la demanda fuera alta, luego no es descartable.
a_2 supera a a_1 si la demanda es baja y supera a a_3 y a_4 si la demanda es alta. No es descartable.
a_3 supera a a_1 y a_2 si la demanda es baja o media, y a a_4 si la demanda es alta. No es descartable.
a_4 es superado por a_3 con demanda media y alta y es igual a a_3 con demanda baja. Este proceso sí es descartable, nunca da mejor resultado que a_3.
Podemos eliminar la cuarta fila de nuestra tabla de posibles procesos y resultados. Como los niveles de demanda son aleatorios, las utilidades devienen en aleatorias también y podemos calcular la utilidad esperada con cada proceso:

Proceso	Utilidad según Nivel de demanda			
	D. Baja	D. Media	D. Alta	
Probabilidad	0.1	0.5	0.4	Utilidad esperada
a_1	70	120	200	70×0.1+120×0.5+200×0.4=87
a_2	80	120	180	80×0.1+120×0.5+180×0.4=80
a_3	100	125	160	100×0.1+125×0.5+160×0.4=74

El proceso a_1 genera mayor utilidad esperada o promedio. A largo plazo es el proceso más conveniente.

2.6 Función Generatriz de Momentos y Cambio de Variable

2.6.1 Función Generatriz de Momentos

Definición
Si X es v.a., se define el "k-ésimo Momento Poblacional" denotado m_k mediante $m_k = E(X^k)$ $k = 0,1,2,...$

Definición
Si X es v.a., se define la Función Generatriz de Momentos de X, denotada $M_X(t)$ mediante $M_X(t) := E(e^{tX})$ donde t es variable no aleatoria o variable matemática definida en un entorno de 0

Nótese que $M_X(0) = 1$

Proposición 1
$M_X^{(k)}(0) = E(X^k) = m_k$, si existe el valor esperado.

En efecto, recordemos primero que $e^Z = \sum_{k=0}^{\infty} \frac{Z^k}{k!} \quad \forall \quad Z \quad real$. En particular

$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k = 1 + \frac{t}{1!} X + \frac{t^2}{2!} X^2 + \frac{t^3}{3!} X^3 + ...$ y tomando valor esperado tenemos

$M_X(t) = E(e^{tX}) = 1 + \frac{t}{1!} E(X) + \frac{t^2}{2!} E(X^2) + \frac{t^3}{3!} E(X^3) + ...$

Derivando $M_X(t)$ con respecto a *t*

$M_X^{(1)}(t) = E(X) + 2\frac{t^1}{2!} E(X^2) + 3\frac{t^2}{3!} E(X^3) + ... = E(X) + \frac{t^1}{1!} E(X^2) + \frac{t^2}{2!} E(X^3) + ...$. Si evaluamos en *t*=0 llegamos a $M_X^{(1)}(0) = E(X)$.

Derivando dos veces $M_X(t)$ con respecto a *t*

$M_X^{(2)}(t) = E(X^2) + 2\frac{t}{2!} E(X^3) + ... = E(X^2) + \frac{t}{1!} E(X^3) + ...$. Evaluando en *t*=0 llegamos a $M_X^{(2)}(0) = E(X^2)$.

Así, inductivamente se llega al resultado general

Proposición 2
Sean X e Y son dos variables aleatorias entonces $M_X(t) = M_Y(t) \Leftrightarrow F_X = F_Y$

Ejemplo 49
Si $P_X(x) = (\frac{1}{2})^x \qquad x = 1,2,3,...$ Hallar $M_X(t)$ y μ_X

Solución:
$M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} (\frac{1}{2})^x = \sum_{x=1}^{\infty} (\frac{e^t}{2})^x = \frac{(e^t/2)}{1-(e^t/2)}$ (aplicando la serie geométrica y tomando *t* de modo que $(e^t/2) < 1$).

Derivando con respecto a t: $M^{'}(t)=\dfrac{(e^t/2)}{[1-(e^t/2)]^2}$ y evaluando en t=0 se obtiene

$$M^{'}(0)=E[X]=\mu_X=\frac{(1/2)}{[1-(1/2)]^2}=2$$

2.6.2 Cambio de Variable

El problema del Cambio de Variable es*: Dada la v.a.* X *e* $Y=H(X)$*, hallar la distribución de* Y *a partir de la distribución de* X .

Hay varias alternativas de solución (una de ellas es usar $M_X(t)$). Nosotros exploraremos el caso en que H tiene inversa:

Sea $G_Y(y)$ la distribución acumulativa de Y y supongamos H^{-1} creciente. Entonces $G_Y(y)=P(Y\le y)=P(H(X)\le y)=P(X\le H^{-1}(y))=F_X(H^{-1}(y))$, donde F_X es la distribución acumulativa de X . Conociendo $G_Y(y)$ podemos obtener la función de densidad $g_Y(y)$ o de probabilidad $P_Y(y)$ de Y mediante derivaciones o restas según sea el caso.

El caso en que H^{-1} es decreciente se trata de manera análoga. El método anteriormente usado se conoce como Método de la Distribución Acumulativa y se puede ampliar al caso en que H no tiene inversa.

Ejemplo 50

En el caso de X =Precio de venta en una privatización, sea $Y=H(X)=\sqrt{X}+1$
Hallar la función de densidad de Y .

Solución

- Primero especifiquemos el rango R_Y de Y :
  Como $\mathbf{0<X\le 1}$ $\Rightarrow 0<\sqrt{X}\le 1$ y sumando 1 a ambos lados de la desigualdad tenemos $1<\sqrt{X}+1\le 2\Rightarrow 1<Y\le 2$ es el rango de la v.a. Y
- Ahora, sea $y\in]1,2]$ y sea $G_Y(y)$ la distribución acumulativa de Y . Entonces
  $G_Y(y)=P(Y\le y)=P(\sqrt{X}+1\le y)=P(\sqrt{X}\le y-1)=P(X\le (y-1)^2)=F_X((y-1)^2)$.

  Derivando con respecto a y obtenemos $g(y)=G_Y^{'}(y)=F_X^{'}((y-1)^2)\times 2(y-1).$

  Ya vimos que $f_X(x)=2x$, por tanto $g(y)=2(y-1)^2\times 2(y-1)=4(y-1)^3$ y así tenemos que $g(y)=4(y-1)^3 \quad 1<y\le 2$ es la función de densidad de Y .

2.6.3 Cálculo del valor esperado por desarrollo asintótico

Dada una v.a. X y una función $H(X)$, para hallar el valor esperado $E[H(X)]$ podemos aplicar la definición o en ciertos casos las propiedades del operador valor esperado $E[\bullet]$. La última alternativa es muy útil pero no siempre es aplicable, ya sea porque la función $H(X)$ no es lineal o porque es lineal por tramos.

Por ejemplo, en el caso de X =Precio de venta en una privatización, con función de densidad $f_X(x)=2x \quad si \quad 0<x\le 1$, ya vimos que $E(X)=2/3$ pero en cambio

$E(X^3)=\int_0^1 x^3 2xdx = 2\left[\frac{x^5}{5}\right]_0^1 = \frac{2}{5} = 0.4 \neq (2/3)^3$, o sea $E(X^3) \neq [E(X)]^3$ ***y en general <u>no se cumple</u> que*** $E[H(X)] = H(E(X))$

Sin embargo, cuando podemos expresar o descomponer $H(X)$ como suma (posiblemente infinita) de funciones, es posible obtener el valor <u>aproximado de</u> $E[H(X)]$

Propiedad

Sea X variable aleatoria con $E(X)=\mu$ y $V(X)=\sigma^2$. Sea $H(X)$ función al menos dos veces diferenciable en $X=\mu$. Entonces se cumplen:

a) $E[H(X)] \cong H(\mu) + \frac{H''(\mu)}{2}\sigma^2$

b) $V[H(X)] \cong [H'(\mu)]^2 \sigma^2$

En efecto, si desarrollamos $H(X)$ en serie de Taylor alrededor de $X=\mu$ y hasta el término cuadrático, sabemos que

$H(X) = H(\mu) + (X-\mu)H'(\mu) + \frac{(X-\mu)^2 H''(\mu)}{2} + R$ donde R es un residuo.

Si tomamos valor esperado:

$$E[H(X)] = E[H(\mu)] + E[(X-\mu)H'(\mu)] + E[\frac{(X-\mu)^2 H''(\mu)}{2}] + E[R] = H(\mu) + \frac{H''(\mu)}{2}E[(X-\mu)^2] + E[R]$$

pues $E[(X-\mu)H'(\mu)] = H'(\mu)E(X-\mu) = 0$. Si además consideramos despreciable al residuo R (o sea consideramos $R=0$), obtenemos el resultado $E[H(X)] \cong H(\mu) + \frac{H''(\mu)}{2}\sigma^2$.

Análogamente, para obtener una aproximación a la varianza $V[H(X)]$, desarrollemos $H(X)$ en serie de Taylor alrededor de $X=\mu$, pero hasta el término lineal o de primer grado:

$H(X) = H(\mu) + (X-\mu)H'(\mu) + R_2$ donde ahora R_2 representa el residuo.

Tomando sólo los dos primeros términos escribimos $H(X) \cong H(\mu) + (X-\mu)H'(\mu)$ y aplicando la varianza y sus propiedades:

$$V[H(X)] \cong V[H(\mu) + (X-\mu)H'(\mu)] = V[(X-\mu)H'(\mu)] = [H'(\mu)]^2 V[(X-\mu)] = [H'(\mu)]^2 V(X) = [H'(\mu)]^2 \sigma^2$$

Ejemplo 51

En el caso del precio de venta X en la privatización, sea $H(X) := X^3$. Aquí tenemos $H'(X) = 3X^2$ y $H''(X) = 6X$. Ya vimos que $\mu = E(X) = \frac{2}{3}$ y $\sigma^2 = \frac{1}{18}$. Luego como $H(\mu) = \mu^3 = \frac{8}{27}, H'(\mu) = 3\mu^2 = \frac{4}{3}$ y $H''(\mu) = 6\mu = 4$, aplicando la propiedad de aproximación $E[X^3] \cong \frac{8}{27} + \frac{\frac{4}{3}}{2} \times \frac{1}{18} = \frac{9}{27} = \frac{1}{3} = 0.33$ y

$V[X^3] \cong \left[\frac{4}{3}\right]^2 \times \frac{1}{18} = \frac{8}{81} = 0.099$