Elementos de Probabilidad y Estadística

Notas de curso

Joaquín Ortega Sánchez

18 de mayo de 2009

Índice general

Capítulo 1

Espacios de Probabilidad

1.1. Introducción

El objetivo de la Teoría de Probabilidad es desarrollar y estudiar modelos matemáticos para experimentos cuyos resultados no pueden predecirse.

FALTA

1.2. Espacio Muestral. Eventos.

Cada resultado posible de un experimento aleatorio será llamado *evento elemental* y el conjunto de los eventos elementales será el *espacio muestral*. Usualmente, denotaremos con Ω el espacio muestral, y mediante ω los eventos elementales (o puntos de Ω).

Veamos algunos ejemplos de experimentos aleatorios y sus espacios muestrales asociados.

1. En una fábrica se toma uno de los artículos producidos y se prueba para determinar si es defectuoso. En este caso podemos considerar $\Omega = \{B, D\}$, donde B indica bueno y D defectuoso. Si en cambio se extraen n artículos y se prueban, podemos considerar $\Omega = \{(\epsilon_1, \epsilon_2, \dots, \epsilon_n) : \epsilon_i = 0 \text{ ó } 1; i = 1, \dots, n\}$ donde $\epsilon_i = 0$ indica que el i-ésimo artículo es bueno y $\epsilon_i = 1$ indica que es defectuoso. Es decir, Ω es el conjunto de n-uplas o vectores de dimensión n de ceros y unos. En este caso Ω consta de 2^n eventos elementales y, en particular, $\sum_{i=1}^{n} \epsilon_i$ representa en número de objetos defectuosos del evento elemental $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$.

2. En un punto de una carretera contamos el número de vehículos que pasan durante un cierto lapso de tiempo. En este caso podemos tomar $\Omega = \{0, 1, 2, \dots\}$, es decir el conjunto de los enteros no-negativos. Podemos, sin embargo, tomar otros conjuntos como espacio muestral en este caso. Por ejemplo, si sabemos que el número de vehículos considerados no supera los mil, podemos considerar $\Omega_1 = \{n : 0 \leq n \leq 1.000\}$, aunque no necesariamente del hecho de que Ω_1 sea subconjunto de Ω, se concluye que la descripción del experimento aleatorio mediante Ω_1 sea mas simple que la que se obtiene usando Ω.

3. En una sucesión de cálculos realizados con una computadora, observamos los primeros k dígitos no tomados en cuenta al truncar los resultados de las operaciones en una cierta cifra decimal. En este caso podemos tomar como espacio muestral $\Omega = \{(a_1, \dots, a_k) : a_i \in \mathbb{Z},\ 0 \leq a_i \leq 9\}$

4. En una fábrica de componentes electrónicos se eligen varios de ellos al azar y se conecta cada uno de ellos hasta que se daña, observando en cada caso el tiempo de duración. Si se trata de un solo componente podemos tomar
$$\Omega = \{t : t \in \mathbb{R},\ t \geq 0\}$$
es decir, el conjunto de números reales no-negativos. Si se consideran n componentes, podemos tomar
$$\Omega = \{(t_1, t_2, \dots, t_n) : t_i \in \mathbb{R},\ t_i \geq 0\}.$$

5. Se lanza un dado repetidamente y se cuenta el número de lanzamientos hasta que salga el 6 por primera vez. En este caso el espacio muestral es el conjunto de los números naturales:
$$\Omega = \{1, 2, 3, \dots\}.$$

6. Se mide la presión y temperatura en una estación meteorológica. Aquí,
$$\Omega = \{(p, t) : p > 0,\ t \in \mathbb{R}\}.$$

7. Se escoge un punto al azar lanzando un dardo a un disco de radio un metro. En este caso el espacio muestral es el conjunto de puntos del plano que estan dentro de la circunferencia de radio 1:
$$\Omega = \{(x, y) : x^2 + y^2 \leq 1\}.$$

En la práctica, al realizar un experimento nos interesa con frecuencia, saber si algún subconjunto de Ω ha ocurrido. A estos subconjuntos los llamaremos *eventos* o *sucesos*. Por ejemplo, en el caso 1 podemos estar interesados en el subconjunto: "entre los n artículos extraídos hay d defectuosos", es decir, en el subconjunto de Ω definido por
$$\{(\epsilon_1, \dots, \epsilon_n) : \epsilon_i = 0 \text{ ó } 1, \sum_1^n \epsilon_i = d\}.$$
En el caso 3 nos interesará saber, por ejemplo, si la primera cifra no tomada en cuenta al truncar es mayor o igual que 5, o sea,
$$\{(a_1, \dots, a_k)\} : 0 \leq a_i \leq 9,\ a_1 \geq 5\}.$$
Análogamente, en la situación planteada en 6, nos interesarán eventos del tipo: "la presión está comprendida entre p_1 y p_2 y la temperatura entre t_1 y t_2", es decir
$$\{(p_i, t_i) : p_1 \leq p \leq p_2,\ t_1 \leq t \leq t_2\}.$$
Estamos interesados, por lo tanto, en considerar familias de subconjuntos de Ω, es decir, familias $\mathcal{A}$ de eventos. Diremos que un evento $A \in \mathcal{A}$ ocurre al realizar un experimento aleatorio cuyo resultado es el evento elemental ω, si $\omega \in A$.

Veamos que condiciones debe cumplir la familia de eventos $\mathcal{A}$. En primer lugar

a. $\boxed{\Omega \in \mathcal{A}}$ es decir que al realizar el experimento algo ocurre. A Ω lo llamaremos *evento cierto*.

Si A es un evento, pediremos que "no ocurre A" también sea un evento, es decir

b. $\boxed{A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}}$ donde $A^c = \Omega - A = \{\omega : \omega \in \Omega,\ \omega \notin A\}$ es el complemento de A.

Finalmente, la familia $\mathcal{A}$ también debe satisfacer que si $A_1, A_2, \dots, A_n, \dots$ son eventos, "ocurre alguno de los A_n" también es un evento, o sea

c. $\boxed{A_n \in \mathcal{A} \ (n = 1, 2, \dots) \ \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}}$

Definición 1.1 Una familia $\mathcal{A}$ de subconjuntos de Ω que satisface las condiciones a, b y c se llama una σ-álgebra de subconjuntos o partes de Ω.

En adelante supondremos, por lo tanto, que las familias de eventos son σ-álgebras. Las siguientes son consecuencias inmediatas de la definición:

1. El conjunto vacío, $\emptyset$, es un evento, ya que $\emptyset = \Omega^c$.
2. $A_1, A_2, \dots A_k \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{k} A_n \in \mathcal{A}$. Basta considerar $A_{n+1} = A_{n+2} = \cdots = \emptyset$ y aplicar 1. y c.
3. $A_1, A_2, \dots, A_n, \dots \in \mathcal{A} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$. En efecto, por las leyes de de Morgan,

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c$$

y basta ahora aplicar b y c.

Ejemplos.

8 Para cualquier conjunto Ω, la σ-álgebra más sencilla es la σ-álgebra trivial $\mathcal{T} = \{\Omega, \emptyset\}$. La mayor σ-álgebra de subconjuntos de Ω es $\mathcal{P}(\Omega)$, el conjunto de partes de Ω, es decir, la colección de todos los subconjuntos de Ω. Cualquier otra σ-álgebra debe contener a $\mathcal{T}$ y estar contenida en $\mathcal{P}(\Omega)$.

Si Ω es finito o numerable usaremos como σ-álgebra a $\mathcal{P}(\Omega)$.

9 *Muestreo con reposición.* De la producción de una fábrica se extrae un artículo al azar y se determina si es bueno o defectuoso (B o D, respectivamente). Se devuelve este artículo al stock y se extrae de nuevo al azar un artículo, que puede ser el mismo. Esta operación se repite una vez más, de modo que en total se extraen tres.

El espacio muestral es:

$$\Omega = \{BBB, BBD, BDB, DBB, BDD, DBD, DDB, DDD\}$$

Observamos que hay 2^3 eventos elementales, ya que en cada una de las tres extracciones hay dos resultados posibles. Consideramos los siguientes eventos:

A_1: "El segundo artículo resultó bueno"

A_2: "Se obtuvo un solo defectuoso en las tres extracciones".

A_3: "No hubo defectuosos".

Los eventos definidos son:

$$A_1 = \{BBB, BBD, DBB, DBD\} \quad A_2 = \{BBD, BDB, DBB\} \quad A_3 = \{BBB\}$$

El número de eventos elementales incluidos en A_1 es 2^2 ya que el resultado de la segunda extracción está fijo. El evento A_2 contiene 3 puntos muestrales, ya que hay tres lugares posibles para el objeto defectuoso en la muestra. Podemos ahora combinar estos eventos utilizando operaciones de conjuntos. Tenemos, por ejemplo,

$$A_1 \cap A_2 = \{BBD, DBB\}$$
$$A_1^c \cup A_2^c = \{BBB, BDB, BDD, DBD, DDB, DDD\}$$
$$A_1 \cap A_2^c = \{BBB, DBD\}$$

10 *Muestreo sin reposición.* De una población de N artículos entre los cuales hay n defectuosos, se extraen sucesivamente r sin reposición y se cuenta el número de los defectuosos en la muestra. El espacio muestral contiene todos los subconjuntos de r elementos tomados entre los N dados.

1.3. Espacios de Probabilidad.

Definición 1.2 Sean Ω un espacio muestral y $\mathcal{A}$ una familia de eventos de Ω, es decir, una σ-álgebra de subconjuntos de Ω. Estamos interesados en asignar a cada evento $A \in \Omega$ un número real $P(A)$, que llamaremos la *probabilidad* de A, de modo tal que se cumplan las siguientes condiciones:

1. $\boxed{P(A) \geq 0 \text{ para todo } A \in \Omega}$ La probabilidad de un evento cualquiera es un número real no negativo.

2. $\boxed{P(\Omega) = 1}$ El evento cierto tiene probabilidad igual a 1.

Si $A_n \in \mathcal{A}$ para $n = 1, 2, \dots$ son eventos disjuntos dos a dos, es decir, tales que $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces

3. $\boxed{P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)}$

Una terna $(\Omega, \mathcal{A}, P)$, formada por un espacio muestral Ω, una familia $\mathcal{A}$ de eventos y una probabilidad P se llama un espacio de probabilidad.

El problema de cómo definir la función P, o sea, de cómo asignar una probabilidad a cada evento, debe ser resuelto de acuerdo a las condiciones concretas de cada experimento aleatorio en consideración.

1.4. Algunas Consecuencias de la Definición.

Veamos a continuación algunas consecuencias de la definición anterior. Usaremos la notación $A + B$ para indicar la unión de los conjuntos A y B cuando ellos son disjuntos.

(1) $P(\emptyset) = 0$.

En efecto, consideremos $A_1 = \Omega$ y $A_i = \emptyset$, $i = 2, 3, \dots$ Entonces $A_i \in \mathcal{A}$ cualquiera que sea i y además si $i \neq j$ se tiene $A_i \cap A_j = \emptyset$. Resulta

$$P(\Omega) = P\left(\sum_{i=1}^{\infty} A_i\right) = P(\Omega) + \sum_{i=2}^{\infty} P(A_i).$$

Luego

$$\sum_{i=2}^{\infty} P(A_i) = 0$$

y como $P(A_i) \geq 0$ para todo i se tiene que $P(A_i) = 0$ para $i \geq 2$. En consecuencia $P(\emptyset) = 0$.

(2) $A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Basta considerar $A_i = \emptyset,\ i \geq 3$ y aplicar la condición 3 de la definición de espacio de probabilidad. De manera similar se pude demostrar que P es finitamente aditiva: Si $A_1, \ldots, A_n$ son disjuntos dos a dos entonces

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k).$$

(3) $P(A^c) = 1 - P(A)$.

Como $A^c \cup A = \Omega$ y $A^c \cap A = \emptyset$ se tiene

$$P(A^c) + P(A) = 1.$$

(4) $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$.

Como $A_2 = A_1 + (A_2 \cap A_1^c)$ resulta

$$P(A_2) = P(A_1) + P(A_2 \cap A_1^c)$$

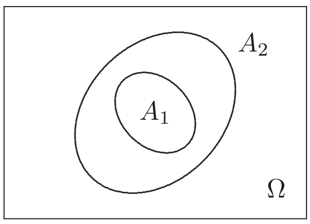


Figura 1.1

y en consecuencia

$$P(A_1) \geq P(A_2) \text{ ya que } P(A_2 \cap A_1^c) \geq 0$$

(5) $P(A) \leq 1$ para todo $A \in \mathcal{A}$.

Esto es consecuencia inmediata del punto anterior al considerar que $A \subset \Omega$.

(6) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

En efecto, considerando que

$$A_1 \cup A_2 = A_1 + (A_2 \cap A_1^c) \quad \text{y} \quad A_2 = (A_1 \cap A_2) + (A_1^c \cap A_2)$$

después de aplicar (2) a ambas igualdades y restar resulta la proposición (6).

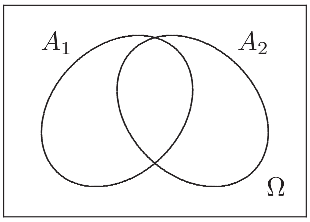


Figura 1.2

(7) $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots \Rightarrow P\left(\bigcup_{n=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$.
Sean

$$B_1 = A_1 \quad \text{y} \quad B_n = A_n \cap A_{n-1}^c \ \text{ si } n > 1,$$

resulta

$$\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} B_i$$

y entonces

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\sum_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(\sum_{i=1}^{n} B_i) = \lim_{n\to\infty} P(A_n)$$

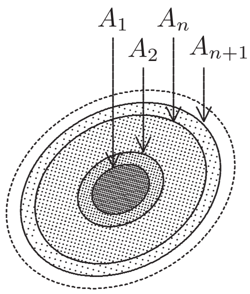


Figura 1.3

(8) $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \Rightarrow P\left(\bigcap_{n=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$.
Como la sucesión $\{A_n^c\}$ es creciente, usando (7) obtenemos

$$P\left(\bigcap_{n=1}^{\infty} A_i\right) = P\left(\left(\bigcup_{n=1}^{\infty} A_i^c\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} A_i^c\right)$$

$$= 1 - \lim_{n\to\infty} P(A_n^c) = 1 - \lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n).$$

(9) $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$.
Para ver esto apliquemos (6) a los eventos $A\cup B$ y C, obteniendo

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

y de manera similar

$$\begin{aligned}P(A\cup B)&=P(A)+P(B)-P(A\cup B)\\ P((A\cup B)\cap C)&=P((A\cap C)\cup(B\cap C))\\ &=P(A\cap C)+P(B\cap C)-P(A\cap B\cap C).\end{aligned}$$

Reemplazando las dos últimas expresiones en la primera obtenemos el resultado.

(10) $P(\bigcup_{i=1}^n A_i)=\sum_{i=1}^n P(A_i)-\sum_{i<j=2}^n P(A_i\cap A_j)+\cdots+(-1)^{n+1}P(A_1\cap\cdots\cap A_n)$.
Para demostrar esta proposición procedemos por inducción completa en n siguiendo las mismas líneas que en la anterior, que corresponde al caso $n=3$. Para $n=2$ es la propiedad (6).

Suponemos entonces que el resultado es cierto para n y queremos deducir que también lo es para $n+1$. ¿Qué significa que el resultado es cierto para n? Significa que

$$P\left(\bigcup_{i=1}^n A_i\right)=\sum_{k=1}^n(-1)^{k+1}\sum_{1\le i_1<i_2<\ldots<i_k\le n}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}) \tag{1.1}$$

Queremos deducir de (1.1) que también es válida una fórmula análoga para

$$P\left(\bigcup_{i=1}^{n+1} A_i\right).$$

Pongamos entonces $B=\bigcup_{i=1}^n A_i$ y apliquemos la propiedad (6) a

$$\begin{aligned}P\left(\bigcup_{i=1}^{n+1} A_i\right)&=P(B\cup A_{n+1})=P(B)+P(A_{n+1})-P(B\cap A_{n+1})\\ &=P\left(\bigcup_{i=1}^{n} A_i\right)+P(A_{n+1})-P(\bigcup_{i=1}^{n}(A_i\cap A_{n+1})).\end{aligned} \tag{1.2}$$

El primero de estos tres términos lo reemplazamos utilizando (1.1) y el último también sólo que, en lugar de cada A_i ponemos $A_i\cap A_{n+1}$. Observemos que es lo que nos queda. En primer lugar,

$$P(A_1)+\cdots+P(A_n)+P(A_{n+1}),$$

los primeros n provenientes del primer sumando en (1.2) y el último del segundo sumando. En segundo lugar

$$-\sum_{1\le i_1<i_2\le n}P(A_{i_1}\cap A_{i_2})-\sum_{i=1}^n P(A_i\cap A_{n+1})=-\sum_{1\le i_1<i_2\le n+1}P(A_{i_1}\cap A_{i_2}).$$

Aquí el primer sumando viene del primero de (1.2) y el segundo, del tercero de (1.2). De la misma manera, para $k\le n$, queda una suma de la forma

$$\begin{aligned}&(-1)^{k+1}\sum_{1\le i_1<i_2<\ldots<i_k\le n}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})\\ &-(-1)^k\sum_{1\le i_1<i_2<\ldots<i_{k-1}\le n}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_{k-1}}\cap A_{n+1})\\ &=(-1)^{k+1}\sum_{1\le i_1<i_2<\ldots<i_k\le n+1}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}).\end{aligned}$$

Finalmente, para $k = n+1$, no tenemos ningún término en el primer sumando de (1.1) y tenemos uno sólo en el tercero que es:

$$(-1)^{n+2} P(A_1 \cap A_2 \cap \cdots \cap A_n \cap A_{n+1}).$$

Juntando todos los términos resulta

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{k=1}^{n+1} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n+1} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

1.5. Ejemplos y Aplicaciones.

1.5.1. Probabilidades en Espacios Finitos.

Sean $\Omega = \{\omega_1, \ldots, \omega_m\}$ un conjunto finito y $\mathcal{A} = \mathcal{P}(\Omega)$ la familia de todos los subconjuntos de Ω. Elijamos m números reales $p_i,\ i = 1, 2, \ldots m$, tales que

$$\begin{cases} p_i \geq 0 \quad \text{para todo } i \\ \displaystyle\sum_{i=1}^{m} p_i = 1 \end{cases}$$

Poniendo $P(\omega_i) = p_i$ $(i = 1, 2, \ldots m)$, queda definida la probabilidad para todo evento $A \in \mathcal{A}$ mediante la asignación

$$P(A) = \sum_{\omega_i \in A} p_i.$$

Un caso particular de interés es aquel en el cual $p_i = 1/m$ para todo i, y ahora si A tiene n elementos

$$P(A) = \frac{n}{m},$$

es decir que si todos los eventos elementales son igualmente probables, la probabilidad de un evento A es el cociente entre el número de elementos que pertenecen a A y el número total de elementos de Ω. Esta definición se conoce como la *definición clásica* y fue propuesta, entre otros, por Laplace. En la sección 1.7 incluimos algunos comentarios al respecto.

En una situación como la descrita, en la cual todos los resultados posibles del experimento tienen la misma probabilidad de ocurrir, el problema de calcular la probabilidad de un evento se reduce a contar cuántos resultados posibles tiene el experimento y cuántos de estos pertenecen al evento que nos interesa. En el próximo capítulo estudiaremos algunas técnicas combinatorias que facilitan estos cálculos.

En un problema específico, podemos determinar si los resultados posibles tienen la misma probabilidad por consideraciones de simetría sobre el experimento que estamos considerando. Por ejemplo, si se trata del lanzamiento de un dado, en principio no hay razones para suponer que alguna cara tenga mayor o menor probabilidad de ocurrir que las demás, y por lo tanto asumimos como modelo que todos los resultados son equiprobables. Algo similar sucede con el lanzamiento de una moneda, el juego de ruleta o la extracción de una carta de un paquete que ha sido bien barajeado.

Por supuesto que en la práctica esto puede no ser cierto: el dado puede no ser perfectamente simétrico, o la ruleta puede estar desbalanceada y favorecer ciertos resultados. Para determinar si este es el caso existen procedimientos estadísticos que permiten contrastar la hipótesis de simetría, pero por los momentos no nos ocuparemos de este problema.

Veamos algunos ejemplos.

1. De los números del 1 al 10 escogemos tres al azar, en orden y sin reposición. ¿Cuál es la probabilidad de obtener 1, 2 y 3, en este orden?

   En este problema podemos describir el espacio muestral como el conjunto de todos los vectores de tres componentes tomadas de los enteros del 1 al 10, sin repetir ninguna componente.

   $$\Omega = \{(a, b, c) : 1 \leq a, b, c \leq 10, \text{ distintas}\}.$$

   Como estamos muestreando al azar, todos los vectores del espacio tienen la misma probabilidad.

   El evento que nos interesa corresponde a un resultado particular, el vector $(1, 2, 3)$. Por lo tanto tenemos que contar cuantos elementos hay en Ω para saber cuál es la probabilidad de cada uno de ellos. La primera componente del vector la podemos escoger de 10 maneras. Para la segunda sólo tenemos 9 posibilidades, porque no podemos repetir el número que ocupa la primera componente. Finalmente, para la tercera hay 8 posibilidades. Por lo tanto tenemos

   $$10 \times 9 \times 8 = 720$$

   puntos en el espacio muestral. Como todos tienen la misma probabilidad, la respuesta al ejemplo es $1/720$.

2. Si los números del ejemplo anterior se escogen con reposición ¿Cuál es la probabilidad de obtener 1, 2 y 3, en este orden?

   En este caso el espacio muestral incluye vectores con componentes repetidas:

   $$\Omega = \{(a, b, c) : 1 \leq a, b, c \leq 10\}.$$

   Para cada componente tenemos ahora 10 posibles valores, de modo que el espacio tiene $10^3 = 1,000$ puntos. Como todos tienen la misma probabilidad, la respuesta en este caso es $1/1,000 = 0.001$.

3. Si lanzamos dos dados, ¿Cuál es la probabilidad de que la suma sea 7?

   Vamos a suponer, para facilitar el razonamiento, que lanzamos un dado primero y luego el otro. Por lo tanto un espacio muestral adecuado para este experimento es el conjunto de pares ordenados formados con los enteros del 1 al 6, con reposición:

   $$\Omega = \{(a, b), 1 \leq a, b \leq 6\}.$$

   En este caso todos los eventos elementales de Ω tienen la misma probabilidad: $1/36$. Los resultados que tienen componentes cuya suma es 7 son

   $$(1, 6); \quad (2, 5); \quad (3, 4); \quad (4, 3); \quad (5, 2); \quad (6, 1).$$

   Por lo tanto la probabilidad de que la suma de los dados sea 7 es

   $$6 \times \frac{1}{36} = \frac{1}{6}.$$

   En este ejemplo podemos considerar otro espacio muestral: el conjunto de las sumas posibles

   $$\Omega' = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

   El problema para usar este espacio como base para nuestro análisis es que sus elementos no son equiprobables. Por ejemplo, para tener una suma de 2, ambos dados tienen que salir 1, lo cual tiene probabilidad $1/36$, y acabamos de ver que la probabilidad de que la suma sea 6 es $1/6$.

4. Si lanzamos dos monedas, ¿cuál es la probabilidad de obtener un águila y un sol?

   Este problema lo hemos incluido para resaltar una dificultad importante que se ejemplifica con el razonamiento de D'Alembert, famoso matemático francés del siglo XVIII, quien argumentó que sólo hay tres casos posibles en esta situación:

   $$\text{(1) dos aguilas,} \qquad \text{(2) dos soles,} \qquad \text{(3) un águila y un sol,}$$

   y concluyó que la probabilidad de obtener una cara y un sello es 1/3. Como hemos visto, el último caso en realidad debe separarse en dos:

   (3a) La primera moneda es águila y la segunda es sol.

   (3b) La primera moneda es sol y la segunda es águila.

   Esto es obvio si lanzamos una moneda tras otra y no simultáneamente, o si las monedas son distinguibles. Por lo tanto la respuesta correcta es $2/4 = 1/2$. Hacemos una observación importante sobre este caso en la sección 1.7.

5. Si lanzamos una moneda dos veces y una de las veces sale águila ¿Cuál es la probabilidad de que el otro lanzamiento haya sido sol?

   Para este ejemplo el espacio muestral es

   $$\Omega = \{SS, SA, AS, AA\}$$

   y todos los resultados tienen igual probabilidad de ocurrir. Si sabemos que uno de los lanzamientos fue A, nos quedan tres resultados posibles y de ellos en dos casos el otro lanzamiento es S. Por lo tanto la probabilidad es 2/3.

   La situación sería distinta si nos dicen que el primer lanzamiento resultó A, pues en este caso el segundo tiene dos posibilidades A y S con igual probabilidad, y la respuesta en este caso sería que la probabilidad es 1/2.

1.5.2. Probabilidades en Espacios Numerables.

Un caso similar al desarrollado en la section anterior se presenta tomando como Ω un conjunto infinito numerable:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m, \dots\}, \quad \mathcal{A} = \mathcal{P}(\Omega) \quad \text{y} \quad P(A) = \sum_{\omega_i \in A} p_i,$$

donde los números p_i verifican

$$\begin{cases} p_i \geq 0 \quad \text{para todo } i \\ \displaystyle\sum_{i=1}^{\infty} p_i = 1 \end{cases}$$

Claramente en este caso no es posible que los p_i sean todos iguales, ya que de ser así no pueden satisfacer las condiciones anteriores. En el capítulo 3 consideraremos en más detalle estos espacios y los del ejemplo anterior.

Veamos un ejemplo.

1. Lanzamos una moneda hasta que salga 'Aguila' por primera vez. Los resultados posibles de este experimento son los números naturales: $\Omega = \mathbb{N}$. La probabilidad de obtener 'Aguila' en el primer lanzamiento es 1/2. La probabilidad de salga 'Sol' en el primer lanzamiento y 'Aguila' en el segundo es $(1/2) \times (1/2) = 1/4$. La probabilidad de tener 'Sol' dos veces y luego 'Aguila' es 1/8 y así sucesivamente. Vemos que la probabilidad de obtener 'Aguila' por primera vez en el n-ésimo lanzamiento es

$p_n = 1/2^n$. Tenemos que verificar que esta asignación define una probabilidad y para esto es necesario que

$$\sum_{n=1}^{\infty} p_n = 1.$$

Recordamos la fórmula para una serie geométrica:

$$1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r} \tag{1.3}$$

y multiplicando ambos lados por r obtenemos

$$r + r^2 + r^3 + r^4 + \cdots = \frac{r}{1-r} \tag{1.4}$$

para $-1 < r < 1$.

Si ponemos $r = 1/2$ en (1.4) obtenemos que la suma $\sum p_n$ vale 1. Además de comprobar que p_n define una probabilidad sobre Ω, este resultado muestra que con probabilidad 1 obtendremos un 'Aguila' en un número finito de lanzamientos, o equivalentemente, que la probabilidad de no obtener nunca 'Aguila' en una sucesión de lanzamientos de una moneda balanceada es 0.

Sea ahora A el evento 'la primera Aguila se obtiene en un número par de lanzamientos'. Tenemos que $A = \{2, 4, 6, \dots\}$ y

$$P(A) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots .$$

Poniendo $r = 1/4$ en la ecuación (1.4) obtenemos que

$$P(A) = \frac{1/4}{1 - 1/4} = \frac{1}{3},$$

de modo que la probabilidad de que la primera 'Aguila' salga en un número par de lanzamientos es $1/3$ y en un número impar, $2/3$.

1.5.3. Otros Ejemplos

(1) *Muestreo con Reposición.* Retomemos el ejemplo 1.2.9 sobre el muestreo con reposición, donde

$$\Omega = \{BBB, BBD, BDB, DBB, BDD, DBD, DDB, DDD\}$$

y sea $\mathcal{A} = \mathcal{P}(\Omega)$. Supongamos que la proporción de defectuosos en la población es $p = n/N$, donde n es el número de defectuosos en el total N de artículos en el stock. Por lo tanto, la proporción de buenos en la población es $1 - p = q$.

Consideremos el evento elemental $\{DDD\}$. Para asignarle la probabilidad correspondiente razonamos así: en cada una de las extracciones hay n formas posibles de elegir un defectuoso. En total resultan n^3 posibilidades de obtener los tres defectuosos y N^3 elecciones posibles de una terna cualquiera. Asignamos al evento $\{DDD\}$ la probabilidad

$$P(\{DDD\}) = \frac{n^3}{N^3} = p^3$$

y análogamente

$$\begin{gathered} P(\{BBB\}) = q^3, \\ P(\{BDD\}) = P(\{DDB\}) = P(\{DBD\}) = p^2 q, \\ P(\{BBD\}) = P(\{BDB\}) = P(\{DBB\}) = p q^2. \end{gathered}$$

Se verifica que

$$P(\Omega) = p^3 + 3p^2q + 3pq^2 + q^3 = (p+q)^3 = 1.$$

Calculemos la probabilidad del evento A: "se obtiene al menos un defectuoso en la muestra". Como A es el complemento del evento A^c: "no se obtiene ningún defectuoso en la muestra", resulta

$$P(A) = 1 - P(A^c) = 1 - q^3.$$

Consideremos ahora la siguiente situación que se presenta en problemas vinculados a control de calidad. Supongamos que se ignora la proporción p de defectuosos en la población y estamos interesados en tener una estimación de ese valor. Extraemos una muestra de tres artículos entre los cuales hay uno solo defectuoso.

Analicemos la probabilidad del evento: "se obtiene un solo defectuoso en la muestra", según diversos valores de p, como indica el cuadro siguiente:

p	$3pq^2$
0.1	0.243
0.2	0.384
0.3	0.441
0.4	0.432
0.5	0.375
0.6	0.288
0.7	0.189
0.8	0.096
0.9	0.027

Si tuviéramos que seleccionar uno de estos valores para p, una opción posible sería admitir aquél que haga mayor la probabilidad del evento que ocurrió efectivamente, o sea 0,3.

Utilizando este criterio, y aceptando como posibles valores de p todos los números reales entre 0 y 1, adoptamos como estimación aquél que haga máxima la probabilidad $3pq^2 = 3p(1-p)^2$ del evento que efectivamente ocurrió. Este criterio de estimación se llama de *"máxima verosimilitud"*. Para maximizar esta función

$$L(p) = 3p(1-p)^2$$

calculamos su derivada

$$L'(p) = 3(1-p)(1-3p)$$

que se anula en $p = 1,\ p = 1/3$.

El gráfico de la función $L(p)$ es el que se indica en la figura 2.4, y el máximo para $p \in [0,1]$ está en $p = 1/3$. Tomamos, por lo tanto, como estimación $\hat{p} = 1/3$, valor que obviamente se adecúa a lo que indica la intuición inmediata, dado que en la muestra de tres resultó uno defectuoso.

(2) *Error de Redondeo.* Consideremos nuevamente el caso del error de redondeo. Supongamos que se trunca el resultado de una operación aritmética en la parte entera, es decir, que en lugar del número real no negativo x se toma su parte entera $[x]$, esto es, el mayor entero que no supera x. El planteo es esencialmente el mismo si se trata del truncamiento en cualquier cifra decimal.

El error cometido al truncar es $x - [x]$, que podemos considerar como un evento elemental del intervalo $[0,1) = \Omega$, tomado como espacio muestral.

Con frecuencia – como veremos al examinar este problema mas adelante – estaremos interesados en asignar al espacio muestral Ω una probabilidad uniforme en el siguiente sentido: intervalos de igual longitud deben tener igual probabilidad. No es difícil probar que una tal probabilidad P debe verificar

$$P([a,b)) = b - a \tag{1.5}$$

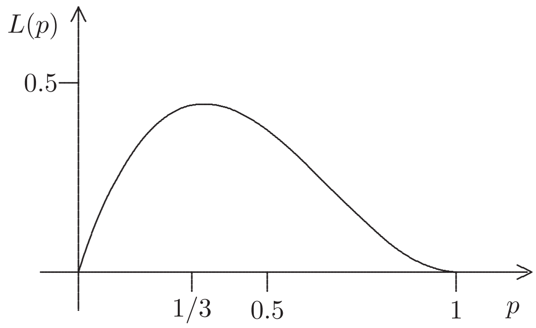


Figura 1.4

cualquiera que sea el intervalo $[a, b), 0 \leq a < b < 1$. Una manera de hacerlo es la siguiente: si P tiene esa propiedad, para n natural se tiene

$$P\left(\left[0, \frac{1}{n}\right)\right) = P\left(\left[\frac{1}{n}, \frac{2}{n}\right)\right) = \cdots = P\left(\left[\frac{n-1}{n}, 1\right)\right)$$

y como la suma de estas n probabilidades es $P(\Omega) = 1$ resulta que cada una de ellas es $1/n$.

Si m y n son enteros positivos, $m < n$, resulta que

$$P\left(\left[0, \frac{m}{n}\right)\right) = P\left(\left[0, \frac{1}{n}\right)\right) + \cdots + P\left(\left[\frac{m-1}{n}, \frac{m}{n}\right)\right) = \frac{m}{n}.$$

Si x es un número real cualquiera perteneciente al intervalo $(0, 1)$, consideremos dos sucesiones de números racionales

$$\frac{m_k}{n_k} < x < \frac{m'_k}{n'_k}, \quad \frac{m_k}{n_k} \to x, \quad \frac{m'_k}{n'_k} \to x, \quad k \to \infty$$

y se tiene

$$\frac{m_k}{n_k} = P\left(\left[0, \frac{m_k}{n_k}\right)\right) \leq P([0, x)) \leq P\left(\left[0, \frac{m'_k}{n'_k}\right)\right) = \frac{m'_k}{n'_k}.$$

Pasando al límite para $k \to \infty$ resulta

$$x \leq P([0, x)) \leq x \Rightarrow P([0, x)) = x,$$

o sea que la probabilidad de cada intervalo es su longitud.

Como familia de eventos podemos tomar la menor σ-álgebra que contiene a los intervalos, llamada σ-álgebra de Borel, que denotamos $\mathcal{B}$, y se puede probar que existe efectivamente una probabilidad P definida para la familia de eventos $\mathcal{B}$, que satisface (1.5), es decir, que asigna probabilidades iguales a intervalos de longitudes iguales.

Determinemos ahora la probabilidad del evento

$$A : \text{“ La primera cifra truncada es 9 ”}.$$

Resulta

$$P(A) = P([0.9\ ,\ 1)) = 1 - 0.9 = 0,1.$$

¿Cuál es la probabilidad de que la segunda cifra truncada sea 9?

Este evento es

$$B = [0.09,\ 0.1) \cup [0.19,\ 0.2) \cup \cdots \cup [0.99,\ 1)$$

y su probabilidad es

$$P(B) = \overbrace{\frac{1}{100} + \cdots + \frac{1}{100}}^{10 \text{ veces}} = 0.1.$$

(3) Un dado está cargado de modo tal que la probabilidad de que salga la cara k es proporcional a k. Hallar la probabilidad de cada uno de los eventos:

a. El resultado de arrojar el dado es un número par.

b. El resultado es menor que 6.

Denotemos por p_k la probabilidad de que ocurra la cara k ($k = 1, 2, 3, 4, 5, 6$). Lo que establece el enunciado es que existe una constante C tal que $p_k = Ck$. Como $p_1 + p_2 + \cdots + p_6 = 1$, se deduce que

$$C(1 + 2 + \cdots + 6) = 1 \quad \Rightarrow \quad 21C = 1 \quad \Rightarrow \quad C = \frac{1}{21} \quad \Rightarrow \quad p_k = \frac{k}{21}$$

Resolvamos ahora a y b.

a. La probabilidad de obtener una cara par es

$$p_2 + p_4 + p_6 = \frac{12}{21} = \frac{4}{7}.$$

b. La probabilidad de obtener un resultado menor que 6 es

$$p_1 + p_2 + p_3 + p_4 + p_5 = \frac{15}{21} = \frac{5}{7}.$$

◀

(4) *El problema de los cumpleaños.* ¿Cuál es la probabilidad de que entre r personas al menos dos cumplan años el mismo día? (Se supone que la duración del año es de 365 días). ¿Cuál es el menor valor de r para el cual esta probabilidad es superior a 1/2?

Tomamos como espacio muestral el conjunto de todas las r-uplas de fechas posibles:

$$\Omega = \{(f_1, f_2, \dots, f_r) \ : \ 1 \leq f_i \leq 365, \ i = 1, \dots, r\}.$$

y la hipótesis natural es que todas las r-uplas son igualmente probables.

Llamemos A el evento de que entre los r individuos seleccionados, no hay dos que cumplan el mismo día, es decir que

$$A = \{(f_1, \dots, f_r) : 1 \leq f_i \leq 365, \text{ los } f_i \text{ son diferentes 2 a 2}\}$$

La pregunta es cuál es la probabilidad de que no ocurra A, es decir

$$P(A^c) = 1 - P(A)$$

y como todos los eventos elementales de Ω son igualmente probables,

$$P(A) = \frac{\sharp A}{\sharp \Omega}.$$

Por los resultados del capítulo anterior con $N = 365$ obtenemos que

$$\sharp\Omega = N^k \qquad \sharp A = N(N-1)\cdots(N-r+1)$$

y por lo tanto

$$P(A^c) = 1 - \frac{N(N-1)\cdots(N-r+1)}{N^r} = 1 - \left(1 - \frac{1}{N}\right)\dots\left(1 - \frac{r-1}{N}\right).$$

Para acotar esta probabilidad utilizamos la desigualdad

$$1 - x \le e^{-x}$$

válida para todo $x \in \mathbb{R}$, que puede ser demostrada usando un desarrollo de MacLaurin de orden 2 o verificando que la figura 2.5 es correcta.

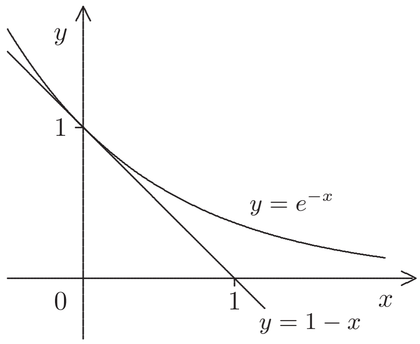


Figura 1.5

Obtenemos

$$P(A^c) > 1 - e^{\frac{1}{N}+\frac{2}{N}+\cdots+\frac{r-1}{N}} = 1 - e^{-\frac{r(r-1)}{2N}}$$

Para $r = 23$ y $N = 365$ obtenemos $P(A^c) > 0.50000175$. Así, en un grupo de 23 personas, con probabilidad mayor que 1/2, hay dos personas que cumplen años el mismo dia, lo cual es bastante sorprendente. Para $r = 30$ la probabilidad es superior a 0.696 y para $r = 50$, superior a 0.965. ◀

(5) Si la probabilidad de encontrar un artículo defectuoso en una población es $p = n/N$, donde N es el número de elementos de la población y n el de defectuosos, y realizamos muestreo con reposición extrayendo un artículo cada vez, calcular la probabilidad de encontrar el primer defectuoso en la m-ésima extracción. Si llamamos p_m a esta probabilidad, verificar que $\sum_{m=1}^{\infty} p_m = 1$.

Veremos ahora una solución al ejercicio con los elementos de que disponemos. Más adelante podremos tratarlo de manera más simple, utilizando conceptos que aún no hemos introducido. Comencemos por $m = 1$; p_1 es la probabilidad de extraer un defectuoso en la primera extracción, que es claramente

$$p_1 = \frac{n}{N} = p.$$

Sea ahora $m > 1$. El evento A_m: "el primer defectuoso es extraído en la m-ésima extracción", se escribe como

$$A_m = B_{m-1} \setminus B_m$$

donde B_m es el evento de que en las primeras m extracciones no hemos encontrado artículos defectuosos. La relación anterior expresa que "encontrar un defectuoso por primera vez en la m-ésima extracción" es lo mismo que "no extraer defectuosos en las $m - 1$ primeras pero si en las m primeras".

Como $B_m \subset B_{m-1}$ se tiene que $P(A_m) = P(B_{m-1}) - P(B_m)$. Por otra parte

$$P(B_m) = \frac{(N-n)^m}{N^m} = (1-p)^m$$

y, por lo tanto, deducimos que

$$p_m = P(A_m) = (1-p)^{m-1} - (1-p)^m \tag{1.6}$$

$$= (1-p)^{m-1}(1-(1-p)) = p(1-p)^{m-1}. \tag{1.7}$$

En resumen, la fórmula

$$p_m = p(1-p)^{m-1}$$

vale para todo $m \geq 1$. Además, como $p > 0$,

$$\sum_{m=1}^{\infty} p(1-p)^{m-1} = p\sum_{m=0}^{\infty}(1-p)^m = \frac{p}{1-(1-p)} = 1.$$

Aquí hemos usado la suma de la serie geométrica:

$$\sum_{m=0}^{\infty} x^m = \frac{1}{1-x}, \qquad \text{válida para } |x| < 1.$$

◀

1.6. La Paradoja de Bertrand.

En 1889 L. F. Bertrand propuso el siguiente problema: Tenemos un triángulo equilátero inscrito en un círculo. ¿Cuál es la probabilidad de que la longitud de una cuerda escogida al azar sea mayor que el lado del triángulo inscrito?

Este problema se sale de las situaciones que hemos estado considerando, pues no se trata de un problema sobre un espacio de probabilidad finito. Sin embargo, vamos a tratar de darle respuesta, intentando usar los mismos principios.

Primera Respuesta: Podemos pensar que la cuerda que vamos a seleccionar tiene un extremo fijo en el punto A y el otro extremo puede ser cualquier punto de la circunferencia. Sea ABC el triángulo inscrito y DAE la tangente a la circunferencia en el punto A.

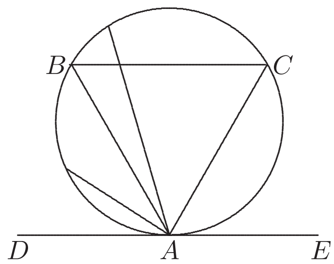


Figura 1.6

Cualquier cuerda que esté dentro del ángulo BAC de 60° es mayor que el lado del triángulo. Cualquier cuerda que esté dentro de alguno de los ángulos BAD o CAE es menor. Ambos ángulos miden también 60°. En resumen, las cuerdas que tienen un extremo fijo en A están en el ángulo DAE que mide 180°. De éstas, las que están dentro del ángulo BAC, que mide 60°, son mayores que el lado del triángulo, el resto son menores. Por lo tanto, la probabilidad buscada es

$$\frac{60}{180} = \frac{1}{3}.$$

Segunda Respuesta: Toda cuerda es perpendicular a un diámetro, que pasa por su punto medio. Podemos pensar que para seleccionar al azar una cuerda, podemos seleccionar inicialmente el diámetro al cual va a ser perpendicular, y luego escogiendo un punto del diámetro, tenemos el punto medio de la cuerda con lo cual ésta queda determinada. Supongamos que la cuerda que escogemos al azar es perpendicular al diámetro

AK, y sobre este diámetro dibujamos la altura del triángulo, como se ve en la figura 2.7. Es fácil mostrar que la distancia del centro del círculo a cualquier lado del triángulo es igual a la mitad del radio del círculo. En particular, OM es la mitad del radio OK, o también un cuarto del diámetro AK. Colocamos el punto N sobre el diámetro de modo que la distancia ON sea igual a OM. En la gráfica hemos dibujado con trazo discontinuo la cuerda paralela al lado BC del triángulo.

Es claro que las cuerdas que cortan al diámetro AK en algún punto del intervalo MN son mayores que el lado del triángulo. La cuerda aleatoria puede pasar por cualquier punto de AK, las que pasan por puntos en el intervalo MN, cuya longitud es la mitad de AK, son mayores que el lado del triángulo. Por lo tanto la probabilidad buscada es $1/2$.

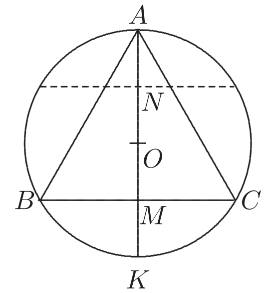


Figura 1.7

Tercera Respuesta: En la figura 2.8 hemos dibujado una circunferencia inscrita en el triángulo equilátero. Como hemos dicho en la solución anterior, el radio de esta circunferencia es la mitad del radio de la circunferencia original. En la figura observamos que si DE es una cuerda cuya distancia del centro es mayor que OM, entonces DE es más corta que BC mientras que si FG es una cuerda cuya distancia al centro es menor que OM, entonces FG es mayor que BC.

Observamos ahora que la distancia de una cuerda al centro de la circunferencia es en realidad la distancia del punto medio de la cuerda al centro de la circunferencia. La cuerda seleccionada al azar puede tener como punto medio a cualquier punto del círculo inicial, y los puntos medios de las cuerdas que son mayores que BC están en el círculo pequeño. En consecuencia la probabilidad de que la cuerda escogida al azar sea mayor que un lado del triángulo es el cociente entre las áreas del círculo pequeño y del círculo grande. Si llamamos r el radio del círculo pequeño, el cociente entre las áreas es

$$\frac{\pi r^2}{\pi(2r)^2} = \frac{\pi r^2}{4\pi r^2} = \frac{1}{4}.$$

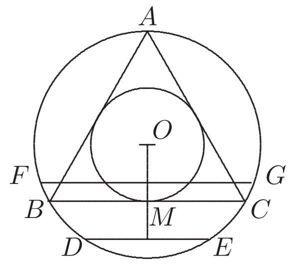


Figura 1.8

¡Hemos obtenido entonces tres respuestas distintas para el problema planteado por Bertrand! Esto parece paradójico, pero no lo es. El punto es que el planteamiento del problema es impreciso: ¿Qué quiere decir escoger una cuerda al azar? Hemos dado tres interpretaciones distintas, en primer lugar suponemos que escoger una cuerda al azar consiste en fijar un extremo de la cuerda y luego seleccionar el ángulo que hace la cuerda con la tangente a la circunferencia en el punto escogido, de manera que todos los ángulos son igualmente probables. En segundo lugar, escogimos un diámetro y sobre él, de manera uniforme, un punto, que es el punto medio de la cuerda. En tercer lugar escogemos un punto del círculo de manera uniforme. Este punto es el punto medio de la cuerda. En cada ocasión estamos considerando un espacio de probabilidad muestral distinto y una probabilidad distinta para medir los sucesos que nos interesan. Siendo así, no es sorprendente que hayamos obtenido respuestas distintas.

La paradoja de Bertrand nos señala el riesgo de usar con demasiada libertad la expresión 'al azar'. Situaciones como esta, en las cuales parece haber un problema de ambigüedad e incertidumbre, tuvieron un efecto negativo en el desarrollo de la Teoría de Probabilidades.

1.7. Comentarios y Algo de Historia.

1.- En primer lugar mencionamos a d'Alembert, quien apareció en el problema 4 de la sección 1.5.1, en relación a un error que hoy parece elemental. Jean le Rond d'Alembert nació en París el 16 de noviembre de 1717, murió en esa misma ciudad el 29 de octubre de 1783, y fue uno de los grandes matemáticos del siglo XVIII. Su error no es difícil de entender: si lanzamos dos monedas idénticas, sólo podemos distinguir tres resultados, los que mencionó d'Alembert. Es un poco mas difícil ver que no son igualmente probables, y hay que asignarles una probabilidad distinta.

Llama la atención que no haya habido ningún intento de verificación experimental de su parte. Después de todo, estamos tratando de hacer un modelo matemático de una situación real y bastan unas cuantas repeticiones del experimento para darse cuenta que los tres resultados posibles no ocurren con igual frecuencia.

Hay una situación física, sin embargo, que corresponde al modelo propuesto por d'Alembert. Para explicarla vamos a considerar una situación equivalente: la repartición de fichas en cajas. En el problema de repartir k fichas en n cajas podemos considerar que las cajas están identificadas por números (o letras o símbolos cualesquiera) y cada ficha asignada a una caja es equivalente a etiquetar la ficha con el número de la caja. Lanzar dos monedas es equivalente a colocar dos fichas en dos caja, una llamada *Aguila* y otra llamada *Sol*.

En este esquema, una repartición de k fichas en n cajas equivale a tomar una muestra de tamaño k de los números $1, 2, \ldots, n$ y esto podemos hacerlo de acuerdo a varios tipos de condiciones, por ejemplo, con o sin reposición, permitiendo que en cada caja haya cualquier número de fichas o a lo sumo una, considerando que las fichas son distinguibles o no, etc. Si no hay restricciones en el número de fichas que puede haber en cada caja, las fichas son distinguibles y hacemos el muestreo con reposición hay n^k muestras posibles. En la física de partículas, si las cajas son niveles de energía y las fichas son partículas, la hipótesis de Maxwell y Boltzmann es que estos n^k arreglos son igualmente probables.

Si no podemos colocar más de una ficha en cada caja, el número de arreglos ordenados es V_k^n. Si no consideramos el orden de las fichas hay $\binom{n}{k}$ y para el caso de las partículas y niveles de energía, la hipótesis de Fermi y Dirac es que estos arreglos son igualmente probables.

Una tercera situación permite un número ilimitado de fichas en cada caja pero sin distinguir las fichas. Es el caso propuesto por d'Alembert para los dos lanzamientos de una moneda. Hay ahora $\binom{n+k-1}{k}$ arreglos posibles y la hipótesis de Bose y Einstein dice que son equiprobables para el caso de partículas y niveles de energía.

2.- Nuestra segunda mención es para Gerolamo Cardano (1501 - 1576), quien es famoso por su contribución a la solución de la ecuación cúbica. Cardano, quien era médico de profesión, creía firmemente en la Astrología y se dice que hizo el horóscopo de Eduardo VI de Inglaterra cuando este tenía dieciseis años, concluyendo que el Rey viviría por encima del promedio, aunque después de los 55 años era muy probable que sufriera de varias enfermedades. Eduardo VI murió poco después de que Cardano hiciese su horóscopo.

Cardano murió en Roma el 21 de septiembre, tres días antes de cumplir 75 años, tal como había predicho, y se dijo que había dejado de comer para asegurarse de que la predicción de su propia muerte fuese correcta.

Cardano escribió el primer libro sobre juegos de azar, alrededor de 1550 pero publicado sólo en 1663: *De ludo alea* (el libro de los juegos de azar), que puede ser descrito como un manual para jugadores. En su libro usa con frecuencia la definición clásica de probabilidades que formulamos en la sección 1.5.1 y que atribuimos a Laplace.

3.- El nacimiento de la Teoría de Probabilidades se asocia usualmente con la correspondencia entre Pierre de Fermat (1601 - 1665) y Blaise Pascal (1623 - 1662). Como partero funjió Antoine Gombauld, Chevalier de Méré, un noble francés con interés en los juegos de azar y las apuestas. De Méré consultó a Pascal, quien escribió a Fermat, comenzando una correspondencia que ejercería una profunda influencia en el desarrollo posterior de las probabilidades. En las palabras de Poisson: "*un problema sobre juegos de azar, planteado a un austero Jansenista* [1] *por un hombre de mundo, fue el origen del cálculo de probabilidades.*"

Pierre de Fermat fue abogado de profesión y magistrado en la corte de Toulouse. Hizo contribuciones importantes en geometría, teoría de números y en los orígenes del cálculo, pero es conocido principalmente por el '*último teorema de Fermat*': la ecuación $x^n + y^n = z^n$ no tiene soluciones enteras (x, y, z) para $n \geq 3$. Este teorema ha sido demostrado recientemente, tras el esfuerzo de numerosos matemáticos, por A. Wiles.

Veamos un ejemplo del tipo de problemas que consideraron en su correspondencia. En una carta de 1.654 Pascal escribe a Fermat sobre un problema planteado por De Méré:

"... no tengo tiempo de enviarle la explicación del problema que M. De Méré encontró difícil. Es muy inteligente, pero no es un geómetra (lo cual, como usted sabe, es un grave defecto) y no puede siquiera comprender que una recta matemática es infinitamente divisible: está convencido de que una recta está compuesta de un número finito de puntos, y nunca he podido convencerlo de lo contrario. Si usted puede lograrlo, le hará un gran servicio.

Me decía haber encontrado algo erróneo con los números porque:

Si uno trata de obtener un seis con un dado, la ventaja de hacerlo en cuatro lanzamientos es como 671 a 625. Si uno trata de obtener un doble seis con dos dados, es desventajoso hacerlo en 24 lanzamientos. Sin embargo 24 es a 36 (que es el número de caras de dos dados) como 4 es a 6 (que es el número de caras de un dado).

Esto le pareció asombroso y dijo en voz alta que las proposiciones no eran consistentes y la aritmética era contradictoria; pero usted seguramente es lo suficientemente instruido para reconocer la falla en su razonamiento."

Este problema estaba basado en un juego de moda en la época, en el cual la casa apuesta pagando uno a uno, a que un jugador lance al menos un seis en cuatro lanzamientos de un dado. Como dice Pascal, este juego es ligeramente favorable a la casa en la proporción 671 a 625. Para ver esto observamos que la probabilidad de que haya al menos un 6 en cuatro lanzamientos de un dado es

$$1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} = \frac{671}{1296} = 0.517$$

El problema que preocupaba a De Méré se refería a un juego similar, lanzando un par de dados: Por qué no es favorable a la casa apostar que el jugador obtendrá al menos un doble seis en 24 lanzamientos de un par de dados? Su razonamiento fue el siguiente: si lanzo un dado hay seis resultados posibles, mientras que si lanzo dos hay 36 $(= 6 \times 6)$. Por lo tanto, si con cuatro lanzamientos el juego con un dado es favorable, con $6 \times 4 = 24$ lanzamientos el juego con dos dados también debe serlo. En realidad, la probabilidad de que haya al menos un doble seis en 24 lanzamientos de dos dados es

$$1 - \left(\frac{35}{36}\right)^2 4 = 0.4913.$$

Si p representa la probabilidad de un resultado favorable en un lanzamiento, $1 - (1 - p)^n$ representa la probabilidad de obtener al menos un resultado favorable en n lanzamientos. Como vemos en esta expresión, dividir p por un número no es equivalente a multiplicar el exponente n por ese mismo número.

4.- Pierre Simon Laplace nació en Normandía el 23 de marzo de 1749 y vivió a través de la Era Napoleónica. En una época que muchos consideran la edad de oro de la ciencia francesa, Laplace fue el científico más ilustre de Francia. Al morir, el 5 de marzo de 1827, Poisson lo alabó como *"el Newton de Francia"*. Elegido a la Academia de Ciencias en 1773, fue profesor en la Escuela Militar, de la Escuela Normal, Ministro del Interior (aunque sólo por seis semanas, antes de ser reemplazado por el hermano de Napoleón) y Canciller del Senado. Fue nombrado Marqués por Luis XVIII.

[1] doctrina de Jansenio, que exageraba las ideas de San Agustín.

Sus principales intereses fueron la astronomía y las probabilidades, lo cual se refleja en sus dos obras fundamentales: *Tratado de Mecánica Celeste* (cuatro volúmenes, 1799 - 1805) y *Teoría Analítica de Probabilidades* (1812). Laplace creía en el determinismo de los sistemas físicos y por lo tanto pensaba que no puede haber probabilidad en el mundo material. La probabilidades surgen de nuestro conocimiento y nuestra ignorancia. La teoría del azar *"consiste en reducir todos los sucesos del mismo tipo a un cierto número de casos igualmente posibles, es decir, aquellos sobre los cuales estamos igualmente indecisos sobre su existencia"*.

5.- La definición clásica no es original de Laplace. Leibniz la menciona en 1678 y algunos piensan que es el primero en usarla. Gottfried Wilhelm Leibniz (1646 - 1716), filósofo racionalista y matemático alemán, tuvo un gran interés por las probabilidades, publicó en 1666 la primera monografía sobre la teoría de combinaciones (*Ars Combinatoria*) y aunque no hizo ninguna contribución formal de importancia a la teoría de probabilidades, fue el primero en intentar su axiomatización. Fue un testigo de excepción del surgimiento de las probabilidades y conoció a todos los protagonistas, excepto a Pascal.

6.- La definición clásica aparece también en 1705 en los trabajos de Jacques Bernoulli. Los Bernoulli son, sin duda, la familia mejor conocida en la historia de las matemáticas. Unos doce de ellos han hecho contribuciones en matemáticas y al menos cinco trabajaron en probabilidades. Jacques (también conocido como Jacob o James) nació en Basilea, Suiza, el 27 de diciembre de 1654. Para 1684 Jacques, y su hermano Jean (también conocido como Johann o John) habían desarrollado por su cuenta el cálculo diferencial, a partir de indicaciones publicadas por Leibniz, y eran reconocidos como matemáticos de primera línea. Posteriormente trabajaron sobre cálculo integral, curvas y problemas de minimización. Jacques fue profesor de la Universidad de Basilea desde 1687 hasta su muerte, mientras que su hermano fue profesor en Groninga desde 1695 hasta 1705, año en que reemplazó a su hermano como profesor en Basilea.

Los hermanos Bernoulli no fueron colaboradores, mas bien fueron rivales y en sus últimos años no tuvieron ningún contacto directo. Al morir Jacques en 1705, dejó una cantidad de trabajos sin publicar, algunos incompletos, en diversos temas de matemáticas, que fueron editados por su sobrino Nicolás y publicados en 1713.

El más importante de ellos trataba sobre probabilidades y fue publicado bajo el nombre de *Ars Conjectandi*. En su *Ensayo Filosófico sobre las Probabilidades*, Laplace incluye una *Nota Histórica Sobre el Cálculo de Probabilidad* en la cual dice lo siguiente sobre el libro de Bernoulli:

"... En este trabajo Bernoulli avanza la teoría de probabilidades mucho más de lo que lo hizo Huygens: da una teoría general de combinaciones y series y las aplica a varios problemas difíciles relacionados con el azar. El trabajo también es notable por la precisión y sutileza de sus observaciones, y por su aplicación de la fórmula binomial a este tipo de problemas, así como por la demostración de un teorema que dice que, si aumentamos, ilimitadamente, el número de observaciones y experimentos, el cociente de los diversos resultados tiende al cociente de sus respectivas probabilidades, y si hacemos suficientemente grande el número de experimentos, este cociente se acerca tanto como queramos al cociente de las probabilidades. Este teorema es muy útil para deducir a partir de observaciones, las leyes y causas asociadas con diversos fenómenos. Bernoulli, con razón, consideró la demostración de este teorema como de gran importancia, y dice haberla pensado durante un período de veinte años. "

Ejercicios

1. Sean A_1, A_2 y A_3 eventos de un espacio muestral. Expresar mediante uniones, intersecciones y complementos los siguientes eventos:

   a. Los tres eventos ocurren.
   b. Ocurre sólo A_1.
   c. Ocurren A_1 y A_2, pero no A_3.
   d. Ocurre al menos uno de los tres eventos.
   e. No ocurre ninguno.
   f. Ocurren al menos dos.
   g. Ocurren dos y no más.

2. Expresar como uniones disjuntas a) $A_1 \cup A_2$; b) $A_1 \cup A_2 \cup A_3$; c) $\bigcup_{i=1}^{n} A_i$.

3. Sea $\Omega = \{AAA, AAS, ASA, SAA, ASS, SAS, SSA, SSS\}$. Describa con palabras los siguientes eventos y calcule sus probabilidades:

a. $B = \{AAA, AAS, ASA, ASS\}$;
b. $C = \{AAA, SSS\}$.
c. $D = \{AAS, ASA, SAA\}$;
d. $E = \{AAS, ASA, SAA, ASS, SAS, SSA\}$.

4. El evento $A \setminus B$ quiere decir que A ocurre pero B no. Demuestre que las operaciones de unión, intersección y complemento se pueden expresar usando sólo esta operación.

5. Demuestre que para cualesquiera conjuntos $A,\ B,\ C$ y D, la siguiente proposición es cierta:
$$(A \cup B) \setminus (C \cup D) \subseteq (A \setminus C) \cup (B \setminus D).$$

6. Sean A y B dos eventos en un espacio de probabilidad. Demuestre que $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$.

7. Sean A y B eventos con probabilidades $P(A) = 3/4$ y $P(B) = 1/3$. Demuestre que $1/12 \leq P(A \cap B) \leq 1/3$, y dé ejemplos que demuestren que los extremos se pueden alcanzar. Halle las cotas correspondientes para $P(A \cup B)$.

8. Sea A y B dos eventos en un espacio de probabilidad. Usando $P(A)$, $P(B)$ y $P(A \cap B)$ exprese las siguientes probabilidades: (a) Para $k = 0, 1, 2$, P(exactamente k de los eventos A y B ocurren). (b) Para $k = 0, 1, 2$, P(al menos k de los eventos A y B ocurren). (c) Para $k = 0, 1, 2$, P(a lo sumo k de los eventos A y B ocurren).

9. Sea A, B y C tres eventos en un espacio de probabilidad. Usando $P(A)$, $P(B)$, $P(C)$, $P(A \cap B)$, $P(A \cap C)$, $P(B \cap C)$ y , $P(A \cap B \cap C)$ exprese las siguientes probabilidades: (a) Para $k = 0, 1, 2, 3$, P(exactamente k de los eventos A y B ocurren). (b) Para $k = 0, 1, 2, 3$, P(al menos k de los eventos A y B ocurren). (c) Para $k = 0, 1, 2, 3$, P(a lo sumo k de los eventos A y B ocurren).

10. Evalúe las probabilidades mencionadas en el ejercicio 8 si (a) $P(A) = P(B) = 1/3$, $P(A \cap B) = 1/6$. (b) $P(A) = P(B) = 1/3$, $P(A \cap B) = 1/9$. (c) $P(A) = P(B) = 1/3$, $P(A \cap B) = 0$.

11. Evalúe las probabilidades mencionadas en el ejercicio 9 si (a) $P(A) = P(B) = P(C) = 1/3$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/9$, $P(A \cap B \cap C) = 1/27$. (b) $P(A) = P(B) = P(C) = 1/3$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = 0$.

12. Sean $A,\ B$ y C tres eventos. Demuestre las siguientes propiedades:
    a. $P(A \cap B) + P((A \setminus B) \cup (B \setminus A)) + P(A^c \cap B^c) = 1$.
    b. $P(A^c \cap B^c) + P(A) + P(A^c \cap B) = 1$.
    c. $P(A^c \cap B^c \cap C^c) + P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C) = 1$.
    d. $P(A \cap B) - P(A)P(B) = P(A^c \cap B^c) - P(A^c)P(B^c)$.

13. Suponga que $P(A) \geq 0.9,\ P(B) \geq 0.8$ y $P(A \cap B \cap C) = 0$, demuestre que $P(C) \leq 0.3$.

14. Demuestre que si $A \cap B \cap C = \emptyset$, entonces
$$P((A \cup B) \cap (B \cup C) \cap (C \cup A)) = P(A \cap B) + P(B \cap C) + P(C \cap A).$$

15. Sea D el evento 'exactamente uno de los eventos $A,\ B$ y C ocurre'. Exprese $P(D)$ en términos de $P(A), P(B), P(C), P(A \cap B), P(A \cap C), P(B \cap C)$ y $P(A \cap B \cap C)$.

16. Demuestre que
    a. $\min\{1, P(A) + P(B)\} \geq P(A \cup B) \geq \max\{P(A), P(B)\}$
    b. $\min\{P(A), P(B)\} \geq P(A \cap B) \geq \max\{0, P(A) + P(B) - 1\}$
    c. $P(\cap_1^n A_i) \geq \sum_1^n P(A_i) - (n-1)$.

17. La condición de σ-aditividad para una medida de probabilidad es equivalente a otras propiedades. Pruebe que es equivalente a las proposiones (a) y (b):

    a. Si $A_1 \subseteq A_2 \subseteq \cdots$ es una sucesión creciente de eventos y $A = A_1 \cup A_2 \cup \cdots$ entonces $P(A) = \lim_{n\to\infty} P(A_n)$.

    b. Si $A_1 \supseteq A_2 \supseteq \cdots$ es una sucesión decreciente de eventos y $A = A_1 \cap A_2 \cap \cdots$ entonces $P(A) = \lim_{n\to\infty} P(A_n)$.

18. Una caja contiene n bolas rojas y n bolas blancas. Se extraen dos bolas al azar. ¿Cuál es el espacio muestral para este experimento? ¿Cuál es la probabilidad de que las dos bolas tengan colores distintos. Halle la probabilidad p_n de que las bolas sean del mismo color y evalúe $\lim_{n\to\infty} p_n$.

19. Una caja contiene 10 bolas negras y 5 bolas rojas. Se extraen 3 bolas al azar, con reposición. Calcular:

    a. La probabilidad de que sean 2 negras y una roja.

    b. La probabilidad de que sean las tres negras.

    c. Repetir el ejercicio anterior suponiendo que la extracción es sin reposición.

20. Se extraen dos cartas sucesivamente de un juego de 52 cartas. Halle la probabilidad de que la segunda carta sea mayor que la primera.

21. Se lanzan al aire simultáneamente tres monedas balanceadas. Calcular:

    a. La probabilidad de obtener 3 caras.

    b. La probabilidad de obtener por lo menos 2 caras.

22. Lanzamos una moneda balanceada cuatro veces. Calcule la probabilidad de los siguientes eventos:

    a. Ocurren al menos tres Aguilas.

    b. Ocurren exactamente tres Aguilas.

    c. Ocurren al menos tres Aguilas consecutivas.

    d. Ocurren exactamente tres Aguilas consecutivas.

23. Se lanzan dos dados. Calcule la probabilidad de los siguientes eventos: a) obtenemos el mismo número en ambos dados; b) la suma es 7 u 11; c) los números son primos relativos, d) la suma es impar; e) el producto es impar; f) un número divide al otro.

24. Se realiza un test de conocimientos con 11 preguntas a contestar por *sí* o *no*. Se da por aprobada la prueba si se contestan correctamente al menos 6 de las 11 preguntas. ¿Cuál es la probabilidad de aprobar el examen contestando al azar?

25. Sean P_1, P_2 dos medidas de probabilidad definidas sobre la misma σ-álgebra $\mathcal{F}$ y sea $0 \leq \alpha \leq 1$. Demuestre que $\alpha P_1 + (1-\alpha)P_2$ también es una medida de probabilidad sobre $\mathcal{F}$. Generalice el resultado a n medidas de probabilidad.

26. a. Sea $p_i = a/i^2$ para $i \in \mathbb{N}$. Halle el valor de a para que p_i defina una probabilidad.

    b. Sea $p_i = b/i^2$ para $i = \pm 1, \pm 2, \ldots$. Halle el valor de b para que p_i defina una probabilidad.

27. Para comenzar un cierto juego es necesario lanzar un 6 con un dado.

    a) ¿Cuál es la probabilidad de lanzar el primer 6 en el tercer intento?

    b) ¿Cuál es la probabilidad de necesitar más de tres intentos?

    c) ¿Cuántos lanzamientos hacen falta para que la probabilidad de haber lanzado un 6 sea al menos 0.95?

    d. ¿Cuál es la probabilidad de que el primer 6 ocurra en un número par de lanzamientos?

28. Sea $\mathcal{F}$ una σ-álgebra de eventos en un espacio finito. Demuestre que $\mathcal{F}$ no puede contener exactamente 6 eventos. ¿Qué enteros pueden ser el cardinal de $\mathcal{F}$?

29. Sean: $\Omega = [0,1]$, $\mathcal{B}$ la familia de conjuntos de Borel y P la probabilidad definida en el ejemplo 6 de la sección 2.4.

    a. Probar que $P(\{\omega\}) = 0$, donde $\{\omega\}$ es el subconjunto de Ω que consta sólo del punto ω. (Verificar previamente que $\{\omega\} \in \mathcal{B}$).

    b. Sean $Q = \{\omega : \omega \in [0,1] \text{ es racional}\}$ e $I = \{\omega : \omega \in [0,1] \text{ es irracional}\}$. Probar que $P(Q) = 0$ y $P(I) = 1$.

30. Se lanza reiteradamente una moneda balanceada. ¿Cuál es la probabilidad de que ocurran 4 caras antes que dos sellos?

31. Se lanzan cuatro dados y se multiplican los números que se obtienen. ¿Cuál es la probabilidad de que este producto sea divisible por 5? ¿Cuál es la probabilidad de que el último dígito en el producto sea 5?

32. Antonio y Bruno acuerdan una competencia de esgrima en una serie de mangas de modo que el primero en ganar dos mangas seguidas gana el combate. Antonio tiene probabilidad p de ganar una manga y Bruno probabilidad $q = 1 - p$. ¿Cuál es la probabilidad de que la competencia termine al cabo de k mangas?

33. En una caja tenemos n bolas con los números del 1 al n. Sea D_r el evento: 'se extrae una bola al azar y el número es divisible por r'. Halle $P(D_3)$, $P(D_4)$, $P(D_3 \cup D_4)$ y $P(D_3 \cap D_4)$ y obtenga los límites de estas probabilidades cuando $n \to \infty$.

34. Definimos la función d sobre $\mathcal{F} \times \mathcal{F}$ por $d(A,B) = P(A\Delta B)$.

    a. Demuestre que para cualesquiera eventos A, B y C

    $$d(A,B) + d(B,C) - d(A,C) = 2P(A \cap B^c \cap C) + P(A^c \cap B \cap C^c)$$

    b. ¿Cuándo vale $d(A,B) = 0$?

    c. Sea $A_1, A_2, \dots$ una sucesión no-decreciente de eventos: $A_i \subseteq A_j$ para $i \leq j$. Demuestre que para $i \leq j \leq k$,

    $$d(A_i, A_k) = d(A_i, A_j) + d(A_j, A_k).$$

35. ¿Cuándo son ciertas las siguientes relaciones?

    a. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

    b. $A \cap (B \cap C) = (A \cap B) \cap C$

    c. $A \cup (B \cup C) = A \setminus (B \setminus C)$

    d. $(A \setminus B) \setminus C = A \setminus (B \setminus C)$

    e. $A\Delta(B\Delta C) = (A\Delta B)\Delta C$

    f. $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

    g. $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

36. Sea $\mathcal{F}$ una σ-álgebra de subconjuntos de Ω y sea $B \in \mathcal{F}$. Demuestre que $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ también es una σ-álgebra de subconjuntos de Ω.

37. Sea $A_n, n \geq 1$ una sucesión de eventos con $P(A_n) = 1$ para todo n. Demuestre que $P(\cap_{n=1}^{\infty} A_n) = 1$.

38. En el juego de 'craps' el jugador lanza dos dados. Si el resultado es 7 u 11, el jugador gana. Si es 2, 3 ó 12, pierde. Si es cualquier otro resultado k, continua lanzando hasta obtener un 7, en cuyo caso pierde, o k, en cuyo caso gana. ¿Cuál es un espacio muestral adecuado para este juego? ¿Cuál es la probabilidad de ganar? ¿Cuál es la probabilidad de ganar en el primero o segundo lanzamiento? ¿Cuál es la probabilidad de ganar si el primer lanzamiento es 6?

39. Sea $A_n, n \geq 1$ una sucesión de eventos. Definimos

$$B_n = \bigcup_{m=n}^{\infty} A_m, \qquad C_n = \bigcap_{m=n}^{\infty} A_m.$$

Es claro que $C_n \subseteq A_n \subseteq B_n$. Las sucesiones (B_n) y (C_n) son decreciente y creciente, respectivamente, con límites

$$\lim B_n = B = \bigcap_{n\geq 1} B_n = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m, \qquad \lim C_n = C = \bigcup_{n\geq 1} C_n = \bigcup_{n\geq 1} \bigcap_{m\geq n} A_m$$

El evento B se conoce como el *límite superior* de la sucesión (A_n) y se denota $\limsup A_n$ mientras que C se conoce como el *límite inferior* de la sucesión (A_n) y se denota $\liminf A_n$. Demuestre que (a) $B = \{\omega \in \Omega : \omega \in A_n$ para infinitos valores de $n\}$. (b) $C = \{\omega \in \Omega : \omega \in A_n$ para todos los valores de n salvo por un número finito $\}$.

40. Con las definiciones del ejercicio anterior, decimos que la sucesión (A_n) converge a $A = \lim_n A_n$ si $B = C(= A)$. Suponga qu esto es ciertoy demuestre (a) A es un evento, es decir, $A \in \mathcal{F}$, (b) $P(A_n) \to P(A)$.

Capítulo 2

Teoría Combinatoria

La Teoría Combinatoria es la rama de las matemáticas que se ocupa del estudio de las formas de contar. Aparte del interés que tiene en sí misma, la combinatoria tiene aplicaciones de gran importancia en otras áreas, y en particular a la Teoría de Probabilidades.

2.1. Dos Principios Básicos.

Comencemos por considerar algunos problemas sencillos.

Problema 1. *En una tienda hay cinco modelos de camisa y tres de pantalón. ¿Cuántos conjuntos distintos de pantalón y camisa podemos comprar?*

► La camisa la podemos elegir de cinco maneras distintas. Para cada una de ellas podemos escoger el pantalón de tres maneras distintas. Por lo tanto hay $5 \times 3 = 15$ maneras de escoger un pantalón y una camisa. ◄

Problema 2. *Las ciudades A, B, y C están conectadas según lo muestra la figura 2.1: hay seis caminos de A a B y cuatro de B a C. ¿De cuántas maneras podemos ir de A a C?*

► Para cada camino que escojamos entre A y B podemos escoger cuatro para continuar hasta C. Como hay seis caminos entre A y B la respuesta es $6 \times 4 = 24$.

Problema 3. *El conjunto $A = \{a_1, a_2, \dots, a_k\}$ tiene k elementos mientras que $B = \{b_1, b_2, \dots, b_n\}$ tiene n. ¿Cuántos elementos tiene el producto cartesiano $A \times B$?*

► El producto cartesiano $A \times B$ está formado por todos los pares ordenados (a, b) donde el primer elemento, a, está en A y el segundo, b, está en B. Para cada uno de los k elementos de A que tomemos como primer miembro del par hay n posibilidades para escoger el segundo a partir de los elementos de B. Por lo tanto tendremos $k \times n$ pares ordenados. ◄

Los tres problemas anteriores tienen características similares: Se trata de escoger dos elementos, cada uno de un conjunto distinto y queremos contar el número de maneras de hacer esto. El resultado general puede enunciarse de la siguiente manera:

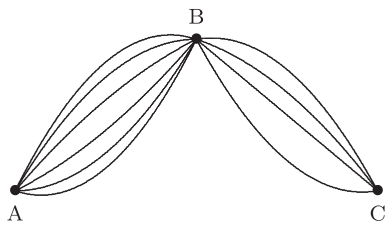


Figura 2.1

Principio de Multiplicación. *Si tenemos dos conjuntos de k y n elementos, respectivamente, y queremos escoger dos elementos de modo que uno sea del primero y el otro del segundo, esto lo podemos hacer de $k \times n$ maneras.*

El principio de multiplicación puede ser aplicado reiteradamente:

Problema 4. *En la tienda del problema 1 hay también cuatro modelos distintos de zapatos. ¿De cuántas maneras podemos escoger un conjunto de camisa, pantalón y zapatos?*

▶ Podemos ahora comenzar con cualquiera de los 15 conjuntos de camisa y pantalón del problema 1. Hay cuatro maneras de completarlo escogiendo un par de zapatos. Por lo tanto el número de posibles conjuntos de camisa, pantalón y zapatos es $15 \times 4 = 60$. ◀

Problema 5. *Una costurera tiene tres botones, cinco agujas y ocho tipos de hilo. ¿De cuántas maneras puede escoger un objeto de cada tipo?*

▶ $3 \times 5 \times 8 = 120$. ◀

Veamos ahora otro tipo de problema.

Problema 6. *Si además de las ciudades A, B y C del problema 2 tenemos una cuarta ciudad D conectada con las anteriores de la manera que indica la figura 1.2, ¿De cuántas maneras podemos ahora viajar de A a C?*

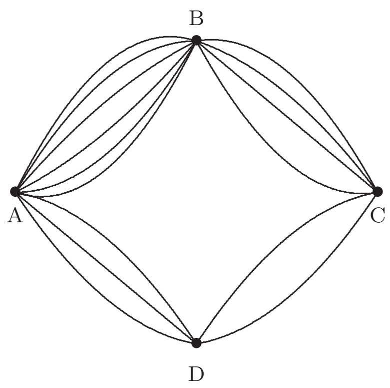


Figura 2.2

▶ Podemos ir de A a C pasando por B o por D. Sabemos por el problema 2 que hay 24 maneras de ir de A a C pasando por B. Por el Principio de Multiplicación hay $3 \times 2 = 6$ maneras de ir de A a C pasando por D. Por lo tanto, en total hay $24 + 6 = 30$ maneras de viajar de A a C. ◀

Problema 7. *Una persona visita dos tiendas con intención de comprar un pantalón. En la primera tienda hay seis modelos diferentes y para cada uno hay tres colores. En la segunda hay diez modelos y cuatro colores para cada modelo. ¿Entre cuantos pantalones tiene que escoger la persona?*

▶ En la primera tienda hay $6 \times 3 = 18$ mientras que en la segunda hay $10 \times 4 = 40$. Para hallar el total de pantalones tenemos que sumar estos dos números, y obtenemos $18 + 40 = 58$. ◀

Vemos que en ambos problemas hay dos situaciones que son excluyentes: Para ir de A a C pasamos por B o por D, pero no por ambos. El pantalón lo compramos en la primera tienda o en la segunda, pero no en ambas. Cuando se presenta una situación de este tipo, el número total de soluciones se obtiene sumando las soluciones bajo las distintas alternativas. Este resultado se puede enunciar de la siguiente manera:

Principio de Suma. *Si una situación puede ocurrir de k maneras distintas y una segunda situación excluyente de la primera puede ocurrir de n maneras, entonces existen $k + n$ maneras en las cuales puede ocurrir la primera o la segunda situación.*

El principio de suma también puede ser aplicado reiteradamente.
Problema 8. *En una tienda hay cinco modelos de pantalón, ocho de camisa y cuatro de zapatos. ¿Cuántas maneras hay de comprar dos objetos con nombres distintos?*

▶ Hay tres casos posibles: Compramos pantalón y camisa; pantalón y zapatos o camisa y zapatos. Es fácil calcular el número de maneras de cada caso: $5 \times 8 = 40$ para el primero, $5 \times 4 = 20$ para el segundo y $8 \times 4 = 32$ para el tercero. En total hay $40 + 20 + 32 = 92$ maneras de comprar dos objetos con nombres distintos. ◀

Problema 9. *¿Cuántos números de a lo sumo tres cifras se pueden formar con los dígitos 3, 4, 7 y 8?*

▶ Los números que vamos a formar pueden tener una, dos o tres cifras. Veamos por separado cuantos hay de cada tipo y luego sumamos los resultados, de acuerdo al principio de la suma. Es claro que de una cifra hay 4. En el caso de dos cifras la primera puede ser cualquiera de los cuatro dígitos, y la segunda también. Por lo tanto hay $4 \times 4 = 16$ números de dos cifras. De manera similar, hay $4 \times 4 \times 4 = 64$. En total tenemos $4 + 16 + 64 = 84$ números de tres o menos cifras formados con los dígitos 3, 4, 7 y 8. ◀

2.2. Número de subconjuntos de un conjunto finito.

Sea $C = \{c_1, c_2, \dots c_n\}$ un conjunto de n elementos. Denotaremos por $\mathcal{P}(C)$ la familia de todos los subconjuntos de C y lo llamaremos el *conjunto de partes* de C.
Por ejemplo, si $C = \{c_1, c_2, c_3\}$, la familia $\mathcal{P}(C)$ consta de los siguientes conjuntos:

$$\begin{array}{rl} \emptyset & \text{(vacío es un subconjunto de } C\text{)} \\ \{c_1\}; \{c_2\}; \{c_3\} & \text{(subconjuntos con 1 elemento)} \\ \{c_1, c_2\}; \{c_1, c_3\}; \{c_2, c_3\} & \text{(subconjuntos con 2 elementos)} \\ \{c_1, c_2, c_3\} & \text{(subconjunto con 3 elementos)} \end{array}$$

Como vemos, en este ejemplo el número de subconjuntos en $\mathcal{P}(C)$ es igual a 8.
Es importante resaltar que al describir un conjunto no importa el orden en el cual se escriben los elementos que pertenecen a él. Así, por ejemplo, $\{c_1, c_2\}$ es el mismo conjunto que $\{c_2, c_1\}$, y no nos interesa el orden

en el cual aparecen los elementos de cada subconjunto. Sin embargo, a los efectos del razonamiento posterior, supondremos que los elementos del conjunto C están ordenados de alguna manera arbitraria, que es aquélla en la cual los describimos inicialmente.

En el ejemplo anterior, como el conjunto inicial tenía sólo tres elementos, resultó fácil escribir explícitamente los subconjuntos y contarlos, pero en general esto no va a ser posible. Por lo tanto queremos un método que nos permita hallar este número de manera más sencilla. Una posibilidad que resulta práctica para calcular el número de conjuntos de la familia $\mathcal{P}(C)$, que denotaremos $\sharp\mathcal{P}(C)$, es la siguiente. Supongamos entonces que $C = \{c_1, c_2, \dots, c_n\}$, vamos tomando uno a uno todos los elementos de C de manera ordenada y decidimos en cada caso si lo incluimos o no en el subconjunto que construimos.

Podemos pensar, entonces, que construir un subconjunto equivale a asignarle a cada elemento un número: le asignamos el 1 si lo incluimos en el subconjunto y el 0 si no lo incluimos. Es decir, que construir todos los subconjuntos de C es equivalente a construir todas las n-uplas de ceros y unos:

$$(a_1, a_2, \dots, a_n) \qquad (a_i = 0 \text{ ó } 1)$$

donde $a_i = 0$ significa que **no** hemos incluido el elemento c_i en el subconjunto y $a_i = 1$ significa que **sí** lo hemos incluido. Por lo tanto tenemos una correspondencia biunívoca entre $\mathcal{P}(C)$ y el conjunto de n-uplas

$$A_n = \{(a_1, a_2, \dots, a_n) : a_i = 0 \text{ ó } 1\} \ ,$$

correspondencia que asocia a cada subconjunto $M \subset C$ la n-upla que tiene un 1 en el lugar i sí, y sólo sí, $c_i \in M$.

Por ejemplo, en el caso del conjunto $C = \{c_1, c_2, c_3\}$ de 3 elementos, si $M = \{c_1\}$ la terna que le corresponde es $(1, 0, 0)$; si en cambio $M = \{c_2, c_3\}$ la terna que le corresponde es $(0, 1, 1)$ mientras que a $M = \{c_1, c_3\}$ le corresponde $(1, 0, 1)$.

Por lo tanto, basta contar cuántas n-tuplas hay en A_n y esto es sencillo.
Para $n = 1$ es claro que A_n tiene 2 elementos:

$$(0); \ (1)$$

Para $n = 2$ tenemos 4:

$$(0,0); \ (0,1); \ (1,0); \ (1,1)$$

Para $n = 3$ tenemos 8:

$$(0,0,0); \ (0,1,0); \ (1,0,0); \ (1,1,0)$$

$$(0,0,1); \ (0,1,1); \ (1,0,1); \ (1,1,1)$$

y en general, si tenemos la familia A_{n-1}, por cada $(n-1)$-upla que ésta contiene podemos fabricar 2 de A_n, según agreguemos un 0 ó un 1 como última coordenada, y de este modo fabricamos todas las n-uplas de A_n una sola vez. O sea que:

$$\sharp A_n = 2(\sharp A_{n-1}) \quad (n \geq 2) \ ,$$

donde $\sharp A_n$ representa el número de elementos del conjunto A_n. Un sencillo argumento de inducción nos dice que

$$\sharp A_n = 2^n$$

y por lo tanto

$$\sharp\mathcal{P}(C) = 2^n.$$

2.3. Variaciones con Repetición.

Problema 10. *Lanzamos una moneda tres veces. ¿Cuántas sucesiones distintas de 'aguilas' y 'soles' podemos obtener?*

▶ Para cada lanzamiento hay dos resultados posibles. Para cada resultado posible del primer lanzamiento hay dos del segundo, lo cual da 2×2 combinaciones para los dos primeros. Para cada una de estas hay otros dos resultados posibles del tercero. En total hay $2 \times 2 \times 2 = 2^3 = 8$ sucesiones distintas. ◀

Problema 11. *¿Cuántos números de exactamente cuatro cifras se pueden formar con los dígitos impares?*

▶ Tenemos cinco dígitos impares: 1, 3, 5, 7 y 9. La cifra que corresponde a las unidades puede ser cualquiera de estas cinco. Lo mismo para las decenas, las centenas y las unidades de mil. Por lo tanto hay $5 \times 5 \times 5 \times 5 = 5^4 = 625$ números de cuatro cifras, todas impares. ◀

Problema 12. *¿Cuántas palabras de tres letras (con o sin sentido) pueden formarse con las letras de la palabra AZUL?*

▶ Para cada una de las letras de la palabra que queremos formar tenemos cuatro que podemos escoger. Por lo tanto hay $4^3 = 64$ palabras. ◀

Los tres problemas anteriores tienen características similares. Utilizando los m elementos de un conjunto C (los *cinco* dígitos impares, los *dos* resultados de lanzar una moneda, las *cuatro* letras de la palabra AZUL), queremos formar sucesiones de longitud n (*cuatro*, *tres* y *cuatro*, respectivamente) **permitiendo que los elementos se repitan** y queremos contar el número de maneras de hacer esto. El resultado es m^n. Veamos cómo se puede deducir ésto en general.

Consideremos un conjunto de m elementos con la notación $C = \{c_1, c_2, \dots, c_m\}$. Veamos el conjunto de n-uplas o vectores de dimensión n que podemos formar con los elementos del conjunto C, permitiendo que los elementos se repitan, es decir,

$$X_n = \{(c_{i_1}, c_{i_2}, \dots, c_{i_n}) : c_{i_j} \in C,\ j = 1, \dots, n\}$$

Por ejemplo, el conjunto A_n considerado en la sección 2.2 de las n-uplas de ceros y unos corresponde a tomar $C = \{0, 1\}$. Si en cambio $C = \{0, 1, 2\}$ y $n = 3$, entonces X_n consiste de las siguientes ternas:

$$\begin{array}{ccccccccc}
(0,0,0); & (0,0,1); & (0,0,2); & (0,1,0); & (0,1,1); & (0,1,2); & (0,2,0); & (0,2,1); & (0,2,2) \\
(1,0,0); & (1,0,1); & (1,0,2); & (1,1,0); & (1,1,1); & (1,1,2); & (1,2,0); & (1,2,1); & (1,2,2) \\
(2,0,0); & (2,0,1); & (2,0,2); & (2,1,0); & (2,1,1); & (2,1,2); & (2,2,0); & (2,2,1); & (2,2,2)
\end{array}$$

Hay que tener en cuenta que, al contrario de lo que sucede en el caso de los subconjuntos, el orden en el cual aparecen las componentes es determinante para las n-uplas. Así, el par (c_1, c_2) es distinto a (c_2, c_1).

Para calcular el número de elementos de X_n, llamado *variaciones (o arreglos) con repetición de m elementos tomados de n en n*, procedemos exactamente igual que en la sección anterior, cuando contamos el número de n-uplas de ceros y unos, sólo que ahora, en lugar de ceros y unos, la n-upla está formada a partir de los elementos de C, que son m. Repitiendo el razonamiento anterior resulta que

$$\sharp X_n = m^n.$$

Problema 13. *Si lanzamos un dado cuatro veces, ¿cuántos resultados posibles hay?*

▶ Para cada lanzamiento hay seis resultados posibles. Como lanzamos el dado cuatro veces el resultado es $6^4 = 1.296$.

Si usamos la notación anterior, $C = \{1, 2, 3, 4, 5, 6\}$, $m = 6$ y $n = 4$. ◀

Problema 14. *En una cuadra hay cinco casas. Hay tres colores para escoger la pintura de cada una de ellas. ¿De cuantas maneras puede pintarse el conjunto de las cinco?*

▶ $3^5 = 243$. ◀

2.4. Variaciones sin Repetición.

Veamos ahora otro tipo de problemas.

Problema 15. *Entre los once jugadores de un equipo de fútbol hay que escoger un capitán y su suplente. ¿Cuántas maneras hay de hacer esto?*

▶ Cualquiera de los once jugadores puede ser seleccionado capitán. Hecho esto, cualquiera de los diez que quedan puede ser su suplente. Por lo tanto hay 11×10 maneras de hacerlo. ◀

La diferencia en este caso está en que la selección del capitán modifica el conjunto a partir del cual podemos seleccionar su suplente, ya que el capitán no puede ser su propio suplente. Por lo tanto, la selección del capitán y su suplente no son independientes, como ocurría en la sección anterior.

Problema 16. *Se colocan veinte tarjetas numeradas de 1 a 20 en una bolsa para rifar tres premios. ¿De cuántas maneras se pueden repartir los premios?*

▶ El primer premio puede ser cualquiera de los veinte números. Seleccionado éste, el segundo puede ser cualquiera de los 19 restantes, y el tercero cualquiera de los 18 que quedan luego de seleccionar primero y segundo. En total hay $20 \times 19 \times 18 = 6840$. ◀

De nuevo, a medida que vamos seleccionando cada número premiado, el conjunto a partir del cual podemos escoger el siguiente cambia.

Veamos cómo podemos calcular este número en general. Consideremos de nuevo un conjunto de m elementos con la notación $C = \{c_1, c_2, \dots, c_m\}$. Veamos ahora el conjunto de n-uplas o vectores de dimensión n que podemos formar con los elementos del conjunto C, impidiendo que los elementos se repitan, es decir, cuando consideramos el conjunto

$$Y_n = \{(c_{i_1}, c_{i_2}, \dots, c_{i_n}) : c_{i_j} \in C,\ j = 1, \dots, n,\ c_{i_j} \text{ distintos 2 a 2}\}.$$

El número de elementos de Y_n se llama las *variaciones (o arreglos) de m elementos tomados de n en n* y se denota V_n^m. Con frecuencia decimos arreglos sin repetición, o simplemente variaciones. Cuando no digamos nada se sobreentenderá que son sin repetición.

Por ejemplo, supongamos que $C = \{c_1, c_2, c_3, c_4\}$ de modo que $m = 4$ y sea $n = 3$. Es fácil verificar que la lista siguiente contiene todos los elementos de Y_n sin que figuren repetidos:

$$\begin{array}{llllll}
(c_1,c_2,c_3); & (c_1,c_2,c_4); & (c_1,c_3,c_2); & (c_1,c_3,c_4); & (c_1,c_4,c_2); & (c_1,c_4,c_3) \\
(c_2,c_1,c_3); & (c_2,c_1,c_4); & (c_2,c_3,c_1); & (c_2,c_3,c_4); & (c_2,c_4,c_1); & (c_2,c_4,c_3) \\
(c_3,c_1,c_2); & (c_3,c_1,c_4); & (c_3,c_2,c_1); & (c_3,c_2,c_4); & (c_3,c_4,c_1); & (c_3,c_4,c_2) \\
(c_4,c_1,c_2); & (c_4,c_1,c_3); & (c_4,c_2,c_1); & (c_4,c_2,c_3); & (c_4,c_3,c_1); & (c_4,c_3,c_2)
\end{array}$$

En consecuencia se observa que $V_3^4 = 24$.

Para obtener una fórmula general para V_n^m procedemos inductivamente en n. Antes que nada observamos que necesariamente se tiene que $n \leq m$, ya que si $n > m$, cualquier n-upla de elementos de C tendrá elementos repetidos. Comencemos con $n = 1$. Es claro que tenemos m 1-uplas que son:

$$(c_1);\ (c_2);\ \dots\ (c_m)$$

y por lo tanto

$$V_1^m = m.$$

Supongamos ahora que $n = 2$. Tenemos:

$$\begin{array}{llll}
(c_1,c_2); & (c_1,c_3); & \dots & (c_1,c_m) \\
(c_2,c_1); & (c_2,c_3); & \dots & (c_2,c_m) \\
\vdots & \vdots & \ddots & \vdots \\
(c_m,c_1); & (c_m,c_2); & \dots & (c_m,c_{m-1})
\end{array}$$

que son $m(m-1)$ pares que se obtienen agregando a cada uno de los m elementos de C colocados en primer término, uno de los $(m-1)$ elementos restantes (¡recordar que no hay repeticiones!). Por lo tanto

$$V_2^m = m(m-1).$$

Para tener una fórmula general para V_n^m, procedemos inductivamente en n, ya que el razonamiento anterior puede generalizarse sin dificultad como sigue:

Supongamos que tenemos todas las $(n-1)$-uplas (sin repetición). ¿Cómo fabricamos las n-uplas sin repetición? Tomamos una $(n-1)$-upla y le agregamos al final uno de los $(m-(n-1))$ elementos de C que no figuran en ella, de modo que, por cada $(n-1)$-upla podemos fabricar $(m-(n-1))$ n-uplas. De esta forma hemos fabricado todas las n-uplas de Y_n sin repetir ninguna. Por lo tanto

$$V_n^m = (m-n+1)V_{n-1}^m \quad (n \leq m). \tag{2.1}$$

Como ya vimos que $V_1^m = m$, deducimos de (1-1) que

$$V_n^m = m(m-1)\cdots(m-n+1) = \frac{m!}{(m-n)!} \tag{2.2}$$

donde $m! = m \times (m-1) \times \cdots \times 2 \times 1$ se conoce como m factorial. En la fórmula (2.2) utilizamos la convención $0! = 1$ (cuando $m = n$).

Problema 17. *En una carrera de fórmula 1 participan 26 corredores. Los cinco primeros ganan puntos según la posición que ocupen (9 puntos al primero, 6 al segundo, etc.) ¿De cuántas maneras pueden repartirse los puntos?*

► $V_5^{26} = 7,893,600.$ ◄

2.5. Permutaciones.

Un caso particular de variaciones son las *permutaciones*, que corresponden a la situación $m = n$. En este caso $V_m^m = m! = m(m-1)(m-2)\cdots 2\cdot 1$. Observamos que ahora las m-uplas contienen **todos** los elementos de C, sin repetición, dispuestos en todos los órdenes posibles.

Por ejemplo, si $m = n = 3$ las permutaciones son:

$$(c_1, c_2, c_3);\ (c_1, c_3, c_2);\ (c_2, c_1, c_3);\ (c_2, c_3, c_1);\ (c_3, c_1, c_2);\ (c_3, c_2, c_1).$$

Claramente $V_3^3 = 6$.

También se emplea con frecuencia para las permutaciones la notación

$$P_m = V_m^m = m!$$

Problema 18. *¿De cuántas maneras podemos colocar cuatro bolas de distintos colores en fila?*

► La primera puede ser cualquiera de las cuatro. La segunda, cualquiera de las tres restantes, etc. La respuesta es $4 \times 3 \times 2 \times 1 = 4! = 24$. ◄

Problema 19. *¿Cuántas palabras, con o sin sentido, pueden obtenerse usando todas las letras de la palabra* PRENSA*?*

► Como la palabra no tiene letras repetidas, la respuesta es $6! = 720$. Más adelante nos encontraremos la situación de palabras con letras repetidas. ◄

2.6. Combinaciones.

Problema 20. *De un grupo de treinta estudiantes queremos escoger dos para participar en una competencia. ¿De cuántas maneras podemos hacerlo?*

▶ El primer estudiante del par puede ser cualquiera de los treinta y, una vez escogido éste, el segundo puede ser cualquiera de los veintinueve restantes. Pero de esta manera hemos contado cada pareja dos veces, cuando A es el primero y B el segundo, y cuando B es el primero y A el segundo. Por lo tanto tenemos que dividir este número entre dos. La respuesta es $\frac{30\times 29}{2} = 435$. ◀

Problema 21. *De un grupo de veinticinco libros queremos escoger tres para leer durante las vacaciones. ¿De cuántas maneras podemos hacer esto?*

▶ Hacemos un razonamiento similar al del problema anterior. Primero contamos cuantos tríos ordenados de libros podemos formar y luego dividimos entre el número de ordenamientos posibles de cada trío. El número de tríos ordenados son las variaciones de 25 elementos tomados de 3 en 3: $V_3^{25} = 25 \times 24 \times 23 = 13.800$. Cada trío lo podemos ordenar de $3! = 6$ maneras. Por lo tanto la respuesta es

$$\frac{V_5^{25}}{3!} = \frac{13.800}{6} = 2,300.$$

◀

Problema 22. *En un juego de dominó, ¿de cuántas maneras podemos escoger una mano?*

▶ Una mano consiste de siete piedras sin importar su orden. La primera puede ser cualquiera de las 28 que forman el juego. Escogida ésta, hay 27 para escoger la segunda, luego 26 para la tercera, y así sucesivamente hasta escoger las siete. En total: $V_7^{28} = 28\times 27\times 26\times 25\times 24\times 23\times 22 = 5.967.561.600$. Pero cada mano ha sido contada varias veces, dependiendo del orden en el cual la escogimos. Por lo tanto tenemos que dividir por el número de maneras de ordenar una mano, que es $7! = 5040$, y la respuesta es

$$\frac{V_7^{28}}{7!} = \frac{5,967,561,600}{5040} = 1,184,040$$

◀

Veamos cómo podemos resolver este tipo de problemas en general. Consideramos nuevamente un conjunto $C = \{c_1, c_2, \dots, c_m\}$ con m elementos. Llamamos *combinaciones de m elementos tomados de n en n* al número de subconjuntos de C que constan de n elementos. Se entiende que $0 \leq n \leq m$ y se denota dicho número por

$$\binom{m}{n} \quad \text{o también} \quad C_n^m.$$

Ya sabemos calcular el número de n-uplas ordenadas V_n^m que se pueden formar con los elementos de C. Es claro que cada subconjunto de C con n elementos da lugar a $n!$ n-uplas ordenadas - tantas como maneras tenemos de ordenar los n elementos del subconjunto - y por lo tanto

$$V_n^m = \binom{m}{n} \times n! \tag{2.3}$$

Reemplazando V_n^m por su valor (fórmula (2.2)), resulta

$$\binom{m}{n} = \frac{m!}{(m-n)!n!}. \tag{2.4}$$

Observamos que $\binom{m}{0} = 1$ para cualquier valor de m. Los números $\binom{m}{n}$ se conocen como *números combinatorios*. Estudiaremos algunas de sus propiedades más adelante. Veamos primero algunos problemas.

Problema 24. *En una práctica de Baloncesto el entrenador quiere escoger un equipo de cinco entre los treinta jugadores que están entrenando. ¿De cuántas maneras puede hacerlo?*

▶ $\binom{30}{5} = \frac{30!}{25!\,5!} = \frac{17,100,720}{120} = 142,506$ ◀

Problema 25. *Un estudiante tiene seis libros y otro tiene nueve. ¿De cuántas maneras pueden intercambiar tres libros?*

▶ El primer estudiante puede escoger tres libros de $\binom{6}{3}$ maneras mientras que el segundo puede hacerlo de $\binom{9}{3}$. Por lo tanto, el número de intercambios posibles es $\binom{6}{3}\binom{9}{3} = 1.120$. ◀

Problema 26. *Hay dos niñas y siete niños en un grupo de nadadores. Se quiere escoger un equipo de cuatro de modo que al menos uno de los nadadores sea niña. ¿De cuántas maneras se puede hacer esto?*

▶ Tenemos dos posibilidades: puede haber una o dos niñas en el equipo. En este último caso los dos varones pueden escogerse de $\binom{7}{2}$. Si hay sólo una niña, la podemos escoger de dos maneras, mientras que a los tres niños restantes los podemos escoger de $\binom{7}{3}$. En total tenemos $\binom{7}{2} + 2\binom{7}{3} = 91$ equipos posibles. ◀

Problema 27. *En el juego de KINO cada cartón tiene 15 números escogidos del 1 al 25. ¿Cuántos cartones hay?*

▶ Como no nos importa en orden en el cual escogemos los 15 números la respuesta es el número combinatorio $\binom{25}{15} = \frac{25!}{15!\,10!} = 3,268,760$.

Una observación importante es que seleccionar los 15 números que están en el cartón es equivalente a seleccionar los 10 que **no** están. Por lo tanto la respuesta también es el número combinatorio $\binom{25}{10} = \frac{25!}{10!\,15!} = 3,268,760$. Esta es una propiedad general que enunciamos a continuación. ◀

Problema 28. *Tenemos tres bolas indistinguibles y 20 cajas. ¿De cuántas maneras podemos colocar las bolas en las cajas de modo que no haya más de una bola en cada caja?*

▶ Podemos enumerar las cajas del 1 al 20 y ahora el problema se reduce a seleccionar subconjuntos de tres elementos del conjunto $\{1, 2, \dots, 20\}$, que representan las cajas que van a estar ocupadas. Ya sabemos que esto lo podemos hacer de

$$\binom{20}{3} = 1,140$$

maneras distintas. ◀

El problema anterior muestra que si tenemos objetos de dos tipos y queremos colocar k objetos de tipo 1 y $n-k$ de tipo 2 en fila, tenemos $\binom{n}{k}$ maneras de hacerlo, pues podemos pensar que los lugares de la fila están numerados y que el problema consiste en contar el número de subconjuntos de k elementos del conjunto $\{1, 2, \dots, n\}$.

2.6.1. Propiedades.

Propiedad 1. $\binom{m}{n} = \binom{m}{m-n}$.
Demostración. A partir de la definición tenemos

$$\binom{m}{n} = \frac{m!}{n!(m-n)!} = \binom{m}{m-n}.$$

Como ejercicio, dé una demostración sin calcular, utilizando solamente la definición de combinación.

Propiedad 2. (Relación de Pascal)

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n} \quad (1 \le n \le m-1). \tag{2.5}$$

Demostración. Tenemos un conjunto C de m elementos y queremos contar el número de subconjuntos de n elementos que tiene. Ya sabemos que este número es $\binom{m}{n}$ pero vamos a calcularlo de otra manera. Sea $c_1 \in C$ un elemento de C, contamos en primer lugar los subconjuntos de C de n elementos que **tienen** a c_1. Esto es equivalente a contar los subconjuntos de $n-1$ elementos del conjunto $C \setminus \{c_1\}$, que son $\binom{m-1}{n-1}$. En segundo lugar contamos los subconjuntos de C de n elementos que **no tienen** al elemento c_1. Como c_1 no puede estar en el subconjunto, tenemos que elegir a partir de los $m-1$ elementos restantes de C. Esto da $\binom{m-1}{n}$ subconjuntos. Aplicando ahora el Principio de Suma tenemos

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}.$$

2.7. El Triángulo de Pascal.

La propiedad 2 sirve para construir un arreglo de números con propiedades útiles e interesantes, que se conoce como el triángulo de Pascal. Supongamos que para un cierto valor de m conocemos los valores de todos los números combinatorios de la forma $\binom{m}{n}$, $0 \leq n \leq m$, entonces la relación (2.5) nos permite calcular los valores de los números combinatorios $\binom{m+1}{n}$, $0 \leq n \leq m+1$:

$$\binom{m+1}{n} = \binom{m}{n-1} + \binom{m}{n}.$$

Por lo tanto, de manera recursiva podemos obtener todos los números combinatorios. Comenzamos con $\binom{0}{0} = 1$, al cual colocamos en el centro de la página. Los siguientes dos son $\binom{1}{0} = 1$ y $\binom{1}{1} = 1$, que colocamos debajo, a ambos lados del 1 que habíamos colocado inicialmente, de modo que éste quede en el centro del espacio que separa los dos números nuevos, como se ve en la figura 2.3.

$$\begin{array}{ccc} & 1 & \\ 1 & & 1 \end{array}$$

Figura 2.3

Para $m = 2$ tenemos en primer lugar los dos números combinatorios de los extremos, que corresponden a $n = 0$ y $n = 2$, i.e. $\binom{2}{0} = 1$ y $\binom{2}{2} = 1$, que colocamos debajo de los anteriores, como se ve en la figura 2.4. Aún cuando es fácil calcular el número combinatorio $\binom{2}{1}$ directamente, vamos a hacerlo usando la fórmula (1.5): $\binom{2}{1} = \binom{1}{0} + \binom{1}{1} = 1 + 1 = 2$. Si colocamos este número en el centro de la tercera fila observamos que su valor es la suma de los dos números que se encuentran sobre él:

$$\begin{array}{ccccc} & & 1 & & \\ & 1 & & 1 & \\ 1 & & 2 & & 1 \end{array}$$

Figura 2.4

Veamos como se construye la fila que corresponde a $m = 3$. Los extremos ambos valen 1: $\binom{3}{0} = \binom{3}{3} = 1$. El resto de los espacios los llenamos sumando en cada caso los dos valores que se encuentran por encima del espacio en cuestión: $\binom{3}{1} = 1 + 2 = 3$, $\binom{3}{2} = 2 + 1$.

Si continuamos este proceso inductivamente obtenemos el triángulo que se indica en la figura 2.6, conocido como triángulo de Pascal.

La fila j tiene $j+1$ números, que corresponden a los números combinatorios $\binom{j}{i}$, para $0 \leq i \leq j$, es decir que cada fila comienza por el número combinatorio $\binom{j}{0}$. Observamos, en consecuencia, que el número que

1

1 1

1 2 1

1 3 3 1

Figura 2.5

j = 0	1
j = 1	1 1
j = 2	1 2 1
j = 3	1 3 3 1
j = 4	1 4 6 4 1
j = 5	1 5 10 10 5 1
j = 6	1 6 15 20 15 6 1
j = 7	1 7 21 35 35 21 7 1
j = 8	1 8 28 56 70 56 28 8 1
j = 9	1 9 36 84 126 126 84 36 9 1
.	. . . .
.	. . . .

Figura 2.6: El Triángulo de Pascal

aparece en el lugar $i+1$ de la fila j, es el número combinatorio $\binom{j}{i}$, por ejemplo, para hallar $\binom{7}{4}$ buscamos el lugar 5 de la fila 7 obtenemos $\binom{7}{4} = 35$.

Otra manera de construir el triángulo es la siguiente. Cambiamos los números por puntos o nodos, como se indica en la figura 2.7.

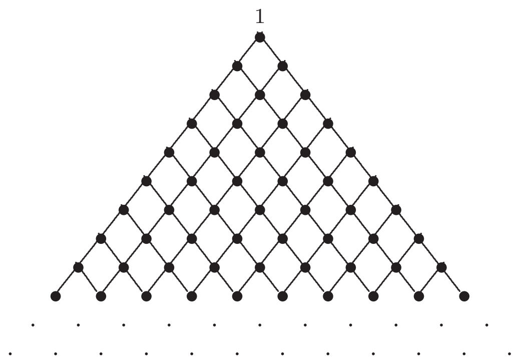


Figura 2.7

Escribimos un 1 sobre el vértice superior, y luego, sobre cada nodo, el número de maneras que hay para llegar a este punto a partir del vértice superior, moviéndonos únicamente hacia abajo. El resultado es el triángulo de Pascal.

Veamos una propiedad interesante del triángulo de Pascal. Si evaluamos la suma de los números en cada fila obtenemos 1, 2, 4, 8, 16, 32, 64, 128, etc. Parece natural la conclusión de que la suma de la n-ésima fila es 2^n. Esto es cierto y podemos probarlo por inducción. Sabemos que es cierto para las primeras filas. Para probar el paso inductivo observamos que cada número de la n-ésima fila es sumando para formar *dos* números

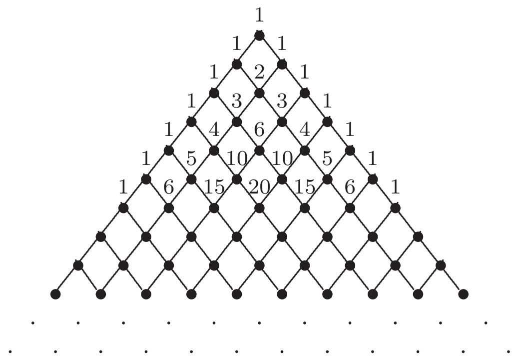


Figura 2.8

de la siguiente fila: los que están por debajo de él, a ambos lados. Por lo tanto la suma de los números de la fila $n+1$ es dos veces la suma de los números de la fila anterior. Esto completa el paso inductivo.

Si escribimos esta relación explícitamente obtenemos la siguiente identidad:

$$\binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{m-1} + \binom{m}{m} = 2^m. \tag{2.6}$$

En realidad, ya hemos visto una demostración combinatoria de esta identidad. El lado derecho representa el número de subconjuntos de un conjunto con m elementos. Por otro lado, el número combinatorio $\binom{m}{n}$ representa el número de subconjuntos de n elementos que se pueden formar a partir de un conjunto de m elementos. La identidad anterior dice que el número total de subconjuntos es igual a la suma de los subconjuntos de 0 elementos más el número de subconjuntos de 1 elemento más ... más el número de subconjuntos de m elementos.

2.8. El Binomio de Newton.

Queremos encontrar ahora una fórmula para la expresión $(a+b)^m$ para valores generales de m. Aún cuando este no es un problema de combinatoria, tiene una solución que está estrechamente ligada a los números combinatorios y al triángulo de Pascal.

Escribamos los valores de esta expresión para los primeros valores de m:

$$\begin{aligned}
(a+b)^0 &= 1, \\
(a+b)^1 &= a+b, \\
(a+b)^2 &= a^2 + 2ab + b^2, \\
(a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3.
\end{aligned}$$

Observamos que los coeficientes de las expresiones que están del lado derecho corresponden a los valores del triángulo de Pascal. Esto sugiere la siguiente identidad:

$$\begin{aligned}
(a+b)^m &= \binom{m}{0}a^m + \binom{m}{1}a^{m-1}b + \binom{m}{2}a^{m-2}b^2 + \cdots + \binom{m}{m-1}ab^{m-1} + \binom{m}{m}b^m. \\
&= \sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k}.
\end{aligned}$$

Haremos la demostración de esta fórmula por inducción completa en m. Observe que el segundo miembro contiene $(m+1)$ sumandos. Para $m=1$, queda

$$(a+b)^1 = \binom{1}{0} a^1 b^0 + \binom{1}{1} a^0 b^1 = a+b,$$

que es obviamente correcto.
Supongamos entonces que la fórmula es correcta para m, e intentemos probar que también lo es para $(m+1)$. Tenemos

$$\begin{aligned}(a+b)^{m+1} &= (a+b)(a+b)^m = (a+b)\sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k}\\ &= \sum_{k=0}^{m}\binom{m}{k}a^{k+1}b^{m-k} + \sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k+1}\\ &= a^{m+1} + \sum_{k=0}^{m-1}\binom{m}{k}a^{k+1}b^{m-k} + b^{m+1} + \sum_{k=1}^{m}\binom{m}{k}a^k b^{m-k+1}. \end{aligned} \tag{2.7}$$

Haciendo un cambio en el índice de la suma $j = k+1$ obtenemos que el segundo sumando en la expresión anterior se puede escribir

$$\sum_{k=0}^{m-1}\binom{m}{k}a^{k+1}b^{m-k} = \sum_{j=1}^{m}\binom{m}{j-1}a^j b^{m-j+1}.$$

Vamos a reemplazar esta expresión en (2.7), pero para mantener la uniformidad en la expresión y simplificarla más fácilmente, usaremos el índice k en lugar de j. Obtenemos

$$\begin{aligned}(2.7) &= a^{m+1} + \sum_{k=1}^{m}\binom{m}{k-1}a^k b^{m-k+1} + b^{m+1} + \sum_{k=1}^{m}\binom{m}{k}a^k b^{m-k+1}\\ &= a^{m+1} + \sum_{k=1}^{m}\left[\binom{m}{k-1} + \binom{m}{k}\right]a^k b^{m-k+1} + b^{m+1}.\end{aligned}$$

Aplicando ahora la propiedad 2:

$$\binom{m}{k-1} + \binom{m}{k} = \binom{m+1}{k}$$

y reemplazando resulta

$$(a+b)^m = a^{m+1} + \sum_{k=1}^{m}\binom{m+1}{k}a^k b^{m-k+1} + b^{m+1} = \sum_{k=0}^{m+1} a^k b^{m-k+1}$$

que muestra que la fórmula es correcta cuando el exponente es $m+1$. Por el principio de inducción sabemos entonces que la fórmula es válida para todo m.

Como caso particular de la fórmula del binomio de Newton podemos obtener de nuevo la identidad (1.6). Basta tomar $a = b = 1$ para obtener

$$2^m = \sum_{k=0}^{m}\binom{m}{k}.$$

2.9. Coeficientes Multinomiales

Los coeficientes binomiales cuentan el número de maneras que tenemos de colocar en fila n objetos de dos tipos de modo que haya k objetos de tipo 1 y $n-k$ de tipo 2, donde $0 \leq k \leq n$. Sabemos que hay

$$\binom{n}{k}$$

maneras de hacerlo.

Supongamos ahora que tenemos m tipos distintos de objetos y queremos colocar en fila k_1 objetos de tipo 1, k_2 de tipo 2, $\dots$, k_m de tipo m, con $k_1 + k_2 + \cdots + k_m = n$ y $0 \leq k_i \leq n$ para $i = 1, 2, \dots, m$. ¿De cuantas maneras podemos hacer esto?

Supongamos que podemos distinguir los n objetos, no solo los de distintos tipo sino también los del mismo tipo. Entonces tendríamos $n!$ arreglos posibles de los n objetos. Pero como en realidad los objetos de Tipo 1 son indistinguibles, cualquier permutación de estos objetos produce resultados que son indistinguibles. Como hay $k_1!$ arreglos de los objetos de tipo 1 en la fila, debemos dividir $n!$ por $k_1!$. Otro tanto ocurre con los objetos de tipo 2: Si los permutamos, obtenemos filas que son indistinguibles, de modo que también hay que dividir por $k_2!$. De manera similar debemos dividir por $k_3!$, $k_4!, \dots, k_m!$, de modo que el número de filas es

$$\frac{n!}{k_1!k_2!\cdots k_m!}.$$

Usaremos la notación

$$\binom{n}{k_1, k_2, \dots, k_m}$$

para estos números, que llamaremos *coeficientes multinomiales.*

Por ejemplo, si tenemos las letras a, b, b, c, c ¿De cuántas maneras podemos ordenarlas? La respuesta es

$$\binom{5}{1, 2, 2} = 30.$$

Existe una relación entre los coeficentes multinomiales y el desarrollo de expresiones del tipo $(x_1 + \cdots + x_m)^n$ que es similar a la de los coeficientes binomiales y el binomio de Newton: El número $\binom{n}{k_1,\dots,k_m}$ es el coeficiente de $x_1^{k_1} \cdots x_m^{k_m}$ en el desarrollo de $(x_1 + \cdots + x_m)^n$.

Para ver esto consideremos cómo se forma un término como $x_1^{k_1} \cdots x_m^{k_m}$ en esta multiplicación. Si consideramos la expresión

$$(x_1 + \cdots + x_m)(x_1 + \cdots + x_m) \cdots (x_1 + \cdots + x_m) \tag{2.8}$$

los distintos términos se obtienen seleccionando una variable de cada factor en (2.8). Hay n factores y queremos seleccionar k_1 veces a x_1, a x_2, k_2 veces y así sucesivamente hasta x_n que lo queremos seleccionar k_n veces, y ya sabemos que esto lo podemos hacer de

$$\binom{n}{k_1, k_2, \dots, k_m} \tag{2.9}$$

maneras distintas. Por lo tanto (2.9) es el coeficiente del término $x_1^{k_1} \cdots x_m^{k_m}$ en el desarrollo de $(x_1 + \cdots + x_m)^n$.

Como ejemplo podemos escribir el desarrollo del cubo de la suma de tres términos:

$$(a + b + c)^3 = a^3 + b^3 + c^3 + 3a^2b + 3a^2c + 3b^2a + 3b^2c + 3c^2a + 3c^2b + 6abc.$$

2.10. Problemas Resueltos.

1. *¿Cuántas palabras (con o sin sentido) pueden formarse usando todas las letras de la palabra "REMAR"?*

► Esta palabra contiene dos veces la letra R y todas las demás son diferentes. Supongamos por un momento que estas dos letras son distinguibles: R_1 y R_2. En este caso hay $5! = 120$ palabras diferentes, pero en realidad dos palabras que puedan obtenerse intercambiando R_1 y R_2 son idénticas. Por lo tanto las 120 palabras se dividen en pares de palabras idénticas, de modo que la respuesta es $120/2 = 60$. ◄

2. *¿Cuántas palabras de (con o sin sentido) pueden formarse usando todas las letras de la palabra "SABANA"?*

► Esta palabra contiene tres veces la letra A. Supongamos de nuevo que estas letras son distinguibles: A_1 A_2 y A_3. En este caso hay $6! = 720$ palabras diferentes, pero en realidad dos palabras que puedan obtenerse intercambiando las letras A_i son idénticas y esto podemos hacerlo de $3! = 6$ maneras diferentes. Por lo tanto las 720 palabras se dividen en grupos de 6 palabras idénticas, de modo que la respuesta es $720/6 = 120$. ◄

3. *¿Cuántas palabras (con o sin sentido) pueden formarse usando todas las letras de la palabra "INTENCION"?*

► Esta palabra contiene tres veces la letra N, dos veces la letra I y las otras son distintas. Si pensamos de nuevo que estas letras son distinguibles, tenemos 9! palabras. Como en realidad las letras I son idénticas, el número de palabras se reduce a $9!/2!$, y ahora si recordamos que las N también son distinguibles nos quedan $9!/(2! \times 3!) = 30,240$ palabras. ◄

4. *¿De cuántas maneras pueden sentarse cinco personas en cinco sillas?*

► En este caso nos interesa el número de permutaciones de cinco elementos, ya que podemos pensar que las sillas están numeradas y el problema es equivalente a ordenar el conjunto de personas. Por lo tanto la respuesta es $5! = 120$. ◄

5. *¿De cuántas maneras pueden sentarse cinco personas en cinco sillas alrededor de una mesa circular, si consideramos que todas las rotaciones de una posición son equivalentes?*

► Obsérvese que se puede elegir arbitrariamente la silla para la primera persona (a menos de rotar simultáneamente a todo el mundo, hasta que esta primera persona quede sentada en esa silla). Es fácil ver, entonces, que el número de disposiciones posibles es el número de maneras de sentarse las 4 personas restantes en las 4 sillas que quedan, es decir $4! = 24$.

El mismo razonamiento dice que, si en lugar de 5 personas y 5 sillas, son n, el resultado es $(n-1)!$. ◄

6. *¿Cuántos números de seis cifras tienen al menos una cifra par?*

► Los números que tienen al menos una cifra par son aquellos que tienen una, dos, tres, ... seis cifras pares. Por lo tanto tendríamos que contar el número de elementos de cada uno de estos conjuntos y luego sumarlos. Resulta más sencillo en esta situación, contar cuantos números no satisfacen la condición (es decir, cuantos no tienen ninguna cifra par) y restar éste del total de los números de seis cifras. Hay $9 \times 10^5 = 900,000$ números de seis cifras (la primera cifra no puede ser 0, por eso la respuesta no es 10^6) de los cuales $5^6 = 15.625$ no tienen ninguna cifra par. Entonces hay $900,000 - 15,625 = 884,375$ números de seis cifras con al menos una de ellas par. ◄

7. *Supongamos que tenemos 3 cajas numeradas y 4 bolas indistinguibles. ¿De cuántas maneras podemos colocar las 4 bolas en las tres cajas?*

▶ Procedamos gráficamente, representando cada caja como la parte limitada por dos barras verticales. Por ejemplo, en la configuración de la figura tenemos 3 bolas en la primera caja, 0 bolas en la segunda y 1 bola en la tercera.

$$\begin{array}{ccc} |\bigcirc\ \bigcirc\ \bigcirc| & \quad & |\bigcirc| \\ 1 & 2 & 3 \end{array}$$

Un instante de reflexión muestra que tenemos la situación siguiente: hay 4 bolas y 4 barras, que se disponen una a continuación de la otra en 8 lugares y cada configuración queda determinada por el lugar que asignamos a las dos barras centrales entre los 6 lugares que podemos elegir para ellas, teniendo en cuenta que las barras de los extremos están obligadas a quedar allí. Es decir, el problema se reduce a decidir de cuántas maneras se pueden colocar dos barras en seis espacios, y esto es $\binom{6}{2}$. ◀

8. *Generalizar el ejemplo anterior al caso en que tenemos b bolas y n cajas.*

▶ El resultado que se obtiene aplicando el mismo razonamiento es:

$$\binom{n+b-1}{n-1} = \binom{n+b-1}{b}$$

◀

9. *¿De cuántas maneras podemos escoger las cuatro manos en un juego de dominó?*

▶ Por el resultado del problema 22 en la sección 2.6 sabemos que hay 1,184,040 maneras de escoger una mano de siete piedras. Veamos de cuantas maneras podemos escoger las cuatro manos que forman el punto de inicio de una partida de dominó. Una vez escogida la primera mano, nos quedan 21 piedras para escoger la segunda, y por lo tanto hay $\binom{21}{7} = 116,280$. Para la tercera tenemos $\binom{14}{7} = 3,432$ y la última esta formada por las siete piedras que quedan. Por el principio de multiplicación la respuesta es el producto de estos números:

$$1,184,040 \times 116,280 \times 3,432 \times 1 = 472,518,347,558,400$$

◀

10. *Un cartón de 'BINGO' esta formado por cinco columnas de números, una para cada letra de la palabra BINGO. Cada columna tiene cinco números, excepto la central, correspondiente a la letra N, que tiene sólo cuatro ya que el lugar central no participa en el juego. Los cinco números de la primera columna se escogen entre los primeros 15 enteros:* $1, 2, \dots, 15$, *los de la segunda entre los enteros* $16, 17, \dots, 30$ *y así sucesivamente hasta la última columna, para la cual escogemos los cinco números entre* $61, 62, \dots 75$. *¿Cuántos cartones distintos de BINGO hay, suponiendo que el orden en el cual aparecen los números en cada columna no es importante?*

▶ Veamos en primer lugar, de cuantas maneras podemos escoger los cinco números de la primera columna, suponiendo que el orden de los números que aparecen en ella no importa. En este caso tenemos que escoger un subconjunto de 5 números a partir de un conjunto que tiene 15, de modo que la respuesta es el número combinatorio $\binom{15}{5} = 3,003$. Los resultados para las columnas correspondientes a las letras I, G y O son iguales, la única diferencia está en la columna central, correspondiente a la letra N. En este caso sólo escogemos cuatro números a partir de los 15 disponibles, de modo que la respuesta es $\binom{15}{4} = 1,365$

Ahora, por el principio de multiplicación, debemos multiplicar estos resultados para obtener la respuesta:

$$3,003^4 \times 1,365 = 111,007,923,832,370,565$$

◀

11. *En el problema anterior consideramos que el orden de los números en el cartón de BINGO no importaba. Esto es cierto si estamos jugando a cartón lleno, pero en muchos otros juegos, como las cuatro esquinas, la 'X', etc. sí importa el lugar en el cual aparece cada número. Si tomamos en cuenta el orden, ¿cuántos cartones distintos de BINGO hay?*

▶ Podemos aprovechar el cálculo que realizamos en el problema anterior si observamos que tenemos que multiplicar el resultado anterior por el número de maneras de ordenar los números que aparecen en cada cartón. Este ordenamiento debemos hacerlo respetando las restricciones de cada columna, es decir, en la primera sólo pueden aparecer números comprendidos entre 1 y 15, en la segunda entre 16 y 30, etc. Por lo tanto, para cada una de las columnas correspondientes a las letras B, I, G, y O tenemos cinco números y $5! = 120$ órdenes posibles. En la columna central hay sólo cuatro números y por lo tanto $4! = 24$ maneras de ordenarlos. En conclusión debemos multiplicar el resultado del problema anterior por $120^4 \times 24 = 4.976.640.000$:

$$111.007.923.832.370.565 \times 4.976.640.000 = 5,52446474061129 \times 10^{26}$$

◀

12. *Demostrar la identidad*

$$\binom{m}{n} = \sum_{k=n-1}^{m-1} \binom{k}{n-1} \tag{2.10}$$

e interpretar en base al triángulo de Pascal.

▶ Comenzamos a partir de la propiedad 2 de los números combinatorios:

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n} \tag{2.11}$$

que forma la base del triángulo de Pascal. Usamos esta propiedad para escribir el segundo sumando del lado derecho como

$$\binom{m-1}{n} = \binom{m-2}{n-1} + \binom{m-2}{n}.$$

Si sustituimos esta expresión en (2.11) y repetimos este procedimiento obtenemos

$$\begin{aligned}
\binom{m}{n} &= \binom{m-1}{n-1} + \binom{m-1}{n} \\
&= \binom{m-1}{n-1} + \binom{m-2}{n-1} + \binom{m-2}{n} \\
&= \binom{m-1}{n-1} + \binom{m-2}{n-1} + \binom{m-3}{n-1} + \binom{m-3}{n} \\
&= \dots
\end{aligned}$$

y el resultado final es

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-2}{n-1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

y teniendo en cuenta que $\binom{n}{n} = \binom{n-1}{n-1}$ obtenemos la expresión (2.10). Para ver qué interpretación tiene ésto en relación al triángulo de Pascal veamos un caso particular: $m = 7, n = 4$. La relación anterior nos dice que

$$35 = \binom{7}{4} = \binom{6}{3} + \binom{5}{3} + \binom{4}{3} + \binom{3}{3} = 20 + 10 + 4 + 1.$$

Si observamos la ubicación de estos números en el triángulo obtenemos

m = 0	1
m = 1	1 1
m = 2	1 2 1
m = 3	1 3 3 1
m = 4	1 4 6 4 1
m = 5	1 5 10 10 5 1
m = 6	1 6 15 20 15 6 1
m = 7	1 7 21 35 35 21 7 1
m = 8	1 8 28 56 70 56 28 8 1
m = 9	1 9 36 84 126 126 84 36 9 1

Figura 2.9

13. *Dos cajas contienen $2n$ bolas cada una, numeradas de 1 hasta $2n$. Se selecciona un conjunto de n bolas de cada caja. Calcular el número de maneras de hacer ésto de modo que ambos conjuntos tengan, a lo sumo, una bola con el mismo número.*

▶ Utilicemos la siguiente notación:

$\{i_1, \dots, i_n\}$ es el conjunto de bolas extraídas de la primera caja.

$\{j_1, \dots, j_n\}$ es el conjunto de bolas extraídas de la segunda caja.

Es claro que los elementos de $\{i_1, \dots, i_n\}$ son 2 a 2 distintos, y lo mismo sucede para $\{j_1, \dots, j_n\}$. Observamos, aunque no forma parte de la pregunta, que el total de extracciones posibles de la primera caja es $\binom{2n}{n}$ y que, por cada una de éstas, podemos tener $\binom{2n}{n}$ extracciones de la segunda caja, de modo que el total de extracciones de parejas de conjuntos $\{i_1, \dots, i_n\}$, $\{j_1, \dots, j_n\}$ es

$$\binom{2n}{n}^2.$$

Veamos ahora la respuesta a la pregunta formulada. El número de maneras de que ambos conjuntos contengan, a lo sumo, una bola con el mismo número, es la suma del número de maneras de que no contengan ningún número en común más el número de maneras de que contengan en común exactamente un número.

¿Cuántas extracciones podemos efectuar, de tal modo que $\{i_1, \dots, i_n\}$ y $\{j_1, \dots, j_n\}$ no contengan ningún número en común?

Tenemos libertad en la selección de $\{i_1, \dots, i_n\}$, que puede ser hecha de $\binom{2n}{n}$ maneras. Por cada elección de $\{i_1, \dots, i_n\}$ en cambio, hay una sola elección de $\{j_1, \dots, j_n\}$ que produce el efecto deseado de que ambos conjuntos no contengan ningún número en común, que es la elección de los n números del conjunto total $\{1, 2, \dots, 2n\}$ que no figuran en $\{i_1, \dots, i_n\}$. En consecuencia tenemos

$$\binom{2n}{n} \tag{2.12}$$

maneras de elegir los subconjuntos $\{i_1, \dots, i_n\}$ (de la primera caja) y $\{j_1, \dots, j_n\}$ (de la segunda) de modo que no tengan ningún elemento en común

¿Cuántas extracciones podemos efectuar, de tal modo que $\{i_1, \dots, i_n\}$ y $\{j_1, \dots, j_n\}$ contengan exactamente un número en común?

Nuevamente, hay $\binom{2n}{n}$ maneras de elegir la extracción $\{i_1, \dots, i_n\}$ de la primera caja. Hecha ésta, debemos contar cuántas formas tenemos de extraer $\{j_1, \dots, j_n\}$ de modo que en este último conjunto figure uno y solo un elemento de $\{i_1, \dots, i_n\}$. Para ello procedemos así: elegimos el elemento de $\{i_1, \dots, i_n\}$ que debe figurar, para lo cual tenemos n alternativas. Luego elegimos los $(n-1)$ elementos restantes de $\{j_1, \dots, j_n\}$ entre los n elementos que quedan en la segunda caja cuando excluimos los de $\{i_1, \dots, i_n\}$, y esto lo podemos hacer de

$$\binom{n}{n-1} = n$$

maneras.

Resumiendo, por cada extracción de la primera caja tenemos n^2 maneras de hacer una extracción de la segunda que tenga exactamente un elemento en común con la hecha en la primera. Recordando que hay $\binom{2n}{n}$ maneras de extraer de la primera, resulta que la respuesta a nuestra segunda interrogante es que hay

$$\binom{2n}{n} n^2 \tag{2.13}$$

maneras de extraer $\{i_1, \dots, i_n\}$ y $\{j_1, \dots, j_n\}$ de modo que tengan un elemento común. Sumando (1.5) y (1.6) tenemos el resultado final

$$\binom{2n}{n}(1 + n^2).$$

◀

2.11. Aplicaciones a Probabilidad

Podemos ahora aplicar las técnicas de conteo que hemos desarrollado en este capítulo para calcular probabilidad en el caso 'clásico', es decir, cuando tenemos un espacio muestral finito y todos los elementos del espacio tienen igual probabilidad de ocurrir. Veamos un par de ejemplos importantes.

2.11.1. Muestreo con Reposición. La Distribución Binomial.

Retomemos el ejemplo del capítulo anterior sobre el muestreo con reposición, pero en lugar de considerar muestras de tres elementos, consideramos muestras de m elementos. Tenemos una población de N objetos de los cuales n son defectuosos. Igual que antes podemos calcular las probabilidades correspondientes, siempre admitiendo que son igualmente probables de ser extraídos todos los conjuntos ordenados de m elementos buenos y defectuosos.

Sea $p_{k,m} (0 \leq k \leq m)$ la probabilidad de extraer exactamente k defectuosos entre los integrantes de la muestra. Sea $N_{k,m}$ el número de maneras posibles en las cuales se pueden extraer k defectuosos en una muestra de m. Entonces

$$N_{k,m} = \binom{m}{k} n^k (N-n)^{m-k}$$

porque podemos elegir de $\binom{m}{k}$ formas los k lugares que ocupan los defectuosos en la muestra, en cada uno de ellos poner cualquiera de los n defectuosos que hay en la población, y colocar además cualquiera de los $N - n$ no defectuosos en cada uno de los $m - k$ lugares restantes.

Dado que el número total de muestras posibles de m elementos (es decir, de eventos elementales) es N^m, resulta que

$$p_{k,m} = \frac{N_{k,m}}{N^m} = \binom{m}{k} \left(\frac{n}{N}\right)^k \left(1 - \frac{n}{N}\right)^{m-k},$$

o sea

$$p_{k,m} = \binom{m}{k} p^k (1-p)^{m-k}$$

puesto que todas las muestras posibles se han supuesto igualmente probables. Esta función de probabilidad se conoce como la *Distribución Binomial* con parámetros m y p. El nombre de la distribución viene del Binomio de Newton, que estudiamos en la sección 2.8. Podemos usar la fórmula del binomio para demostrar que, efectivamente, la expresión anterior define una probabilidad. Tenemos que verificar que, si sumamos $p_{k,m}$ sobre todos los valores posibles de k, el resultado es 1.

$$\begin{aligned} \sum_{k=0}^{m} p_{k,m} &= \sum_{k=0}^{m} \binom{m}{k} p^k (1-p)^{m-k} \\ &= (p + (1-p))^m = 1. \end{aligned}$$

Ensayos de Bernoulli y la Distribución Binomial. Si realizamos una serie de experimentos con dos resultados posibles, que podemos llamar éxito y fracaso, o 1 y 0, y éxito o 1 tiene probabilidad p de ocurrir, tenemos una situación similar a la que acabamos de analizar. Este tipo de experimento se como un *experimento de Bernoulli* con probabilidad p de éxito. Si realizamos m experimentos de este tipo, la probabilidad de obtener k éxitos está dada por la distribución binomial $p_{k,m}$ que acabamos de definir.

2.11.2. Muestreo sin Reposición. La Distribución Hipergeométrica.

De una población de N artículos entre los cuales hay n defectuosos, se extraen sucesivamente r sin reposición y se cuenta el número de los defectuosos en la muestra. El espacio muestral contiene todos los subconjuntos de r elementos tomados entre los N dados, es decir, $\binom{N}{r}$ puntos muestrales.

Consideremos el evento: "en la muestra hay exactamente s defectuosos", donde $s \leq n,\ s \leq r$. Queremos calcular el número de puntos de este evento, para lo cual observamos que los defectuosos se pueden elegir de $\binom{n}{s}$ formas diferentes, y los no defectuosos de $\binom{N-n}{r-s}$ formas diferentes. Dado que cualquier elección de s defectuosos se puede combinar con cualquier elección de $r-s$ no defectuosos, tenemos en total

$$\binom{n}{s}\binom{N-n}{r-s}$$

muestras posibles en las que hay exactamente s defectuosos.

La probabilidad de obtener exactamente s defectuosos en una muestra de tamaño r tomada en una población de N objetos con n defectuosos es

$$p(s) = \frac{\binom{n}{s}\binom{N-n}{r-s}}{\binom{N}{r}} \qquad s \leq r,\ s \leq n,$$

admitiendo que todas las extracciones posibles son igualmente probables.

Para obtener esta fórmula observemos que hay $\binom{N}{r}$ muestras de tamaño r sin reposición, hay $\binom{n}{s}$ maneras de escoger s defectuosos entre los n que hay en la población y por cada una de ellas hay $\binom{N-n}{r-s}$ maneras de escoger los $r-s$ objetos en buen estado. Esta función de probabilidad se conoce como la *Distribución Hipergeométrica*.

Aplicaremos este modelo a la estimación de N, situación que se presenta frecuentemente en la práctica. Supongamos, por ejemplo, que se desea estimar el total N de peces en un lago. Podemos proceder del siguiente modo: extraemos una muestra de tamaño n y marcamos los peces antes de reintegrarlos al lago. Posteriormente se extrae una muestra, que puede ser del mismo tamaño n, y contamos el número de peces marcados. Se supone que no ha variado el número total de peces y que todos tienen la misma probabilidad de salir en la segunda muestra. Veamos cómo procedemos para estimar N suponiendo que en la segunda muestra se obtuvieron s peces marcados.

Usando el método de máxima verosimilitud, introducido en el capítulo anterior, consideramos la función

$$L(N) = \frac{\binom{n}{s}\binom{N-n}{n-s}}{\binom{N}{n}} \qquad s \leq n,$$

que representa la probabilidad de obtener s peces marcados en una muestra de tamaño n, si el tamaño de la población es N y hay n peces marcados, y determinamos N tal que $L(N)$ sea máximo. Para ello, como se trata de una función discreta, no podemos usar los métodos del Cálculo y consideramos una comparación entre valores sucesivos de la función L para determinar el valor que la maximiza.

Consideramos entonces

$$\frac{L(N)}{L(N-1)} = \frac{\binom{N-1}{n}\binom{N-n}{n-s}}{\binom{N}{n}\binom{N-n-1}{n-s}} = \frac{(N-n)^2}{N(N-2n+s)}.$$

Si el cociente es mayor que 1, resulta $L(N) > L(N-1)$. Para que esto ocurra es necesario que

$$(N-n)^2 = N^2 - 2nN + n^2 > N(N-2n+s) = N^2 - 2nN + sN$$

simplificando esta expresión obtenemos $n^2 > sN$. En consecuencia, para $N < n^2/s$ se tiene $L(N) > L(N-1)$ y la desigualdad cambia de sentido si $N > n^2/s$. Por lo tanto, el valor de N que maximiza a $L(N)$ es el mayor entero que no supera a n^2/s. En particular, si $n = 1,000$ y $s = 100$ resulta

$$\hat{N} = \frac{10^6}{10^2} = 10,000.$$

2.12. Problemas Resueltos

1. De los 38 números de una ruleta (del 1 al 36, y los números 0 y 00), 18 son rojos. ¿Cuál es la probabilidad de que en cinco juegos un número rojo gane exactamente dos veces?

▶ Este es un caso de muestreo con reposición: Tenemos 38 números para escoger y en cada juego puede ganar cualquiera de ellos. Si realizamos cinco juegos hay 38^5 resultados posibles. Para contar de cuantas maneras puede salir un número rojo exactamente dos veces observamos que hay $\binom{5}{2}$ maneras de escoger los juegos en los cuales ganan números rojos. En cada uno de ellos puede ganar cualquiera de los 18 números rojos que tiene la ruleta, lo que nos da un total de 18^2 posibilidades, y en cada uno de los juegos en los cuales no gana un número rojo podemos colocar cualquiera de los 20 números restantes, para un total de 20^3. Tenemos entonces que hay $\binom{5}{2}18^2 20^3$ maneras en las cuales pueden resultar exactamente dos números rojos ganadores en cinco juegos. La probabilidad que buscamos es, por lo tanto

$$\binom{5}{2}\frac{18^2 20^3}{38^5} = \binom{5}{2}\left(\frac{18}{38}\right)^2\left(\frac{20}{38}\right)^3.$$

Hemos podido utilizar los resultados de la sección 2.11.1 para resolver este problema de manera más sencilla. Vimos allí que la probabilidad de obtener exactamente k defectuosos en una muestra de tamaño m realizada con reposición si la proporción de defectuosos en la población es p sigue una distribución binomial de parámetros m y p:

$$\binom{m}{k}p^k(1-p)^{m-k}.$$

Este es exactamente el resultado que acabamos de obtener con $m = 5$, $k = 2$ y $p = 18/38$. ◀

2. Se lanza un dado seis veces. ¿Cuál es la probabilidad de que los resultados sean todos distintos?

► Hay 6^6 resultados posibles. De ellos 6! corresponden a tener en cada lanzamiento un resultado posible. Por lo tanto la probabilidad que buscamos es

$$\frac{6!}{6^6} = \frac{5}{324} = 0.0154$$

◄

3. Si colocamos al azar n fichas en n cajas, ¿Cuál es la probabilidad de que cada caja tenga exactamente una ficha?

► Numeramos las cajas de 1 a n. Distribuir las fichas en las n cajas es equivalente a asignarle a cada ficha el número de la caja en la cual la colocamos. A cada ficha podemos asignarle cualquiera de los n números, y como hay n fichas, tenemos n^n distribuciones posibles. ¿Cuántas de ellas corresponden a tener exactamente una ficha en cada caja? Para que esto sea cierto, al hacer una lista de los números que hemos asignado a las fichas deben estar todos los números de 1 a n sin que haya ninguno repetido. Esto se puede hacer de $n!$ maneras. Por lo tanto la probabilidad que buscamos es

$$\frac{n!}{n^n}.$$

◄

4. Si colocamos al azar n fichas en m cajas, ¿Cuál es la probabilidad de que ninguna caja tenga más de una ficha?

► En primer lugar observamos que si hay más fichas que cajas, es imposible hacer una distribución sin que haya al menos una caja con más de una ficha. En este caso la probabilidad buscada es 0. Supongamos, entonces que $m \geq n$. Si adoptamos el mismo procedimiento que en el problema anterior, asignándole a cada ficha el número de la caja que le toca, vemos que hay m^n distribuciones posibles. Veamos cuantas hay sin que ninguna caja tenga más de una ficha. Esto equivale a escoger una muestra de n números sin reposición de los m que corresponden a los números de las cajas. Esto lo podemos hacer de V_n^m maneras, de modo que la probabilidad que buscamos es 0 si $m < n$ y

$$\frac{V_n^m}{m^n} = \frac{m(m-1)\dots(m-n+1)}{m^n} \quad \text{si } m \geq n.$$

◄

5. Si colocamos al azar n fichas en m cajas, ¿Cuál es la probabilidad de que la primera caja tenga exactamente k fichas?

► Para que esto ocurra, a exactamente k de las fichas tenemos que asignarle el número 1, y un número distinto de 1 al resto de las fichas. Las fichas a las cuales vamos a asignarle el número 1 las podemos escoger de $\binom{n}{k}$ maneras. A cada una de las $n-k$ fichas restantes podemos asignarle cualquiera de los $m-1$ números que no son 1, y esto lo podemos hacer de $(m-1)^{n-k}$ maneras. La probabilidad que buscamos es

$$\frac{\binom{n}{k}(m-1)^{n-k}}{m^n}.$$

◄

6. Tenemos 100 cartas numeradas del 1 al 100. Se mezclan las cartas y luego se van volteando una a una. ¿Cuál es la probabilidad de que la carta con el número j aparezca en el j-ésimo lugar al ser volteada?

► En este problema podemos pensar que el resultado de voltear las 100 cartas es un arreglo de 100 números entre 1 y 100, sin repetir ninguno:

$$(a_1, a_2, \dots, a_{100}), \quad 1 \leq a_i \leq 100, \quad i = 1, \dots, 100.$$

¿Cuántos arreglos podemos formar? Hay uno por cada permutación de los números de 1 a 100, por lo tanto hay 100! resultados posibles. ¿Cuántos de ellos tienen una j en el lugar j? Si fijamos el número que ocupa este lugar, quedan 99 números para distribuir en los 99 lugares restantes. Hay 99! maneras de hacer esto. Por lo tanto la probabilidad buscada es

$$\frac{99!}{100!} = \frac{1}{100}.$$

En general, si en lugar de 100 números hay n, la probabilidad de que en el j-ésimo lugar aparezca la carta con el número j es $1/n$.

◀

7. En el problema anterior, ¿Cuál es la probabilidad de que ninguna carta aparezca en su lugar?

▶ Llamemos A_j el evento que ocurre si la $j$ésima carta aparece en su lugar. Vimos en el ejercicio anterior que la probabilidad de este evento es $1/100$. Queremos calcular la siguiente probabilidad:

$$P(A_1^c \cap \cdots \cap A_{100}^c) = 1 - P(A_1 \cup \cdots \cup A_{100})$$

y para obtener esta última probabilidad podemos usar el principio de inclusión-exclusión que vimos en la sección 1.4 del primer capítulo:

$$P(A_1 \cup \cdots \cup A_{100}) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{101} P(A_1 \cap \cdots \cap A_{100}). \tag{2.14}$$

El primer término de esta ecuación es 1 porque hay 100 sumandos y cada unos de ellos vale $1/100$. Para el segundo término tenemos que si dos números específicos quedan fijos simultáneamente, el resto se puede permutar de $(100-2)!$ maneras, de modo que

$$P(A_i \cap A - j) = \frac{(100-2)!}{100!} = \frac{1}{100 \cdot 99}.$$

Nos falta contar el número de términos en la segunda suma, que es el número de maneras de seleccionar dos números enteros del 1 al 100, y esto lo podemos hacer de $\binom{100}{2}$ maneras distintas. Por lo tanto el segundo término de la ecuación (2.14) es

$$-\binom{100}{2} \frac{1}{100 \cdot 99} = -\frac{100 \cdot 99}{2!} \frac{1}{100 \cdot 99} = -\frac{1}{2!}.$$

Para el tercer término tenemos que para cualesquiera i, j, k fijos,

$$P(A_i \cap A_j \cap A_k) = \frac{(100-3)!}{100!} = \frac{1}{100 \cdot 99 \cdot 98}$$

y hay $\binom{100}{3}$ términos en la suma, de modo que el tercer término de (2.14) es

$$\binom{100}{3} \frac{1}{100 \cdot 99 \cdot 98} = \frac{100 \cdot 99 \cdot 98}{3!} \frac{1}{100 \cdot 99 \cdot 98} = -\frac{1}{3!}.$$

Continuando este razonamiento y sustituyendo los términos obtenidos en (2.14) obtenemos

$$P(A_1 \cup \cdots \cup A_{100}) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots - \frac{1}{100!} \approx 0.367879$$

y la probabilidad de que ningún número quede en su lugar es

$$1 - P(A_1 \cup \cdots \cup A_{100}) = \frac{1}{2!} - \frac{1}{3!} - \cdots + \frac{1}{100!} \approx 0.632121$$

Si en lugar de tener 100 cartas tenemos n, la probabilidad de que haya al menos un número en su lugar es

$$P(A_1 \cup \cdots \cup A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!},$$

y si llamamos p_n a la probabilidad de que ningún número quede en su lugar:

$$p_n = 1 - P(A_1 \cup \cdots \cup A_n) = \frac{1}{2!} - \frac{1}{3!} - \cdots + (-1)^n \frac{1}{n!}. \tag{2.15}$$

Recordemos el desarrollo de la función exponencial en serie de potencias:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \cdots.$$

Si ponemos $x = -1$ obtenemos

$$e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} + \cdots \approx 0.3678794 \tag{2.16}$$

y observamos que p_n converge a e^{-1}. Es interesante observar que los términos de la serie en (2.16) alternan de signo y como los denominadores son factoriales, que crecen muy rápidamente, las serie converge también muy rapidamente. Con $n = 5$, $p_n = 0.3666667$ y con $n = 10$, $p_n = 0.3678795$.

◀

8. Lanzamos un dado repetidamente hasta obtener el primer seis. ¿Cuál es la probabilidad de obtenerlo en el k-ésimo lanzamiento?

▶ Llamemos X al lugar en el cual ocurre el primer seis. Queremos calcular la probabilidad del evento $\{X = k\}$ y para ello observamos que hay 6^k resultados posibles de lanzar un dado k veces. Para que el primer seis ocurra en el lanzamiento k es necesario que en los primeros $k-1$ lanzamientos nunca ocurra un seis, y que en el k-ésimo ocurra un seis. Lo segundo sólo puede suceder de una manera, pero para lo primero tenemos 5^{k-1} posibilidades, ya que para cada uno de los $k-1$ lanzamientos hay 5 resultados posibles. Por lo tanto

$$P(X = k) = \frac{5^{k-1}}{6^k} = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6}.$$

Este resultado puede generalizarse de la siguiente manera: supongamos que realizamos una sucesión de experimentos cada uno de los cuales tiene dos resultados posibles: *éxito* y *fracaso* ó '1 ' y '0 ' con probabilidades respectivas p y $q = 1-p$. Llamemos X al lugar en el cual ocurre el primer éxito, ¿cuál es la probabilidad de que el primer éxito ocurra en el lugar k? La situación es similar a la del lanzamiento del dado si llamamos *éxito* a obtener un seis en el lanzamiento y ponemos $p = 1/6$. El resultado general, cuya demostración veremos más adelante, es

$$P(X = k) = q^{k-1}p.$$

Esta función de probabilidad se conoce como la *Distribución Geométrica* de parámetro p. Observamos que X puede tomar como valor cualquier entero positivo.

◀

2.13. Resumen.

Cuando escogemos elementos que pertenecen a un conjunto decimos que realizamos *un muestreo* y podemos hacerlo de acuerdo a diversos criterios: Con reposición de los elementos al conjunto antes de hacer la siguiente selección, o sin reposición. Podemos también tener en cuenta el orden en el cual hacemos la selección o no. Esto nos da cuatro posibilidades:

- **Muestreo con orden y con reposición:** Variaciones con repetición. Si queremos seleccionar k elementos de un conjunto de tamaño n, con reposición y en orden, lo podemos hacer de n^k maneras.

- **Muestreo con orden y sin reposición:** Variaciones. Si queremos seleccionar k elementos de un conjunto de tamaño n, sin reposición y en orden, es necesario que $k \leq n$ y lo podemos hacer de $V_k^n = n(n-1)\cdots(n-k+1)$ maneras. El caso particular $k = n$ se reduce a contar los posibles órdenes de los n elementos del conjunto, se conoce como las permutaciones de n y $V_n^n = n!$.

- **Muestreo sin orden y sin reposición:** Combinaciones. Si queremos seleccionar k elementos de un conjunto de tamaño n, sin reposición y sin orden, es necesario que $k \leq n$ y lo podemos hacer de $\binom{n}{k} = \frac{n!}{n!(n-k)!}$ maneras.

- **Muestreo sin orden y con reposición:** Este caso no tiene un nombre particular y es el más complicado de los cuatro, pero ya lo hemos encontrado anteriormente, en los problemas 7 y 8 de la sección 2.10. Si queremos seleccionar k elementos de un conjunto de tamaño n, con reposición y sin orden, ésto lo podemos hacer de $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$ maneras. Veamos esto: podemos pensar que a cada elemento del conjunto le asignamos una caja, de modo que en total tenemos n cajas. Tomamos ahora k bolas y las repartimos en las cajas, sin importar si hay cajas vacías o cajas con más de una bola. El número de bolas que hay en una caja es representa el número de veces que hemos seleccionado ese elemento en la muestra. Como vimos en el problema 8 de la sección 2.10, ésto efectivamente se puede hacer de $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$ maneras.

2.14. Comentarios y Algo de Historia.

1.- Blaise Pascal nació en Clermont, Francia, el 19 de junio de 1623. Publicó a los diecisiete años el brillante "*Ensayo Sobre las Cónicas*", se interesó posteriormente en el trabajo experimental de Torricelli y cultivó un considerable interés por la física experimental. Desarrolló una máquina calculadora, similar a las cajas registradoras de hace algunos años. Escribió, posteriormente, el "*Tratado del Triángulo Aritmético*" sobre los números combinatorios. Este triángulo, conocido ahora como el Triángulo de Pascal y que estudiamos en la sección 2.7, era conocido por los matemáticos europeos desde hacía al menos un siglo. Aparece, entre otros, en el trabajo de Stifel y Tartaglia. El crédito de Pascal no está en su descubrimiento sino en haber realizado un estudio sistemático y elegante de sus propiedades. Menos conocido, pero más importante, es el hecho de que Pascal introdujo la inducción matemática y aplicó esta técnica para demostrar resultados sobre los coeficientes binomiales.

El triángulo de Pascal era conocido por el poeta y matemático árabe Omar Khayam unos 550 años antes de Pascal y también aparece en el *Precioso Espejo de los Cuatro Elementos* escrito hacia 1.300 por el matemático chino Chu Shi Kei.

2.- Isaac Newton nació el 25 de diciembre de 1642 en Woolsthorpe Manor, Inglaterra. En 1661 ingresó en Trinity College de la Universidad de Cambridge, donde estudió con Isaac Barrow. En un manuscrito de 1665 presentó la fórmula para el desarrollo binomial con cualquier potencia y describió las ideas fundamentales de su método de *fluentes* y *fluxiones*, un método equivalente al Cálculo de Leibniz.

Su trabajo más importante fue "*Principios Matemáticos de la Filosofía Natural* ", publicado en Londres en 1.687. En este importante tratado se presenta la Ley de la Gravitación Universal, y las bases de la Mecánica Clásica, cuyos principios dominaron la física de los siglos XVIII y XIX.

Newton no trabajó en el área de probabilidad. La única contribución conocida aparece en lista de ejercicios (25) y fue un problema que le propuso Samuel Pepys y que respondió correctamente, aunque luego Pepys no se mostró muy dispuesto a aceptar su respuesta.

3.- El problema 7 de la sección 2.12 es un problema clásico en la historia de las probabilidades y fue propuesto inicialmente por Pierre Rémond de Montmort (1678-1719) en su tratado sobre probabilidades *Essay d'analyse sur les jeux de hazard* publicado en 1708. El juego original consistía en trece bolas numeradas que se sacaban en sucesión de una caja, y por esa razón se llamaba *treize* o trece. También era conocido como *rencontres* o coincidencias.

Ejercicios

1. ¿Cuántas diagonales tiene un polígono convexo de n lados?

2. a. ¿De cuántas maneras podemos escoger un comité de tres personas en un grupo de 20?

   b. ¿De cuántas maneras podemos escoger un presidente, un secretario y un tesorero?

3. Hay N niñas y N niños en una fiesta.

   a. ¿Cuántas parejas de baile de sexos distintos pueden formarse?

   b. ¿De cuántas maneras se pueden colocar en una fila de modo que los sexos se alternen?

4. Un examen tiene 12 preguntas que pueden ser respondidas *cierto* o $falso$. Si un estudiante decide responder *cierto* a seis de ellas, ¿de cuántas maneras puede hacerlo?

5. Con las letras de la palabra LIBRO, ¿Cuántas palabras de cinco letras o menos (con o sin sentido) pueden formarse? ¿Cuántas de ellas no tienen letras repetidas?

6. Calcule cuantas palabras con o sin sentido pueden formrse con las letras de las siguientes palabras.

   CUAUTITLAN, CUERAMARO, TLALNEPLANTLA TLACOQUEMECATL.

7. Se disponen 5 bolas blancas y 5 negras en 3 cajas numeradas. ¿Cuántas maneras hay de hacerlo?

8. ¿Cuántos números se pueden formar usando todos los dígitos 1, 2, 2, 3, 3, 3, 4, 4, 4, 4?

9. En una mesa de un restaurant seis personas ordenan arrachera, tres piden enchiladas, dos piden pollo y uno pide pasta.

   a. ¿De cuántas maneras pueden servirse los 12 platillos de modo que todos reciban lo que ordenaron?

   b. ¿De cuántas maneras pueden servirse de modo que nadie reciba lo que ordenó?

10. Una mano de POKER consiste de cinco cartas tomadas de un juego de barajas. ¿De cuántas maneras se puede obtener

    a. una escalera (cinco cartas en orden, sin importar la pinta; el As puede terminar la escalera pero no comenzarla)?

    b. un trío?

    c. un par?

    d. dos pares?

    e. un par y un trio (full house)?

    f. Halle la probabilidad de los eventos anteriores.

11. ¿Cuál es la probabilidad de que cuatro personas seleccionadas al azar hayan nacido en diferentes dias de la semana?

12. Se disponen en fila 2 bolas blancas y 6 bolas negras de modo que no haya dos bolas blancas consecutivas (la figura indica una manera posible). ¿Cuántas maneras hay de hacerlo?

    ● ● ○ ● ○ ● ●●

13. ¿De cuántas maneras pueden sentarse cinco hombres y cinco mujeres en una mesa redonda de modo que no haya dos hombres sentados uno al lado del otro?

14. ¿De cuántas maneras pueden sentarse cinco hombres y ocho mujeres en una mesa redonda si los hombres se sientan todos juntos?

15. Seleccionamos cuatro niños al azar y sin reposición de una familia que tiene exactamentes dos varones. La probabilidad de no escoger ningún varón es la mitad de seleccionar ambos ¿Cuántos niños en total hay en la familia?

16. ¿Cuántos números de cinco cifras tienen todas sus cifras de igual paridad (todas pares o todas impares)?

17. En una mesa rectángular los anfitriones se sientan en los extremos. ¿De cuántas maneras se pueden sentar

    a. seis invitados, tres a cada lado?

    b. cuatro mujeres y cuatro hombres, sentados cuatro a cada lado de modo que no haya dos personas del mismo sexo juntas?

    c. ocho invitados, cuatro a cada lado de la mesa, de modo que dos invitados específicos se sienten juntos?

18. ¿De cuántas maneras podemos escoger cuatro cartas de distintas pintas y distintos valores a partir de un juego de 52 cartas?

19. ¿Cuál es la probabilidad de que una mano de bridge (13 cartas) tenga los cuatro aces?

20. ¿Cuántas biyecciones hay de A a B, si ambos conjuntos tienen n elementos?

21. Determine los enteros n tales que $n! > 2^n$.

22. Un restaurante ofrece un menú con las siguientes posibilidades: cuatro sopas para escoger una, dos platillos principales, para escoger uno, dos acompañantes a escoger entre tres tipos de papas, tres tipos de vegetales y una ensalada, Cuatro postres para escoger uno y una bebida de tres.

    a. ¿Cuántas comidas diferentes se pueden ordenar si sólo hay un tipo de papa en cada orden, suponiendo que no se omite ningún tiempo?

    b. ¿Cuántas comidas diferentes se pueden ordenar si sólo hay un tipo de papa en cada orden, y se omite un tiempo distinto del plato principal?

23. Clara dice que es capaz de distinguir Pepsi Cola de Coca Cola por el sabor 75 % de las veces. Pedro piensa que Clara sólo está adivinando. Para determinar quién tiene la razón Clara debe probar 10 vasos en cada uno de los cuales hay alguna de las dos gaseosas, que ha sido seleccionada al azar lanzando una moneda. Clara gana si acierta 7 o más veces. Halle la probabilidad de que Clara gane si en realidad lo que dice es cierto. Halle la probabilidad de que Clara gane si en realidad está adivinando.

24. En un grupo de 12 personas hay dos de apellido Pérez. Si no importa el orden, ¿de cuántas maneras se pueden escoger siete personas a) sin restricciones? b) si se deben incluir los dos Pérez? c) sin incluir ningún Pérez? d) si sólo un Pérez se incluye? e) si al menos un Pérez se incluye? f) si a lo sumo un Pérez se incluye?

25. a. Halle la probabilidad de obtener al menos un seis al lanzar 6 dados. b. Halle la probabilidad de obtener al menos dos seises al lanzar 12 dados. c. Halle la probabilidad de obtener al menos tres seises al lanzar 18 dados. Compare los tres resultados.

26. a. Se colocan 3 bolas numeradas en 3 cajas numeradas. Hallar el número de maneras de hacerlo de modo que: (i) Al menos una caja quede vacía. (ii) Exactamente una caja quede vacía.

    b. Repetir el cálculo hecho en a. cuando hay n bolas y n cajas.

27. Las claves o Números de Identificación Personal (NIP) para las tarjetas bancarias tienen usualmente 4 cifras. Si una computadora asigna estos números al azar, ¿Cuál es la probabilidad de que las cuatro cifras sean diferentes? ¿Cuál es la probabilidad de que haya exactamente dos dígitos repetidos?

28. Los dados para jugar poker consisten de 5 dados con los símbolos $\{9, 10, J, Q, K, A\}$. Al lanzar estos cinco dados ¿Cuál es la probabilidad de obtener un 'full house' (un trío y una pareja? ¿Cuál es la probabilidad de obtener dos pares?

29. Una caja contiene ocho bolas, dos rojas, dos azules, dos blancas y dos negras. Las bolas se separan al azar en dos grupos de cuatro bolas cada uno ¿Cuál es la probabilidad de que cada conjunto tenga una bola de cada color?

30. Una caja contiene 100 bolas numeradas del 1 al 100. Se seleccionan al azar dos bolas sin reposición ¿Cuál es la probabilidad de que la suma de los números sea par?

31. Un amigo te invita jugar Aguila o Sol lanzando su moneda, pero tu piensas que la moneda no está balanceada. Le propones jugar de la siguiente manera: lanzas la moneda dos veces, si el resultado es AS tu ganas, si es SA tu amigo gana y si es AA o SS ninguno gana y se vuelve a lanzar la moneda hasta lograr dos resultados distintos. Supón que la probabilidad de que la moneda salga A es p. Halla la probabilidad de AS, SA, AA, SS con dos lanzamientos de la moneda. Usando esto demuestra que la probabilidad de ganar en el nuevo esquema es $1/2$.

32. Un medicamento tiene una efectividad desconocida p. Para estimar p se aplica el medicamento a n pacientes y se encuentra que resultó efectivo en m de ellos. El *pricipio de máxima verosimilitud* dice que debemos estimar p seleccionando el valor que maximiza la probabilidad de que ocurra lo que ocurrió en el experimento. Suponga que el experimento puede considerarse una serie de ensayos de Bernoulli con probabilidad de éxito p y demuestre que el estimador de máxima verosimilitud para p es m/n.

33. Veintidos caballos de madera distintos se van a colocar en un carrusel en dos círculos concéntricos. ¿De cuántas maneras se puede hacer esto si cada círculo debe tener 11 caballos y cada uno de ellos debe estar al lado de otro caballo?

34. Demuestre que a partir de un conjunto de n elementos se pueden formar 2^{n-1} subconjuntos con un número par de elementos.

35. Sea $\mathcal{A}$ la colección de subconjuntos de $\{1, 2, 3, \ldots, n\}$ que tienen tamaño par (por convención $\emptyset$ tiene tamaño par), y sea $\mathcal{B}$ la colección de los subconjuntos de tamaño impar. Establezca una biyección de $\mathcal{A}$ a $\mathcal{B}$, lo cual demuestra que ambos tienen la misma cardinalidad.

36. Demuestre que $\binom{n}{k}$ y $\binom{2n}{2k}$ tienen la misma paridad, es decir, ambos son pares o ambos son impares.

37. Demuestre que hay infinitas filas del triángulo de Pascal que consisten unicamente de números impares.

38. Sea a un número del Triángulo de Pascal. Demuestre que la suma de los números del triángulo que se encuentran dentro del paralelogramo limitado por los lados del triángulo y las diagonales que pasan por a (ver figura 2.10) es igual a $a - 1$

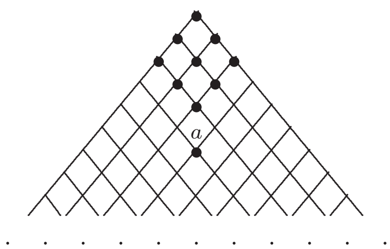


Figura 2.10

39. Demuestre las siguientes identidades

$$\binom{n}{0} - \binom{n}{1} + \cdots + (-1)^n \binom{n}{n} = 0, \qquad \binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}.$$

40. Sea x un elemento de un conjunto A de tamaño $2n$. Cuente los subconjuntos de A de n elementos que incluyen a x y los que lo excluyen. Use esto para demostrar que $\binom{2n}{n} = 2\binom{2n-1}{n-1}$

41. Sea $m = \binom{n}{2}$, demuestre que $\binom{m}{2} = 3\binom{n+1}{4}$.

42. Demuestre las siguientes identidades

$$a.\ \sum_{k=1}^{n} 2^{k-1} = 2^n - 1, \qquad b.\ \sum_{k=1}^{n-1} k = \binom{n}{2}, \qquad c.\ \sum_{k=1}^{n} k(n-k) = \binom{n}{3}.$$

43. **Fórmula de Van der Monde.** Demuestre que para m, n enteros y $r \leq m \wedge n$,

$$\sum_k \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}.$$

44. Demuestre que $\binom{n}{k} = \sum_{r=k}^{n} \binom{r-1}{k-1}$.

45. Considere todas las poligonales $(n, S_n), n \geq 0$ que parten del origen – es decir, que $S_0 = 0$ – y que, en cada paso, saltan una unidad hacia arriba o hacia abajo. Dicho de otra manera, $S_n = X_1 + X_2 + \cdots + X_n$, donde cada X_i vale 1 ó -1.

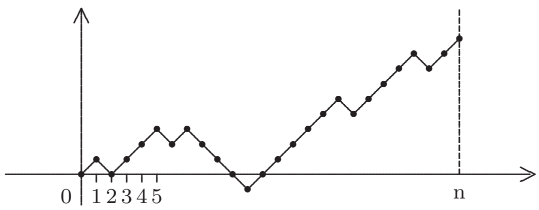


Figura 2.11

a. ¿Cuántas poligonales podemos construir en el intervalo de tiempo $[0, n]$?

b. ¿Cuántas poligonales satisfacen $S_n = 0$?

c. Usamos la notación $N_{n,h} = \sharp\{$ poligonales tales que $S_n = h\}$. Sea k un entero positivo y ℓ un entero no-negativo. Probar que

$$N_{n,k+\ell} = \sharp\{ \text{ poligonales tales que } S_n = k - \ell, \text{ y para algún } m \leq n \text{ se tiene } S_m = k\}$$

d. Sea k un entero positivo, probar que

$$\sharp\{\text{poligonales tales que } S_n = 0, \max_{1 \leq m \leq n} S_n \geq k\} = N_{n,2k}.$$

46. Agrupando un conjunto de n^2 puntos de dos maneras distintas, de una prueba combinatoria de $n^2 = 2\binom{n}{2} + n$.

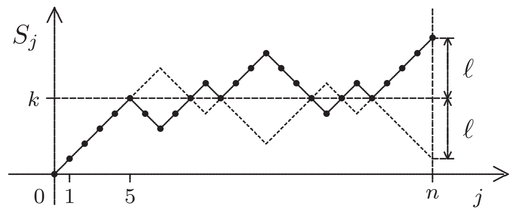


Figura 2.12

47. Demuestre que $\binom{k}{j}\binom{n}{k} = \binom{n}{j}\binom{n-j}{k-j}$ contando los elementos de un conjunto de dos maneras distintas.

48. Consideremos un tablero cuadriculado como el de la figura 2.13, con las columnas numeradas y las filas indicadas por letras. Supongamos que un punto se mueve sobre los nodos de modo que en cada movimiento se puede dirigir hacia adelante, a la izquierda o a la derecha, pero nunca se devuelve. El punto comienza en la intersección $J12$ viniendo de la calle I. ¿Cuántas rutas distintas hay para

    a. llegar a la intersección $L8$ después de 6 movimientos?

    b. regresar a $J12$ después de cuatro movimientos?

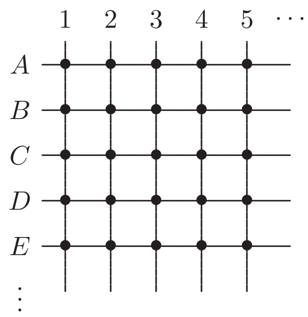


Figura 2.13

49. En el tablero del problema anterior queremos ir de $B2$ a $J10$ en el menor número posible de movimientos ¿Cuántas rutas posibles hay?

50. ¿Cuántos rectángulos (de cualquier tamaño se pueden formar usando los segmentos de una retícula con m rectas horizontales y n verticales. En la figura 2.14, $m = 4, n = 6$.

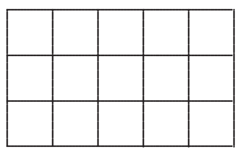


Figura 2.14

Capítulo 3

Probabilidad Condicional e Independencia

3.1. Introducción.

Consideremos una población de 20 estudiantes entre los cuales hay 14 que estudian Medicina y 6 que estudian Ingeniería. De esta población se escogen sin reposición dos estudiantes al azar y se consideran los eventos:

E_1: "El primer estudiante seleccionado estudia Medicina".

E_2: "El segundo estudiante seleccionado estudia Medicina".

El espacio muestral que consideramos consiste de la colección de todos los pares ordenados

$$(a_i, a_j); \quad (a_i, b_k); \quad (b_k, a_i); \quad (b_k, b_h)$$

donde los a_i son estudiantes de Medicina y los b_j son de Ingeniería, $i \neq j; k \neq h$; $i, j \leq 14; h, k \leq 6$. El número de eventos elementales es 20×19.

La siguiente tabla de doble entrada indica el número de puntos muestrales correspondientes a la partición de Ω según los eventos E_1, E_2 y sus complementos. En la última fila aparecen los totales correspondientes a cada columna y en la última columna los correspondientes a cada fila.

	E_2	E_2^c	
E_1	14×13	14×6	14×19
E_1^c	6×14	6×5	6×19
	14×19	6×19	20×19

Tabla 3.1

Utilizando este cuadro podemos calcular fácilmente las probabilidades de eventos tales como

$$P(E_1 \cap E_2) = \frac{14 \times 13}{20 \times 19}; \qquad P(E_1) = \frac{14 \times 19}{20 \times 19}; \qquad P(E_1^c \cap E_2) = \frac{6 \times 14}{20 \times 19}.$$

Si suponemos que el primer estudiante estudia Medicina, ¿cuál es la probabilidad de que el segundo también?

En este caso vemos, a partir de la tabla, que hay 14×19 resultados posibles, de los cuales 14×13 son favorables al evento E_2 y por lo tanto la probabilidad que deseamos calcular es

$$\frac{14 \times 13}{14 \times 19} = \frac{(14 \times 13)/(20 \times 19)}{(14 \times 19)/(20 \times 19)} = \frac{P(E_1 \cap E_2)}{P(E_1)}.$$

La probabilidad que hemos calculado se llama "*probabilidad condicional de E_2 dado E_1*" y se denota $P(E_2|E_1)$.

Observamos que $P(E_2) = \frac{14 \times 19}{20 \times 19} = \frac{7}{10}$ no coincide con $P(E_2|E_1) = \frac{13}{19}$. Al saber que ocurrió E_1 disponemos de cierta información adicional que modifica nuestro espacio muestral: la nueva población, para la segunda extracción, no coincide con la original, ya que sólo quedan 13 estudiantes de Medicina de un total de 19 estudiantes posibles.

Notemos además que si las extracciones se realizan con reposición esto no ocurre, ya que el resultado de la primera extracción no nos da ninguna información sobre la segunda. En este caso se tiene:

$$P(E_2|E_1) = P(E_2) = \frac{7}{10}.$$

Definición 3.1 Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $B \in \mathcal{A}$ tal que $P(B) > 0$. Definimos una nueva función $P(\cdot|B)$ de $\mathcal{A}$ en $\mathbb{R}$ mediante

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{para todo } A \in \mathcal{A}.$$

Esta función $P(\cdot|B)$ que hemos definido es una probabilidad; en efecto:

1. $P(A|B) \geq 0$ para todo $A \in \mathcal{A}$.

2. $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

3. Sean $A_1, A_2, \ldots$ conjuntos disjuntos en $\mathcal{A}$, entonces

$$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{P(B \cap \bigcup_{i=1}^{\infty} A_i)}{P(B)} &= \frac{P(\bigcup_{i=1}^{\infty} B \cap A_i)}{P(B)} \\ = \frac{\sum_{i=1}^{\infty} P(B \cap A_i)}{P(B)} &= \sum_{i=1}^{\infty} P(A_i|B). \end{aligned}$$

$P(\cdot|B)$ se llama probabilidad condicional dado B.

Dos propiedades elementales de la probabilidad condicional son las siguientes:

- Si A y B son disjuntos, entonces $P(A|B) = 0$. En efecto,

$$A \cap B = \emptyset \Rightarrow P(A \cap B) = 0 \quad \text{y} \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = 0.$$

- Si $B \subset A$ entonces $P(A|B) = 1$ ya que en este caso $P(A \cap B) = P(B)$.

Ejemplos

1. Se lanzan dos dados hasta que la suma de los puntos sea 7 u 8. Si sale 7 gana el jugador A, si sale 8 gana B. ¿Cuál es la probabilidad de que gane A?

► Vamos a resolver este problema de dos maneras. Supongamos primero que al cabo de n lanzamientos gana A. Esto quiere decir que en los $n-1$ primeros lanzamientos no salió ni 7 ni 8, y que en el n-ésimo se obtuvo 7. Este evento tiene probabilidad

$$p_n = \frac{1}{6}\left(\frac{25}{36}\right)^{n-1}.$$

De esta manera, la probabilidad de que A gane es:

$$\begin{aligned} P(A) = \sum_{n=1}^{\infty} p_n &= \sum_{n=1}^{\infty} \frac{1}{6}\left(\frac{25}{36}\right)^{n-1} \\ &= \frac{1}{6}\sum_{n=0}^{\infty}\left(\frac{25}{36}\right)^{n} \\ &= \frac{1}{6}\,\frac{1}{1-\frac{25}{36}} = \frac{6}{11}. \end{aligned}$$

Otra forma de resolver el problema es observar que la probabilidad de que gane A es la probabilidad condicional de que la suma de puntos sea 7, dado que el juego termina en un cierto lanzamiento, es decir, dado que la suma es 7 u 8 (hay que observar que la probabilidad de que el juego no termine es cero) o sea

$$P(A) = P(\{7\}|\{7,8\}) = \frac{P(\{7\})}{P(\{7,8\})} = \frac{\frac{6}{36}}{\frac{11}{36}} = \frac{6}{11}.$$

◄

2. Se lanza un dado dos veces.

a) Si la suma de los resultados es 8, ¿cuál es la probabilidad de que el primer lanzamiento haya resultado en $k,\ 1 \le k \le 6$?

b) Si el primer lanzamiento resultó en 3, ¿cuál es la probabilidad de que el segundo sea $k,\ 1 \le k \le 6$?

c) Si el primer lanzamiento resultó en 3, ¿cuál es la probabilidad de que la suma de ambos lanzamientos sea 7?

► Sea X el resultado del primer lanzamiento e Y el del segundo. Sabemos que $P(X = k) = P(Y = k) = 1/6,\ 1 \le k \le 6$.

a) Queremos calcular

$$P(X = k|X + Y = 8) = \frac{P((X = k) \cap (X + Y = 8))}{P(X + Y = 8)}.$$

Veamos primero la probabilidad en el denominador. Observamos que hay 5 resultados cuya suma es ocho de un total de 36 resultados posibles, que corresponden a los pares ordenados $(2, 6)$; $(3, 5)$; $(4, 4)$; $(5, 3)$; $(6, 2)$ y por lo tanto la probabilidad en el denominador vale $5/36$. Por otro lado, la probabilidad en el numerador vale 0 si tomamos $k = 1$. Para $2 \le k \le 6$ hay un solo resultado del segundo lanzamiento para el cual la suma es ocho: $Y = 8 - k$ y en consecuencia la probabilidad en el numerador es $1/36$. Finalmente tenemos

$$P(X = k|X + Y = 8) = \begin{cases} 1/5 & \text{si } 2 \le k \le 6, \\ 0 & \text{si } k = 1. \end{cases}$$

b) Ahora nos interesa calcular

$$P(Y = k|X = 3) = \frac{P((Y = k) \cap (X = 3))}{P(X = 3)}.$$

Sabemos que $P(X = 3) = 1/6$. Para calcular la probabilidad del numerador observamos que de un total de 36 resultados posibles, sólo uno corresponde al evento $(Y = k) \cap (X = 3)$ y por lo tanto esta probabilidad es $1/36$. En consecuencia

$$P(Y = k|X = 3) = \frac{1/36}{1/6} = \frac{1}{6}.$$

Este resultado es igual a $P(Y = k)$ y concuerda con lo que uno esperaría intuitivamente, ya que saber el resultado del primer lanzamiento no debe afectar en manera alguna el resultado del segundo.

c) Nos interesa ahora

$$P(X + Y = 7|X = 3) = \frac{P((Y + Y = 7) \cap (X = 3))}{P(X = 3)}$$

pero

$$(X + Y = 7) \cap (X = 3) = (3 + Y = 7) \cap (X = 3) = (Y = 4) \cap (X = 3),$$

por lo tanto

$$P(X + Y = 7|X = 3) = \frac{P((Y = 4) \cap (X = 3))}{P(X = 3)}$$

y por el resultado de la parte b de este ejercicio sabemos que esta probabilidad es $1/6$. ◀

3. Consideremos ahora una situación que se presenta con frecuencia en casos de controles masivos aplicados en prevención médica y control de calidad de productos.

   Para controlar una cierta enfermedad en una población donde la proporción de enfermos es p se usa un determinado examen médico para detectar enfermos. Se sabe que la probabilidad de que al aplicar el examen a un enfermo lo muestre como tal es de 0.90, y que la probabilidad de que el examen aplicado a una persona sana la señale como enferma es 0.01. Calcular la probabilidad de que una persona esté realmente enferma si el examen médico lo mostró como tal.

▶ Para responder esta pregunta elegimos al azar una persona en la población y consideramos los eventos

E: "la persona está enferma".

R: "el examen la detecta como enferma".

Queremos calcular

$$P(E|R) = \frac{P(E \cap R)}{P(R)}.$$

Sabemos que

$$\begin{aligned} P(E) &= p, \\ P(R|E) &= \frac{P(E \cap R)}{P(E)} = 0.90, \\ P(R|E^c) &= \frac{P(E^c \cap R)}{P(E^c)} = 0.01. \end{aligned}$$

A partir de las dos primeras igualdades obtenemos

$$P(E \cap R) = 0.90p$$

y a partir de la tercera

$$P(E^c \cap R) = 0.01(1 - p).$$

Por lo tanto

$$\begin{aligned} P(R) &= P(R \cap E) + P(R \cap E^c) \\ &= 0.90p + 0.01(1-p), \end{aligned}$$

y en consecuencia

$$P(E|R) = \frac{0.90p}{0.90p + 0.01(1-p)} = \frac{90p}{89p+1}.$$

En particular, si $p = \frac{1}{30}$ resulta

$$P(E|R) \simeq 0.76.$$

Llama la atención que esta probabilidad no sea muy próxima a 1, como uno esperaría. Analizando el comportamiento de $P(E|R)$ como función de p (ver figura 3.1) observamos que si la proporción p de enfermos en la población es pequeña, el método de control masivo es insuficiente, dado que $P(E|R)$ está lejos de 1. Por ejemplo, si $p = 0.001$, resulta $P(E|R) \simeq 0.083$.

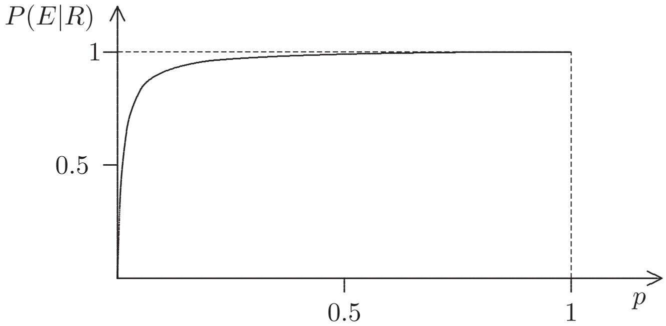


Figura 3.1

Los datos de este ejemplo han sido elegidos arbitrariamente, pero observamos, sin embargo, que cualquier método de control que pueda implicar errores en la clasificación considerada, está sujeto a dificultades como la que hemos encontrado anteriormente, que requieren un análisis previo a su aplicación. ◀

4. Estudiemos la distribución del tiempo de vida de un aparato eléctrico bajo la hipótesis de que el aparato no envejece, sino que se destruye por una perturbación aleatoria. Lo que queremos decir es que el objeto en cuestión cumple con la siguiente condición: dado que el aparato está en funcionamiento en el instante s, se comporta respecto a su tiempo de vida futuro a partir de ese instante como si hubiera comenzado a funcionar en s, es decir que

$$P(T > s + h | T > s) = P(T > h), \qquad h \geq 0,$$

o sea

$$\frac{P(T > s + h)}{P(T > s)} = P(T > h).$$

Si llamamos $\varphi(t) = P(T > t)$, la igualdad anterior se puede escribir

$$\varphi(s + h) = \varphi(s)\varphi(h) \tag{3.1}$$

para cualquier par de números positivos s y h. Si φ no es identicamente nula veamos que

$$\varphi(t) = e^{-\lambda t} \qquad \text{donde } t \geq 0, \lambda \geq 0.$$

► De (3.1), si $s = h = 0$, resulta
$$\varphi(0) = (\varphi(0))^2$$
de donde $\varphi(0) = 0$ ó $\varphi(0) = 1$. Ahora bien,
$$\varphi(0) = 0 \Rightarrow \varphi(s) = \varphi(0)\varphi(s) = 0, \text{ para todo } s \geq 0$$
y tenemos una solución trivial.

Consideremos ahora el caso $\varphi(0) = 1$. A partir de (3.1) deducimos
$$\varphi(ns) = \varphi(s + s + \cdots + s) = (\varphi(s))^n, \text{ para todo } s \geq 0.$$
Si $s = 1/n$ resulta
$$\varphi(1/n) = (\varphi(1))^{\frac{1}{n}}$$
poniendo ahora $s = n/m$, se tiene
$$\varphi(n/m) = (\varphi(1/m))^n = (\varphi(1))^{\frac{n}{m}},$$
es decir que
$$\varphi(t) = (\varphi(1))^t \qquad \text{si } t \in \mathbb{Q},\ t \geq 0.$$
Falta mostrar que la relación anterior es cierta para cualquier número real positivo t. Sean (t_k), (t'_k) dos sucesiones de racionales positivos tales que
$$t_k < t < t'_k, \quad t_k \to t, \quad t'_k \to t,$$
entonces, dado que la función φ es monótona decreciente, resulta
$$(\varphi(1))^{t_k} = \varphi(t_k) \leq \varphi(t) \leq \varphi(t'_k) = (\varphi(1))^{t'_k}$$
haciendo $k \to \infty$ tenemos
$$(\varphi(1))^t \leq \varphi(t) \leq (\varphi(1))^t$$
y por lo tanto $\varphi(t) = (\varphi(1))^t$. Poniendo $\varphi(1) = e^{-\lambda}$ obtenemos
$$\varphi(t) = e^{-\lambda t}$$
donde λ debe ser positivo para que φ sea decreciente. Concluimos entonces que $P(T > t) = e^{-\lambda t}$ y decimos que en este caso que T tiene distribución exponencial.

Como ejemplo numérico de esta situación supongamos que el tiempo de vida T medido en horas de una cierta marca de focos tiene una distribución exponencial con parámetro $\lambda = 0.001$.

a) Calcule la probabilidad de que un foco de esta marca dure 2000 h.

b) Calcule la probabilidad de que dure 2000 h. dado que ha durado 1000 h.

a) $P(T \geq 2000) = e^{-0.001 \times 2000} = e^{-2} \approx 0.13533$.

b) $P(T \geq 2000 | T \geq 1000) = \frac{P((T\geq 2000)\cap(T\geq 1000))}{P(T\geq 1000)} = \frac{P(T\geq 2000)}{P(T\geq 1000)} = \frac{e^{-2}}{e^{-1}} = e^{-1} \approx 0.367879$.

◄

5. Consideremos las familias que tienen dos niños y supongamos que varones y hembras son igualmente probables. Si escogemos al azar una familia y en ella hay un hijo varón, ¿cuál es la probabilidad de que el otro hijo sea también varón?

► En este problema el espacio muestral es

$$\Omega = \{(v,v); (v,h); (h,v); (h,h)\}$$

donde v es varón, h es hembra y el orden del par es el orden de los nacimientos. Cada uno de los puntos tiene probabilidad $1/4$.

Al considerar este problema uno podría llegar a la conclusión de que, como ambos sexos son igualmente probables, la probabilidad que buscamos es $1/2$. Veremos, sin embargo, que esto no es cierto. Sea

$$A : \text{“uno de los hijos es varón''}$$
$$B : \text{“ambos hijos son varones''}$$

entonces $B \subset A$ y $A \cap B = B$. Por lo tanto

$$P(B|A) = \frac{P(B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3},$$

que es la respuesta correcta al problema. ◄

6. Veamos un problema parecido al anterior. Consideremos de nuevo las familias que tienen dos hijos. Si escogemos un niño al azar entre estas familias y resulta ser varón, ¿cuál es la probabilidad de que el otro niño en la familia también sea varón?

► En este ejercicio, una representación apropiada del espacio muestral es la siguiente:

$$\Omega' = \{h_v, h_h, v_h, v_v\}$$

donde los puntos del espacio muestral no son ahora las familias sino los hijos en las familias y h_v : “hembra que tiene un hermano”, h_h : “hembra que tiene una hermana”, v_h : “varón que tiene una hermana” y v_v : “varón que tiene un hermano”. Sea

$$A' : \text{“ el niño escogido es varón''}$$
$$B' : \text{“ el niño escogido tiene un hermano''}$$

entonces

$$A' \cap B' = \{v_v\}$$

y por lo tanto

$$P(B'|A') = \frac{P(A' \cap B')}{P(A')} = \frac{1/4}{1/2} = \frac{1}{2}.$$

◄

7. **El Problema de Selección Optima.** Vamos a considerar la siguiente situación: tenemos n objetos del mismo tipo (por ejemplo, televisores, teléfonos celulares, etc.) y suponemos que si nos dan cualquier par de objetos, siempre podemos decidir cuál es mejor, sin que hayan empates. Queremos seleccionar el mejor de ellos con las siguientes restricciones: nos ofrecen los objetos uno a uno al azar y en cada caso debemos decidir si lo aceptamos o no. Si lo aceptamos tenemos que quedarnos con ese objetos. Si no lo aceptamos se nos muestra el siguiente, pero los objetos rechazados no pueden ser escogidos posteriormente.

   Adoptamos la siguiente regla general: ‘nunca escogemos un objeto que sea de calidad inferior a los que hemos rechazado previamente’. Entonces podemos escoger el primer objeto y detener la búsqueda, o rechazamos el primer objeto y examinamos los posteriores con la esperanza de obtener alguno mejor a lo largo de la sucesión. Por supuesto, es posible que con esta estrategia nunca seleccionemos el mejor objeto, o que lo rechacemos inicialmente y nunca podamos hacer una selección. Por otro lado, si el número de objetos es grande, es razonable rechazar el primero con la esperanza de encontrar uno mejor posteriormente.

Supongamos que, siguiendo la regla descrita anteriormente, seleccionamos el i-ésimo objeto, de modo que este objeto debe ser mejor que todos los anteriores. ¿Cuál es la probabilidad de que el objeto que escogemos sea realmente el mejor de todos?

► Sea B el evento que el i-ésimo objeto sea el mejor de los observados hasta ese momento y sea A el evento que ocurre si el i-ésimo objeto es el mejor de los n objetos. Queremos calcular $P(A|B)$, para lo cual necesitamos $P(B)$ y $P(A \cap B)$. Tenemos que $A \subset B$, de modo que $A \cap B = A$ y $P(A \cap B) = P(A)$. Por hipótesis los objetos se presentan al azar, de modo que todos los posibles órdenes tienen la misma probabilidad. Por lo tanto $P(B)$ es la probabilidad de que un objeto dado, el mejor de todos, ocupe el lugar i. Como hay $i!$ permutaciones de i objetos y hay $(i-1)!$ permutaciones si fijamos el lugar de un objeto, la probabilidad es

$$P(B) = \frac{(i-1)!}{i!} = \frac{1}{i}.$$

De manera similar $P(A)$ es la probabilidad de que en una permutación hecha al azar de n objetos distinguibles, un objeto dado, el mejor, ocupe el i-ésimo lugar, de modo que

$$P(A) = \frac{(m-1)!}{m!} = \frac{1}{m}.$$

Por lo tanto la probabilidad condicional es

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{i}{m}.$$

◄

3.2. Resultados Básicos sobre Probabilidad Condicional.

Estudiaremos en esta sección algunas propiedades sencillas pero fundamentales de la probabilidad condicional.

Proposición 3.1 *Para cualquier colección finita de eventos $A_1, \dots, A_n$ se tiene*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

siempre que $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$.

Demostración. Como

$$P(A_1) \geq P(A_1 \cap A_2) \geq \cdots \geq P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$$

todas las probabilidades condicionales que aparecen están bien definidas. Si escribimos explícitamente el segundo miembro de la igualdad obtenemos

$$P(A_1)\frac{P(A_1 \cap A_2)}{P(A_1)} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})}$$

y simplificando se obtiene el primer miembro de la igualdad. ■

Como ejemplo consideremos el siguiente problema.

1. Se extraen sin reposición tres cartas de un juego completo de barajas. Calcule la probabilidad de que ninguna sea trébol.

▶ Sea A_i: "la i-ésima carta no es trébol". Queremos calcular

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = \frac{39}{52}\frac{38}{51}\frac{37}{50} = 0.4135$$

◀

Una de las propiedades más útiles de la probabilidad condicional es la Ley de la Probabilidad Total, que demostramos a continuación.

Proposición 3.2 (Ley de la Probabilidad Total) *Sea $B_1, B_2, \dots$ una familia finita o numerable de conjuntos disjuntos dos a dos cuya unión es Ω. Entonces, para cualquier evento A,*

$$P(A) = \sum_i P(B_i)P(A|B_i)$$

donde la suma se toma sobre todos los índices i para los cuales $P(B_i) > 0$.

Demostración.

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(A \cap (\cup_i B_i)) \\ &= P(\cup_i (A \cap B_i)) = \sum_i P(A \cap B_i) \\ &= \sum_i P(B_i)P(A|B_i). \end{aligned}$$

■

Veamos ahora algunos ejemplos.

2 En una bolsa se colocan n tarjetas con nombres escritos en ellas y se extraen dos, sucesivamente y sin reposición. Si $m < n$ de los nombres son de mujer, calcule la probabilidad de que el segundo nombre extraído sea de mujer.

▶ Sea A el evento cuya probabilidad deseamos calcular y F: "el primer nombre extraído es femenino". Los eventos F y F^c son disjuntos y forman una partición del espacio muestral. En consecuencia

$$P(A) = P(A|F)P(F) + P(A|F^c)P(F^c)$$

y

$$P(F) = \frac{m}{n}; \quad P(F^c) = \frac{n-m}{n}; \quad P(A|F) = \frac{m-1}{n-1}; \quad P(A|F^c) = \frac{m}{n-1}.$$

Por lo tanto

$$\begin{aligned} P(A) &= \frac{(m-1)}{(n-1)}\frac{m}{n} + \frac{m}{(n-1)}\frac{(n-m)}{n} \\ &= \frac{m}{(n-1)n}(m-1+n-m) = \frac{m}{n}. \end{aligned}$$

◀

3 Cada vez que el repartidor de una pizzeria regresa a buscar los pedidos para repartir, se encuentra que pueden haber entre 1 y 5 encargos esperando, y cada una de estas posibilidades tiene la misma probabilidad. Si, en promedio, la mitad de los clientes le da propina, calcule la probabilidad de que obtenga al menos dos propinas en un viaje.

▶ Sea A: "obtiene al menos dos propinas en el viaje" y sean B_i para $i = 1, \dots, 5$ los eventos "hay i encargos por repartir al iniciar el viaje", respectivamente. Estos conjuntos B_i forman una partición del espacio muestral y por lo tanto

$$P(A) = \sum_{i=1}^{5} P(A|B_i)P(B_i).$$

Sabemos que $P(B_i) = 1/5$ y además, si comienza con i encargos, el número de propinas tiene distribución binomial con parámetros $n = i$ y $p = 1/2$ (ver ejemplo 2.4.4). Las probabilidades condicionales son:

$$\begin{aligned}
P(A|B_1) &= 0, \qquad P(A|B_2) = \left(\frac{1}{2}\right)^2, \\
P(A|B_3) &= P(2 \text{ éxitos en 3 intentos}) + P(3 \text{ éxitos en 3 intentos}) \\
&= \binom{3}{2}\frac{1}{2^3} + \binom{3}{3}\frac{1}{2^3} = \frac{1}{2}, \\
P(A|B_4) &= \left[\binom{4}{2} + \binom{4}{3} + \binom{4}{4}\right]\frac{1}{2^4}, \\
P(A|B_5) &= \left[\binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5}\right]\frac{1}{2^5},
\end{aligned}$$

y finalmente obtenemos $P(A) = \frac{9}{20}$. ◀

4 **La Ruina del Jugador.** Un jugador con capital inicial k juega en un casino una sucesión de juegos hasta lograr un capital m o hasta que se arruine. Si la probabilidad de ganar en cada juego es p ¿Cuál es la probabilidad de que se arruine?

▶ La probabilidad de ruina depende de k y de m. Supongamos que m está fijo y sea u_k la probabilidad de ruina del jugador si al comienzo tiene un capital de k unidades. Si el jugador gana en el primer juego, la probabilidad de que se arruine es ahora u_{k+1}, mientras que si pierde, la probabilidad de ruina es u_{k-1}.

Sea G_1 el evento "el jugador gana el primer juego", mientras que R es el evento "el jugador se arruina". Entonces

$$P(R|G_1) = u_{k+1}, \qquad P(R|G_1^c) = u_{k-1}.$$

Por lo tanto,

$$\begin{aligned}
P(R) &= P(R|G_1)P(G_1) + P(R|G_i^c)P(G_1^c) \\
&= pP(R|G_1) + qP(R|G_1^c),
\end{aligned}$$

es decir

$$u_k = pu_{k+1} + qu_{k-1}, \qquad 1 \le k \le m-1,$$

con las condiciones iniciales

$$u_0 = 1, \qquad u_m = 0.$$

Para resolver esta ecuación introducimos las diferencias $\delta_k = u_k - u_{k-1}$ para $k = 1, 2, \dots, m$. Como $p + q = 1$ tenemos

$$u_k = (p+q)u_k = pu_{k+1} + qu_{k-1}$$

de donde obtenemos

$$p(u_{k+1} - u_k) = q(u_k - u_{k-1})$$

es decir

$$p\delta_{k+1} = q\delta_k \quad \Rightarrow \quad \delta_{k+1} = \frac{q}{p}\delta_k$$

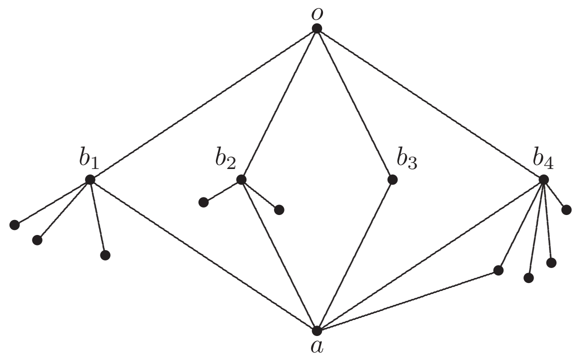


Figura 3.2

Iterando esta relación obtenemos que

$$\delta_{k+1} = \Big(\frac{q}{p}\Big)^{k-1}\delta_1$$

Para regresar a las u_k observamos que $\delta_1 = u_1 - u_0 = u_1 - 1$ y $\delta_1 + \delta_2 + \cdots + \delta_k = u_k - 1$. Por lo tanto,

$$\begin{aligned} u_k &= 1 + \delta_1 + \delta_2 + \cdots + \delta_k \\ &= 1 + \delta_1 + \Big(\frac{q}{p}\Big)\delta_1 + \cdots + \Big(\frac{q}{p}\Big)^{k-1}\delta_1 \\ &= 1 + \Big[1 + \Big(\frac{q}{p}\Big) + \cdots + \Big(\frac{q}{p}\Big)^{k-1}\Big]\delta_1 \end{aligned}$$

Falta determinar δ_1. Como $u_m = 0$,

$$0 = u_m = 1 + \Big[1 + \Big(\frac{q}{p}\Big) + \cdots + \Big(\frac{q}{p}\Big)^{m-1}\Big]\delta_1$$

y obtenemos

$$\delta_1 = \frac{-1}{1 + \big(\frac{q}{p}\big) + \cdots + \big(\frac{q}{p}\big)^{m-1}}$$

de modo que

$$\begin{aligned} u_k &= 1 - \frac{1 + \big(\frac{q}{p}\big) + \cdots + \big(\frac{q}{p}\big)^{k-1}}{1 + \big(\frac{q}{p}\big) + \cdots + \big(\frac{q}{p}\big)^{m-1}} \\ &= \begin{cases} 1 - \frac{k}{m} = \frac{m-k}{m} & \text{si } p = q = 1/2 \\ 1 - \frac{1-(q/p)^k}{1-(q/p)^m} = \frac{(q/p)^k - (q/p)^m}{1-(q/p)^m} & \text{si } p \neq q \end{cases} \end{aligned}$$

◀

5 Un caminante sale del punto o y escoge uno de los caminos ob_1, ob_2, ob_3, ob_4 al azar (ver figura 3.2). En cada uno de los cruces siguientes de nuevo escoge al azar. ¿cuál es la probabilidad de que el caminante llegue al punto a?

▶ Sea $B_k,\ k = 1, 2, 3, 4$, el evento "el caminante pasa por el punto b_k", estos eventos forman una partición de Ω y además son equiprobables. Por lo tanto $P(B_k) = 1/4,\ k = 1, 2, 3, 4$.

Si el caminante llega a b_1, hay cuatro caminos que puede escoger y sólo uno de ellos lo lleva a a, de donde la probabilidad condicional de llegar a a pasando por b_1 es $1/4$. Si A es el evento "el caminante llega a a", tenemos

$$P(A|B_1) = \frac{1}{4}.$$

De manera similar obtenemos

$$P(A|B_2) = \frac{1}{3}; \quad P(A|B_3) = 1; \quad P(A|B_4) = \frac{2}{5}.$$

Por lo tanto, la probabilidad de llegar a a es

$$P(A) = \sum_{i=1}^{4} P(A|B_i)P(B_i) = \frac{1}{4}\left(\frac{1}{4} + \frac{1}{3} + 1 + \frac{2}{5}\right) = \frac{119}{240}.$$

◀

6 ¿Cuál es la probabilidad de obtener 6 números distintos al lanzar 6 dados simétricos?

▶ Consideremos los siguientes eventos:

E_1: "se obtiene cualquier resultado en el primer lanzamiento".

E_2: "el segundo dado es distinto al primero".

E_3: "el tercer dado es distinto a los dos primeros".

y así sucesivamente. Entonces tenemos:

$$\begin{aligned} P(E_1) &= 1 \\ P(E_2|E_1) &= 5/6 \\ P(E_3|E_1 \cap E_2) &= 4/6 \\ &\vdots \\ P(E_6|E_1 \cap E_2 \cap \cdots \cap E_5) &= 1/6 \end{aligned}$$

y usando la proposición 3.1

$$P(E_1 \cap E_2 \cap \cdots \cap E_6) = 1 \times \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} = \frac{6!}{6^6}.$$

◀

7 Se extraen dos cartas al azar de un juego de barajas. Calcule la probabilidad de que las cartas sean un As y un diez.

▶ Consideremos los siguientes eventos:

A_1 : "la primera carta es un As" $\quad B_1$: "la primera carta es un diez"

A_2 : "la segunda carta es un As" $\quad B_2$: "la segunda carta es un diez"

C : "se extraen un As y un diez".

Tenemos

$$\begin{aligned} P(C) &= P(B_2|A_1)P(A_1) + P(A_2|B_1)P(B_1) \\ &= \frac{4}{51} \times \frac{4}{52} + \frac{4}{51} \times \frac{4}{52} \\ &= \frac{8}{663}. \end{aligned}$$

◀

3.3. El Teorema de Bayes

El otro resultado fundamental sobre probabilidad condicional se conoce como el Teorema de Bayes y es útil en situaciones en las cuales se conocen probabilidades de la forma $P(A|B_i)$ y $P(B_i)$ y se desea determinar $P(B_i|A)$.

Proposición 3.3 (Teorema de Bayes) *Sea $B_1, B_2, \dots$ una partición finita o numerable de Ω y sea A un evento cualquiera con $P(A) > 0$. Entonces*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}.$$

Demostración. Por la definición de probabilidad condicional tenemos:

$$P(B_i|A) = \frac{(A \cap B_i)}{P(A)} \tag{3.2}$$

y

$$P(A \cap B_i) = P(A|B_i)P(B_i). \tag{3.3}$$

Por la proposición 3.2 tenemos

$$P(A) = \sum_j P(A|B_j)P(B_j). \tag{3.4}$$

Usando (3.3) y (3.4) en (3.2) obtenemos el resultado. ■

Ejemplos.

1. De cien pacientes en un hospital con una cierta enfermedad se escogen diez para usar un tratamiento que aumenta la probabilidad de sanar de 0.50 a 0.75. Si posteriormente un doctor encuentra a un paciente curado, ¿cuál es la probabilidad de que haya recibido el tratamiento?

► Sean

C: "el paciente está curado".

T: "el paciente recibió el tratamiento".

A partir de la información dada obtenemos,

$$P(T) = \frac{10}{100} = 0.1; \qquad P(T^c) = \frac{90}{100} = 0.9;$$
$$P(C|T) = 0.75; \qquad P(C|T^c) = 0.50.$$

Usando el Teorema de Bayes

$$\begin{aligned} P(T|C) &= \frac{P(C|T)P(T)}{P(C|T)P(T) + P(C|T^c)P(T^c)} \\ &= \frac{(0.75)(0.1)}{(0.75)(0.1) + (0.5)(0.9)} = \frac{1}{7}. \end{aligned}$$

◄

2. Tres cajas contienen dos monedas cada una. En la primera, C_1, ambas son de oro; en la segunda, C_2, ambas son de plata y en la tercera, C_3, una es de oro y otra es de plata. Se escoge una caja al azar y luego una moneda también al azar. Si la moneda es de oro, ¿cuál es la probabilidad de que venga de la caja que contiene dos monedas de oro?

► Sabemos que $P(C_i) = 1/3$. Sea O: "se escoge una moneda de oro". Usando el Teorema de Bayes,

$$\begin{aligned} P(C_1|O) &= \frac{P(O|C_1)P(C_1)}{P(O|C_1)P(C_1) + P(O|C_2)P(C_2) + P(O|C_3)P(C_3)} \\ &= \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3}} = \frac{2}{3} \end{aligned}$$

Si uno considera el problema sin mucho cuidado podría razonar de la siguiente manera: luego de sacar una moneda de oro sólo hay dos cajas de las cuales pudo provenir, y como estamos escogiendo las cajas al azar, la probabilidad de que venga de C_1 es $1/2$. El problema con este razonamiento es que es más probable sacar una moneda de oro de C_1 que de C_3, y el argumento no toma esto en cuenta. ◄

3. Tres enfermedades distintas y excluyentes A, B y C producen el mismo conjunto de síntomas H. Un estudio clínico muestra que las probabilidades de contraer las enfermedades son 0.01; 0.005 y 0.02 respectivamente. Además, la probabilidad de que el paciente desarrolle los síntomas H para cada enfermedad son 0.90; 0.95 y 0.75, respectivamente. Si una persona enferma tiene los síntomas H, ¿cuál es la probabilidad de que tenga la enfermedad A?

► Sabemos lo siguiente:

$$\begin{aligned} P(A) &= 0.01, \qquad P(B) = 0.005, \qquad P(C) = 0.02 \\ P(H|A) &= 0.9, \qquad P(H|B) = 0.95, \qquad P(H|C) = 0.75 \end{aligned}$$

Usando el Teorema de Bayes tenemos

$$\begin{aligned} P(A|H) &= \frac{P(H|A)P(A)}{P(H|A)P(A) + P(H|B)P(B) + P(H|C)P(C)} \\ &= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.95 \times 0.005 + 0.75 \times 0.02} = \frac{9}{28.75} = 0.313. \end{aligned}$$

◄

4. Un estudiante responde una pregunta de un examen de múltiple selección que tiene cuatro respuestas posibles. Suponga que la probabilidad de que el estudiante conozca la respuesta a la pregunta es 0.8 y la probabilidad de que adivine es 0.2. Si el estudiante adivina, la probabilidad de que acierte es 0.25. Si el estudiante responde acertadamente la pregunta, ¿cuál es la probabilidad de que el estudiante realmente supiera la respuesta?

► Definamos los siguientes eventos:

C: "el estudiante conoce la respuesta".

A: "el estudiante responde acertadamente".

Queremos calcular $P(C|A)$ y sabemos que

$$P(C) = 0.8; \qquad P(A|C^c) = 0.25; \qquad P(A|C) = 1.$$

Usando el Teorema de Bayes

$$\begin{aligned} P(C|A) &= \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)} \\ &= \frac{0.8}{0.8 + (0.25)(0.2)} = \frac{0.8}{0.85} = 0.941. \end{aligned}$$

◄

3.4. Eventos Independientes

En los ejemplos que hemos considerado en relación al concepto de probabilidad condicional, hemos visto que en ciertos casos, cuando la ocurrencia del evento A no aporta información respecto a la ocurrencia o no del evento B, se tiene

$$P(B|A) = P(B).$$

En este caso decimos que B es independiente de A. Ahora bien, como

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

podemos expresar la relación anterior en la forma siguiente, que es simétrica en ambos sucesos:

$$P(A \cap B) = P(A)P(B). \tag{3.5}$$

Observamos además que (3.5) es válida aún cuando A o B sean sucesos de probabilidad nula. Usaremos esta relación como definición.

Definición 3.2 Dos eventos A y B son independientes si vale (3.5).

Es fácil verificar que Ω y $\emptyset$ son independientes de cualquier evento.

Proposición 3.4 *Si A y B son independientes, también lo son: a) A y B^c, b) A^c y B^c.*

Demostración. Veamos b), la demostración de a) es similar.

$$\begin{aligned}
P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B)\\
&= 1 - P(A) - P(B) + P(A \cap B)\\
&= 1 - P(A) - P(B) + P(A)P(B)\\
&= [1 - P(A)][1 - P(B)]\\
&= P(A^c)P(B^c).
\end{aligned}$$

■

Ejemplo

1. Un lote de diez objetos contiene cuatro defectuosos y seis en buen estado. Se extraen dos objetos sucesivamente y sin reposición. Sean los eventos D_1: "el primer objeto es defectuoso" y D_2: "el segundo objeto es defectuoso". ¿Son independientes estos eventos? ¿Qué sucede si los objetos se extraen con reposición?

▶ En el primer caso podemos calcular $P(D_2)$ de la siguiente manera:

$$\begin{aligned}
P(D_2) &= P(D_2|D_1)P(D_1) + P(D_2|D_1^c)P(D_1^c)\\
&= \frac{3}{9}\frac{4}{10} + \frac{4}{9}\frac{6}{10} = \frac{36}{90} = \frac{2}{5}.
\end{aligned}$$

Por otro lado,

$$P(D_2|D_1) = \frac{3}{9} \neq \frac{2}{5} = P(D_2)$$

de modo que D_1 y D_2 no son independientes.

En cambio, si los objetos se extraen con reposición tenemos $P(D_1) = P(D_2) = 4/10$, mientras que $P(D_1 \cap D_2) = (4/10)^2$ y los eventos son independientes. ◀

La definición de independencia se puede generalizar a una familia cualquiera de eventos:

Definición 3.3 Sea $\mathcal{C} = \{A_i, i \in I\}$ una familia de eventos. Diremos que los eventos A_i son independientes si para cualquier colección finita de eventos $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ de $\mathcal{C}$, se cumple:

$$P(\bigcap_{j=1}^{n} A_{i_j}) = \prod_{j=1}^{n} P(A_{i_j}). \tag{3.6}$$

En este caso decimos también que $\mathcal{C}$ es una familia de eventos independientes.

Observemos que en la definición *sólo* intervienen colecciones finitas de eventos de $\mathcal{C}$, pero intervienen *todas* las colecciones finitas. Por ejemplo, si la familia consta de tres eventos, no es suficiente verificar (3.6) para las parejas de eventos. En efecto, sea el experimento de lanzar dos dados y $\mathcal{C} = \{A_1, A_2, A_3\}$ con

$$\begin{aligned} &A_1 : \text{“se obtiene un 6 en el primer dado''}. \\ &A_2 : \text{“se obtiene un 1 en el segundo dado''}. \\ &A_3 : \text{“la suma de los dos dados es 7''}. \end{aligned}$$

Claramente,

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{6}$$

y

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{6} \times \frac{1}{6}.$$

Pero

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{36} \neq \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6},$$

de modo que los tres eventos no son independientes.

Ejemplos.

2 Si A y B son eventos independientes y la probabilidad de que ambos ocurran es 0.16, mientras que la probabilidad de que ninguno ocurra es 0.36, calcule $P(A)$ y $P(B)$.

▶ Sabemos que $P(A \cap B) = P(A)P(B) = 0.16$ y $P((A \cup B)^c) = 0.36$, de donde obtenemos

$$\begin{aligned} P(A \cup B) &= 0.64 \\ &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - 0.16. \end{aligned}$$

Por lo tanto tenemos el siguiente par de ecuaciones

$$\begin{aligned} P(A) + P(B) &= 0.8 \\ P(A)P(B) &= 0.16 \end{aligned}$$

de donde se obtiene que $P(A) = P(B) = 0.4$.

◀

3 ¿Cuál es la probabilidad de obtener tres 6 al lanzar ocho dados?

▶ El problema es equivalente a calcular la probabilidad de obtener tres 6 al lanzar ocho veces el mismo dado. Sea E_i el evento “obtenemos un 6 en el i-ésimo lanzamiento”. Calculemos primero la probabilidad de que los primeros tres sean 6 y el resto no, es decir,

$$P(E_1 \cap E_2 \cap E_3 \cap E_4^c \cap E_5^c \cap E_6^c \cap E_7^c \cap E_8^c)$$

y por independencia esto es
$$\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

Es fácil ver que esta es la probabilidad de cualquier conjunto de ocho lanzamientos de un dado en el cual haya exactamente tres 6. Como hay $\binom{8}{3}$ conjuntos de este tipo, tenemos que la probabilidad deseada es
$$\binom{8}{3}\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

◀

4 Una muestra de tamaño 4 se extrae con reposición de una bolsa que contiene 6 bolas, de las cuales 4 son blancas. Sea A el evento "la primera bola es blanca" y B: "la tercera bola es blanca". ¿Son independientes estos eventos? ¿Qué sucede si el muestreo se hace sin reposición?

▶ En el primer caso (muestreo con reposición) tenemos
$$P(A) = \frac{4}{6}, \quad P(B) = \frac{4}{6} \text{ y } P(A \cap B) = \left(\frac{4}{6}\right)^2,$$
de modo que los eventos son independientes.

En el segundo caso, de nuevo $P(A) = 4/6$ y para calcular $P(B)$ podemos proceder como en el ejemplo 3.2.2 obteniendo $P(B) = 4/6$. Para calcular $P(A \cap B)$ consideremos el evento C definido por "la segunda bola es blanca". Entonces
$$\begin{aligned} P(A \cap B) &= P(A \cap B \cap C) + P(A \cap B \cap C^c) \\ &= P(A)P(C|A)P(B|A \cap C) + P(A)P(C^c|A)P(B|A \cap C^c) \\ &= \frac{4}{6}\frac{3}{5}\frac{2}{4} + \frac{4}{6}\frac{2}{5}\frac{3}{4} = \frac{2}{5} \end{aligned}$$
y en este caso los eventos no son independientes. ◀

5 Demuestre: si $P(B|A) = P(B|A^c)$ entonces A y B son independientes.

▶ La relación se puede escribir
$$\frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A^c)}{P(A^c)}.$$

Multiplicando por $P(A^c)$ ambos lados obtenemos
$$\begin{aligned} P(B \cap A^c) &= \frac{(1 - P(A))P(B \cap A)}{P(A)} \\ &= \frac{P(B \cap A)}{P(A)} - P(B \cap A), \end{aligned}$$
de donde
$$\frac{P(B \cap A)}{P(A)} = P(B \cap A) + P(B \cap A^c) = P(B)$$
es decir
$$P(A \cap B) = P(A)P(B).$$

◀

6 El primer hijo que tiene una mujer es un niño hemofílico. La mujer, en cuya historia familiar no aparecen casos de hemofilia, tiene dudas sobre tener un segundo hijo, pero piensa que su hijo no heredó la hemofilia de ella y que su enfermedad es debida a una mutación. Por lo tanto, la probabilidad de que un segundo hijo tenga hemofilia es la de que nuevamente la enfermedad venga de una mutación, y esto es un número pequeño, m (digamos $m = 10^{-5}$). Calcule cuál es en realidad la probabilidad de que el segundo hijo tenga hemofilia si el primero nació hemofílico.

► Definamos los siguientes eventos:

$$\begin{aligned} &A_1 : \text{“la madre es portadora de la enfermedad''}. \\ &A_2 : \text{“el primer hijo es hemofílico''}. \\ &A_3 : \text{“el segundo hijo es hemofílico''}. \end{aligned}$$

Un varón normal tiene un par de cromosomas XY y tiene hemofilia si y sólo si, en lugar del cromosoma X tiene un cromosoma X', que lleva un gen que produce la enfermedad. Sea m la probabilidad de que un cromosoma X mute en un cromosoma X'.

La madre tiene dos cromosomas X y el evento A_1 ocurre sólo si al menos uno de estos dos cromosomas X es mutante, lo cual ocurre con probabilidad

$$P(A_1) = 1 - (1 - m)^2 = 2m - m^2 \approx 2m$$

donde hemos supuesto que las mutaciones ocurren en forma independiente, y hemos desechado el término m^2 por ser mucho menor que $2m$.

Si la madre es portadora de la enfermedad y uno de sus cromosomas es X', su hijo tendrá probabilidad $1/2$ de heredar el cromosoma X, es decir

$$P(A_2|A_1) = P(A_2^c|A_1) = \frac{1}{2}$$

Si en cambio la madre no lleva la enfermedad, su hijo será hemofílico si su cromosoma X sufre una mutación:

$$P(A_2|A_1^c) = m.$$

Además, por independencia, tenemos que

$$\begin{aligned} P(A_3|A_1 \cap A_2) &= P(A_3^c|A_1 \cap A_2) = P(A_3|A_1 \cap A_2^c) \\ &= P(A_3^c|A_1 \cap A_2^c) = \frac{1}{2}, \\ P(A_3|A_1^c \cap A_2) &= P(A_3|A_1^c \cap A_2^c) = m. \end{aligned}$$

Queremos calcular

$$P(A_3|A_2) = \frac{P(A_2 \cap A_3)}{P(A_2)}$$

pero

$$\begin{aligned} P(A_2 \cap A_3) &= P(A_1 \cap A_2 \cap A_3) + P(A_1^c \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) + P(A_1^c)P(A_2|A_1^c)P(A_3|A_2 \cap A_1^c) \\ &= 2m\left(\frac{1}{2}\right)^2 + (1 - 2m)m^2 \\ &= \frac{m}{2} + m^2 - 2m^3 \approx \frac{m}{2} \end{aligned}$$

y además

$$\begin{aligned} P(A_2) &= P(A_2|A_1)P(A_1) + P(A_2|A_1^c)P(A_1^c) \\ &= \frac{1}{2}2m + m(1-2m) \\ &\approx 2m. \end{aligned}$$

Por lo tanto

$$P(A_3|A_2) \approx \frac{m/2}{2m} = \frac{1}{4}.$$

◀

7 *Muestreo con Reposición.* Analicemos de nuevo, desde el punto de vista de la independencia, el muestreo con reposición de una población con dos tipos de individuos A y B en proporciones p y $1-p$ respectivamente. Deseamos calcular la probabilidad de extraer r del tipo A cuando tomamos una muestra de tamaño n, $(r \le n)$ extrayendo uno a la vez.

▶ Supongamos que la extracción dió como resultado que los primeros r fueron del tipo A y los $n-r$ restantes del tipo B. ¿Cómo podemos calcular la probabilidad de que esto ocurra? Para ello, llamemos A_i al evento "en la i-ésima extracción obtuvimos un individuo de tipo A". En virtud de la independencia,

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_r \cap A_{r+1}^c \cap \cdots \cap A_n^c) &= \prod_{i=1}^{r} P(A_i) \prod_{i=r+1}^{n} P(A_i^c) \\ &= p^r(1-p)^{n-r} \end{aligned}$$

Si observamos ahora que cualquiera sea el orden en que extraigamos los r individuos de tipo A y los $n-r$ de tipo B, la probabilidad es la misma, tendremos que la probabilidad buscada es

$$\binom{n}{r} p^r(1-p)^{n-r}, \qquad 0 \le r \le n,$$

es decir, la distribución binomial que habíamos encontrado anteriormente. ◀

8 Consideremos de nuevo la población del ejemplo anterior y calculemos la probabilidad de obtener por primera vez un individuo de tipo A en la n-ésima extracción.

▶ Sea A_i el evento "en la i-ésima extracción se obtiene un individuo de tipo A" y sea $\mathcal{C} = \{A_i, i \ge 1\}$. $\mathcal{C}$ es una familia de eventos independientes. También son familias de eventos independientes las obtenidas a partir de $\mathcal{C}$ al sustituir cualesquiera conjuntos A_i por sus complementos (ver la proposición 3.4). Entonces

$$\begin{aligned} p_n &= P(\text{ el primero de tipo } A \text{ es el } n\text{-esimo}) \\ &= P(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) = (1-p)^{n-1}p, \end{aligned}$$

que es la distribución geométrica que vimos en el capítulo anterior.

Las probabilidades p_n que hemos calculado determinan la distribución de un "tiempo de vida" discreto, es decir, del número de repeticiones de un experimento hasta que ocurra un evento determinado. Es de interés observar que las probabilidades p_n pueden ser utilizadas como aproximación discreta de la situación que hemos estudiado anteriormente en relación a la distribución del tiempo de vida de aparatos que "no envejecen" (ver ejemplo 3.1.4).

Sea T el tiempo de vida de un aparato de esa clase. Para calcular $P(T > t)$ dividimos el intervalo $[0,t]$ en n intervalos iguales $I_1,\ I_2, \dots, I_n$ (ver figura 3.3).

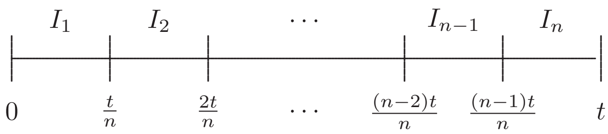


Figura 3.3

Si A_k: "el aparato no se daña en el intervalo I_k", es claro que

$$\{T > t\} = \bigcap_{k=1}^{n} A_k.$$

Ahora bien, la hipótesis de "no envejecimiento" se traduce en el hecho de que los eventos $A_1, \ldots, A_n$ son independientes, es decir que el hecho de que no ocurran desperfectos en los intervalos $I_1, \ldots, I_k$ no suministra información sobre la probabilidad de daños en el intervalo I_{k+1}.

Si suponemos que cuando la longitud h de un intervalo es pequeña, la probabilidad es aproximadamente proporcional a h, o más precisamente, que

$$P(\text{se daña en } [0, h]) = \lambda h + o(h)$$

donde λ es la constante de proporcionalidad y $o(h)$ es una función de h que satisface $\frac{o(h)}{h} \to 0$ $(h \to 0)$, resulta que

$$\begin{aligned} P(T > t) &= P(\cap_{k=1}^{n} A_k) = \prod_{k=1}^{n} P(A_k) \\ &= \left(1 - \frac{\lambda t}{n} + o(\frac{t}{n})\right)^n \\ &\to e^{-\lambda t} \qquad (n \to \infty) \end{aligned}$$

y volvemos a encontrar la distribución exponencial. ◀

9 Al probar independientemente dos aparatos electrónicos iguales hasta su inutilización, se obtuvieron los siguientes valores para la duración de vida: 21 horas y 30.2 horas. Suponiendo que el aparato no sufre envejecimiento, esto es, deterioro progresivo, tenemos dos observaciones correspondientes a la distribución

$$P(T > x) = e^{-\lambda x} \qquad \text{para cierto } \lambda.$$

¿Cómo obtener un valor razonable de λ, esto es, *una estimación de* λ?

▶ Naturalmente, el valor "razonable" de λ depende de qué consideremos "razonable", o sea, del criterio de estimación que se adopte. Como ya hemos visto el criterio de máxima verosimilitud, trataremos de utilizarlo.

En una primera etapa simplificaremos el problema. Supongamos que sólo observamos el primer aparato hasta su destrucción, es decir, 21 horas. Además, consideremos que en realidad no medímos exactamente las 21 horas, sino que sólo observamos que el aparato quedó inútil entre las 21 y 21+h horas de vida ($h > 0$). Tenemos

$$P(21 < T < 21 + h) = \int_{21}^{21+h} \lambda e^{-\lambda t} dt.$$

De acuerdo al criterio adoptado, debemos maximizar esta probabilidad. Como h está fijo, esto es equivalente a maximizar lo siguiente:

$$\frac{1}{h} P(21 < T < 21 + h) = \frac{1}{h} \int_{21}^{21+h} \lambda e^{-\lambda t} dt$$

pero si h es pequeño, esto es aproximadamente igual a

$$\lambda e^{-21\lambda}$$

debido al Teorema del Valor Medio del Cálculo Integral. Por lo tanto, el problema se reduce a maximizar la función $\lambda e^{-\lambda t}$ en el punto $t = 21$. Obtenemos

$$\hat{\lambda} = \frac{1}{21}.$$

Después de estas consideraciones resulta fácil considerar el caso original. Usando la independencia, debemos maximizar el producto

$$\frac{1}{h}\ P(21 < T < 21 + h)\ \frac{1}{k}\ P(30.2 < T < 30.2 + k) \simeq \lambda e^{-21\lambda}\ \lambda e^{-30.2\lambda}$$

si h y k son pequeños. Se obtiene

$$\hat{\lambda} = \frac{2}{21 + 30.2}.$$

Resulta sencillo ahora generalizar la aplicación anterior al caso de n observaciones independientes $t_1, t_2, \dots, t_n$. Queremos obtener el valor de $\hat{\lambda}$ (si es único) que maximice

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

como

$$L'(\lambda) = \left(\frac{n}{\lambda} - \sum_{i=1}^{n} t_i \right) L(\lambda)$$

resulta que un valor posible para el máximo en el intervalo $(0,\infty)$ es

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{t}}$$

donde $\bar{t}$ es el promedio de las n observaciones. Por otra parte, la función $L(\lambda)$ es positiva para $\lambda \in (0, \infty)$, pero $L(\lambda) \to 0$ cuando $\lambda \to 0^+$ y $L(\lambda) \to 0$ cuando $\lambda \to \infty$, de modo que si consideramos un intervalo cerrado $[a, A]$ con $a > 0$ suficientemente pequeño y A suficientemente grande, $L(\lambda)$ tiene necesariamente un máximo en este intervalo y lo alcanza en $\hat{\lambda}$ ya que

$$L(\hat{\lambda}) = \left(\frac{n}{\sum_{i=1}^{n} t_i} \right)^n e^{-n}$$

es estrictamente positivo y finito a menos que $t_1 = t_2 = \cdots = t_n = 0$ y esto sólo puede suceder con probabilidad 0. ◀

10 Decimos que A atrae a B si $P(B|A) > P(B)$ y A repele a B si $P(B|A) < P(B)$. (a) Demuestre que si B atrae a A entonces A atrae a B y B^c repele a A. (b) Un documento que buscamos está en alguno de los n folders de un cajón y B_j respresenta el evento de que esté en el j-ésimo folder, donde $P(B_j) = b_j > 0$. Por otro lado F_j representa el evento de que en una búsqueda rápida en el folder j no encontremos el documento, y $P(F_j|B_j) = a_j < 1$. Demuestre que B_j y F_j son mutuamente repelentes, pero que F_j atrae a B_i para $i \neq j$.

▶ a) Como B atrae a A, usando la definición tenemos que $P(A \cap B) > P(A)P(B)$. Dividiendo ahora por $P(A)$ obtenemos $P(B|A) > P(B)$. Por otro lado

$$P(A|B^c)P(B^c) = P(A) - P(A|B)P(B) < P(A)(1 - P(B)) = P(A)P(B^c),$$

y al dividir por $P(B^c) > 0$ obtenemos que B^c repele a A.

b) Por el teorema de Bayes,

$$P(B_j|F_j) = \frac{P(F_j|B_j)P(B_j)}{\sum_i P(F_i|B_i)P(B_i)} = \frac{a_j b_j}{1 - b_j + a_j b_j}$$

ya que, para $i \neq j$, $P(F_j|B_i) = 1$. Por lo tanto

$$P(B_j) - P(B_j|F_j) = \frac{b_j(1-b_j)(1-a_j)}{1 - b_j + a_j b_j} > 0,$$

de modo que F_j repele a B_j. Además, para $i \neq j$

$$P(B_i|F_j) - P(B_i) = \frac{b_i}{1 - b_j + a_j b_j} - b_i = \frac{b_i b_j (1-a_j)}{1 - b_j + a_j b_j} > 0,$$

◀

11 **El Problema de los Puntos** Dos jugadores que llamaremos A y B juegan una serie de juegos independientes y en cada uno ambos tienen igual probabilidad de ganar. Ambos apuestan una cierta cantidad C y el acuerdo es que el primero en ganar n juegos gana la apuesta. Sin embargo, la serie de juegos se interrumpe antes de que alguno de los jugadores haya ganado n juegos. La pregunta es ¿cómo debe repartirse el dinero de la apuesta?

▶ Este problema, conocido como el problema de los puntos, tiene una larga historia, fue considerado por Pascal y Fermat en su correspondencia y resuelto por ambos. La idea de Pascal es que el monto de la apuesta debe repartirse en proporción a la probabilidad que cada jugador tiene de ganar si la serie de juegos continuara.

Veamos un caso extremo para comenzar. Supongamos que $n = 5$ y la serie de juegos se detiene cuando A ha ganado 4 juegos y B no ha ganado ninguno. En esta situación la única manera en que B puede ganar es si gana 5 partidas seguidas. Como los juegos son independientes y la probabilidad de ganar en cada juego es $1/2$, la probabilidad de que B gane es $(1/2)^5 = 1/32$. Por lo tanto, de acuerdo a Pascal B debería recibir $1/32$ del total de la apuesta y A debería recibir el resto.

Veamos otro caso particular. Supongamos que A necesita un juego para ganar mientras que B necesita 3. Para decidir quién gana A y B deben jugar a lo sumo tres juegos más. Veamos cuáles son los resultados posibles de tres juegos:

$$\begin{array}{cccc} \underline{\text{AAA}} & \underline{\text{AAB}} & \underline{\text{ABB}} & \text{BBB} \\ & \underline{\text{ABA}} & \underline{\text{BAB}} & \\ & \underline{\text{BAA}} & \underline{\text{BBA}} & \end{array}$$

En siete de ellos, los que están subrayados, A gana al menos un juego antes de que B gane tres. Como todas las series de tres juegos tienen la misma probabilidad $1/8$, A gana con probabilidad $7/8$ y A debe llevarse $7/8$ partes de la apuesta.

Se puede argumentar que si A gana la primera partida, la serie termina allí pues ya se sabe que A es el ganador, de modo que en realidad sólo tendríamos las siguientes posibilidades:

$$\underline{\text{A}} \qquad \underline{\text{BA}} \qquad \underline{\text{BBA}} \qquad \text{BBB}$$

Sin embargo, estos eventos no son igualmente probables: las probabilidades respectivas son $1/2$, $1/4$, $1/8$ y $1/8$. Como en los tres primeros casos A gana la serie, la probabilidad de que esto ocurra es

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

que es el resultado que obtuvimos anteriormente. La ventaja del primer enfoque es que facilita el cálculo pues como los eventos son equiprobables, sólo hay que contar en cuantos casos gana A. Este es el enfoque que vamos a tomar para resolver el caso general, que formulamos a continuación.

Si A ha ganado $n-a$ juegos y B ha ganado $n-b$, el problema se puede se puede plantear de manera equivalente como sigue: ¿Cuál es la probabilidad de que A gane a juegos antes de que B gane b juegos?

Llamemos $u(a,b)$ la probabilidad que queremos calcular. Para que el evento que nos interesa ocurra, A y B deben jugar a lo sumo $a+b-1$ juegos. Por lo tanto la probabilidad de que A gane la serie es la probabilidad de que A gane al menos a juegos en una serie de $a+b-1$. La probabilidad de que A gane exactamente k juegos en esta serie tiene una distribución binomial con $a+b-1$ ensayos y probabilidad de éxito en cada ensayo $1/2$, esto es

$$\binom{a+b-1}{k}\Big(\frac{1}{2}\Big)^{a+b-1}$$

Como $u(a,b)$ es la probabilidad de que A gane al menos a juegos en una serie de $a+b-1$ tenemos

$$u(a,b)=\sum_{k=a}^{a+b-1}\binom{a+b-1}{k}\Big(\frac{1}{2}\Big)^{a+b-1}=\sum_{k=0}^{b-1}\binom{a+b-1}{k}\Big(\frac{1}{2}\Big)^{a+b-1}.$$

◀

12 **Cadenas de Markov.** En un proceso de fabricación se sabe que una máquina produce siempre la misma cantidad C de artículos, que se dividen en dos categorías, A y B. La máquina puede operar en dos estados, E_1 y E_2, y las probabilidades respectivas de producir un artículo de tipo A son p_1 y p_2, donde $p_2 > p_1$. Las condiciones del proceso determinan que la máquina debe operar en un mismo estado toda una jornada.

Si la máquina trabaja un día en el estado $E_i,\ i=1,2$, al día siguiente hay que determinar si va a operar en el mismo estado o en el otro. El problema que vamos a enfocar es el de regular el proceso. Tomamos las decisiones del siguiente modo:

- Si la máquina opera un día en el estado E_1, observamos la cantidad de artículos del tipo A producidos. Sea r esta cantidad. Si $r \le s$, donde s es un número fijo, al día siguiente se utilizará la máquina en el estado E_2. Si en cambio $r > s$, se la mantiene en el estado E_1 al día siguiente. Es decir, si el número de artículos de categoría A es muy pequeño, cambiamos al estado E_2, donde tenemos mayor probabilidad ($p_2 > p_1$) de producir artículos de categoría A. Pero si el número de artículos de categoría A es grande, decidimos que la máquina opere en el mismo estado E_1.
- Análogamente, si una jornada la máquina fue utilizada en el estado E_2, para decidir el estado en el cual trabajará al día siguiente, observamos el número r de artículos de tipo A producidos. Si $r \le t$, donde t ha sido fijado previamente, al día siguiente se mantiene en el estado E_2, en el cual tenemos mayor probabilidad de producir artículos de categoría A. Si en cambio $r > t$, cambiamos al estado E_1.

Este esquema de decisiones nos dice que si estamos en el estado E_1, al día siguiente nos mantenemos en este estado o lo cambiamos al estado E_2, según ciertas probabilidades a y $1-a$, que podemos calcular a partir de la distribución binomial, conocidos $C,\ p_1$ y s. De manera similar, si estamos en el estado E_2, lo mantenemos o no en la jornada siguiente según ciertas probabilidades b y $1-b$, calculadas también con la distribución binomial conocidos $C,\ p_2$ y t.

Llamando $E_i^n,\ i=1,2;\ n=1,2,\dots$ el suceso "en el n-ésimo día la máquina funciona en el estado E_i" tenemos

$$P(E_1^{n+1}|E_1^n)=a \qquad P(E_2^{n+1}|E_1^n)=1-a \tag{3.7}$$

$$P(E_2^{n+1}|E_2^n)=b \qquad P(E_1^{n+1}|E_2^n)=1-b \tag{3.8}$$

para cualquier $n \geq 1$.

Observamos que las probabilidades de pasar de un estado a otro dependen sólamente del estado en el cual la máquina operó el día previo, y no de los restantes días pasados, cualquiera sea el número $n+1$.

Representemos las probabilidades de cambio de estado por una tabla:

Dia $n+1$

		E_1	E_2
Dia n	E_1	a	$1-a$
	E_2	$1-b$	b

Tabla 3.2

de donde obtenemos la matriz

$$M = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}.$$

Observamos que las sumas de las filas de M tienen el valor constante 1. Una matriz con esta propiedad, y para la cual $a_{ij} \geq 0$, se llama *estocástica* o *de Markov*. Podemos, ahora, calcular probabilidades de interés, como por ejemplo la probabilidad de pasar del estado E_i al estado E_j en m pasos. Sea

$$p_{ij}^{(m)} = P(\text{pasar del estado } E_i \text{ al } E_j \text{ en } m \text{ pasos})$$

donde $i, j = 1, 2$.

Ahora bien, considerando el estado ocupado en el primer paso, esta probabilidad se puede expresar:

$$\begin{aligned} p_{ij}^{(m)} &= p_{i1}^{(1)} p_{1j}^{(m-1)} + p_{i2}^{(1)} p_{2j}^{(m-1)} \\ &= \sum_{k=1}^{2} p_{ik}^{(1)} p_{kj}^{(m-1)} \end{aligned}$$

Los términos $p_{ij}^{(1)}$ para i fijo, son los términos de la i-ésima fila de M, y los términos $p_{ij}^{(1)}$ para j fijo, son los términos de la columna j de M.

Si $m = 2$, de modo que $p_{kj}^{(m-1)} = p_{kj}^{(1)}$, obtenemos

$$p_{11}^{(2)} = p_{11}^{(1)} p_{11}^{(1)} + p_{12}^{(1)} p_{21}^{(1)} \qquad p_{12}^{(2)} = p_{11}^{(1)} p_{12}^{(1)} + p_{12}^{(1)} p_{22}^{(1)}$$

$$p_{21}^{(2)} = p_{21}^{(1)} p_{11}^{(1)} + p_{22}^{(1)} p_{21}^{(1)} \qquad p_{22}^{(2)} = p_{21}^{(1)} p_{12}^{(1)} + p_{22}^{(1)} p_{22}^{(1)}$$

de donde se observa que los términos $p_{ij}^{(2)}$ son los términos de la matriz M^2, esto es

$$M^2 = \begin{pmatrix} a^2 + (1-a)(1-b) & a(1-a)(1-b)b \\ a(1-b) + b(1-b) & (1-b)(1-a) + b^2 \end{pmatrix}$$

donde M^2 resulta también estocástica.

Procediendo por inducción obtenemos que M^m es estocástica y tiene como términos las probabilidades $p_{ij}^{(m)}$ de pasar del estado i al j en m pasos.

Si al comenzar el proceso, el primer día del año, partimos del estado E_i, al cabo de $m+1$ días estaremos en el estado E_j con cierta probabilidad. Es fácil calcular esta probabilidad porque, para $m = 1$ (esto es, en el segundo día), las probabilidades de estar en E_1 y E_2 están dadas por

$$(1, 0)\begin{pmatrix} a & 1-b \\ 1-b & b \end{pmatrix} = (a,\ 1-b)$$

es decir, la primera fila de M, y en general, procediendo por inducción, tendremos para el $(m-1)$-ésimo día:

$$(1, 0)M^m.$$

Si en el primer día del proceso estamos en E_1 o E_2 con probabilidades p y $1-p$, determinadas por algún procedimiento azaroso, al cabo de m pasos, es decir el día $m+1$, tendremos para cada estado las probabilidades determinadas por el vector

$$(p,\ 1-p)M^m$$

Se nos presenta ahora el problema de calcular las potencias de M, y observar si cuando $m \to \infty$ resulta algún valor límite para M^m. Ahora bien, como M^m es estocástica, es suficiente calcular los elementos de la diagonal de la matriz M^m. Sea $a_m = p_{11}^{(m)},\ a_1 = a,\ b_m = p_{22}^{(m)},\ b_1 = b$. Tenemos

$$\begin{aligned} a_{m+1} &= a_m a + (1-a_m)(1-b) \\ b_{m+1} &= b_m b + (1-b_m)(1-a) \end{aligned}$$

Estudiemos la primera de estas igualdades recurrentes, ya que la segunda se obtiene cambiando a por b y a_m por b_m. Resulta

$$\begin{aligned} a_{m+1} - a_m &= a_m a + (1-a_m)(1-b) - a_m \\ &= a_m a + (1-a_m)(1-b) - a_{m-1}a - (1-a_{m-1})(1-b) \\ &= (a_m - a_{m-1})a - (1-b)(a_m - a_{m-1}) \end{aligned}$$

y si llamamos $\Delta_m = a_{m+1} - a_m$ obtenemos

$$\begin{aligned} \Delta_m &= (-1+b+a)\Delta_{m-1} \\ &= (-1+b+a)^2\Delta_{m-2} \\ &= (-1+b+a)^{m-1}\Delta_1 \end{aligned}$$

Supongamos por el momento que $0 < b+a < 2$, de modo que $|-1+b+a| < 1$. Obtenemos entonces

$$\begin{aligned} a_{m+1} &= \Delta_m + \Delta_{m-1} + \cdots + \Delta_1 + a \\ &= a + \Delta_1 \sum_{j=0}^{m-1} (-1+b+a)^j \\ &= a + \frac{\Delta_1(-1+b+a)^m - 1}{-2+b+a} \end{aligned}$$

y como

$$\begin{aligned} \Delta_1 = a_2 - a_1 &= a^2 + (1-a)(1-b) - a \\ &= (a-1)(-1+b+a) \end{aligned}$$

entonces

$$a_{m-1} = a + (a-1)(-1+b+a)\frac{(-1+b+a)^m - 1}{-2+b+a}$$

y de manera similar

$$b_{m+1} = b + (b-1)(-1+b+a)\frac{(-1+b+a)^m - 1}{-2+b+a}$$

y los demás términos de la matriz se obtienen como $1 - a_{m+1}$ y $1 - b_{m+1}$.

En particular, haciendo $m \to \infty$ obtenemos

$$\begin{aligned} a_m \to a_\infty &= a + \frac{(a-1)(-1+b+a)}{2-a-b} \\ &= \frac{1-b}{2-a-b} \end{aligned}$$

y la matriz M^m converge a M^∞ donde

$$M^\infty = \begin{pmatrix} \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \\ \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \end{pmatrix}$$

como se verifica fácilmente luego de algunos cálculos sencillos.

Por consiguiente, si las probabilidades de partir de E_1 o E_2 el primer día eran p y $1-p$, tenemos en el límite

$$(p, 1-p)\begin{pmatrix} \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \\ \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \end{pmatrix} = \left(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}\right)$$

Observamos que el resultado obtenido, que es la probabilidad límite de estar en el primer o segundo estado, es el mismo cualquiera haya sido el vector inicial de probabilidades $(p, 1-p)$.

La distribución límite

$$\left(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}\right)$$

se llama *estacionaria* porque tiene la siguiente propiedad:

$$\left(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}\right) M = \left(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}\right)$$

y por lo tanto es invariante en cualquier número de pasos, o sea que si el día inicial la distribución estacionaria nos da las probabilidades de partir del estado E_1 o del estado E_2, cualquier otro día tendremos la misma distribución de probabilidades.

Resta por estudiar los casos $a+b=2$ y $a+b=0$. Si $a+b=2$, necesariamente $a=b=1$ y resulta

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{de modo que} \quad M^m = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad M^\infty = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

pero este es un caso sin interés porque la máquina permanece siempre en el estado del cual parte. Por otro lado, si $a+b=0$, es decir $a=b=0$ resulta

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

y no existe distribución límite porque saltamos con probabilidad 1 de un estado al otro en cada paso.

Señalamos también, que resulta intuitivo que obtenemos de este modo una regulación efectiva del número de artículos de categoría A producidos, al menos a la larga, o sea, después de un número grande de días de aplicar el sistema de decisiones. ◀

Ejercicios.

1. Sea A, B y C eventos con probabilidad estrictamente positiva. Demuestre las siguientes relaciones:

   a. $P(A^c \cup B^c) = 1 - P(B)P(A|B)$
   b. $P(A \cap B|B \cup C) = P(A \cap B|B)P(B|B \cup C)$
   c. $P(B \cap C|A) = P(C|A)P(B|A \cap C)$ si $P(A \cap C \neq 0$
   d. $P(A|B)P(B|C)P(C|A) = P(B|A)P(C|B)P(A|C)$
   e. $\frac{P(A|A\cup B)}{P(B|A\cup B)} = \frac{P(A)}{P(B)}$

2. Demuestre: $\frac{P(B^c|A)}{P(B)} + \frac{P(A^c)}{P(A)} = \frac{P(A^c|B)}{P(A)} + \frac{P(B^c}{P(B)}$.

3. Un detector de mentiras muestra una señal positiva (indicando una mentira) 10 % de las veces que el sujeto dice la verdad y 94 % de las veces que miente. Si dos personas son sospechosas de un crimen que se sabe ha cometido uno solo de ellos, y ambos dicen ser inocentes, ¿cuál es la probabilidad de que una señal positiva del detector corresponda al culpable?

4. Se obtiene una muestra de cuatro bolas a partir de una bolsa que contiene doce, de las cuales ocho son blancas. Si el muestreo es sin reposición, halle la probabilidad de que la tercera bola sea blanca, si sabemos que la muestra tiene tres bolas blancas. ¿Que sucede si el muestreo se hace con reposición?

5. Se lanza un par de dados simétricos. Calcule la probabilidad de que la suma sea 7 dado que:

   a. La suma es impar,
   b. La suma es mayor que 6,
   c. El resultado del primer dado fue impar,
   d. El resultado del segundo dado fue par,
   e. El resultado de al menos un dado fue impar,
   f. Los dos dados tuvieron el mismo resultado,
   g. Los dos dados tuvieron distintos resultados,
   h. La suma de los dos dados fue 13.

6. Una bolsa contiene cuatro bolas blancas y dos negras y otra contiene tres de cada color. Se escoge una bolsa al azar y luego se selecciona una bola, también al azar. ¿Cuál es la probabilidad de que sea blanca?

7. Una caja contiene 10 focos, cuatro malos y seis buenos. Los focos se prueban de la siguiente manera: se extraen al azar y se prueban sin reemplazarlos. Este proceso se repite hasta localizar los cuatro en mal estado. ¿Cuál es la probabilidad de que el último en mal estado se identifique en la quinta prueba? ¿y en la décima?

8. Cierta vacuna brinda protección parcial contra una enfermedad, de modo que una persona vacunada tiene probabilidad 0.4 de contraer la enfermedad, mientras que para una persona no vacunada esta probabilidad es de 0.8. Si 75 % de la población está vacunada, ¿Cuál es la probabilidad de que una persona que tiene la enfermedad haya sido vacunada?

9. Luego de una serie de pruebas para evaluar un nuevo tipo de examen para detectar cáncer, se ha determinado que 97 % de los pacientes cancerosos de un hospital reaccionan positivamente, mientras que sólo 5 % de aquellos que no tienen cáncer muestran un resultado positivo. Si 2 % de los pacientes del hospital tienen cáncer, ¿cuál es la probabilidad de que un paciente elegido al azar que reacciona positivamente al examen realmente tenga cáncer?

10. Suponga que 5 % de los hombres y 25 de cada 10.000 mujeres son daltónicos. Se escoge un daltónico al azar, ¿cuál es la probabilidad de que sea hombre?

11. Una ferretería tiene tres cajas con igual cantidad de tornillos. Dos de las cajas contienen 5 % de tornillos defectuosos y la otra contiene 10 %. Se escogen dos cajas al azar y se mezclan y de ella se extraen cinco tornillos, uno de los cuales es defectuoso, ¿cuál es la probabilidad de que una de las cajas usadas en la mezcla haya sido la que contenía 10 % de defectuosos?

12. Tres sucursales de una tienda tienen 8, 12, y 14 empleados de los cuales 4, 7 y 10 son mujeres, respectivamente.

    a. Se escoge una sucursal al azar y de ella se escoge un empleado. Si éste es una mujer, ¿cuál es la probabilidad de que ella trabaje en la sucursal con 12 empleados?

    b. Si se escoge un segundo empleado de la misma sucursal, ¿cuál es la probabilidad de que se escoja una mujer?

13. Las señales telegráficas "*punto*" y "*raya*" se envían en proporción 3:4. Debido a ciertas condiciones que causan una transmisión muy errática, un punto se transforma en una raya con probabilidad $1/4$, y una raya en un punto con probabilidad $1/3$. Si se recibe un punto, ¿cuál es la probabilidad de que haya sido enviado un punto?

14. En una bolsa hay cinco bolas blancas y tres negras y en otra hay tres blancas y siete negras. Se escoge una bolsa al azar y se selecciona una bola, ¿cuál es la probabilidad de que sea blanca?

15. La caja I contiene 50 tornillos y 70 clavos. La caja II contiene 40 tornillos y 20 clavos.

    a. Calcule la probabilidad de extraer un tornillo si se selecciona una caja al azar y luego se extrae un objeto.

    b. Calcule la probabilidad de extraer un tornillo si se mezclan los contenidos de ambas cajas y luego se extrae un objeto.

    c. Si se selecciona una caja al azar y luego se extrae un objeto que resulta ser un tornillo, ¿cuál es la probabilidad de que provenga de la caja I?

16. Lanzamos una moneda repetidamente hasta que obtener sol diez veces. a) ¿Cuál es la probabilidad de no haber obtenido dos águilas en sucesión para ese momento? b) ¿Cuál es la probabilidad de no haber obtenido dos soles en sucesión para ese momento?

17. Se extrae una bola de una caja que contiene cuatro blancas y dos negras. Si la bola es blanca se la deja fuera de la bolsa, mientras que si es negra se la vuelve a colocar dentro. Extraemos luego otra bola. Sea A el evento "la primera bola es blanca" y $B =$ "la segunda bola es blanca". Diga si las siguientes proposiciones son ciertas o falsas:

    a. $P(A) = 2/3$ b. $P(B) = 3/5$
    c. $P(B|A) = 3/5$ d. $P(A|B) = 9/4$
    e. Los eventos A y B son disjuntos.

18. Lanzamos una moneda tres veces y consideramos los siguientes eventos: A: el primer lanzamiento es águila, B: el segundo lanzamiento es sol, C: el tercer lanzamiento es águila, D: los tres lanzamientos son iguales, E: hay exactamente un águila en los tres lanzamientos.

    a. ¿Cuáles de los siguientes eventos son independientes?

    (i) A, B (ii) A, D (iii) A, E (iv) D, E.

    b. ¿Cuáles de los siguientes tríos de eventos son independientes?

    (i) A, B, C (ii) A, B, D (iii) C, D, E.

19. Sean A y B dos eventos disjuntos. Demuestre que si A y B son independientes, alguno de los dos tiene probabilidad 0.

20. De un ejemplo de tres eventos A, B, C tales que $P(A \cap B \cap C) = P(A)P(B)P(C)$ pero $P(A^c \cap B^c \cap C^c) \neq P(A^c)P(B^c)P(C^c)$.

21. Sea A, B y C eventos independientes y $P(C) \neq 0$. Demuestre:

    a. $P(A \cap B|C) = P(A|C)P(B|C)$. b. $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.

    c. $P(A|B \cap B) = P(A)$ siempre que $P(B \cap C) \neq 0$.

22. Sean $G = \{1,2,3\}$, $H = \{4,5,6\}$ Lanzamos dos dados y sean los eventos A : 'el primer dado cae en H, B :'El segundo dado cae en H, C : un dado cae en G y el otro en H, D : el total es cuatro, E : el total es cinco y F : el total es siete. ¿Cuáles de las siguientes proposiciones son ciertas?

a. A y F son independientes.
b. A y D son independientes.
c. A y E son independientes.
d. $P(A \cap B \cap C) = P(A)P(B)P(C)$.
e. A y C son independientes.
f. C y E son independientes.
g. $P(A \cap C \cap E) = P(A)P(C)P(E)$.
h. A, C y E son independientes.

23. a. De un ejemplo de tres eventos A, B y C tales que $P(A \cap B \cap C) = P(A)P(B)P(C)$ pero $P(A^c \cap B^c \cap C^c) \neq P(A^c)P(B^c)P(C^c)$.

b. Demuestre que si $P(A \cap B \cap C) = P(A)P(B)P(C)$; $P(A^c \cap B \cap C) = P(A^c)P(B)P(C)$; $P(A \cap B^c \cap C) = P(A)P(B^c)P(C)$ y $P(A \cap B \cap C^c) = P(A)P(B)P(C^c)$ entonces A, B y C son independientes.

24. Demuestre que si
$$\frac{P(A)}{P(A \cap B)} + \frac{P(B)}{P(A \cap B)} = \frac{1}{P(A)} + \frac{1}{P(B)},$$
entonces A y B son independientes.

25. Lanzamos un dado cinco veces, a. Si el dado sale 1 al menos una vez, ¿cuál es la probabilidad de que salga 1 al menos dos veces? b. Si el primer lanzamiento due 1, ¿cuál es la probabilidad de que salga 1 al menos dos veces?

26. Lanzamos una moneda repetidamente hasta que sol haya ocurrido diez veces. a) ¿Cuál es la probabilidad de que para ese momento no hayan ocurrido dos águilas seguidas? b) ¿Cuál es la probabilidad de que para ese momento no hayan ocurrido dos soles seguidas?

27. Sea $\Omega = \{a, b, c, d, e\}$ y $P(a) = P(b) = 1/8$, $P(c) = P(d) = P(e) = 3/16$. Sean $A = \{a, d, e\}$, $B = \{a, c, e\}$ y $C = \{a, c, d\}$. Demuestre que $P(A \cap B \cap C) = P(A)P(B)P(C)$ pero ningún par de eventos son independientes.

28. Se toma una muestra de cinco transistores producidos por una máquina que en promedio produce 20 % de transistores defectuosos.

a. Calcule la probabilidad de que si seleccionamos un transistor de la muestra, éste resulte defectuoso.

b. Suponga que seleccionamos al azar un transistor de la muestra y resulta defectuoso, ¿cuál es la probabilidad de que un segundo transistor seleccionado al azar resulte defectuoso?

29. Si $A_1, \dots, A_n$ son eventos independientes, muestre que
$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - (1 - P(A_1))(1 - P(A_2)) \cdots (1 - P(A_n)).$$

30. a. Si sabemos que una mano de poker tiene al menos tres aces ¿cuál es la probabilidad de que tenga los cuatro? b. Si sabemos que una mano de poker tiene los aces de corazón, trébol y diamante ¿cuál es la probabilidad de que también tenga el as de pica? c. Halle la probabilidad de que una mano de poker tenga los dos aces negros dado que tiene al menos tres aces.

31. En un closet hay tres pares de medias blancas y dos de media negras. Las medias con todas nuevas y del mismo estilo y tamaño. Si seleccionamos dos medias al azar, ¿cuál es la probabilidad de que formen un par? Resuelva el mismo problema cambiando media por zapatos.

32. En un closet tenemos seis pares diferentes de zapatos. Si sacamos cinco zapatos al azar ¿Cuál es la probabilidad de obtener al menos un par de zapatos?

33. Una muestra de tamaño 4 se extrae con reposición de una bolsa que contiene 6 bolas, de las cuales 4 son blancas. Sea A el evento "exactamente una de las dos primeras bolas extraídas es blanca" y sea $B =$ "la cuarta bola es blanca". ¿Son A y B independientes? ¿Qué sucede si el muestreo se realiza sin reposición?

34. Considere de nuevo el ejercicio anterior y definamos C como el evento "exactamente dos de las bolas extraídas son blancas" ¿Son A, B y C independientes? ¿Son B y C independientes?

35. Un fabricante de motos inscribe tres corredores en una carrera. Sea $A_i, i = 1, 2, 3$ el evento definido por "el i-ésimo corredor finaliza la carrera en alguno de los tres primeros lugares". Si los eventos A_1, A_2 y A_3 son independientes y $P(A_1) = P(A_2) = P(A_3) = 0,1$ calcule las probabilidades de los siguientes eventos:

    a. Ninguno de los corredores termina entre los tres primeros.

    b. Al menos uno termina entre los tres primeros.

    c. Al menos dos terminan entre los tres primeros.

    d. Todos terminan entre los tres primeros.

36. Sean A y B eventos independientes tales que con probabilidad $1/6$ ocurren simultáneamente, y con probabilidad $1/3$ ninguno de ellos ocurre. Halle $P(A)$ y $P(B)$. ¿Están determinadas de forma única estas probabilidades?

37. ¿Cuál es el menor valor de n para el cual la probabilidad de obtener al menos un 6 en una serie de n lanzamientos de un dado es mayor que $3/4$?

38. Los eventos $A_1, A_2, \ldots$ son independientes y $P(A_j) = p,\ j = 1, 2, \ldots$ Halle el menor valor de n para el cual $P(\cup_1^n A_k) \geq p_0$ donde p_0 es un número fijo.

39. Una máquina consiste de 4 componentes conectados en paralelo, de modo que la máquina falla sólo si los cuatro componentes fallan. Supongamos que los componentes son independientes entre sí. Si los componentes tienen probabilidades 0.1; 0.2; 0.3 y 0.4 de fallar cuando la máquina es encendida, ¿cuál es la probabilidad de que funcione al encenderla?

40. En una fábrica de computadoras un inspector revisa un lote de 20 máquinas y encuentra que 3 de ellas necesitan ajustes antes de empacarlas. Otro empleado descuidado mezcla las computadoras que habían sido revisadas, de modo que el inspector tiene que hacerlo otra vez. a. ¿Cuál es la probabilidad de que no haga falta probar más de 17 computadoras? b. ¿Cuál es la probabilidad de que haga falta probar exactamente 17 computadoras?

41. Supongamos que A y B son independientes, y B y C son independientes.

    a. ¿Son A y C independientes en general?

    b. ¿Es B independiente de $A \cup C$?

    c. ¿Es B independiente de $A \cap C$?

42. Se lanza una moneda balanceada n veces. Muestre que la probabilidad condicional de águila en cualquier lanzamiento específico dado que hay k águilas en n lanzamientos es k/n ($k > 0$).

43. Suponga que A y B son eventos tales que $P(A|B) = P(B|A)$, $P(A \cup B) = 1$ y $P(A \cap B) > 0$. Demuestre que $P(A) > 1/2$.

44. Un arquero acierta al centro de la diana con probabilidad 0.9. a) ¿Cuál es la probabilidad de que logre acertar 8 veces si lanza 10 flechas? b) ¿Si lanza flechas hasta acertar 8 veces al centro de la diana, ¿Cuál es la probabilidad de que necesite a lo sumo lanzar 10 flechas?

45. Considere estos dos sistemas de ensayos de Bernoulli: 1) Lanzamos una moneda y águila es éxito; 2) Lanzamos un dado y 6 es éxito. Para cada uno de estos casos calcule $P(A)/P(B)$, donde A: 'el tercer éxito ocurre en el quinto ensayo', B: 'tres de los primeros cinco ensayos resultan en éxito'. Generalice reemplazando 3 por i y 5 por j.

46. Lanzamos cuatro dados, ¿Cuál es la probabilidad de que el mayor número sea 5 y el menor sea 3?

47. La caja A tiene dos bolas rojas y tres negras. La caja B tiene cinco rojas y una blanca. Se selecciona una bola al azar de la caja A y se coloca en la caja B y luego se escoge una bola al azar de la caja B. a) ¿Cuál es la probabilidad de que ambas bolas sean rojas? b) ¿Cuál es la probabilidad de que la segunda bola sea roja? c) Dado que la segunda bola es roja ¿Cuál es la probabilidad de que la primera también haya sido roja? d) Dado que la segunda bola es blanca ¿cuál es la probabilidad de que la primera haya sido roja?

48. Una fábrica produce 300 automóviles al día. La fábrica compra baterías de dos proveedores. La compañía A le vende 100 baterías al día, de las cuales 99 % funcionan correctamente. Las otras 200 baterías son producidas por la compañía B, de las cuales 5 % son defectuosas. Si seleccionamos un auto al azar de la producción de un día y la batería es defectuosa ¿cuál es la probabilidad de que haya sido producida por la empresa B?

49. Un empleado debe verificar el funcionamiento de una máquina que produce tornillos al inicio del día. Esta máquina necesita repararse una vez cada 10 días, en promedio y cuando necesita repararse, todos los tornillos que produce son defectuosos. Cuando la máquina trabaja adecuadamente, 5 % de los tornillos producidos son defectuosos y aparecen al azar a lo largo de la producción del día. ¿Cuál es la probabilidad de que la máquina esté funcionando bien si a. el primer tornillo que el inspector revisa es defectuoso? b. los dos primeros tornillos que el inspector revisa son defectuosos? c. los tres primeros son defectuosos?

50. Una carta de un juego de naipes se ha perdido. Trece cartas se extraen de las 51 restantes y resultan ser tres diamantes, dos picas, cuatro corazones y cuatro tréboles. Halle la probabilidad de que la carta perdida sea de cada una de las pintas.

51. Resuelva el problema de los puntos cuando los dos jugadores tienen probabilidades p y $q = 1 - p$ de ganar.

52. **Las cajas de cerillos de Banach.** Una persona tiene dos cajas de n cerillos, una en el bolsillo derecho y otra en el izquierdo. Cuando necesita un cerillo escoge una caja al azar hasta que se encuentra una caja vacía. ¿Cuál es la probabilidad de que la otra caja tenga k cerillos? (El celebre matemático polaco Stefan Banach solía reunirse con otros matemáticos en el Café Escocés en Lwów, Polonia, en donde había un cuaderno en el cual se anotaban los problemas planteados y sus soluciones. Esta libreta se conoce como el Libro Escocés. El problema anterior es el último problema incluido en este libro).

53. **La paradoja de Galton.** Si lanzamos tres monedas al menos dos de ellas son iguales, y la tercera tiene probabilidad 1/2 de caer águila o sol, de modo que la probabilidad de que las tres sean iguales es 1/2. En realidad la probabilidad de que las tres monedas sean iguales es 1/4. ¿Qué está mal en el razonamiento anterior?

54. **El modelo de Pólya.** El matemático polaco G. Pólya propuso el siguiente modelo para el proceso de contagio. Comenzamos con una caja que contiene una bola blanca y otra negra. Cada segundo escogemos una bola al azar de la caja, la reemplazamos y añadimos otra del mismo color. Haga una simulación d este proceso con una computadora y trata de hacer una predicción sobre la proporción de bolas blancas luego de un tiempo largo. ¿Es cierto que las proporciones de bolas de un color dado tienen una tendencia a estabilizarse?

55. En una colección de 65 monedas una de ellas tiene dos águilas y el resto son monedas normales. Si seleccionamos una moneda al azar y al lanzarla sale águila seis veces seguidas, ¿Cuál es la probabilidad de que hayamos escogido la moneda con dos águilas?

56. Te dan dos cajas y cincuenta bolas, la mitad blancas y la otra mitad negras. Debs distribuir las bolas en las cajas sin restricciones pero quieres maximizar la probabilidad de obtener una bola blanca si escoges una caja al azar y luego una bola al azar. ¿Cómo debes distribuir las bolas? Justifica tu respuesta.

57. Lanzamos una moneda n veces y obtenemos k soles. Demuestra que la probabilidad condicional de obtener un sol en cualquier lanzamiento específico, dado que hay k soles en total es k/n.

58. **La paradoja de Simpson.** Un fabricante de focos tiene dos plantas. La planta A vende lotes de focos que consisten de 1000 focos regulares y 2000 focos ahorradores. A través de pruebas de control de calidad se sabe que, en promedio, hay 2 focos regulares y 11 ahorradores defectuosos por lote. En la planta B se venden lotes de 2000 focos regulares y 1000 ahorradores, y en promedio hay 5 regulares y 6 ahorradores defectuosos por lote.

    El gerente de la planta A afirma que ellos son más eficientes pues sus tasas de focos defectuosos son $0.2\,\%$ y $0.55\,\%$ mientras que para la otra planta son $0.25\,\%$ y $0.6\,\%$. Por su parte el gerente de la planta B responde diciendo 'cada lote de 3000 focos que producimos contiene 11 focos defectuosos, comparado con 13 defectuosos para los focos producidos por A, de modo que nuestra tasa de $0.37\,\%$ de focos defectuosos es inferior a la de ellos, que es $0.43\,\%$. ¿Quién tiene la razón?

59. El evento A atrae al evento B si $P(B|A) > P(B)$ (ver el ejercicio 3.4.1). a) ¿Es transitiva esta relación? b) Demuestre que si A atrae a B y a C pero repele a $B \cap C$, entonces A atrae a $B \cup C$. c) ¿Es posible que A atraiga a B y a C pero repela a $B \cup C$? d) Demuestre que si $B_1, \dots B_n$ es una partición dl espacio muestral y A atrae algún B_j entonces debe repeler a algún B_i.

60. Demuestre que si $A_1, \dots A_n$ son eventos independientes definidos en un espacio muestral Ω y $0 < P(A_i) < 1$ para todo i, entonces Ω debe tener al menos 2^n puntos.

Capítulo 4

Variables Aleatorias

4.1. Introducción.

A la realización de un experimento aleatorio le hemos asociado un modelo matemático representado por un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde Ω es el conjunto de resultados posibles del experimento, $\mathcal{F}$ es la colección de eventos, de acuerdo a la sección 2.1 y P es una función que le asigna a cada conjunto en $\mathcal{F}$ un número entre 0 y 1 que representa su probabilidad y satisface las condiciones de la sección 2.2.

Con frecuencia estamos interesados en considerar funciones definidas sobre Ω, es decir, correspondencias que asocian a cada evento elemental un cierto valor. Por ejemplo, en los casos de muestreo que hemos mencionado anteriormente, si tomamos una muestra de tamaño n de los objetos producidos en una fábrica, el espacio muestral correspondiente es

$$\Omega = \{(e_1, \dots, e_n) : e_i = 0 \text{ ó } 1,\ i = 1, \dots n\}$$

donde $e_i = 0$ indica que hemos extraído un objeto bueno la i-ésima vez y $e_i = 1$ que hemos extraído uno defectuoso y nos interesa la función

$$(e_1, \dots, e_n) \mapsto \sum_{i=1}^{n} e_i$$

que asocia a cada evento elemental el número de objetos defectuosos que contiene la muestra respectiva.

Análogamente, en el caso del error de redondeo, supongamos que debemos efectuar un cálculo del tipo

$$y = \varphi(x_1, x_2, \dots, x_k) \tag{4.1}$$

donde $\varphi : \mathbb{R}^k \to \mathbb{R}$ y que cada una de las magnitudes $x_1, x_2, \dots, x_k$ se calcula a su vez con un cierto error de redondeo, es decir que en lugar de valores exactos $x_1, x_2, \dots, x_k$ obtenemos valores aproximados $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$, respectivamente:

$$x_i = \bar{x}_i + \delta_i \qquad (i = 1, \dots, k)$$

donde $\delta_1, \delta_2, \dots, \delta_k$ son los errores de redondeo.

Si en lugar de (4.1) calculamos

$$\bar{y} = \varphi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k) = \varphi(x_1 - \delta_1, x_2 - \delta_2, \dots, x_k - \delta_k)$$

cometemos un cierto error

$$y - \bar{y} = \varphi(x_1, x_2, \dots, x_k) - \varphi(x_1 - \delta_1, x_2 - \delta_2, \dots, x_k - \delta_k)$$

que es función del evento elemental $(\delta_1, \delta_2, \dots, \delta_k)$, considerado como elemento del espacio muestral

$$\Omega = \{(\delta_1, \delta_2, \dots, \delta_k) : \delta_i \in \mathbb{R}, \quad (i = 1, \dots k)\}.$$

Del mismo modo el lector puede verificar que en cada uno de los ejemplos que hemos considerado anteriormente, aparecen vinculadas a los problemas en cuestión ciertas funciones de los resultados obtenidos en los experimentos aleatorios, es decir, funciones de los eventos elementales. Estas funciones se llaman *variables aleatorias*.

Con frecuencia va a resultar de interés poder calcular la probabilidad de que una variable aleatoria X tome valores en un intervalo I, es decir, la probabilidad del conjunto

$$\{\omega : X(\omega) \in I\}.$$

Pero esto sólo podemos hacerlo si este conjunto está en $\mathcal{F}$, ya que P está definida únicamente sobre $\mathcal{F}$, y en principio, si X es cualquiera, este conjunto no tiene por qué estar en $\mathcal{F}$. Por lo tanto exigiremos como parte de la definición de variable aleatoria, que para cualquier intervalo $I \subset \mathbb{R}$ el conjunto $\{\omega : X(\omega) \in I\}$ esté en $\mathcal{F}$.

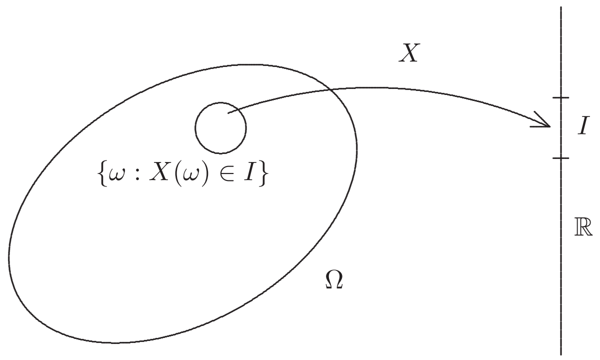


Figura 4.1

Definición 4.1 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Una función

$$X : \Omega \to \mathbb{R}$$

es una variable aleatoria real, o simplemente una variable aleatoria, si se cumple que para cualquier intervalo I en $\mathbb{R}$, el conjunto

$$\{\omega : X(\omega) \in I\}$$

es un evento (es decir, está en $\mathcal{F}$).

Para definir el concepto de variable aleatoria vectorial, introducimos primero el concepto de intervalo en $\mathbb{R}^m$.

Definición 4.2 Un intervalo en $\mathbb{R}^m$ es un conjunto de la forma:

$$\{(x_1, x_2, \ldots, x_m) : x_i \in I_1, x_2 \in I_2, \ldots, x_m \in I_m\}$$

donde $I_1, I_2, \ldots, I_m$ son intervalos de la recta.

Observamos que en $\mathbb{R}^2$, un intervalo no es otra cosa que un rectángulo con lados paralelos a los ejes coordenados, y en $\mathbb{R}^3$, un paralelepípedo de aristas paralelas a los ejes coordenados.

Definición 4.3 Una función

$$Y : \Omega \to \mathbb{R}^m$$

es una variable aleatoria vectorial, o simplemente un vector aleatorio, si se cumple que para cualquier intervalo I en $\mathbb{R}^m$, el conjunto

$$\{\omega : Y(\omega) \in I\}$$

es un evento (es decir, está en $\mathcal{F}$).

Frecuentemente denotaremos al conjunto $\{\omega : X(\omega) \in I\}$ por $\{X \in I\}$ o por $X^{-1}(I)$, y lo llamaremos la preimagen de I por la función X. La definición dice que X es una variable aleatoria si la preimagen por X de cualquier intervalo es un evento.

Por ejemplo, si $\mathcal{F}$ es la familia de todos los subconjuntos de Ω, cualquier función X definida sobre Ω es una variable aleatoria, ya que para cualquier intervalo I

$$X^{-1}(I) = \{\omega : X(\omega) \in I\} \subset \Omega$$

y por lo tanto $X^{-1}(I) \in \mathcal{F}$. Si en cambio $\mathcal{F}$ no es la familia de todos los subconjuntos de Ω, una función definida sobre Ω no tiene por qué satisfacer la definición de variable aleatoria. Como ejemplo consideremos

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

y la función $X(\omega) = \omega$. X no es una variable aleatoria, porque, por ejemplo,

$$X^{-1}\left([0, 3/2]\right) = \{1\} \notin \mathcal{F}.$$

Ejemplos

1. Sea c un número real, la función X definida por $X(\omega) = c$ para todo ω es una variable aleatoria, ya que para cualquier intervalo I,

$$\{\omega : X(\omega) \in I\} = \begin{cases} \Omega, & \text{si } c \in I \\ \emptyset, & \text{si } c \notin I \end{cases}$$

   y tanto Ω como $\emptyset$ son siempre eventos. Esta es una variable aleatoria constante.

2. Sea c un número real y definamos la función $X : \mathbb{R} \to \mathbb{R}$ por $X(\omega) = \omega + c$. En este caso el espacio de probabilidad sobre el cual está definida la función X es $(\mathbb{R}, \mathcal{B}, P)$ donde $\mathcal{B}$ son los conjuntos de Borel y P es alguna probabilidad definida sobre $\mathcal{B}$. Sea I algún intervalo en $\mathbb{R}$,

$$\{\omega : X(\omega) \in I\} = \{\omega : \omega + c \in I\} = \{\omega : \omega \in I - c\} = I - c$$

   donde, si A es un conjunto cualquiera y c una constante, definimos

$$A + c = \{x : x = a + c,\ a \in A\}.$$

   Por lo tanto, la preimagen de cualquier intervalo I por X es otro intervalo que se obtiene trasladando I una distancia $-c$. Pero como todo intervalo es un conjunto de Borel vemos que

$$\{\omega : X(\omega) \in I\} \in \mathcal{B}$$

   y X es una variable aleatoria.

3. Sea A un evento ($A \in \mathcal{F}$). Definimos X por

$$X(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases}$$

Observamos que el evento A ocurre si y sólo si $X(\omega) = 1$. Además si I es un intervalo tenemos que

$$\{\omega : X(\omega) \in I\} = \begin{cases} \Omega, & \text{si } 0, \in I,\ 1 \in I \\ A, & \text{si } 0 \notin I,\ 1 \in I \\ A^c, & \text{si } 0 \in I,\ 1 \notin I \\ \emptyset, & \text{si } 0 \notin I,\ 1 \notin I \end{cases}$$

y por lo tanto este conjunto siempre está en $\mathcal{F}$ y X es una variable aleatoria.

Esta variable aleatoria se conoce como la *variable indicadora* o *función indicadora* de A porque el valor de X nos dice si A ocurrió o no. Las notaciones usuales son $\mathbf{1}_A$ o χ_A.

Recíprocamente, si X es una variable aleatoria sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ que sólo toma los valores 1 y 0, entonces X es la variable indicatriz del evento

$$A = \{\omega : X(\omega) = 1\}$$

4.2. Operaciones con Variables Aleatorias.

Esta sección está destinada a probar que si se efectuan las operaciones usuales con variables aleatorias, se obtienen nuevas funciones que también son variables aleatorias.

Recordemos que la familia de conjuntos de Borel en $\mathbb{R}^m$ es la menor σ-álgebra que contiene a todos los intervalos abiertos. El siguiente lema será de utilidad.

Lema 4.1 *Sea $X : \Omega \to \mathbb{R}^m$ una función. X es una variable aleatoria si y sólo si la preimagen de cualquier conjunto de Borel es un evento, es decir*

$$X^{-1}(B) \in \mathcal{F} \qquad \text{para todo } B \in \mathcal{B} \tag{4.2}$$

donde $\mathcal{B}$ denota la familia de conjuntos de Borel.

Demostración. Puesto que los intervalos son conjuntos de Borel, si se cumple (4.2) se cumple que la preimagen de cualquier intervalo es un evento, y por lo tanto X es una variable aleatoria.

Recíprocamente, supongamos que la preimagen por X de cualquier intervalo es un evento. Tenemos que probar que la preimagen de cualquier conjunto de Borel también es un evento.

Consideremos la familia $\mathcal{D}$ de subconjuntos de $\mathbb{R}^m$ definida por

$$\mathcal{D} = \{D \subset \mathbb{R}^m : X^{-1}(D) \in \mathcal{F}\}$$

o sea que $\mathcal{D}$ es la familia de los conjuntos $D \subset \mathbb{R}^m$ cuya preimagen por X es un evento. Tenemos:

(I) $\mathbb{R}^m \in \mathcal{D}$, ya que $X^{-1}(\mathbb{R}^m) = \Omega \in \mathcal{F}$.

(II) $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D}$ ya que

$$\begin{aligned} X^{-1}(D^c) &= \{\omega :\ X(\omega) \in D^c\} = \{\omega :\ X(\omega) \notin D\} \\ &= \{\omega :\ X(\omega) \in D\}^c = (X^{-1}(D))^c \end{aligned}$$

y este último conjunto está en $\mathcal{F}$ porque $X^{-1}(D) \in \mathcal{F}$ y $\mathcal{F}$ es una σ-álgebra.

(III) Sea $\{D_n, n \geq 1\}$ una sucesión de conjuntos en $\mathcal{D}$, entonces

$$\cup_{n=1}^{\infty} D_n \in \mathcal{F}$$

ya que

$$\begin{aligned} X^{-1}(\cup_{n=1}^{\infty} D_n) &= \{\omega : X(\omega) \in \cup_{n=1}^{\infty} D_n\} \\ &= \cup_{n=1}^{\infty}\{\omega : X(\omega) \in D_n\} \\ &= \cup_{n=1}^{\infty} X^{-1}(D_n) \in \mathcal{F}. \end{aligned}$$

Es decir, hemos demostrado que $\mathcal{D}$ es una σ-álgebra, y como hemos supuesto que contiene a los intervalos abiertos, $\mathcal{D}$ debe contener a la menor σ-álgebra que contiene a los intervalos abiertos, que es justamente $\mathcal{B}$. Por lo tanto

$$B \in \mathcal{B} \Rightarrow B \in \mathcal{D} \Rightarrow X^{-1}(B) \in \mathcal{F}$$

que es lo que queríamos probar. ■

Consideremos ahora la siguiente situación

$$\Omega \xrightarrow{\text{X}} \mathbb{R}^m \xrightarrow{\text{g}} \mathbb{R}$$

es decir que X es una función de Ω en $\mathbb{R}^m$ y g una función de $\mathbb{R}^m$ en $\mathbb{R}$. En Ω tenemos la σ-álgebra de eventos $\mathcal{F}$ y en $\mathbb{R}^m$ la σ-álgebra de conjuntos de Borel $\mathcal{B}$. Definimos la *función compuesta* $Y : \Omega \to \mathbb{R}$ mediante

$$Y(\omega) = g(X(\omega)).$$

Lema 4.2 *Con las notaciones anteriores, si X y g son variables aleatorias, también lo es Y.*

Demostración. Para probar que Y es una variable aleatoria, tenemos que ver que la preimagen de cualquier intervalo I es un evento, es decir, está en $\mathcal{F}$. Tenemos:

$$\begin{aligned} Y^{-1}(I) &= \{Y \in I\} = \{\omega : g(X(\omega)) \in I\} \\ &= \{\omega : X(\omega) \in g^{-1}(I)\} = X^{-1}(g^{-1}(I)). \end{aligned}$$

Dado que g es una variable aleatoria, $g^{-1}(I) \in \mathcal{B}$, y como X también lo es, usando el lema 4.1 obtenemos

$$X^{-1}(g^{-1}(I)) \in \mathcal{F}.$$

■

Observación 4.1 En el lema anterior X es una variable aleatoria vectorial, que toma valores en $\mathbb{R}^m$, y por lo tanto se puede escribir

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

donde cada una de las $X_i,\ i = 1, \dots, m$ es una función

$$X_i : \Omega \to \mathbb{R}$$

y por lo tanto también podemos escribir

$$Y(\omega) = g(X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

Lema 4.3 *Las siguientes condiciones son equivalentes*

1. $\{\omega : X(\omega) \in I\} \in \mathcal{F}$ *para todo intervalo* $I \subset \mathbb{R}$.

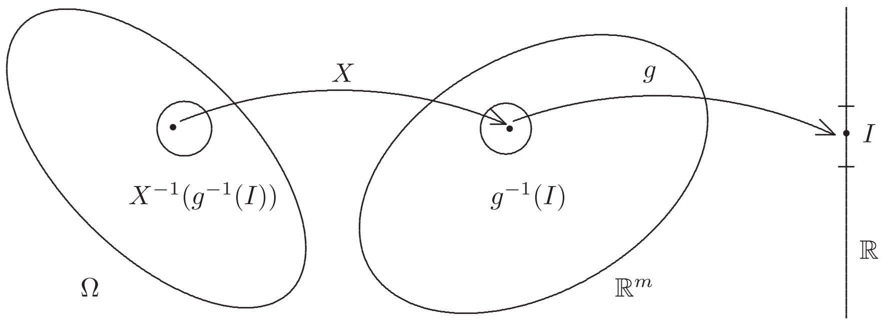


Figura 4.2

2. $\{\omega : X(\omega) < c\} \in \mathcal{F}$ *para todo* $c \in \mathbb{R}$.

3. $\{\omega : X(\omega) \leq c\} \in \mathcal{F}$ *para todo* $c \in \mathbb{R}$.

4. $\{\omega : X(\omega) > c\} \in \mathcal{F}$ *para todo* $c \in \mathbb{R}$.

5. $\{\omega : X(\omega) \geq c\} \in \mathcal{F}$ *para todo* $c \in \mathbb{R}$.

y por lo tanto cualquiera de estas condiciones puede ser utilizada en la definición de variable aleatoria.

Demostración. Como

$$\{X < c\}^c = \{X \geq c\} \qquad \text{y} \qquad \{X > c\}^c = \{X \leq c\}$$

es inmediato que 2 $\Leftrightarrow$ 5 y 3 $\Leftrightarrow$ 4. Veamos que 2 $\Leftrightarrow$ 3. Supongamos que 2 es cierto, como

$$\{X \leq c\} = \bigcap_{n=1}^{\infty} \left\{ X < c + \frac{1}{n} \right\}$$

y cada uno de los conjuntos de la intersección en el segundo miembro está en $\mathcal{F}$ (por 2), concluimos que 3 es cierto. Recíprocamente, si 3 es cierto, como

$$\{X < c\} = \bigcup_{n=1}^{\infty} \left\{ X \leq c - \frac{1}{n} \right\}$$

se obtiene que 2 es cierto.

Hemos visto hasta ahora que las cuatro últimas condiciones son equivalentes entre sí. Veamos ahora que también son equivalentes a la primera. Si 1 es cierta, escribiendo

$$\{X < c\} = \bigcup_{n=1}^{\infty} \{c - n \leq X < c - n + 1\}$$

vemos que los conjuntos que aparecen en la unión están en la σ-álgebra $\mathcal{F}$, de donde concluimos que $\{X < c\} \in \mathcal{F}$. Por lo tanto 2 es cierta, y como consecuencia 3, 4 y 5 también.

Finalmente, supongamos que 2, 3, 4 y 5 son ciertas; es fácil ver que, para cualquier intervalo I, el conjunto $\{X \in I\}$ se puede escribir en base a los conjuntos que aparecen en las condiciones 2, 3, 4 y 5 usando uniones e intersecciones. Por ejemplo, si $I = [a, b)$ entonces

$$\{\omega : X(\omega) \in I\} = \{\omega : a \leq X(\omega) < c\} = \{\omega : a \leq X(\omega)\} \cap \{\omega : X(\omega) < c\}.$$

Por lo que hemos supuesto, los conjuntos del segundo miembro están en $\mathcal{F}$, y por las propiedades de $\mathcal{F}$ concluimos que

$$\{\omega : X(\omega) \in I\} \in \mathcal{F}$$

para todo intervalo $I \subset \mathbb{R}$. ■

Proposición 4.1 *La suma de dos variables aleatorias también es una variable aleatoria.*

Demostración. Observamos que la desigualdad

$$X_1 + X_2 < c$$

es cierta si y sólo si existe algún racional r tal que

$$X_1 < r \qquad \text{y} \qquad r < c - X_2$$

por lo tanto, si $\mathbb{Q}$ es el conjunto de los numeros racionales, podemos escribir

$$\begin{aligned}\{\omega : X_1(\omega) + X_2(\omega) < c\} &= \bigcup_{r\in\mathbb{Q}} \{\omega : X_1(\omega) < r\} \cap \{\omega : r < c - X_2(\omega)\} \\ &= \bigcup_{r\in\mathbb{Q}} \{\omega : X_1(\omega) < r\} \cap \{\omega : X_2(\omega) < c - r\}.\end{aligned}$$

Pero como X_1 y X_2 son variables aleatorias, por el lema 4.3 los conjuntos que aparecen en el segundo miembro están en $\mathcal{F}$, y por lo tanto también está su unión. ■

De manera similar se puede demostrar que el producto, el cociente, etc. de variables aleatorias da como resultado variables aleatorias.

Antes de enunciar el próximo resultado recordamos la definición de función monótona.

Definición 4.4 Decimos que $g : \mathbb{R} \to \mathbb{R}$ es creciente (resp. decreciente) si se cumple que $x_1 < x_2 \Rightarrow g(x_1) \leq g(x_2)$ (resp. $g(x_1) \geq g(x_2)$). Si en lugar de la desigualdad en sentido amplio $(\leq, \geq)$ ponemos en sentido estricto $(<, >)$, decimos que g es estrictamente creciente (resp. decreciente). Decimos que g es monótona cuando es creciente o decreciente.

Proposición 4.2 *Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria y $g : \mathbb{R} \to \mathbb{R}$ una función monótona. Entonces $Y : \Omega \to \mathbb{R}$ definida por $Y(\omega) = g(X(\omega))$ también es una variable aleatoria.*

Demostración. Supondremos que g es creciente. La demostración para el caso decreciente es análoga. Por el lema 4.2 es suficiente demostrar que g es una variable aleatoria, es decir, que para cualquier intervalo $I \subset \mathbb{R}$ se tiene que $\{z \in \mathbb{R} : g(z) \in I\}$ es un conjunto de Borel, pero por el lema 4.3 sabemos que es suficiente demostrar que para cualquier $c \in \mathbb{R}$ se tiene que

$$\{z \in \mathbb{R} : g(z) < c\} \in \mathcal{B}.$$

Consideremos primero un caso particular: supongamos que la función g es continua y estrictamente creciente. Se pueden presentar tres casos:
(A) La función siempre es mayor que c. En este caso el conjunto que nos interesa es vacío y por lo tanto está en $\mathcal{B}$.
(B) La función siempre es menor que c. Entonces el conjunto que nos interesa es $\mathbb{R}$ y está en $\mathcal{B}$.
(C) Existe un único punto $y \in \mathbb{R}$ tal que $g(y) = c$ (ver Figura 4.3) y entonces el conjunto que nos interesa está formado por los puntos que están a la izquierda de y, sin incluir a y, es decir

$$\{z \in \mathbb{R} : g(z) < c\} = (-\infty, y)$$

y este conjunto está en $\mathcal{B}$, con lo cual hemos probado que g es una variable aleatoria.

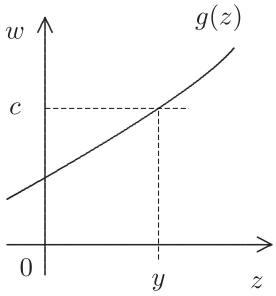


Figura 4.3

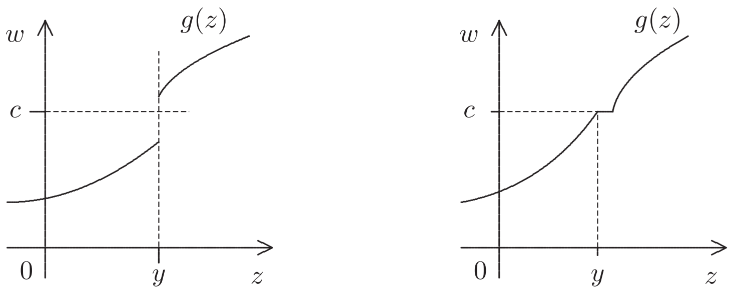


Figura 4.4

En general esto no ocurre para una función monótona cualquiera ya que el punto y puede no existir si la función es discontinua, o puede no ser único, si la función no es estrictamente creciente (Figura 4.4).

Para resolver este problema definimos

$$y = \sup\{z \in \mathbb{R} : g(z) < c\}.$$

Si $y \in \{z \in \mathbb{R} : g(z) < c\}$ entonces

$$\{z \in \mathbb{R} : g(z) < c\} = (-\infty, y] \in \mathcal{B}$$

mientras que si $y \notin \{z \in \mathbb{R} : g(z) < c\}$ entonces

$$\{z \in \mathbb{R} : g(z) < c\} = (-\infty, y) \in \mathcal{B}$$

y esto termina la demostración. ■

4.3. Función de Distribución.

Definición 4.5 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}$ una variable aleatoria. Llamaremos función de distribución de la variable aleatoria X a la función F definida por

$$F(x) = P(\{\omega : X(\omega) \leq x\}) = P(X \leq x).$$

En algunas ocasiones, para resaltar que F es la función de distribución de X, escribiremos F_X en lugar de F.

Proposición 4.3 *Si F es una función de distribución, satisface las siguientes propiedades:*

1. *F es creciente (en sentido amplio).*
2. $\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$
3. *F es continua por la derecha.*

Demostración. Sea F la función de distribución de la variable aleatoria X.

1. Si $x_1 < x_2$ entonces
$$F(x_2) - F(x_1) = P(X \leq x_2) - P(X \leq x_1) = P(x_1 < X \leq x_2) \geq 0.$$

2. Sea $\{x_n\}$ una sucesión decreciente de números reales, $x_n \to -\infty$. Entonces, la sucesión de eventos $\{\omega : X(\omega) \leq x_n\}$ es una sucesión decreciente y además
$$\bigcap_{n=1}^{\infty} \{\omega : X(\omega) \leq x_n\} = \emptyset.$$
Por lo tanto
$$\lim_{n \to -\infty} F(x_n) = \lim_{n \to -\infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\emptyset) = 0.$$
Esto prueba que $\lim_{n \to -\infty} F(x_n) = 0$. Del mismo modo, si $\{x_n\}$ es una sucesión creciente y $x_n \to \infty$, la sucesión de eventos $\{\omega : X(\omega) \leq x_n\}$ es creciente y
$$\bigcup_{n=1}^{\infty} \{\omega : X(\omega) \leq x_n\} = \Omega.$$
Por lo tanto
$$\lim_{n \to \infty} F(x_n) = \lim_{n \to \infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\Omega) = 1.$$
Esto prueba que $\lim_{n \to \infty} F(x_n) = 1$.

3. Para probar que F es continua por la derecha en todo punto, basta probar que si $\{x_n\}$ es una sucesión decreciente que tiende a a, entonces
$$\lim_{n \to \infty} F(x_n) = F(a).$$
Veamos esto:
$$\{X \leq a\} = \bigcap_n \{X \leq x_n\}$$
y puesto que $\{X \leq x_n\}$ es una sucesión decreciente de eventos, resulta
$$\lim_n F(x_n) = \lim_n P(X \leq x_n) = P(X \leq a) = F(a).$$

■

Recíprocamente, si una función $F : \mathbb{R} \to [0, 1]$ satisface las propiedades 1, 2 y 3 se puede demostrar que F es la función de distribución de una variable aleatoria. Para esto basta tomar $\Omega = \mathbb{R}$, $\mathcal{F} =$ la familia de conjuntos de Borel de $\mathbb{R}$ y definir la probabilidad P de modo que
$$P((a, b]) = F(b) - F(a)$$
(la demostración de la existencia de esta probabilidad P escapa al contenido de este texto). Entonces, F es la función de distribución de la variable aleatoria $X(\omega) = \omega$, ya que
$$F_X(x) = P(\{\omega : X(\omega) \leq x\}) = P(\{\omega : \omega \leq x\}) = P((-\infty, x]) = F(x).$$

Llamemos

$$F(a^-) = \lim_{x \uparrow a} F(x),$$

el límite por la izquierda de F en a. Tenemos que $P(X < x) = F(x^-)$ y en consecuencia la siguiente proposición.

Proposición 4.4 *Sea X una variable aleatoria con función de distribución F. Entonces*

$$P(X \in (a, b]) = F(b) - F(a) \tag{4.3}$$

$$P(X \in [a, b]) = F(b) - F(a^-) \tag{4.4}$$

$$P(X \in (a, b)) = F(b^-) - F(a) \tag{4.5}$$

$$P(X \in [a, b)) = F(b^-) - F(a^-) \tag{4.6}$$

Demostración. Veamos la demostración de (4.4), las demás son similares:

$$P(X \in [a, b]) = P(X < b) - P(X \leq a) = F(b^-) - F(a)$$

■

Si ponemos ahora $a = b = x$ en (4.4) obtenemos que

$$P(X = x) = F(x) - F(x^-) \tag{4.7}$$

de modo que la función de distribución es continua en x si y sólo si $P(X = x) = 0$.

4.4. Variables Aleatorias Discretas.

Definición 4.6 Una variable aleatoria X es *discreta* si existe un conjunto finito o numerable de valores $\{x_n\}$ tal que

$$\sum_n P(X = x_n) = 1,$$

es decir, la probabilidad de que dicha variable tome valores fuera del conjunto $\{x_n\}$ es cero.

Por ejemplo, en los casos de muestreo que hemos considerado anteriormente, si

$$\Omega = \{(e_1, \dots, e_n) : e_i = 0 \text{ ó } 1,\ i = 1, \dots, n\}$$

donde $e_i = 0$ indica que la i-ésima extracción ha resultado en un objeto bueno y $e_i = 1$ indica uno defectuoso, la función $X : \Omega \to \mathbb{R}$ definida por

$$X(e_1, e_2, \dots, e_n) = \sum_{i=1}^{n} e_i$$

que representa el total de objetos defectuosos en la muestra, es una variable aleatoria discreta, ya que solo toma valores en el conjunto $\{0, 1, 2, \dots, n\}$.

Para ver como es la función de distribución de una variable aleatoria discreta consideramos

$$p_n = P(X = x_n).$$

Con frecuencia, la sucesión $\{p_n\}$ se denomina la *función de probabilidad* de la variable aleatoria X. Entonces

$$F_X(x) = P(X \leq x) = \sum_{x_n \leq x} p_n$$

donde la suma se extiende a aquellos valores de n para los cuales $x_n \leq x$. La figura 4.5 (a) representa una gráfica típica de función de distribución de una variable aleatoria discreta, mientras que la figura 4.5 (b) representa la función de probabilidad correspondiente.

Ejemplos

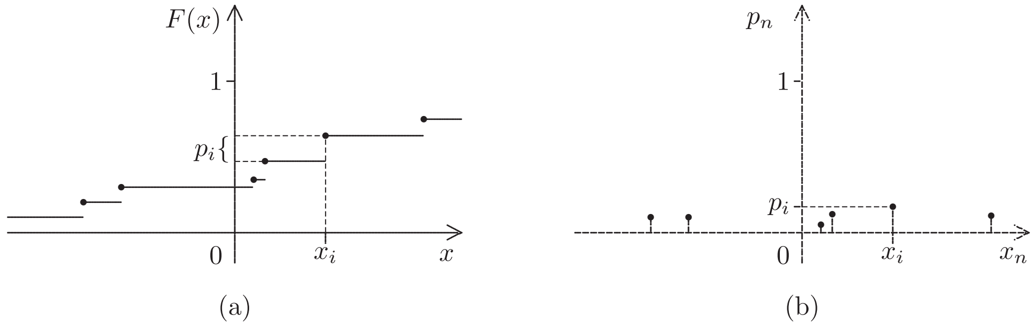


Figura 4.5

1. Lanzamos dos dados y llamamos X a la suma de los resultados. Esta suma puede tomar once valores distintos y si suponemos que los dados son simétricos podemos calcular la probabilidad de cada uno de ellos:

x_i :	2	3	4	5	6	7	8	9	10	11	12
p_i :	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

El gráfico correspondiente a esta función de probabilidad es el siguiente:

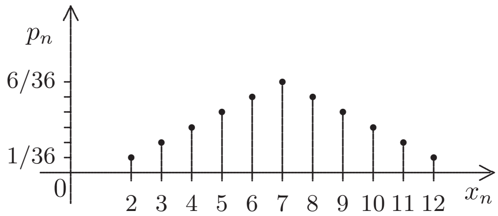


Figura 4.6

La función de distribución de esta variable se puede representar de la siguiente manera:

2. Consideremos una caja que contiene seis fichas numeradas del uno al seis. Se extraen dos fichas con reposición y se observa el mayor de los números. ¿Cómo es la función de probabilidad de esta variable? ¿Cómo es, si el muestreo se realiza sin reposición?

► Estudiemos primero el caso de muestreo con reposición. El espacio muestral correspondiente a este experimento es el conjunto de pares (ω_1, ω_2), donde $\omega_i \in \{1, 2, 3, 4, 5, 6\}$ para $i = 1, 2$. La variable aleatoria $X : \Omega \to \mathbb{R}$ que estamos considerando está definida por

$$X(\omega_1, \omega_2) = \text{máx}\{\omega_1,\ \omega_2\}$$

y toma valores en el conjunto $\{1, 2, 3, 4, 5, 6\}$. Si suponemos que todas las fichas tienen la misma probabilidad de ser extraídas entonces todos los eventos elementales que componen el espacio muestral son igualmente probables y su probabilidad es $1/36$. Es fácil ahora calcular la función de probabilidad de la variable aleatoria X:

x_i :	1	2	3	4	5	6
p_i :	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

y su representación gráfica se presenta en la figura 4.8.

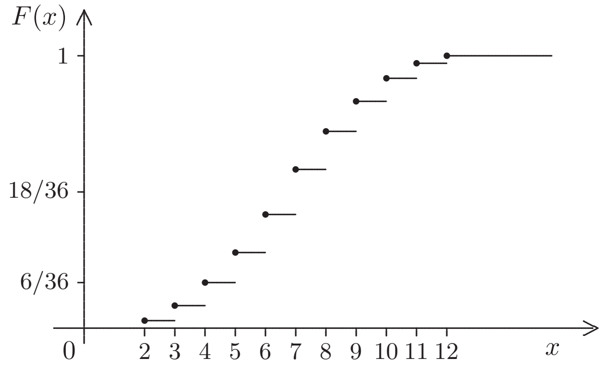


Figura 4.7

Veamos ahora qué sucede si el muestreo se realiza sin reposición. El espacio muestral es ahora el conjunto de pares

$$\{(\omega_1, \omega_2) : \ \omega_i \in \{1, 2, 3, 4, 5, 6\} \text{ y } \omega_1 \neq \omega_2\}$$

La variable

$$X(\omega_1, \omega_2) = \text{máx}\{\omega_1, \ \omega_2\}$$

ahora toma valores en el conjunto $\{2, 3, 4, 5, 6\}$.

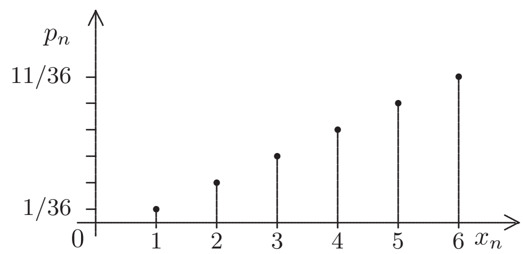


Figura 4.8

Nuevamente suponemos que todas las fichas tienen la misma probabilidad de ser extraídas, de modo que los eventos que componen el espacio muestral tienen probabilidad 1/30. La siguiente tabla representa la función de probabilidad de la variable aleatoria X

x_i :	2	3	4	5	6
p_i :	$\frac{2}{30}$	$\frac{4}{30}$	$\frac{6}{30}$	$\frac{8}{30}$	$\frac{10}{30}$

y su representación gráfica es:

◀

3. Se lanza al aire una moneda repetidamente y se observa en cuál de los lanzamientos la moneda cae aguila por primera vez. Hallar la función de probabilidad de esta variable.

► Si A denota aguila y S sol, cada evento elemental es una sucesión infinita de estos símbolos:

$$\omega = (A, A, A, A, S, A, A, S, S, S, \dots)$$

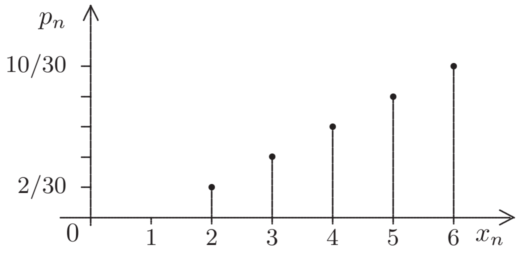


Figura 4.9

y la variable aleatoria que estamos considerando le asigna a cada evento elemental el número correspondiente al lugar de la primera A. Por ejemplo:

$$X(S, S, S, A, A, S, A, \dots) = 4$$

Observamos que X puede tomar como valor cualquier entero positivo, y por independencia podemos calcular su función de probabilidad:

$$P(X = 1) = \frac{1}{2}, \qquad P(X = 2) = \frac{1}{2}\frac{1}{2} = \left(\frac{1}{2}\right)^2,$$

y en general $X = n$ si y sólo si los $n - 1$ primeros lanzamientos resultaron en S y el n-ésimo en A, lo cual tiene probabilidad

$$P(X = n) = \left(\frac{1}{2}\right)^n.$$

Como

$$\sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} 2^{-n} = 1,$$

la variable aleatoria X es discreta y toma valores sobre el conjunto numerable $\{1, 2, 3, \dots\}$. ◀

4. Consideremos ahora la situación en la que elegimos al azar un número en el intervalo $[0, 1)$ y definimos $X(\omega)$ como la tercera cifra después de la coma en el desarrollo decimal de ω. En este caso, los posibles valores de $X(\omega)$ son $\{0, 1, \dots, 9\}$ y cada uno tiene probabilidad 1/10, es decir que

$$p_n = P(X = n) = \frac{1}{10}, \qquad (n = 0, 1, \dots, 9)$$

La gráfica de la función de probabilidad es:

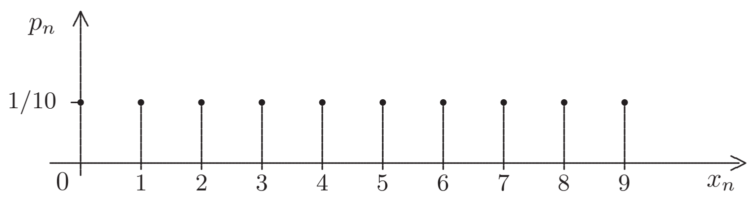


Figura 4.10

A continuación vamos a considerar algunas de las distribuciones discretas más importantes.

4.4.1. La Distribución de Bernoulli

Esta es la distribución más sencilla y corresponde a una variable que toma sólo dos valores: 1 con probabilidad p y 0 con probabilidad $q = 1 - p$.

Si A es un evento y definimos la *función indicadora* de A por

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Esta variable aleatoria vale 1 cuando A ocurre y 0 cuando no, es decir, nos indica cuándo ocurre A. Por lo tanto $\mathbf{1}_A$ tiene distribución de Bernoulli con $p = P(A)$.

4.4.2. La Distribución Uniforme

Una variable aleatoria con valores en el conjunto $\{x_1, x_2, \dots, x_n\}$ tiene *distribución uniforme* si todos los puntos x_i, $1 \leq i \leq n$ tienen la misma probabilidad. Como hay n valores posibles esto quiere decir que

$$P(X = x_i) = \frac{1}{n}.$$

Para un dado convencional tenemos $x_i = i,\ 1 \leq i \leq 6$.

4.4.3. La Distribución Binomial.

Recordemos el caso de muestreo con reposición, en el cual la variable que nos interesa especialmente es el número de defectuosos d_n que contiene una muestra de n elementos, es decir,

$$d_n(e_1, \dots, e_n) = \sum_{i=1}^{n} e_i$$

donde $e_i = 0$ ó 1, según que en la i-ésima extracción hayamos obtenido un artículo bueno o defectuoso, respectivamente. Hemos visto que la función de probabilidad de la variable aleatoria discreta d_n es

$$P(d_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = p_{k,n} \qquad (k = 0, 1, \dots, n)$$

donde p es la probabilidad de obtener un objeto defectuoso en una extracción.

En general, si una variable aleatoria discreta X tiene esta función de probabilidad decimos que X tiene una *distribución binomial* con parámetros n y p. En este caso usaremos la notación $X \sim b(n, p)$

Si $p = 1/2$ la función de probabilidad es simétrica con respecto a $n/2$, ya que en este caso $P(d_n = k) = P(d_n = n - k)$.

Ejemplo.

> Se extraen con reposición cinco cartas de un juego de barajas. Sea X el número de diamantes en la muestra. ¿Cuál es la probabilidad de que haya exactamente dos diamantes entre las cinco cartas? ¿Cuál es la probabilidad de que haya a lo sumo dos diamantes?

▶ Para responder la primera pregunta queremos calcular $P(X = 2)$, y como la probabilidad de obtener un diamante en cada extracción es $1/4$ tenemos:

$$P(X = 2) = \binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 = 0.264$$

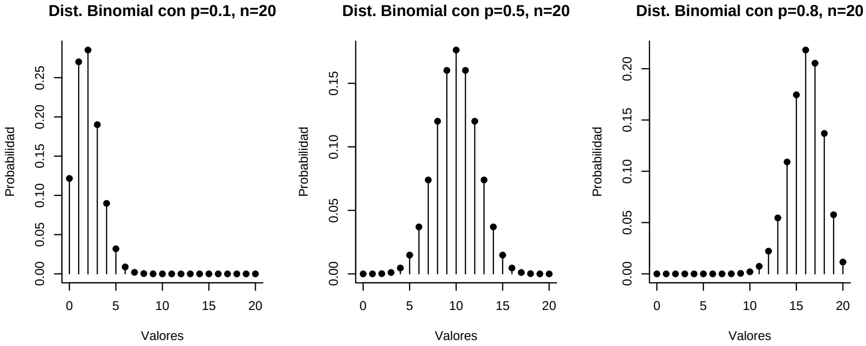


Figura 4.11: Distribución binomial para $n = 20$ y tres valores de p.

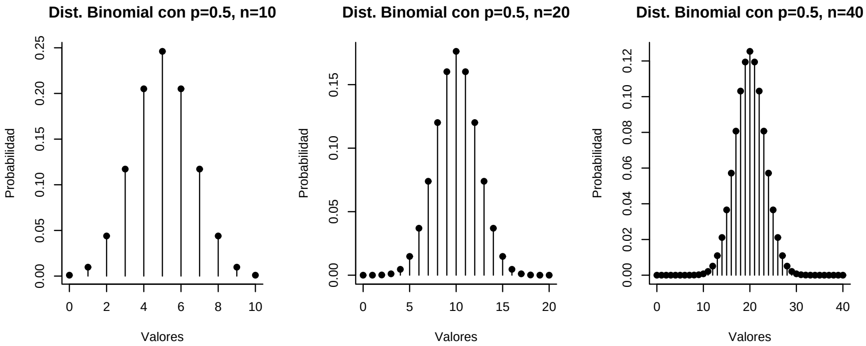


Figura 4.12: Distribución binomial para $p = 0.5$ y tres valores de n.

Para la segunda pregunta tenemos que:

$$\begin{aligned}
P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= \binom{5}{0}\left(\frac{3}{4}\right)^5 + \binom{5}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^4 + \binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 \\
&= 0.237 + 0.396 + 0.264 \\
&= 0.897
\end{aligned}$$

◀

Podemos obtener una relación recursiva entre los términos de la distribución. Si $X \sim b(n,p)$ tenemos

$$\begin{aligned}
P(X = k+1) &= \binom{n}{k+1} p^{k+1}(1-p)^{n-k+1} \\
&= \frac{n!}{(k+1)!(n-k-1)!} p^{k+1}(1-p)^{n-k} \\
&= \frac{n-k}{k+1}\frac{n!}{k!(n-k)!}\Big(\frac{p}{1-p}\Big) p^k (1-p)^{n-k} \\
&= \frac{n-k}{k+1}\Big(\frac{p}{1-p}\Big) P(X = k). \qquad (4.8)
\end{aligned}$$

Podemos usar esta relación comenzando en $P(X = 0) = (1-p)^n$ o en $P(X = n) = p^n$ para calcular los valores de la distribución.

4.4.4. La Distribución de Poisson.

Decimos que la variable aleatoria X tiene *distribución de Poisson* con parámetro λ, $(\lambda > 0)$ si

$$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda} \qquad (n = 0, 1, 2, \dots).$$

Esta relación define efectivamente una función de probabilidad ya que, usando el desarrollo en serie de Taylor de la función exponencial,

$$\sum_{n=0}^{\infty} P(X = n) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1,$$

y es otro ejemplo de una variable aleatoria que toma valores en un conjunto numerable. Usaremos la notación $X \sim \mathcal{P}(\lambda)$

Esta distribución tiene numerosas aplicaciones y gran interés en sí misma, pero además es útil como aproximación a la distribución binomial para n grande y p pequeño, hecho que estudiaremos a continuación.

Consideremos la distribución binomial cuando n crece y p tiende a cero de manera tal que el producto np permanece fijo. La distribución binomial es

$$p_{k,n} = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

multiplicando numerador y denominador por n^k y llamando μ a np obtenemos

$$\begin{aligned} p_{k,n} &= \frac{n(n-1)\cdots(n-k+1)}{n^k\ k!} (np)^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\mu^k}{k!} (1-p)^{n-k} \\ &= \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \frac{\mu^k}{k!} (1-p)^{n-k} \\ &= \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right)}{(1-p)^k} \frac{\mu^k}{k!} (1-p)^n \end{aligned} \tag{4.9}$$

Ahora bien, podemos escribir

$$(1-p)^n = [(1-p)^{-1/p}]^{-np} = [(1-p)^{-1/p}]^{-\mu}$$

pero por la definición de e sabemos que

$$\lim_{z \to 0} (1+z)^{1/z} = e.$$

Por lo tanto, si ponemos $z = -p$ obtenemos

$$\lim_{p \to 0} (1-p)^n = \lim_{p \to 0} [(1-p)^{-1/p}]^{-\mu} = e^{-\mu}.$$

Además

$$\lim_{n \to \infty} \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right)}{(1-p)^k} = 1$$

ya que hemos supuesto que $p \to 0$ cuando $n \to \infty$ y $np = \mu$ permanece constante. Usando estos dos resultados en (4.9) obtenemos

$$\lim_{n\to\infty} p_{k,n} = \frac{e^{-\mu}\mu^k}{k!}.$$

Hemos demostrado el siguiente resultado.

Teorema 4.1 (de Aproximación de Poisson) *Sea* $X_n \sim b(n, p_n)$ *y supongamos que cuando* $n \to \infty$, $p_n \to 0$ *de modo que* np_n *permanece constante y es igual a* μ. *Entonces, cuando* $n \to \infty$

$$p_{k,n} = P(X_n = k) \to e^{-\mu}\frac{\mu^k}{k!}$$

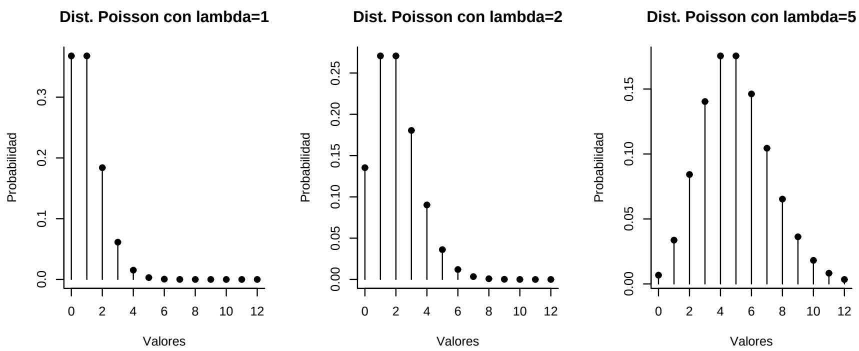


Figura 4.13: Distribución de Poisson para tres valores de λ.

Ejemplos.

1. Las llamadas que se reciben en una central telefónica por minuto tienen distribución de Poisson con parámetro $\mu = 4$. Si la central puede manejar un máximo de seis llamadas por minuto, ¿cuál es la probabilidad de que la central sea insuficiente para atender las llamadas que llegan en un minuto?

▶ Sea X el número de llamadas que se reciben en un período de un minuto. Calculemos primero

$$P(X \le 6) = \sum_{i=0}^{6} P(X = i) = \sum_{i=1}^{6} \frac{e^{-4}4^i}{i!} = 0.889,$$

por lo tanto

$$P(X > 6) = 1 - P(X \le 6) = 1 - 0.889 = 0.11.$$

◀

2. Se toma una muestra de 400 fusibles fabricados usando un procedimiento que, en promedio, produce 1 % de fusibles defectuosos. ¿Cuál es la probabilidad de que, a lo sumo, haya 5 fusibles defectuosos en la muestra?

▶ Sea X el número de fusibles defectuosos en la muestra. Sabemos que X tiene una distribución binomial con $n = 400$, $p = 0.01$ y deseamos calcular

$$P(X \le 5) = \sum_{i=0}^{5} P(X = i) = \sum_{i=0}^{5} \binom{400}{i} (0.01)^i (0.99)^{400-i}$$

y el cálculo de esta suma es trabajoso. Por lo tanto utilizaremos la distribución de Poisson con parámetro

$$\mu = np = 400 \times 0.01 = 4,$$

para aproximar la distribución binomial.

$$\begin{aligned} P(X \leq 5) &= \sum_{i=0}^{5} \frac{e^{-4}4^i}{i!} = e^{-4}\left(1 + 4 + \frac{4^2}{2} + \frac{4^3}{6} + \frac{4^4}{24} + \frac{4^5}{120}\right) \\ &= 0.785. \end{aligned}$$

◀

Para la distribución de poisson también hay una relación recursiva que permite calcular sus valores. Si $X \sim \mathcal{P}(\lambda)$ tenemos

$$\frac{P(X = i+1)}{P(X = i)} = \frac{e^{-\lambda}\lambda^{i+1}/(i+1)!}{e^{-\lambda}\lambda^i/i!} = \frac{\lambda}{i+1}$$

es decir,

$$P(X = i+1) = \frac{\lambda}{i+1} P(X = i), \quad i \geq 0. \tag{4.10}$$

4.4.5. La Distribución Hipergeométrica.

Hemos visto que el problema relativo al número de componentes defectuosos obtenidas al realizar un muestreo al azar con reposición, nos lleva a una variable aleatoria con distribución binomial.

Si realizamos un muestreo al azar pero sin reemplazar los componentes que son extraídos, la variable aleatoria que representa el número de componentes defectuosos tiene una distribución distinta a la binomial, que se conoce como *distribución hipergeométrica*.

Supongamos que en total hay n objetos de los cuales r son de tipo I (por ejemplo, defectuosos) y $n-r$ son de tipo II (por ejemplo, en buen estado). Extraemos un grupo de k elementos de esta población y llamamos X a la variable aleatoria que representa el número de objetos de tipo I en la muestra. Queremos calcular la función de probabilidad de X, es decir:

$$P(X = j)$$

donde j puede ser cualquier entero entre 0 y el menor entre k y r. Para hallar esta probabilidad observamos que el grupo de objetos que hemos escogido tiene j objetos de tipo I y $n-j$ objetos de tipo II. Los de tipo I se pueden escoger de $\binom{r}{j}$ maneras distintas mientras que los de tipo II en $\binom{n-r}{k-j}$ maneras distintas. Como cada selección de j objetos de tipo I se puede combinar con cualquier selección de $(n-j)$ objetos de tipo II tenemos que:

$$P(X = j) = \frac{\binom{r}{j}\binom{n-r}{k-j}}{\binom{n}{k}}.$$

Usando propiedades de los números combinatorios es posible reescribir la fórmula anterior:

$$P(X = j) = \frac{\binom{k}{j}\binom{n-k}{r-j}}{\binom{n}{r}}.$$

Estas probabilidades están definidas sólo si j es menor que r y k, pero si definimos:

$$\binom{a}{b} = 0 \quad \text{cuando } b > a$$

las expresiones anteriores dan $P(X = j) = 0$ cuando $j > r$ ó $j > k$.

Ejemplo.

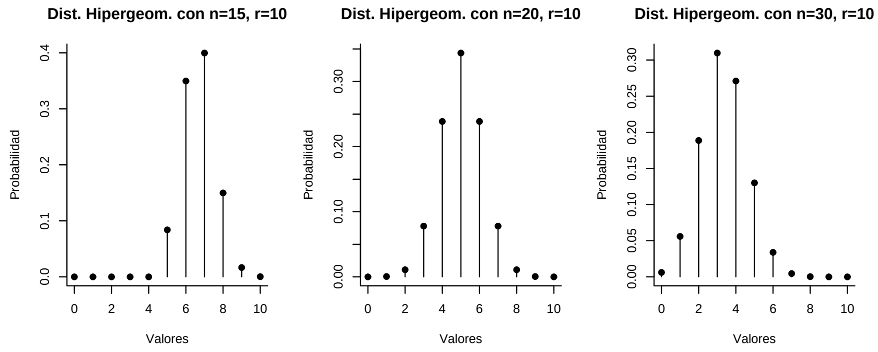


Figura 4.14: Distribución Hipergeométrica para $r = 10$ y tres valores de n.

► Como ilustración consideremos el ejemplo de una población de 100 personas de las cuales 10 tienen miopía. La probabilidad de que haya a lo sumo dos personas miopes en un grupo de 10 escogidos al azar y sin reemplazo es:

$$P(X \leq 2) = \sum_{j=0}^{2} \frac{\binom{10}{j}\binom{90}{10-j}}{\binom{100}{10}} = 0.94$$

◄

4.4.6. La Distribución Geométrica

Consideremos un fusible eléctrico que no se deteriora con el paso del tiempo pero que se quema debido a fallas en la corriente eléctrica que ocurren al azar pero en forma homogénea en el tiempo. El fusible es observado cada día y llamaremos X al número de días que transcurren hasta que el fusible falla, suponiendo que el día cero el fusible es nuevo. Queremos hallar la función de probabilidades de X.

Igual que en el caso del tiempo de vida de un componente electrónico que estudiamos en el ejemplo 3.3.1, la idea de que el fusible no se deteriora con el paso del tiempo se puede expresar con mayor precisión de la manera siguiente: si sabemos que el fusible no ha fallado antes o durante el día n, es decir, $X > n$, entonces la probabilidad de que no falle hasta después del día $n + m$, $P(X > n + m|X > n)$ debe ser igual a la probabilidad de que un fusible nuevo el día n no falle hasta después del día $n + m$.

Como las fallas eléctricas que hacen que el fusible se queme ocurren en forma homogénea en el tiempo, esta probabilidad debe depender solamente del número de días transcurridos, que es m, pero no de n. Por lo tanto tenemos la ecuación

$$P(X > n + m|X > n) = P(X > m)$$

y usando la definición de probabilidad condicional podemos reescribir esta identidad como

$$P(X > n + m) = P(X > n)P(X > m) \quad n, m = 0, 1, 2, \dots \tag{4.11}$$

Si hacemos $n = m = 0$ obtenemos

$$P(X > 0) = (P(X > 0))^2$$

y por lo tanto $P(X > 0) = 0$ ó 1. Si $P(X > 0) = 0$ entonces $P(X = 0) = 1$, lo cual es imposible. Por lo tanto, $P(X > 0) = 1$.

Llamemos $p = P(X = 1)$, entonces

$$P(X > 1) = 1 - p$$

y usando (4.11) con $m = 1$ obtenemos

$$P(X > n + 1) = (1 - p)P(X > n).$$

Iterando en n obtenemos que

$$P(X > n) = (1-p)^n$$

y por lo tanto

$$\begin{aligned} P(X = n) &= P(X > n-1) - P(X > n) \\ &= (1-p)^{n-1} - (1-p)^n \\ &= p(1-p)^{n-1} \end{aligned}$$

para $n \geq 1$.

Definición 4.7 Decimos que la variable aleatoria Y tiene *distribución geométrica* si su función de probabilidad es

$$P(Y = n) = \begin{cases} p(1-p)^{n-1} & n = 0, 1, 2, \dots \\ 0 & \text{para cualquier otro } n \end{cases}$$

donde $0 < p < 1$. Usaremos la notación $X \sim \mathcal{G}(p)$ en este caso.

Observamos que en el ejemplo anterior la variable X tiene distribución geométrica.

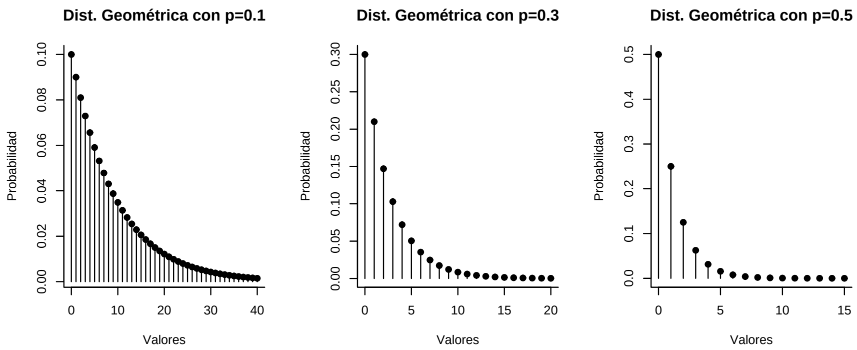


Figura 4.15: Distribución Geométrica para tres valores de p.

4.4.7. La Distribución Binomial Negativa.

Esta distribución también se conoce como la Distribución de Pascal y aparece en el contexto de una sucesión de ensayos de Bernoulli con probabilidad de éxito p, cuando nos hacemos una pregunta similar a la realizada para la distribución geométrica, pero en lugar de preguntar por el número de ensayos necesarios para lograr el primer éxito, preguntamos por el número de ensayos necesarios para lograr k éxitos.

Sea X la variable descrita anteriormente. X vale n si y sólo si el k-ésimo éxito ocurre en el n-ésino ensayo, esto es, en los primeros $n-1$ ensayos hay $k-1$ éxitos y en el n-ésimo ensayo hay un éxito. La probabilidad de esto último es p, mientras que la probabilidad de tener $k-1$ éxitos en $n-1$ ensayos es una distribución binomial:

$$\binom{n-1}{k-1} p^{k-1} q^{n-k}.$$

Como los ensayos son independientes, tenemos que la probabilidad $P(X = n)$ es el producto de las dos expresiones anteriores, es decir,

$$P(X = n) = \binom{n-1}{k-1} p^k q^{n-k}.$$

Ejemplo.

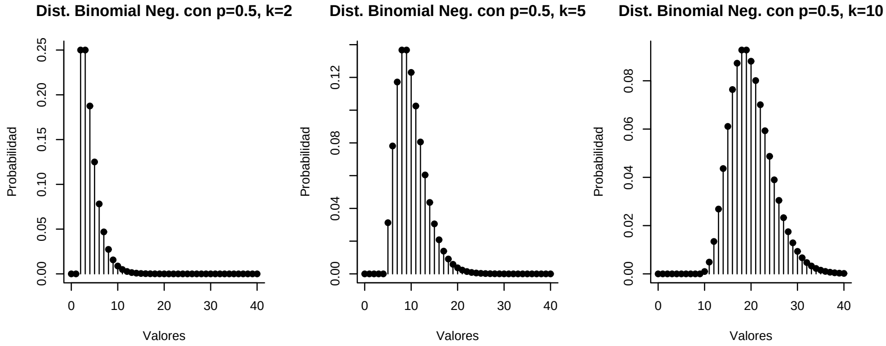


Figura 4.16: Distribución Binomial Negativa con $p = 0.5$ para tres valores de k.

Un pescador va todos los dias al muelle y se queda pescando hasta que hayan pasado dos horas o hasta que logre pescar un pez. Si la probabilidad de que no pesque nada es 0.6, ¿cuál es la probabilidad de que tenga que esperar cinco días para pescar tres peces?

▶ Sea X el número de días necesarios para pescar tres peces. Esta variable tiene distribución binomial negativa con parámetros 3 y 0.4, por lo tanto

$$P(X = 5) = \binom{4}{2}(0.4)^3(0.6)^2 = 0.138$$

◀

4.5. Variables Aleatorias Continuas.

Las variables aleatorias que hemos estudiado en las secciones anteriores típicamente representan el número de objetos que poseen una cierta propiedad, como por ejemplo el número de objetos defectuosos en una muestra de tamaño n.

Hay muchas situaciones en las cuales las variables aleatorias que debemos considerar toman valores continuos en lugar de discretos. Por ejemplo, en el capítulo anterior consideramos el tiempo de vida útil T de una maquinaria y obtuvimos, bajo ciertas condiciones específicas, que esta variable T satisface

$$P(T > x) = e^{-\lambda x}, \qquad \text{para } x > 0 \text{ y algún } \lambda > 0.$$

Esta es una variable aleatoria que puede tomar cualquier valor real positivo, y por lo tanto no está concentrada en un conjunto numerable de valores. Más aún, para cualquier $x > 0$ se tiene que

$$P(T = x) = 0 \tag{4.12}$$

es decir, que la probabilidad de que la variable aleatoria tome cualquier valor fijo es 0.

En este caso, la función de distribución de T es

$$F(x) = P(T \leq x) = 1 - P(T > x) = \begin{cases} 0, & \text{si } x \leq 0 \\ 1 - e^{-\lambda x}, & \text{si } x > 0 \end{cases}$$

que es una función continua cuya gráfica es:

En general diremos que una variable aleatoria X es *continua* si su función de distribución lo es. Por (4.7) esto equivale a pedir

$$P(X = x) = 0, \qquad \text{para todo } x \in \mathbb{R}.$$

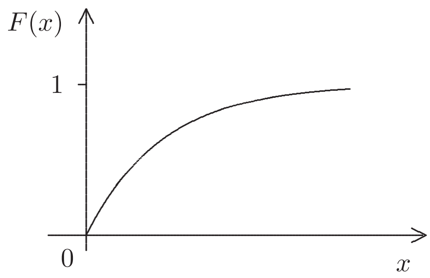


Figura 4.17

Ejemplo.

Consideremos el experimento que consiste en escoger un punto al azar en un disco D de radio R con centro en el origen. Interpretaremos la expresión "al azar" como equivalente a "si A y B son subconjuntos del disco con igual área y ω es el punto que se escoge al azar entonces $P(\omega \in A) = P(\omega \in B)$". Como conclusión, la probabilidad de que el punto escogido esté en un subconjunto A del disco debe ser proporcional al área de A:

$$P(\omega \in A) = C|A|$$

donde C es la constante de proporcionalidad y $|A|$ representa el área del conjunto A. Como

$$P(\omega \in D) = 1 = C|D|$$

obtenemos que

$$C = \frac{1}{|D|} \qquad \text{y} \qquad P(\omega \in A) = \frac{|A|}{|D|}.$$

En el razonamiento anterior hay un punto fundamental que hemos pasado por alto y es el concepto de área de un subconjunto de D. ¿Qué es el área de un subconjunto de $A \subset D$?

Si A es una figura geométrica elemental, como un rectángulo o un círculo, sabemos calcular, con exactitud, su área, pero ¿qué sucede con conjuntos más complicados? Para citar un ejemplo consideremos el conjunto

$$\{\omega \in D : \text{ la primera coordenada de } \omega \text{ es racional}\}.$$

¿Cómo se define en este caso el área del conjunto?

La resolución de este problema requiere herramientas matemáticas de la Teoría de la Medida, que están más allá del nivel de este curso. Nos limitaremos a decir que existe una σ-álgebra $\mathcal{F}$ de subconjuntos de D (o más generalmente de $\mathbb{R}^2$) y una función no-negativa m definida sobre ella que es σ-aditiva y que coincide con la noción de área para todas las figuras elementales.

En particular, si A es un círculo entonces sabemos calcular $P(\omega \in A)$, y esto va a ser suficiente para el ejemplo en cuestión.

Sobre este espacio D definimos la variable X como la distancia del punto escogido al origen y calcularemos su función de distribución. Si $0 \leq x \leq R$, el evento

$$\{\omega : X(\omega) \leq x\}$$

es el disco del plano que está centrado en el origen y tiene radio x. Su área es πx^2. Por lo tanto

$$F(x) = P(X \leq x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}, \qquad 0 \leq x \leq R.$$

Además, si $x < 0$ entonces $P(X \leq x) = 0$ y si $x > R$ entonces $P(X \leq x) = 1$. Por lo tanto

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ x^2/R^2, & \text{si } 0 \leq x \leq R, \\ 1, & \text{si } x > R, \end{cases}$$

que es una función continua, de modo que X es una variable aleatoria continua. La gráfica de F es

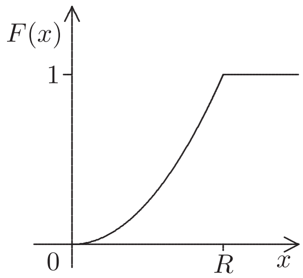


Figura 4.18

◀

Es importante observar que, de acuerdo a las definiciones que hemos dado, hay variables aleatorias que no son discretas ni continuas, es decir, hay variables aleatorias que ni toman únicamente valores en un conjunto numerable, ni tienen funciones de distribución continuas. Por ejemplo, la variable aleatoria correspondiente a la función de distribución que representamos en la figura 4.19 está en esta clase.

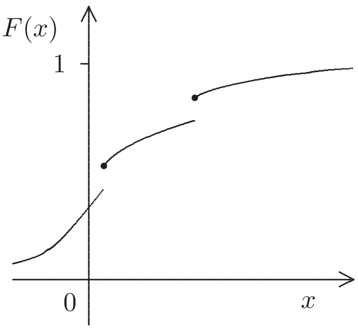


Figura 4.19

4.6. Densidades.

Definición 4.8 Sea X una variable aleatoria con función de distribución F. Decimos que F tiene densidad o es absolutamente continua, si existe una función f no negativa tal que

$$F(x) = \int_{-\infty}^{x} f(t)\ dt, \qquad \text{para todo } x \in \mathbb{R}. \tag{4.13}$$

La función f se llama la densidad de la función de distribución o de la distribución de probabilidad o de la variable aleatoria X.

De la propiedad 2 de la proposición 4.3 resulta que

$$\int_{-\infty}^{\infty} f(t)\ dt = 1. \tag{4.14}$$

Además, si f es continua en un punto x_0, F es derivable en x_0 y

$$F'(x_0) = f(x_0).$$

En efecto,

$$\begin{aligned}\frac{F(x_0+h)-F(x_0)}{h} - f(x_0) &= \frac{1}{h}\left(\int_{-\infty}^{x_0+h} f(t)\ dt - \int_{-\infty}^{x_0} f(t)\ dt\right) - f(x_0)\\ &= \frac{1}{h}\int_{x_0}^{x_0+h} f(t)\ dt - f(x_0)\\ &= \frac{1}{h}\int_{x_0}^{x_0+h} (f(t)-f(x_0))\ dt.\end{aligned}$$

Dado $\varepsilon > 0$, como f es continua, existe $\delta > 0$ tal que

$$|t - x_0| < \delta \Rightarrow |f(t) - f(x_0))| < \varepsilon$$

y en consecuencia, si $|h| < \delta$ entonces

$$\begin{aligned}\left|\frac{F(x_0+h)-F(x_0)}{h} - f(x_0)\right| &\leq \frac{1}{|h|}\int_{x_0}^{x_0+h} |f(t)-f(x_0)|\ dt\\ &\leq \frac{1}{|h|}\int_{x_0}^{x_0+h} \varepsilon\ dt = \varepsilon\end{aligned}$$

y esto demuestra el resultado.

También se tiene que

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(t)\ dt.$$

Geométricamente, por lo tanto, la probabilidad de que la variable aleatoria X pertenezca al intervalo $(a, b]$ es el área comprendida entre la gráfica de la función f, el eje x y las verticales por a y b.

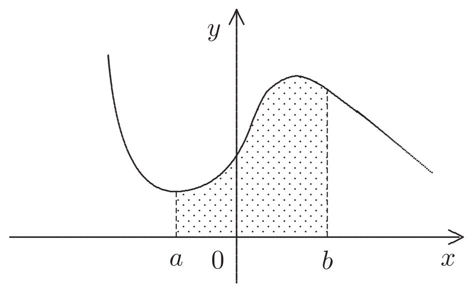


Figura 4.20

La función f puede, en general, ser bastante complicada. A lo largo del curso nos limitaremos a considerar funciones suficientemente regulares como para poder manejarlas con las nociones básicas de cálculo diferencial e integral. Estas funciones f serán continuas, salvo a lo sumo en un número finito de puntos. En cualquier caso, admitimos como resultado que si F tiene densidad, entonces F es una función continua en todo punto, y por lo tanto X es una variable aleatoria continua.

Recíprocamente, si f es una función no-negativa que verifica la condición (4.14), la función F definida por (4.13) es una función de distribución, es decir, que satisface las condiciones 1, 2 y 3 de la proposición 4.3. En efecto, sea $x < y$, entonces

$$F(y) - F(x) = \int_{-\infty}^{y} f(t)\ dt - \int_{-\infty}^{x} f(t)\ dt = \int_{x}^{y} f(t)\ dt \geq 0$$

porque f es no-negativa. Esto muestra que $F(x) \leq F(y)$ y la condición 1 es válida. La condición 2 es inmediata y en cuanto a la continuidad por la derecha consideremos

$$F(x + 1/n) - F(x) = \int_{x}^{x+\frac{1}{n}} f(t)\ dt$$

y esto tiende a cero cuando $n \to \infty$ ya que f es integrable.

De modo que decir que f es una densidad de probabilidad no es otra cosa que decir que f es no-negativa y verifica (4.14).

4.6.1. La Distribución Uniforme.

Definición 4.9 Una variable aleatoria X tiene *distribución uniforme* en el intervalo $[a, b]$ si para cualquier intervalo I contenido en $[a, b]$ se tiene que $P(X \in I)$ es proporcional a la longitud de I. Usaremos la notación $X \sim \mathcal{U}[a, b]$ en este caso.

Hemos considerado anteriormente esta distribución de probabilidad cuando estudiamos el problema del error de redondeo al truncar un número en su parte entera.

Podemos calcular la función de distribución de X,

$$F_X(x) = P(X \in [a, x]) = K(x - a)$$

donde K es la constante de proporcionalidad. Como

$$F_X(b) = P(X \in [a, b]) = 1$$

obtenemos

$$K(b - a) = 1 \qquad \text{de donde} \qquad K = \frac{1}{b - a}.$$

En consecuencia la distribución de probabilidad de X es la siguiente

$$F_X(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } x > b. \end{cases}$$

Esta distribución tiene como densidad la función f_X definida por

$$f_X(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{1}{b-a}, & \text{si } a \leq x \leq b, \\ 0, & \text{si } x > b, \end{cases}$$

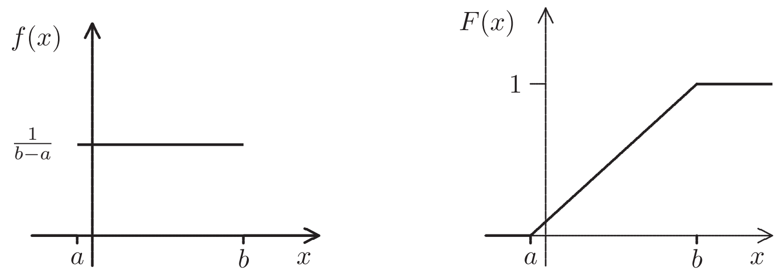


Figura 4.21

ya que se verifica inmediatamente que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt \qquad \text{para todo } x \in \mathbb{R}.$$

En la figura 4.21 representamos las gráficas de estas funciones. Usaremos la notación $X \sim \mathcal{U}[a, b]$

Ejemplo.

Entre las 7 y las 8 de la mañana los trenes salen de cierta estación de metro cada 10 minutos a partir de las 7:03. Calcule la probabilidad de que una persona que llega a la estación tenga que esperar menos de 2 minutos por el tren si la llegada de la persona a la estación tiene distribución uniforme en el intervalo:

i. de 7 a 8 a.m.

ii. de 7:15 a 7:30 a.m.

► Para que una persona espere menos de dos minutos tiene que llegar a la estación en uno de los intervalos de la forma $(t - 2, t)$ donde t es uno de los instantes en los cuales parte un tren.

En el primer caso los intervalos de interés son

$$\begin{array}{lll} (7:01,\ 7:03) & (7:11,\ 7:13) & (7:21,\ 7:23) \\ (7:31,\ 7:33) & (7:41,\ 7:43) & (7:51,\ 7:53) \end{array}$$

Sea B la unión de estos intervalos. Sabemos que la distribución del tiempo de llegada es uniforme en $[7:00, 8:00]$ y deseamos calcular la probabilidad de que $X \in B$. Como la longitud total de B es 12 minutos tenemos

$$P(X \in B) = \frac{\text{longitud de } B}{60} = \frac{12}{60} = \frac{1}{5}.$$

En el segundo caso, usando una notación similar tenemos

$$B = (7:21, 7:23),$$

de modo que

$$P(X \in B) = \frac{2}{15}.$$

◄

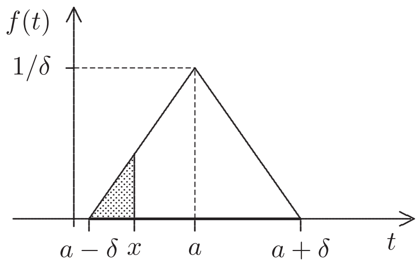


Figura 4.22

4.6.2. La Distribución Triangular.

Una distribución que se presenta con frecuencia en las aplicaciones es aquella que tiene densidad triangular simétrica, como la que se representa en la figura 4.22. Hay un punto a de máxima densidad y puntos $a-\delta$, $a+\delta$ equidistantes de a, entre los cuales toma su valor la variable aleatoria en cuestión. Entre el punto medio y los extremos, la densidad f varía linealmente.

El valor $f(a)$ tiene que ser tal que el área del triángulo sea 1, es decir

$$\int_{-\infty}^{\infty} f(t)\ dt = \delta f(a) = 1 \quad \Rightarrow \quad f(a) = \frac{1}{\delta}.$$

Por lo tanto la función de distribución es

$$\begin{aligned} F(x) &= 0, \quad \text{si } x < a-\delta \\ F(x) &= 1, \quad \text{si } x > a+\delta, \end{aligned}$$

y si $a-\delta \leq x < a$

$$F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_{a-\delta}^{x} f(t)\ dt$$

que es al área del triángulo pequeño indicado en la figura 4.17, es decir

$$\begin{aligned} F(x) &= \frac{1}{2}(x-(a-\delta))f(x) \\ &= \frac{1}{2}(x-(a-\delta))\frac{(x-(a-\delta))}{\delta}f(a) \\ &= \frac{1}{2}\frac{(x-a+\delta)^2}{\delta^2}. \end{aligned}$$

Si $a \leq x < a+\delta$, usando la simetría del triángulo,

$$\begin{aligned} F(x) &= 1 - \frac{1}{2}(a+\delta-x)f(x) \\ &= 1 - \frac{1}{2}\frac{(a+\delta-x)^2}{\delta^2}. \end{aligned}$$

Resumiendo

$$F(x) = \begin{cases} 0, & \text{si } x < a-\delta, \\ \frac{(x-a+\delta)^2}{2\delta^2}, & \text{si } a-\delta \leq x < a, \\ 1 - \frac{(a+\delta-x)^2}{2\delta^2}, & \text{si } a \leq x < a+\delta, \\ 1, & \text{si } x \geq a+\delta. \end{cases}$$

cuya gráfica es

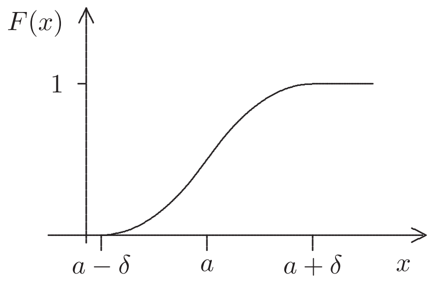


Figura 4.23

4.6.3. La Distribución Exponencial.

Definición 4.10 Decimos que la variable aleatoria X tiene *distribución exponencial* si

$$P(X > x) = e^{-\lambda x} \quad (x \geq 0)$$

donde $\lambda > 0$. Por lo tanto, la función de distribución respectiva es

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - e^{-\lambda x}, & \text{si } x \geq 0. \end{cases}$$

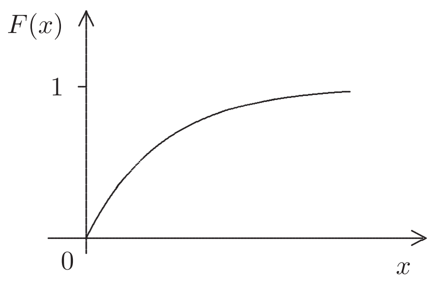


Figura 4.24

y la densidad de esta distribución es

$$f_X(x) = \begin{cases} 0, & \text{si } x < 0 \\ \lambda e^{-\lambda x}, & \text{si } x \geq 0. \end{cases}$$

Usaremos la notación $X \sim \mathcal{E}(\lambda)$. Una propiedad importante de la distribución exponencial es la siguiente: para a y b no negativos

$$P(X > a + b) = P(X > a)P(X > b). \tag{4.15}$$

La verificación de esta propiedad es inmediata a partir de la definición. Una forma equivalente de escribir esta propiedad es

$$P(X > a + b | X > a) = P(X > b), \ a \geq 0, \ b \geq 0, \tag{4.16}$$

que ha sido interpretada en ocasiones anteriores como una formulación precisa de la distribución del tiempo de vida de un objeto que "no envejece" con el paso del tiempo, o de la "falta de memoria" de esta distribución. Mas aún, en el ejemplo 3.1.3 vimos que si (4.16), o equivalentemente (4.15), es cierto entonces se deduce que

$$P(X > x) = e^{-\lambda x}, \quad (x \geq 0)$$

para algún λ positivo. Por lo tanto cualquiera de las relaciones (4.15) o (4.16) caracteriza a la distribución exponencial.

Ejemplos.

1. La distribución exponencial surge, por ejemplo, en el estudio del tiempo de vida de un material radioactivo. Si suponemos que la tasa a la cual decae una masa m de material radioactivo es proporcional a la cantidad de material presente en el instante t, entonces m satisface la ecuación

$$\frac{dm}{dt} = -\lambda m$$

donde λ es una constante que depende del material. La solución de esta ecuación es

$$m = m_0 \ e^{-\lambda t}.$$

En efecto, la derivada de esta función es

$$\frac{dm}{dt} = m_0 e^{-\lambda t}(-\lambda) = -\lambda m,$$

donde m_0 es la cantidad de material en el instante $t = 0$. La proporción de material original que ha decaído en t unidades de tiempo está dada por $(m_0 - m)/m_0$, que puede ser interpretada como la probabilidad de que un átomo seleccionado al azar entre el material original decaiga durante un período de tiempo t. Si X representa la vida de este átomo,

$$F_X(x) = P(X \leq t) = \frac{m_0 - m}{m_0} = 1 - e^{-\lambda t}$$

de modo que X tiene distribución exponencial. ◀

2. Sea X una variable aleatoria con distribución uniforme en $(0, 1)$. Hallar la densidad de

$$Y = \frac{-1}{\lambda} \ln(1 - X) \quad \text{para } \lambda > 0.$$

▶ Sea G la función de distribución de Y, como esta variable sólo toma valores positivos tenemos que $G(y) = 0$ para $y \leq 0$. Para $y > 0$

$$\begin{aligned} G(y) = P(Y \leq y) &= P\left(\frac{-1}{\lambda} \ln(1 - X) \leq y\right) \\ &= P(\ln(1 - X) \geq -\lambda y) = P(1 - X \geq e^{-\lambda y}) \\ &= P(X \leq 1 - e^{-\lambda y}) = 1 - e^{-\lambda y} \end{aligned}$$

y por lo tanto $G'(y) = \lambda e^{-\lambda y}$ para $y \geq 0$ y $G'(y) = 0$ para $y \leq 0$. En consecuencia, la densidad de Y está dada por

$$g(y) = \begin{cases} \lambda e^{-\lambda y} & \text{si } y > 0 \\ 0 & \text{si } y \leq 0 \end{cases}$$

es decir, Y tiene distribución exponencial con parámetro λ. ◀

4.6.4. La Distribución Normal.

La *función de distribución normal* con parámetros μ y σ^2 (cuyo significado veremos más adelante), es aquella que tiene densidad

$$n(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(\frac{-(y-\mu)^2}{2\ \sigma^2}\right).$$

Dado que esta función nunca es negativa, para ver que es efectivamente una densidad de probabilidad debemos probar que

$$\int_{-\infty}^{\infty} n(y;\mu,\sigma^2)\ dy = 1,$$

haciendo el cambio de variables $z = (x-\mu)/\sqrt{2}\sigma$ resulta

$$\begin{aligned}\int_{-\infty}^{\infty} n(y;\mu,\sigma^2)\ dy &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2} \sigma\sqrt{2}\ dz \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz. \end{aligned} \tag{4.17}$$

Una manera de calcular esta última integral es la siguiente. Sea C_r el disco con centro en el origen y radio r y C_r' el disco con el mismo centro y radio $\sqrt{2}r$. Sea D_r el cuadrado con centro en el origen y lado $2r$ (ver figura 4.24).

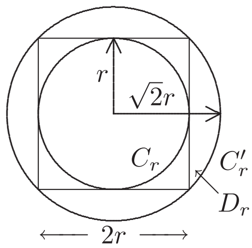


Figura 4.25

Dado que el integrando común en las integrales siguientes no es negativo, tenemos que

$$\iint_{C_r} e^{-(u^2+v^2)} du\ dv \le \iint_{D_r} e^{-(u^2+v^2)} du\ dv \le \iint_{C_r'} e^{-(u^2+v^2)} du\ dv \tag{4.18}$$

y además

$$\iint_{D_r} e^{-(u^2+v^2)} du\ dv = \int_{-r}^{r} e^{-u^2} du \int_{-r}^{r} e^{-v^2} dv = \left(\int_{-r}^{r} e^{-u^2} du\right)^2.$$

Consideremos ahora la integral de la izquierda en (4.18). Pasando a coordenadas polares ρ, θ por medio de la transformación $u = \rho\cos\theta$, $v = \rho \operatorname{sen}\theta$, obtenemos

$$\begin{aligned}\iint_{C_r} e^{-(u^2+v^2)} du\ dv &= \int_0^{2\pi} d\theta \int_0^r e^{-\rho^2} \rho\ d\rho \\ &= 2\pi \left[-\frac{1}{2} e^{-\rho^2} \Big|_0^r \right] \\ &= 2\pi \left[\frac{1}{2}(1-e^{-r^2}) \right] \\ &= \pi(1-e^{-r^2}).\end{aligned}$$

De forma análoga, cambiando r por $\sqrt{2}r$ resulta

$$\iint_{C'_r} e^{-(u^2+v^2)} du\ dv = \pi(1 - e^{-2r^2}).$$

Reemplazando en (4.18)

$$\pi(1 - e^{-r^2}) \leq \left(\int_{-r}^{r} e^{-u^2} du \right)^2 \leq \pi(1 - e^{-2r^2}),$$

haciendo $r \to \infty$

$$\pi \leq \left(\int_{-\infty}^{\infty} e^{-u^2} du \right)^2 \leq \pi$$

y por lo tanto

$$\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}.$$

Sustituyendo en (4.17) resulta

$$\int_{-\infty}^{\infty} n(y; \mu, \sigma^2)\ dy = 1.$$

Si X tiene distribución normal de parámetros μ y σ^2 usaremos la notación $X \sim \mathcal{N}(\mu, \sigma^2)$. En el caso $\mu = 0$ y $\sigma^2 = 1$, la densidad $n(y; 0, 1)$ se conoce como la densidad normal estándar o típica, y se denota usualmente por ϕ, de modo que

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

En la Figura 4.25 representamos la densidad normal para $\mu = 0$ y tres valores de σ: 0.5, 1 y 2. Estas densidades son claramente simétricas respecto al origen. La función de distribución correspondiente a la densidad ϕ se denota usualmente por Φ. Esta distribución no tiene una fórmula sencilla y debe ser calculada numéricamente. Es posible calcular los valores de esta función en `R` usando la función `dnorm` cuyos parámetros son $x,\ \mu = 0,\ \sigma = 1$.

Como ϕ es simétrica respecto al origen tenemos que

$$\begin{aligned} \Phi(-x) &= \int_{-\infty}^{-x} \phi(y)\ dy = \int_{x}^{\infty} \phi(y)\ dy \\ &= \int_{-\infty}^{\infty} \phi(y)\ dy - \int_{-\infty}^{x} \phi(y)\ dy \\ &= 1 - \Phi(x) \end{aligned}$$

de modo que para cualquier valor de x se tiene $\Phi(-x) = 1 - \Phi(x)$ y esta fórmula nos permite obtener el valor de $\Phi(-x)$ a partir del valor de $\Phi(x)$. Por lo tanto bastaconocer los valores de $\Phi(x)$ para $x \geq 0$.

Sea X una variable aleatoria con densidad ϕ y consideremos $Y = \mu + \sigma X$, donde $\sigma > 0$. En la próxima sección demostraremos que Y tiene como densidad a la función

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} \right) = n(y; \mu, \sigma^2) \tag{4.19}$$

y este hecho nos permite calcular cualquier distribución normal a partir de Φ ya que

$$P(Y \leq y) = P(\mu + \sigma X \leq y) = P\left(X \leq \frac{y-\mu}{\sigma} \right) = \Phi\left(\frac{y-\mu}{\sigma} \right).$$

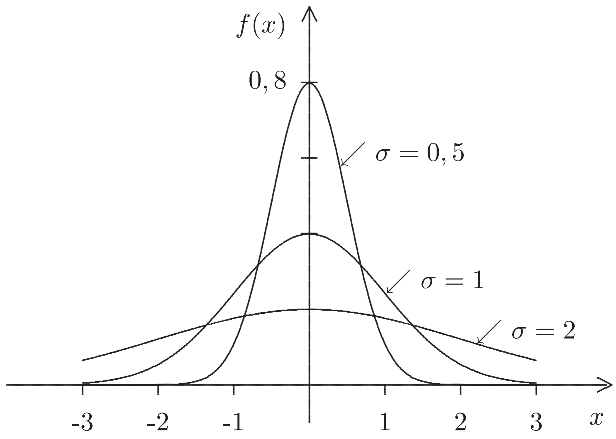


Figura 4.26

Por lo tanto, si Y tiene una distribución con densidad $n(y;\mu,\sigma^2)$ y $a \leq b$ entonces

$$P(a \leq Y \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Por ejemplo, supongamos que Y tiene una distribución normal con parámetros $\mu = 0.5$ y $\sigma = 4$ y queremos calcular la probabilidad de que $-0.5 \leq Y \leq 2.4$:

$$P(-0.5 \leq Y \leq 2.5) = \Phi\left(\frac{1}{2}\right) - \Phi\left(\frac{-1}{4}\right) = 0.6915 - 0.4013 = 0.2902.$$

4.7. Cambio de Variable.

Sea X una variable aleatoria con función de distribución continua y $g : \mathbb{R} \to (a,b)$ una función biyectiva con derivada continua y que no se anula (es decir, $g'(x) \neq 0$, para todo $x \in \mathbb{R}$). En estas condiciones g es una función estrictamente monótona cuyo rango es justamente (a,b). El intervalo (a,b) puede ser también una semirecta o la recta entera. Consideremos la nueva variable aleatoria $Y = g(X)$ en el caso particular cuando g es creciente.

Proposición 4.5 *Sea X una variable aleatoria con función de distribución F_X y sea g una función estrictamente creciente. Definimos $Y = g(X)$ y sea F_Y la función de distribución de esta variable. Entonces*

$$F_Y(y) = F_X(g^{-1}(y)). \tag{4.20}$$

Demostración. Como g es estríctamente creciente los eventos $\{X \leq g^{-1}(y)\}$ y $\{g(X) \leq y\}$ son iguales. Por lo tanto,

$$F_Y(y) = P(Y \leq y) = P(g(Y) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

■

Si g es estríctamente decreciente entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Corolario 4.1 *Sea F una función de distribución estríctamente creciente para los y tales que $0 < F(y) < 1$ y sea $U \sim \mathcal{U}[0,1]$. Entonces la variable $Z = F^{-1}(U)$ tiene distribución F.*

Demostración. La función de distribución de U es $F_U(u) = u$ para $u \in [0,1]$. Entonces

$$F_Z(z) = F_U(F(z)) = F(z) \tag{4.21}$$

de modo que Z tiene función de distribución F. ■

Observación 4.2 El resultado anterior es cierto en general si utilizamos la *inversa generalizada* $F^{\leftarrow}$ de la función F cuando esta no sea estrictamente creciente, que se define por la siguiente expresión:

$$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$$

Por lo tanto, para cualquier función de distribución F, la variable aleatoria $Z = F^{\leftarrow}(U)$ tiene función de distribución F.

Proposición 4.6 *Sea X una variable aleatoria con función de distribución continua cuya distribución de probabilidad tiene densidad f_X y $g : \mathbb{R} \to (a, b)$ una función biyectiva con derivada continua y que no se anula (es decir, $g'(x) \neq 0$, para todo $x \in \mathbb{R}$). Definimos $Y = g(X)$ y sea F_Y la función de distribución de esta variable. Entonces Y tiene densidad*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\frac{1}{g'(g^{-1}(y))}, & \text{si } y \in (a, b), \\ 0, & \text{si } y \notin (a, b), \end{cases}$$

donde g^{-1} denota la función inversa de g.

Demostración. En efecto, sea $a < y_0 < b$ y supongamos que g es creciente,

$$P(Y \leq y_0) = P(X \leq g^{-1}(y_0)) = \int_{-\infty}^{g^{-1}(y_0)} f_X(x)\, dx.$$

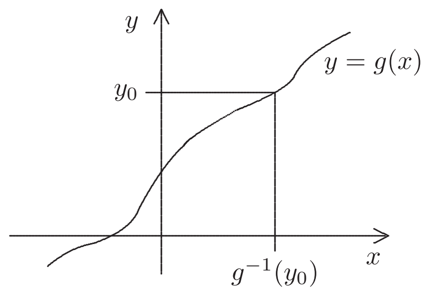


Figura 4.27

Haciendo el cambio de variables $y = g(x)$ en esta integral, obtenemos

$$P(Y \leq y_0)) = \int_a^{y_0} f_X(g^{-1}(y))\frac{1}{g'(g^{-1}(y))} dy.$$

Además, es inmediato que

$$P(Y \leq y_0) = \begin{cases} 0, & \text{si } y_0 \leq a, \\ 1, & \text{si } y_0 \geq b. \end{cases}$$

En cualquier caso resulta

$$P(Y \leq y_0) = \int_{-\infty}^{y_0} f_Y(y)\, dy,$$

donde $f_Y(y)$ es la función indicada anteriormente. ■

Por ejemplo, si $y = g(x) = mx + c$ donde m y c son constantes tales que $m \neq 0$ entonces la variable

$$Y = g(X) = mX + c$$

tiene densidad

$$f_Y(y) = \frac{1}{m} f_X\left(\frac{y-c}{m}\right), \qquad -\infty < y < \infty.$$

En el caso particular de la distribución normal, si $X \sim \mathcal{N}(0,1)$ con densidad $\phi(x)$, la variable $Y = \mu + \sigma X$ tiene como densidad

$$\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{1\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = n(y;\mu,\sigma^2)$$

como afirmamos en (4.19).

4.8. Simulación de Variables Aleatorias.

Los generadores de números aleatorios simulan valores de la distribución $\mathcal{U}[0,1]$. El Corolario 4.1 y la Observación 4.2 nos dan un método para simular una variable aleatoria con función de distribución F: Generamos el valor u de una variable uniforme en $[0,1]$ y evaluamos la inversa generalizada en u: $F^{\leftarrow}(u)$. Sin embargo, dependiendo de la naturaleza de la función de distribución F, es posible que la inversa generalizada tenga una expresión complicada o incluso no sea posible escribirla en términos de funciones elementales, como ocurre en el caso de las variables Gaussianas. Por esta razón hay métodos *ad hoc* que resultan más eficientes en muchos casos.

4.8.1. Variables Discretas

Si queremos simular una variable aleatoria finita X con valores $x_1, \dots, x_n$ y probabilidades respectivas $p_1, \dots, p_n$, podemos dividir el intervalo $[0,1]$ en subintervalos usando las probabilidades p_i:

$$[0, p_1); \qquad [p_1, p_1 + p_2); \qquad [p_1 + p_2, p_1 + p_2 + p_3); \qquad \cdots \qquad \Big[\sum_{j<n} p_j, 1\Big].$$

Ahora generamos una variable U con distribución uniforme en $[0,1]$ y si el valor cae en el i-ésimo intervalo le asignamos a X el valor x_i. Como la probabilidad de que U caiga en el intervalo i es igual a la longitud del intervalo, que es p_i, vemos que

$$P(X = x_i) = p_i, \qquad \text{para } 1 \le i \le n.$$

Este método se conoce como el método de la transformada inversa. Desde el punto de vista computacional es conveniente ordenar los valores según el tamaño de las p_i, colocando estas probabilidades de mayor a menor, porque para identificar el intervalo en cual cae U tenemos que comparar con p_1, luego con $p_1 + p_2$, y así sucesivamente hasta obtener el primer valor menor que U. Ordenar las probabilidad hace que se maximice la probabilidad de que U esté en los primeros intervalos, y esto reduce el número de comparaciones que hay que hacer en promedio para obtener el valor de X.

Este método también funciona para variables discretas con una cantidad infinita de valores. La misma observación sobre el ordenamiento de los valores de las probabilidades es válida.

Distribución de Bernoulli

Un caso particular sencillo es el de la distribución de Bernoulli con probabilidad de éxito p. Para generar un valor de la variable X con esta distribución, generamos U y si $U < p$, $X = 1$ y si no, $X = 0$.

Distribución Uniforme Discreta

Sea X una variable aleatoria que toma valores $\{x_1, x_2, \dots, x_n\}$ con igual probabilidad. Para simular esta distribución generamos un número aleatorio $U \in (0,1]$, dividimos el intervalo $[0,1]$ en n intervalos iguales y le asignamos a la variables el valor x_k si

$$\frac{k-1}{n} < U \le \frac{k}{n}$$

es decir, el valor de la variable es X_k con $k = \lceil Un \rceil$, donde $\lceil a \rceil$ es la función techo y representa el menor entero que es mayor o igual a a.

Distribución Binomial

Una manera sencilla de simular una variable con distribución binomial de parámetros n y p es generar n variables de Bernoulli con probabilidad de éxito p y sumarlas. Esto resulta un poco pesado si n es grande, pero en este caso podemos usar el Teorema Central del Límite, que estudiaremos más adelante.

Otra posibilidad es usar el método de la transformada inversa junto con la relación (4.8) que demostramos anteriormente. Para esto generamos una variable uniforme U y comparamos con $P(X = 0) = (1-p)^n$. Si U es menor que este valor ponemos $X = 0$, en caso contrario multiplicamos $P(X = 0)$ por $pn/(1-p)$ para obtener $P(X = 1)$ y comparamos. Si U es menor que este valor ponemos $X = 1$, en caso contrario repetimos el procedimiento hasta conseguir el valor de X. El algoritmo se puede describir como sigue:

Paso 1: Generamos una variable uniforme U.
Paso 2: Ponemos $a = p/(1-p)$; $b = (1-p)^n$; $c = b$; $i = 0$.
Paso 3: Si $U < c$ ponemos $X = i$ y paramos.
Paso 4: $b = ab(n-i)/(i+1)$; $c = c + b$; $i = i + 1$.
Paso 5: Vamos al paso 3.

Distribución de Poisson

Al igual que para la distribución binomial, la relación (4.10) permite aplicar el método de la transformada inversa para generar la distribución de Poisson. El algoritmo es el siguiente:

Paso 1: Generamos una variable uniforme U.
Paso 2: Ponemos $a = e^{-\lambda}$; $b = a$; $i = 0$.
Paso 3: Si $U < b$ ponemos $X = i$ y paramos.
Paso 4: $a = \lambda a/(i+1)$; $b = b + a$; $i = i + 1$.
Paso 5: Vamos al paso 3.

Distribución Geométrica

Una manera de generar variables con distribución geométrica es generar una sucesión de variables de Bernoulli hasta obtener el primer éxito, es decir, generamos una sucesión de números aleatorios en $[0, 1]$ hasta obtener el primero que sea menor que p. Sin embargo, si p es pequeño esto puede ser lento (toma en promedio $1/p$ pasos). Para evitar esto podemos seguir el método alternativo que describimos a continuación. Sea X una v.a. con distribución geométrica de parámetro p, $0 < p < 1$ y sea u un número aleatorio en $[0, 1]$. Definimos Y como el menor entero que satisface la desigualdad $1 - q^Y \geq u$. Entonces

$$\begin{aligned} P(Y = j) &= P(1 - q^j \geq u > 1 - q^{j-1}) \\ &= q^{j-1} - q^j = q^{j-1}(1-q) = q^{j-1}p, \end{aligned}$$

de modo que Y también tiene una distribución geométrica de parámetro p. Por lo tanto, para generar Y basta resolver la ecuación que la define, es decir,

$$Y = \left\lfloor \frac{\log(1-u)}{\log q} \right\rfloor$$

pero como $1 - u$ y u tienen la misma distribución, podemos usar

$$Y = \left\lfloor \frac{\log(u)}{\log q} \right\rfloor.$$

Distribución Binomial Negativa

Observamos que una variable con distribución binomial negativa de parámetros k y p es la suma de k variables geométricas con parámetro p: una por cada éxito en la sucesión de ensayos. Esto lo veremos con mayor detalle en el próximo cap´itulo. Esta observación es útil para generar variables con esta distribución: si u_j, $j = 1, \dots, k$ son números aleatorios en $[0, 1]$, la siguiente expresión produce el valor de una variable con distribución binomial negativa:

$$\sum_{j=1}^{k} \left\lfloor \frac{\log(u_j)}{\log q} \right\rfloor.$$

4.8.2. Variables Continuas

Si X es una variable continua con función de distribución F invertible, para simular X basta generar una variable uniforme U y poner $X = F^{-1}(U)$. Esto es consecuencia del corolario 4.1. Sin embargo, con frecuencia las funciones de distribución continuas no son invertibles o si lo son, es posible que las inversas no tengan una expresión en términos de funciones elementales. Por esta razón estudiamos a continuación algunas de las distribuciones continuas que hemos considerado anteriormente.

Distribución Uniforme

Si queremos simular la distribución $\mathcal{U}[a, b]$ generamos u uniforme en $[0, 1]$ y usamos la transformación $u \mapsto a + u(b - a)$.

Distribución Exponencial

Para simular variables con distribución exponencial usamos la relación que obtuvimos en la seccion 4.6.3: Si $U \sim \mathcal{U}(0, 1)$ entonces $X = -\ln(1 - U)/\lambda \sim \mathcal{E}(\lambda)$. Observamos ahora que si U tiene distribución uniforme en $(0, 1)$, $1 - U$ también. Por lo tanto, para simular esta distribución a partir de una variable $U \sim \mathcal{U}(0, 1)$ hacemos la transformación $-\ln(U)/\lambda$.

Distribución Normal

La función de distribución normal Φ no se puede escribir en términos de funciones simples, y lo mismo ocurre con su inversa, lo que dificulta la aplicación del método de la transformada inversa. Sin embargo existen otros métodos y uno de los más populares es el de Box-Muller, también conocido como el método polar.

Aún cuando la justificación del método no es complicada, requiere algunos conceptos que no hemos introducido, así que vamos a describir el método sin demostrar que efectivamente lo que obtenemos es el valor de una variable normal. El algoritmo es el siguiente:

Paso 1: Generamos variables uniformes U_1 y U_2.
Paso 2: Ponemos $V_1 = 2U_1 - 1$; $V_2 = 2U_2 - 1$; $S = V_1^2 + V_2^2$.
Paso 3: Si $S > 1$ regresamos al paso 1.
Paso 4: X y Y son variables normales típicas independientes:

$$X = \sqrt{\frac{-2\log S}{S}} V_1, \qquad Y = \sqrt{\frac{-2\log S}{S}} V_1.$$

4.8.3. Generación de Variables Aleatorias en `R`

El lenguaje `R` tiene incorporadas una serie de rutinas para generar variables aleatorias. La sintaxis precisa de la instrucción correspondiente depende de la distribución, pero todas tienen el formato común `r`*dist*, donde *dist* designa la distribución; por ejemplo, para generar valores a partir de la distribución normal usamos `rnorm`.

Según la distribución, puede ser necesario especificar uno o varios parámetros. La tabla que presentamos a continuación presenta las distribuciones más comunes, los parámetros requeridos y sus valores por defecto. `n` representa siempre el tamaño de la muestra.

Distribución	**Función en `R`**
Binomial	`rbinom(n, size, prob)`
Poisson	`rpois(n, lambda)`
Geométrica	`rgeom(n, prob)`
Hipergeométrica	`rhyper(nn, m, n, k)`
Binomial Negativa	`rnbinom(n, size, prob)`
Multinomial	`rmultinom(n, size, prob)`
Uniforme	`runif(,n min=0, max=1)`
Exponencial	`rexp(n, rate=1)`
Gaussiana	`rnorm(n, mean=0, sd=1)`
Gamma	`rgamma(n, shape, scale=1)`
Weibull	`rweibull(n, shape, scale=1)`
Cauchy	`rcauchy(n, location=0, scale=1)`
Beta	`rbeta(n, shape1, shape2)`
t	`rt(n, df)`
Fisher	`rf(n, df1, df2)`
χ^2	`rchisq(n, df)`
Logística	`rlogis(n, location=0, scale=1)`
Lognormal	`rlnorm(n, meanlog=0, sdlog=1)`

Además, `R` tiene la función `sample` que permite obtener muestras con o sin reposición de conjuntos finitos de valores. La sintaxis es

```
sample(x, size, replace = FALSE, prob = NULL)
```

donde

- `x` es el conjunto a partir del cual queremos obtener la muestra, escrito como un vector,
- `size` es el tamaño de la muestra,
- `replace` permite indicar si se permiten repeticiones (`replace = TRUE`) o no y finalmente
- `prob` es un vector de probabilidades si se desea hacer un muestreo pesado y no uniforme.

4.9. Ejemplos.

1. Sea X una variable aleatoria continua con densidad f. Hallar la densidad de la variable aleatoria $Y = X^2$.

► Observamos que como la función $g(x) = x^2$ no es biyectiva, no es posible aplicar los resultados de la sección 4.7. Sean F y G las funciones de distribución de las variables X, Y respectivamente. Es inmediato que $G(y) = 0$ para $y \leq 0$. Para $y > 0$

$$\begin{aligned}
G(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F(\sqrt{y}) - F(-\sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(t)\, dt \\
&= \int_0^{\sqrt{y}} f(t)\, dt + \int_0^{\sqrt{y}} f(-t)\, dt \\
&= \int_0^{\sqrt{y}} (f(t) + f(-t))\, dt
\end{aligned}$$

y haciendo el cambio de variable $t = \sqrt{s}$ obtenemos, para $y > 0$

$$G(y) = \int_0^y \frac{1}{2\sqrt{s}}(f(\sqrt{s}) + f(-\sqrt{s}))\ ds.$$

Por lo tanto, la densidad g de Y es

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y})), & \text{para } y > 0 \\ 0, & \text{para } y \leq 0 \end{cases}$$

Observamos que si la densidad f de X es continua, entonces F es diferenciable y por lo tanto también lo es G, de modo que podemos obtener g directamente, derivando G:

$$g(y) = G'(y) = \frac{d}{dy}(F(\sqrt{y}) - F(-\sqrt{y})) = \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y}))$$

para $y > 0$. ◀

2. Sea X una variable aleatoria con densidad continua f que no se anula. Si F es la distribución de X, definimos la variable aleatoria Y por $Y = F(X)$, es decir, usando la notación de la sección 4.7, $g = F$. Hallar la densidad de Y.

▶ De nuevo, con la notación de la sección 4.7, tenemos que $(a, b) = (0, 1)$, $g(x) = F(x)$ y $g'(x) = f(x)$. Por lo tanto, la densidad de Y es 0 fuera de $(0, 1)$, y si $y \in (0, 1)$ entonces la densidad es

$$f(F^{-1}(y))\frac{1}{f(F^{-1}(y))} = 1.$$

Es decir, Y tiene una distribución uniforme en $(0, 1)$. ◀

3. Decimos que una densidad es *simétrica* si $f(-x) = f(x)$ para todo x. Una variable aleatoria X es *simétrica* si X y $-X$ tienen la misma función de distribución. Demuestre que una variable aleatoria X con densidad es simétrica si y sólo si su densidad f es simétrica.

▶ Supongamos primero que X tiene densidad simétrica f, entonces

$$\begin{aligned} P(-X \leq x) = P(X \geq -x) &= \int_{-x}^{\infty} f(t)\ dt \\ &= \int_{-\infty}^{x} f(-t)\ dt = \int_{-\infty}^{x} f(t)\ dt \\ &= P(X \leq x) \end{aligned}$$

de modo que X y $-X$ tienen la misma distribución. Supongamos ahora que X y $-X$ tienen la misma función de distribución y por lo tanto la misma densidad g. Definimos

$$f(x) = \frac{1}{2}(g(x) + g(-x))$$

Es fácil verificar que esta función es una densidad simétrica. Además

$$\begin{aligned} \int_{-\infty}^{x} f(t)\ dt &= \frac{1}{2}\int_{-\infty}^{x} g(t)\ dt + \frac{1}{2}\int_{-\infty}^{x} g(-t) dt \\ &= \frac{1}{2}\int_{-\infty}^{x} g(t)\ dt + \frac{1}{2}\int_{-x}^{\infty} g(t)\ dt \\ &= \frac{1}{2}P(X \leq x) + \frac{1}{2}P(X \geq -x) \\ &= P(X \leq x) \end{aligned}$$

de donde se obtiene que f es la densidad de X. ◀

Ejercicios

1. Sea X una variable aleatoria con función de probabilidad dada por la siguiente tabla:

$x_i:$	-2	-1	0	1	2	3	4	5
$p_i:$	0.1	0.2	0.15	0.2	0.1	0.15	0.05	0.05

   Calcule las probabilidades de los siguientes eventos:

   a. X es negativa. b. X es par. c. X toma valores entre 1 y 5, ambos inclusive.

   d. $P(X = -2|X \leq 0)$. e. $P(X \geq 2|X > 0)$.

2. Determine el valor de la constante A para que las siguientes sean funciones de probabilidad.

   a. $P(X = i) = \begin{cases} Ai & i = 1, 2, \ldots, n \\ 0 & \text{en otro caso.} \end{cases}$.

   b. $P(X = i) = \begin{cases} A/2^i & i = 1, 2, \ldots, n \\ 0 & \text{en otro caso.} \end{cases}$.

   c. $P(X = i) = \begin{cases} A/3^i & i = 1, 3, 5, 7, \ldots, 2n-1 \\ A/4^i & i = 2, 4, 6, 8, \ldots, 2n \\ 0 & \text{en otro caso.} \end{cases}$.

3. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad con $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \{\emptyset, \{2, 4, 6\}, \{1, 3, 5\}, \Omega\}$, y sean U, V, W funciones definidas en Ω por

$$U(\omega) = 5\omega + 32, \qquad V(\omega) = \begin{cases} 1 & \text{si } \omega \text{ es par,} \\ 0 & \text{si } \omega \text{ es impar,} \end{cases} \qquad W(\omega) = \omega^2.$$

   Determine cuáles de estas funciones son variables aleatorias sobre el espacio de probabilidad Ω.

4. Determine el valor de la constante C para que la siguiente sea una función de probabilidad.

$$P(X = n) = \frac{C}{n(n+1)}, \quad n \in \mathbb{N}.$$

5. ¿Para qué valores de C y α es la función p definida por $p(n) = Cn^\alpha$ para $n \in \mathbb{N}$ una función de probabilidad?

6. Sea X una variable con distribución uniforme en el conjunto $\{1, 2, \ldots, 50\}$. Calcule

   a. $P(X \geq 15)$, b. $P(2.5 < X \leq 43.2)$, c. $P(X > 20|X > 10)$, d. $P(X \leq 435.6|X > 15)$.

7. Sea X una variable aleatoria discreta con función de probabilidad p dada por:

$x_i:$	-2	-1	0	1	2	3
$p_i:$	0.1	0.2	0.15	0.25	0.15	0.15

   Sea Y la variable aleatoria definida por $Y = X^2$. Halle la función de probabilidad de Y. Calcule el valor de la función de distribución de X y de Y en los puntos $1, 3/4$ y $\pi - 3$.

8. Sea X una variable aleatoria con función de distribución F dada por

$$F(x) = \begin{cases} 0 & \text{para } x < 0, \\ \frac{1}{2} & \text{para } 0 \leq x < \frac{1}{4} \\ \frac{3}{4} & \text{para } \frac{1}{4} \leq x < \frac{3}{4} \\ 1 & \text{para } \frac{3}{4} \leq x \end{cases}$$

   Determine la función de probabilidad de X.

9. En un grupo grande de objetos una fracción θ son defectuosos. Si el número de extracciones (con reposición) necesarias para obtener el primer objeto defectuoso es una variable aleatoria X con función de probabilidad $P(X = j) = A(0.95)^{j-1}, \qquad j = 1, 2, \dots$

   a. Calcule el valor de A.

   b. ¿Cuál es la proporción θ de defectuosos?

   c. ¿Cuál es la probabilidad de que sea necesario examinar más de 20 objetos antes de obtener el primer defectuoso?

10. Una caja tiene 10 bolas numeradas del 1 al 10. Seleccionamos dos bolas al azar con reposición de la caja. Sea X el mayor de los dos números, calcule la función de probabilidad de X.

11. Resuelva el problema anterior para el caso de muestreo sin reposición.

12. Para determinar la efectividad de una nueva vacuna contra la gripe se vacunan 10 personas que son observadas por un período de un año. De ellas, 8 no tuvieron gripe durante este lapso. Si se sabe que la probabilidad de no tener gripe en un período de un año es 0.5 ¿cuál es la probabilidad de que 8 o más personas del grupo no hayan sufrido la enfermedad si la vacuna no es efectiva?

13. Considere un cierto defecto en el metabolismo que ocurre en aproximadamente 1 de cada 100 nacimientos. Si cuatro niños nacen en cierto hospital el mismo día, calcule la probabilidad de que

   a. ninguno tenga el defecto.

   b. no más de uno tenga el defecto.

14. El número de carros que cruzan un puente durante un período fijo de tiempo es una variable aleatoria con distribución de Poisson. Si la probabilidad de que ningún carro cruce el puente en este período es 1/4, halle una expresión para la probabilidad de que al menos dos carros lo crucen.

15. Lanzamos un dado hasta que la suma de los resultados sea mayor que 6 y sea X el número de lanzamientos necesarios para conseguir esto. Sea F la función de distribución de esta variable. Determine la función de probabilidad de X y el valor de F para $x = 1, 3$ y 7.

16. En una caja tenemos tres bolas numeradas $1, 2$ y 3. Sacamos tres bolas con reposición y llamamos X_i, $i = 1, 2, 3$ al resultado de la i-ésima extracción. Sea $\overline{X}$ el promedio de estas variables:

$$\overline{X} = (X_1 + X_2 + X_3)/3.$$

   Determine la función de probabilidad de $\overline{X}$. Calcule la probabilidad de que exactamente dos extracciones sean iguales a 3.

17. Un amigo te propone el siguiente juego: Lanzan una moneda hasta que salga sol. Si el número de lanzamientos es par, tú ganas, si es impar, pierdes. ¿Jugarías este juego?

18. Un vendedor de periódicos compra cada periódico por 1.50 y lo vende por 2.50. Los que no vende los regresa al distribuidor y recibe 1.25 por ellos. Supongamos que la distribución de la demanda D es

$$P(D = k) = \frac{e^{-10}10^k}{k!}$$

   Describa la variable aleatoria X que representa su ganancia diaria si compra 10 periódicos cada día.

19. Un llavero tiene cuatro llaves de apariencia similar pero sólo una de ellas abre la puerta de cierta oficina. Se selecciona al azar una llave y se prueba, si no funciona se selecciona al azar una de las restantes y se prueba de nuevo. Sea X el número de llaves que se prueban antes de encontrar la que abre la puerta. Halle su distribución de probabilidad.

20. Sea X una variable aleatoria con distribución de Poisson de parámetro λ. ¿cuál es la probabilidad de que X tome valor par (considerando a cero como par)?

21. Verifique que las siguientes funciones son densidades y obtenga la función de distribución correspondiente.

    a. $f(x) = \begin{cases} \cos x & \text{para } 0 < x < \pi/2 \\ 0 & \text{en otro caso} \end{cases}$ b. $f(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{para } |x| < 1 \\ 0 & \text{en otro caso} \end{cases}.$

22. Sea X una variable aleatoria con valores en $[0,1]$ y función de distribución $F(x) = x^2$. ¿Cuál es la densidad de X? Calcule las siguientes probabilidades:

    a. $P(\frac{1}{4} \le X \le \frac{3}{4})$, b. $P(X > 1/2)$, c. $P(X \le 3/4 | X > 1/2)$.

23. Sea X una variable aleatoria con distribución normal de parámetros $\mu = 12$, $\sigma^2 = 9$. Use `R` para calcular

    a. $P(X > 3)$. b. $P(|X - 12| < 4)$. c. $P(|X - 10| > 2)$.

24. Determine el valor que debe tomar la constante A en cada caso para que las siguientes funciones sean densidad de una función de distribución.

    a. $f(x) = Ae^{-\alpha|x-\theta|}$, $-\infty < x < \infty$, α y θ constantes.

    b. $f(x) = Ax^{\alpha+1}$, $x > x_0 > 0$, α constante.

    c. $f(x) = Ax(1-x)$, $0 \le x \le 1$.

    d. $f(x) = \frac{A}{1+x^2}$, $-\infty < x < \infty$.

25. Sea $f(x) = Cxe^{-x}$, $x > 0$ una densidad.

    a. Determine el valor de C. b. Calcule $P(X < 2)$. c. Calcule $P(2 < X < 3)$.

26. Halle la función de distribución F y su gráfica si la densidad es

    a. $f(x) = 1/2$, $0 \le x \le 2$. b. $f(x) = \begin{cases} x, & \text{si } 0 \le x \le 1 \\ 2 - x, & \text{si } 1 \le x \le 2. \end{cases}$

27. Si $f(x) = \frac{1}{2}e^{-x/2}$, $x > 0$, halle un número x_0 tal que $P(X > x_0) = 1/2$.

28. Sea X una variable aleatoria con distribución exponencial de parámetro $\lambda = 0.5$. Calcule

    a. $P(X > 1)$, b. $P(0.5 < X < 1.5)$, c. $P(X > 2 | X > 1)$.

29. La vida de una máquina, medida en horas, tiene densidad $f(x) = C/x^2$, $x > 100$.

    a. Calcule C. b. Halle la función de distribución. c. Calcule $P(X > 500)$.

30. La temperatura T de cierto objeto, medida en grados Fahrenheit, tiene una distribución normal con parámetros $\mu = 98.6$ y $\sigma^2 = 2$. La temperatura θ medida en grados centígrados está relacionada con T por la fórmula
$$\theta = \frac{5}{9}(T - 32).$$
    Obtenga la distribución de θ.

31. La magnitud v de la velocidad de una molécula con masa m en un gas de temperatura absoluta T es una variable aleatoria que, de acuerdo a la teoría cinética de los gases, posee una distribución de Maxwell con parámetro $\alpha = (2kT/m)^{1/2}$, donde k es la constante de Boltzman. La distribución de Maxwell de parámetro α tiene densidad
$$f(x) = \begin{cases} \frac{4}{\sqrt{\pi}} \frac{1}{\alpha^3} x^2 \exp\left(\frac{x^2}{k^2}\right) & \text{si } x > 0 \\ 0 & \text{si } x \le 0 \end{cases}$$
    ¿Cuál es la densidad de la energía cinética $E = mv^2/2$ de una molécula?

32. Halle la densidad de $Y = e^X$ donde X tiene distribución normal con parámetros μ y σ^2. (Se dice que la variable Y tiene distribución lognormal con parámetros μ y σ^2).

33. Una señal se codifica como una sucesión de ceros y unos para transmitirla digitalmente. Debido a imperfecciones en el canal de transmisión cualquiera de estos dígitos se recibe erroneamente (uno se recibe como cero o cero se recibe como uno) con probabilidad p.

    a. ¿Cuál es la probabilidad de tener al menos un error en una sucesión de n dígitos?

    b. Para reducir la probabilidad de error cada dígito se repite tres veces. cada dígito en el trío puede trasmitirse erroneamente con probabilidad p y tomamos como valor de cada trío al entero que se repita más veces: 001 lo interpretamos como 0.¿Cuál es la probabilidad de que cualquier dígito se reciba erroneamente? ¿Cuál es la probabilidad de tener al menos un error en una sucesión de n dígitos?

34. Dos jugadores A y B llevan a cabo una serie de juegos de manera independiente. La probabilidad de que A gane es p, la de B es q y la probabilidad de un empate es $1-p-q$. La serie termina una vez que alguno de los dos gana una partida. Este es un formato común para eliminatorias de 'muerte súbita'.

    a. ¿Cuál es la probabilidad de que A gane en el n-ésimo juego?

    b. ¿Cuál es la probabilidad de que A gane la serie?

    c. ¿Cuál es la probabilidad de que la serie dure n partidas?

35. Lanzamos un dado repetidamente hasta obtener un seis. Sea A_n el evento que ocurre si el primer seis aparece en el n-ésimo lanzamiento y B el evento que el número de lanzamientos requeridos sea par. Hallar $P(B)$ y $P(A_n|B)$.

36. Sea $X \sim b(n,p)$. Demuestre que $(P(X=k))^2 \geq P(X=k+1)P(X=k-1)$ para todo k.

37. Sea $X \sim b(n,p)$ y $Y \sim b(n,1-p)$, demuestre que $P(X=k) = P(Y=n-k)$. De una interpretación para este resultado.

38. Una fábrica recibe un lote de componentes y los prueba para verificar su funcionamiento. Por cada 100 componentes se prueban 10 y se acepta el lote si a lo sumo un componente falla. ¿Cuál es la probabilidad de aceptar un lote de tamaño 100 que contiene 7 defectuosos?

39. Lanzamos un dado hasta obtener el primer seis y sea T el lanzamiento en el cual esto ocurre. (a) ¿Cuál es la distribución de probabilidad de T? (b) Calcule $P(T > 6)$. (c) Calcule $P(T > 6|T > 3)$.

40. En una sucesión de ensayos de Bernoulli ¿Cuál es la probabilidad de que el primer éxito ocurra luego del quinto ensayo dado que no ha ocurrido en los dos primeros ensayos?

41. Sea X una variable con distribución de Poisson de parámetro $\lambda = 0.3$. Calcule $P(X=0), P(X=1)$ y $P(X > 1)$.

42. En promedio 1 persona en 1,000 tiene un tipo particular de sangre. (a) Hallar la probabilidad de que en una ciudad de 10,000 personas ninguna tenga este tipo de sangre. (b) ¿Cuántas personas hay que examinar para tener una probabilidad mayor a 1/2 de encontrar al menos una persona con este tipo de sangre.

43. Escriba un programa de computación que tenga como entradas n, p, j y si $X \sim b(n,p)$ calcule el valor de $P(X=j)$ y la aproximación de Poisson para este valor.

44. Considere la distribución de Poisson con parámetro λ. Demuestre que el resultado más probable es el entero k tal que $\lambda - 1 \leq k \leq \lambda$. ¿Bajo qué condiciones hay dos valores más probables?

45. Suponga que la probabilidad de haya un accidente importante en una planta eléctrica es de 0.005 en un año. Si un país tiene 100 plantas de este tipo, ¿cuál es la probabilidad de que haya al menos un accidente en un año?

46. Una linea aérea ha determinado que 4 % de los pasajeros que reservan pasajes en una ruta dada no se aparecen al momento del vuelo. En consecuencia han adoptado la política de vender 100 pasajes en un avión que sólo tiene 98 asientos. Si para un vuelo dado hay 100 asientos reservados, halle la probabilidad de que todos los pasajeros que se presentan tengan un asiento disponible.

47. Sea X una variable aleatoria con distribución uniforme en $1 \leq k \leq m$ ¿Cuánto vale $P(X = k|a \leq X \leq b)$? En particular halle $P(X > n + k|X > n)$.

48. Se capturan a miembros de una población de N animales y luego de marcarlos se liberan. Los animales luego son recapturados uno a uno hasta obtener $m \leq a$ animales marcados. Sea X el número de animales capturados hasta obtener m marcados, demuestre que la distribución de esta variable aleatoria está dada por
$$P(X = n) = \frac{a}{N}\binom{a-1}{m-1}\binom{N-a}{n-m}\binom{N-1}{n-1}^{-1}$$
Esta se conoce como la distribución hipergeométrica negativa.

49. Sea X una variable aleatoria discreta con distribución de Poisson de parámetro λ. Demuestre que la probabilidad de que X sea par es $e^{-\lambda}\cosh\lambda$. ¿Cuánto vale la probabilidad de que X sea impar?

50. Si X es una variable aleatoria discreta con distribución geométrica de parámetro p, demuestre que $P(X > k) = (1 - p)^k$.

51. Sea X una variable aleatoria con distribución geométrica de parámetro p. Sea M un entero positivo y definimos
$$Y = \begin{cases} X & \text{si } X < M, \\ M & \text{si } X \geq M, \end{cases}$$
es decir, $Y = \min(X, M)$. Calcule la función de probabilidad de Y.

52. Sea X una variable aleatoria con distribución geométrica de parámetro p. Calcule la función de probabilidad de X^2.

53. Una caja contiene k bolas numeradas del 1 a k. Seleccionamos una muestra aleatoria de tamaño n sin reposición. Sea Y el mayor de los números obtenidos y Z el menor. (a) Calcule la probabilidad $P(Y \leq y)$. (b) Calcule la probabilidad $P(Z \geq z)$.

54. Un grupo de m personas espera por el ascensor en un edificio de 10 pisos. Supongamos que cada una de estas personas escoge su piso de manera independiente de las otras y al azar, de modo que cada persona selecciona un piso con probabilidad $1/10$. Sea S_m el número de veces que el ascensor se detiene. Para estudiar esta variable aleatoria introducimos las variables R_i para $i = 1, \dots, 10$, donde R_i vale 1 si el ascensor se detiene en el piso i y 0 si no.

    a. Cada R_i tiene una distribución de Bernoulli. Demuestre que la probabilidad de éxito p vale $1 - \left(\frac{9}{10}\right)^m$.

    b. Tenemos que $S_m = R_1 + R_2 + \cdots + R_{10}$ ¿Es cierto que $S_m \sim b(10, p)$?

    c. Si $m = 1$ tenemos $P(S_1 = 1) = 1$. Halle las funciones de probabilidad para $m = 2$ y 3.

55. El fabricante de monedas del rey entrega las monedas que manufactura en cajas de 500 monedas y coloca una moneda falsa en cada caja. El rey tiene por costumbre revisar 1 moneda seleccionada al azar en cada caja y revisa 500 cajas cada vez. ¿Cuál es la probabilidad de que encuentre al menos una moneda falsa? ¿Cuál sería si revisa dos monedas de cada caja?

56. Sacamos una mano de trece cartas de un juego de 52. Calcule la probabilidad de que

    a. las pintas se distribuyan 4, 4, 3 y 2 (por ejemplo, cuatro diamantes, cuatro tréboles, tres corazones y dos picas).

    b. las pintas se distribuyan 5, 3, 3 y 2.

57. Tienes un juego de cuatro dados especiales. El primero tiene dos lados con 0 y cuatro lados con 4. El segundo tiene 3 en todos los lados. El tercero tiene cuatro lados iguales a 2 y 6 en los dos lados restantes. El cuarto tiene 1 en tres lados y 5 en los otros tres. Para el juego entre dos personas una escoge el dado que quiere y luego la otra hace lo mismo con los tres restantes. Ambos lanzan su dado y el que saque el mayor resultado gana. Demuestre que no importa cuál dado escoja la primera persona, la segunda siempre puede escoger de modo de tener probabilidad 2/3 de ganar.

58. Escriba un programa de computación para simular n valores de una variable de Bernoulli con $p = 1/3$. Corra el programa para $n = 100;\ 1000;\ 10000$ y en cada caso determine la proporción de los valores que son iguales a 1.

59. Escriba un programa de computación que tenga como entrada la función de probabilidad $p_i, i = 1, \dots n$ y como resultado produzca un valor de la variable con esta función de probabilidad y valores en $\{1, 2, \dots, n\}$.

60. Considere la distribución binomial negativa con parámetros p y k. Verifique la relación

$$P(X = j + 1) = \frac{j(1-p)}{j+1-k} P(X = j).$$

Use esta relación para dar un nuevo algoritmo para generar esta distribución.

61. Dé un método para generar una variable aleatoria tal que

$$P(X = i) = \frac{e^{-\lambda}\lambda^i/i!}{\sum_{i=0}^k e^{-\lambda}\lambda^i/i!}, \qquad i = 0, \dots, k.$$

62. Dé un método para generar una variable aleatoria con distribución triangular.

63. Dé un método para generar una variable aleatoria con función de densidad

$$f(x) = \frac{e^x}{e-1}, \qquad 0 \le x \le 1.$$

64. Dé un método para generar una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{x-2}{2}, & \text{si } 2 \le x \le 3, \\ \frac{2-x/3}{2}, & \text{si } 3 \le x \le 6. \end{cases}$$

65. Use el método de la transformada inversa para generar una variable aleatoria con función de distribución

$$F(x) = \frac{x^2 + x}{2}, \qquad 0 \le x \le 1.$$

Capítulo 5

Distribución Conjunta e Independencia

5.1. Distribución Conjunta de Dos Variables Aleatorias.

Sean X e Y dos variables aleatorias sobre un espacio de probabilidad común $(\Omega, \mathcal{A}, P)$. Llamaremos *función de distribución conjunta* , o simplemente *distribución conjunta*, de X e Y, a la función

$$F(x, y) = P(X \leq x,\ Y \leq y).$$

En algunas ocasiones usaremos $F_{X,Y}(x, y)$ en lugar de $F(x, y)$ para destacar que se trata de la distribución conjunta de X e Y.

La definición anterior indica que $F(x, y)$ es la probabilidad de que el punto (X, Y) pertenezca al cuadrante que queda "abajo y a la izquierda"del punto (x, y), incluyendo el borde, indicado en la figura 5.1 (a).

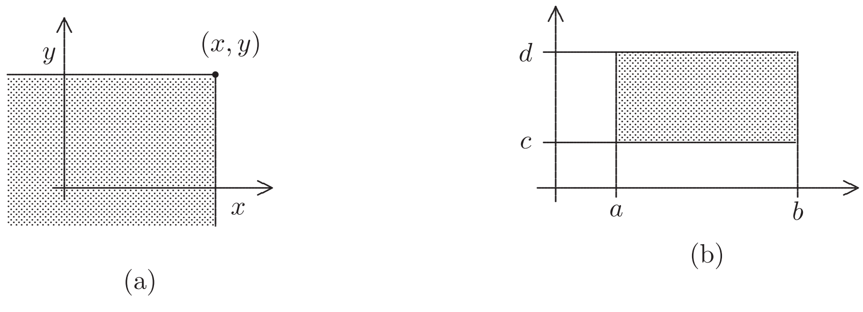


Figura 5.1

De esta manera

$$F(x, y) = P(\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\}).$$

A partir de la definición obtenemos (ver figura 5.1 (b))

$$\begin{aligned} P(a < X \le b,\ c < Y \le d) &= P(X \le b,\ Y \le d) - P(X \le b,\ Y \le c) \\ &\quad - P(X \le a,\ Y \le d) + P(X \le a,\ Y \le c) \\ &= F(b,d) - F(b,c) - F(a,d) + F(a,c) \end{aligned} \tag{5.1}$$

5.1.1. Propiedades

La distribución conjunta de dos variables tiene además las siguientes propiedades:

1. $F(x,y)$ es creciente en cualquiera de las dos variables. Por ejemplo, si $x < x'$ entonces

$$\{\omega : X(\omega) \le x\} \subset \{\omega : X(\omega) \le x'\}$$

y por lo tanto

$$\begin{aligned} F(x,y) &= P(\{\omega : X(\omega) \le x\} \cap \{\omega : Y(\omega) \le y\}) \\ &\le P(\{\omega : X(\omega) \le x'\} \cap \{\omega : Y(\omega) \le y\}) \\ &= F(x',y). \end{aligned}$$

2. $\lim\limits_{x\to-\infty} F(x,y) = \lim\limits_{y\to-\infty} F(x,y) = 0; \quad \lim\limits_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1,$

y como la función F es creciente en ambas variables se deduce que, para cualesquiera x, y,

$$0 \le F(x,y) \le 1.$$

3. $F(x,y)$ es continua por la derecha en cualquiera de las variables.

En contraste con el caso de funciones de distribución unidimensionales, para que una función $F(x,y)$ sea la distribución conjunta de un par de variables aleatorias X e Y, no es suficiente que tenga las tres propiedades que hemos considerado. Por ejemplo, la función

$$F(x,y) = \begin{cases} 0 & \text{si } x+y < 0 \\ 1 & \text{si } x+y \ge 0 \end{cases}$$

toma el valor 0 en los puntos que están debajo de la recta $y = -x$, y el valor 1 para los puntos sobre y por encima de la recta (ver figura 5.2).

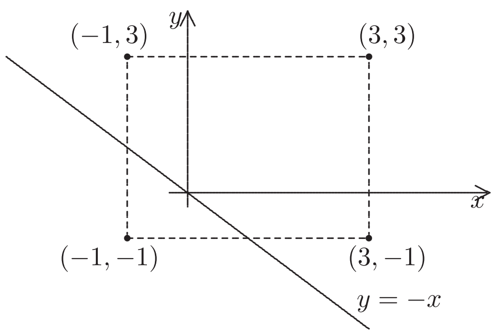


Figura 5.2

La función es creciente, continua por la derecha y satisface la propiedad 2. Sin embargo, si aplicamos la fórmula (5.1) para calcular la probabilidad de que el punto (X,Y) esté en el rectángulo de vértices $(3,3);\ (3,-1);\ (-1,3);\ (-1,-1)$, obtenemos

$$P(-1 < X \le 3, -1 < Y \le 3) = F(3,3) - F(3,-1) - F(-1,3) + F(-1,-1) = -1$$

lo cual es imposible ya que una probabilidad no puede ser negativa. Por lo tanto es necesario añadir la condición de que el segundo miembro de la relación (5.1) no sea negativo para ninguna colección de números $a < b,\ c < d$.

Teorema 5.1 *Una función $F(x,y)$ es la distribución conjunta de un par de variables aleatorias si y sólo si satisface las propiedades 1, 2 y 3 y además para cualesquiera $a < b,\ c < d$,*

$$F(b,d) - F(a,d) - F(b,c) + F(a,c) \ge 0.$$

A partir de la función de distribución conjunta $F_{X,Y}$ de dos variables aleatorias es posible obtener las funciones de distribución F_X y F_Y correspondientes a las variables X e Y. En efecto, para cualquier $x \in \mathbb{R}$ tenemos

$$\begin{aligned} F_X(x) &= P(X \le x) = P(X \le x,\ Y < \infty) \\ &= \lim_{y\to\infty} P(X \le x, Y \le y) \\ &= \lim_{y\to\infty} F(x,y) \end{aligned}$$

y de manera similar, para cualquier $y \in \mathbb{R}$

$$F_Y(y) = \lim_{x\to\infty} F(x,y).$$

Las funciones F_X y F_Y se conocen como las *funciones de distribución marginales* de X e Y, respectivamente.

5.2. Variables Aleatorias Independientes.

Definición 5.1 Se dice que las variables X e Y son *independientes* si cualesquiera sean los intervalos $(a,b]$ y $(c,d]$, se verifica que los eventos

$$\{X \in (a,b]\} \qquad \text{y} \qquad \{Y \in (c,d]\}$$

son independientes, es decir que

$$P(a < X \le b,\ c < Y \le d) = P(a < X \le b)P(c < Y \le d). \tag{5.2}$$

En términos menos precisos, de acuerdo a lo que hemos visto sobre independencia de eventos en el Capítulo 3, esta relación dice que saber que el valor de X está comprendido entre ciertos valores, no arroja información alguna sobre la probabilidad de que Y esté en algún intervalo dado.

Es fácil ver que la condición (5.2) es equivalente a la condición

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \text{para todo } x,y \in \mathbb{R}. \tag{5.3}$$

En efecto, si se cumple (5.2) basta poner $b = x,\ d = y$ y hacer tender $a \to -\infty,\ c \to -\infty$, para obtener (5.3). Recíprocamente, si se cumple (5.3), poniendo $F_{X,Y} = F$ tenemos

$$\begin{aligned} P(a < X \le b,\ c < Y \le d) &= F(b,d) - F(a,d) - F(b,c) + F(a,c) \\ &= F_X(b)F_Y(d) - F_X(a)F_Y(d) - F_X(b)F_Y(c) + F_X(a)F_Y(c) \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c)) \\ &= P(a < X \le b)P(c < Y \le d) \end{aligned}$$

o sea que (5.3) implica (5.2), cualesquiera sean $a,\ b,\ c$ y d.

Las relaciones (5.2) y (5.3) dicen que los eventos $\{X \in B_1\}$ y $\{Y \in B_2\}$ son independientes cuando B_1 y B_2 son intervalos semiabiertos, en el caso de (5.2), y semirectas cerradas a la derecha, en el caso de (5.3). Es posible probar, aunque no lo haremos en este texto, que (5.3), o equivalentemente (5.2), implica que los eventos $\{X \in B_1\}$ y $\{Y \in B_2\}$ son independientes para cualesquiera conjuntos de Borel B_1 y B_2.

5.3. Distribución Conjunta de Variables Aleatorias Discretas.

Si X e Y son variables aleatorias discretas, con funciones de probabilidad respectivas

$$\begin{aligned} P(X = x_i) &= p_i \qquad (i = 1, 2, \dots) \\ P(Y = y_j) &= q_j \qquad (j = 1, 2, \dots) \end{aligned}$$

donde

$$p_i \geq 0, \quad q_j \geq 0, \quad \sum_i p_i = \sum_j q_j = 1,$$

la función de distribución conjunta queda definida por los números

$$r_{ij} = P(X = x_i,\ Y = y_j) \quad (i = 1, 2, \dots; j = 1, 2, \dots).$$

Estas probabilidades deben satisfacer las condiciones

$$p_i = \sum_j r_{ij} \quad (i = 1, 2, \dots) \tag{5.4}$$

$$q_j = \sum_i r_{ij} \quad (j = 1, 2, \dots) \tag{5.5}$$

ya que, por ejemplo,

$$p_i = P(X = x_i) = P\left(\cup_j \{X = x_i,\ Y = y_j\}\right) = \sum_j P(X = x_i,\ Y = y_j) = \sum_j r_{ij}.$$

En este caso $\{r_{ij}\}$ se llama la *función de probabilidad conjunta* y $\{p_i\}$, $\{q_j\}$ son las *funciones marginales de probabilidad*. A partir de $\{r_{ij}\}$, $F_{X,Y}$ se determina mediante

$$F_{X,Y}(x, y) = \sum_{\substack{i: x_i \leq x \\ j: y_j \leq y}} r_{ij}$$

y las variables X e Y son independientes si y sólo si, para todo $i,\ j$ se tiene que

$$r_{ij} = P(X = i,\ Y = j) = P(X = i)P(Y = j) = p_i q_j.$$

Supongamos que con probabilidad 1 las variables aleatorias X e Y toman un número finito de valores n y m respectivamente. La situación queda descrita por el siguiente cuadro:

		Y				
		y_1	y_2	$\cdots\cdots$	y_m	
	x_1	r_{11}	r_{12}	$\cdots\cdots$	r_{1m}	p_1
	x_2	r_{21}	r_{22}	$\cdots\cdots$	r_{2m}	p_2
X	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	x_n	r_{n1}	r_{n2}	$\cdots\cdots$	r_{nm}	p_n
		q_1	q_2	$\cdots\cdots$	q_m	

Tabla 5.1

En la última columna aparecen $p_1,\ p_2, \dots, p_n$, que son las sumas respectivas de las filas (condición (5.4)) y en la última fila $q_1, q_2,\ \dots, q_m$, que son las sumas respectivas de las columnas (condición (5.5)).

Una situación de este tipo aparece en diversos problemas de aplicación. Supongamos que en el proceso de producción de un objeto nos interesan dos magnitudes, por ejemplo, el diámetro y la longitud de un cilindro, la densidad de un producto y la concentración de un componente del mismo, etc. Los valores obtenidos para estas dos magnitudes durante el proceso de producción fluctúan en virtud de diversas causas, algunas de ellas incontrolables, y otras cuyo control requeriría un costo elevado.

Sean α y β dichas magnitudes, procedemos a dividir el rango de variación de ambas en un número finito de secciones que numeramos ordenadamente, de 1 a n para la magnitud α y de 1 a m para la magnitud β. En la figura 5.4 hemos tomado $n = 5,\ m = 4$.

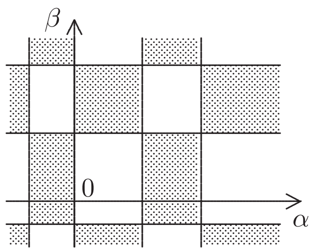


Figura 5.3

Llamemos X a la sección en la cual cae la magnitud α de un objeto e Y a la sección en la cual cae β. Para cada objeto tendremos entonces

$$r_{ij} = P(X = i,\ Y = j) \quad (i = 1, \dots, n;\ j = 1, \dots, m)$$

y tenemos definido un cuadro de doble entrada como el anterior con sus funciones de probabilidad (5.4) y (5.5).

Para estudiar la regulación de un proceso de producción, se extrae una muestra de N objetos producidos y se clasifica como se ha indicado:

		Y				
		1	2	$\cdots\cdots$	m	
	1	N_{11}	N_{12}	$\cdots\cdots$	N_{1m}	P_1
	2	N_{21}	N_{22}	$\cdots\cdots$	N_{2m}	P_2
X	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	n	N_{n1}	N_{n2}	$\cdots\cdots$	N_{nm}	P_n
		Q_1	Q_2	$\cdots\cdots$	Q_m	

Tabla 5.2

N_{ij} es el número de objetos de la muestra tales que el valor de α está en la i-ésima sección y el de β en la j-ésima. A partir de una muestra de este tipo es posible inferir resultados sobre la Tabla 5.1, cuyos elementos en general son desconocidos. Por ejemplo, es interesante saber si las magnitudes consideradas fluctúan independientemente, es decir si

$$P(X = i,\ Y = j) = P(X = i)P(Y = j)$$

o sea

$$r_{ij} = p_i\, q_j$$

para $i = 1, \dots, n;\ j = 1, \dots, m$. Es claro que si ambas magnitudes son independientes, se puede regular el valor de una sin afectar el de la otra, o más precisamente, su distribución de probabilidad, mientras que, por el contrario, cuando no hay independencia, se debe esperar que al regular el valor de una de las variables se modifique la distribución de probabilidad de la otra.

5.4. La Distribución Multinomial.

Consideremos el siguiente ejemplo: se lanza un dado n veces y se cuenta el número X_1 de veces que se obtiene 1 y el número X_2 de veces que se obtiene 2. Supongamos que el dado es simétrico (es decir, que cada cara tiene probabilidad 1/6 en cada lanzamiento) y que los lanzamientos son independientes, entonces la distribución conjunta de X_1 y X_2 está dada por

$$r_{ij} = P(X_1 = i,\ X_2 = j) \quad \text{con} \quad i + j \leq n$$

que se calcula mediante

$$\begin{aligned} r_{ij} &= \frac{n!}{i!\ j!\ (n-i-j)!} \left(\frac{1}{6}\right)^i \left(\frac{1}{6}\right)^j \left(\frac{4}{6}\right)^{n-i-j} \\ &= \frac{n!}{i!\ j!\ (n-i-j)!} \frac{4^{n-i-j}}{6^n}. \end{aligned} \tag{5.6}$$

Una manera de obtener esta expresión es la siguiente: el número de resultados ordenados posibles en los n lanzamientos es 6^n, ya que en cada lanzamiento tenemos 6 resultados posibles, y estos 6^n resultados son igualmente probables. Por lo tanto, basta calcular el número de veces que obtenemos i *unos*, j *dos* y $n-i-j$ caras que no son ni *uno* ni *dos*. Para ello procedemos así: elegimos los i lugares en que ponemos los *unos*, lo cual podemos hacer de $\binom{n}{i}$ formas; entre los $n-i$ lugares que nos quedan, elegimos j lugares donde colocar los *dos*, lo cual podemos hacer de $\binom{n-i}{j}$ maneras, y en los $n-i-j$ lugares que nos quedan, colocamos de todas las maneras posibles caras que no son ni 1 ni 2, lo cual podemos hacer de 4^{n-i-j} formas. En total tendremos

$$\binom{n}{i}\binom{n-i}{j} 4^{n-i-j} = \frac{n!}{i!\ (n-i)!} \frac{(n-i)!}{j!\ (n-i-j)!} 4^{n-i-j}$$

casos favorables. Dividiendo por el número de casos posibles, 6^n, obtenemos (5.6).

Esta distribución es un caso particular de la distribución multinomial. En cuanto a las distribuciones marginales tenemos para $i = 0, 1, \dots, n$,

$$\begin{aligned} p_i = P(X_n = i) &= \sum_{j=0}^{n-i} r_{ij} \\ &= \sum_{j=0}^{n-i} \frac{n!}{i!\ j!\ (n-i-j)!} \frac{4^{n-i-j}}{6^n} \\ &= \frac{n!}{i!(n-i)!} \sum_{j=0}^{n-i} \frac{(n-i)!}{j!(n-i-j)!} \frac{4^{n-i-j}}{6^n} \\ &= \binom{n}{i} \frac{1}{6^n}(1+4)^{n-i} = \binom{n}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{n-i}, \end{aligned}$$

la función de probabilidad binomial $b(n, 1/6)$, que sabemos que corresponde a la variable aleatoria X_1.

Una situación análoga a la anterior se plantea cuando consideramos el siguiente caso de muestreo. Tenemos una población de N individuos clasificados en 3 grupos de tamaños respectivos $N_1,\ N_2,\ N_3$. Los grupos son disjuntos y $N_1 + N_2 + N_3 = N$.

Si tomamos una muestra de tamaño n al azar y con reposición, y denotamos por $X_1,\ X_2,\ X_3$ respectivamente el número de elementos de la muestra que están en la clase 1, 2 ó 3, tenemos

$$P(X_1 = i, X_2 = j, X_3 = n - i - j) = \frac{n!}{i!\ j!\ (n-i-j)!} p_1^i\ p_2^j\ p_3^{n-i-j}$$

para $i + j \leq n$, donde $p_k = N_k/N,\ (k = 1, 2, 3)$ indica la fracción de la población que está en la i-ésima clase. El cálculo es enteramente similar al del ejemplo anterior.

Si en lugar de tres clases tenemos m, con una notación análoga se tiene

$$P(X_1 = i_1, X_2 = i_2, \dots, X_m = i_m) = \frac{n!}{i_1!\ i_2! \cdots i_m!} p_1^{i_1} p_2^{i_2} \cdots p_m^{i_m}$$

con $i_1 + i_2 + \cdots + i_m = n$. La distribución correspondiente se conoce como la *distribución multinomial.*

Si el muestreo se hace sin reposición, en lugar de la probabilidad anterior tenemos

$$P(X_1 = i_1, X_2 = i_2, \dots, X_m = i_m) = \frac{\binom{N_1}{i_1}\binom{N_2}{i_2} \dots \binom{N_m}{i_m}}{\binom{N}{n}}$$

donde $N_1 + N_2 + \cdots + N_m = N,\ n_1 + n_2 + \cdots + n_m = n$, para la función de probabilidad conjunta de $X_1,\ X_2, \dots, X_m$. En este caso, las distribuciones marginales son hipergeométricas en lugar de binomiales.

5.5. Funciones de Variables Aleatorias Independientes

Observamos que si X e Y son variables aleatorias independientes, y g y h son funciones tales que la preimagen de un intervalo es un conjunto de Borel, entonces las variables aleatorias

$$X_1 = g(X) \qquad Y_1 = h(Y)$$

también son independientes. Esto se apoya en el hecho de que siendo $I,\ J$ intervalos

$$\{X_1 \in I\} = \{g(X) \in I\} = \{X \in g^{-1}(I)\} = X^{-1}(g^{-1}(I))$$

y del mismo modo

$$\{Y_1 \in J\} = Y^{-1}(h^{-1}(J)).$$

Como $g^{-1}(I),\ h^{-1}(J)$ son conjuntos de Borel y las variables X e Y son independientes, los eventos $\{X_1 \in I\},\ \{Y_1 \in J\}$ resultan ser independientes.

Esto es cierto si, por ejemplo, las funciones g y h son continuas o monótonas o tienen una cantidad numerable de discontinuidades.

5.6. Suma de Variables Aleatorias Independientes.

Sean X e Y dos variables aleatorias discretas con función de probabilidad conjunta

$$r_{ij} = P(X = x_i, Y = y_j) \qquad (i = 1, 2, \dots; j = 1, 2, \dots).$$

Si queremos calcular la función de probabilidad de la variable aleatoria discreta

$$S = X + Y$$

es inmediato que la misma se obtiene mediante

$$P(S = s_k) = \sum_{x_i + y_j = s_k} r_{ij} \tag{5.7}$$

donde la suma se extiende a todas las parejas de índices (i, j) tales que $x_i + y_j = s_k$. En particular, si X e Y son independientes con funciones de probabilidad respectivas

$$P(X = x_i) = p_i; \qquad P(Y = y_j) = q_j,$$

la fórmula (5.7) se reduce a

$$P(S = s_k) = \sum_{x_i + y_j = s_k} p_i \, q_j.$$

Ejemplo.

Consideremos, por ejemplo, el caso de la distribución binomial, ejemplificada mediante el modelo más sencillo de control de calidad, consistente en sucesivas extracciones independientes con reposición y con probabilidad p de extraer un objeto defectuoso en cada ocasión.

Ponemos $X_i = 0$ ó 1 respectivamente, si extraemos un objeto bueno o defectuoso en la i-ésima extracción, y denotamos por d_n el número de defectuosos al cabo de n extracciones. Entonces

$$d_n = d_{n-1} + X_n \tag{5.8}$$

donde las variables aleatorias discretas d_{n-1} y X_n son independientes. La función de probabilidad de d_n se obtiene usando (5.7) y (5.8):

$$p_{n,k} = P(d_n = k) = \sum_{j=0}^{k} P(d_{n-1} = j) P(X_n = k - j).$$

En la suma intervienen a lo sumo dos términos significativos, ya que

$$P(X_n = 1) = p, \quad P(X_n = 0) = 1 - p \quad \text{y} \quad P(X_n \neq 0 \text{ ó } 1) = 0.$$

Se obtiene

$$\begin{aligned} &p_{1,1} = p \qquad p_{1,0} = 1 - p \\ &p_{n,n} = p \; p_{n-1,n-1} \qquad p_{n,0} = (1 - p) p_{n-1,0} \\ &p_{n,k} = p \; p_{n-1,k-1} + (1 - p) p_{n-1,k} \quad (k = 1, 2, \dots, n - 1) \end{aligned}$$

Utilizando estas igualdades y procediendo por inducción se obtiene la fórmula conocida para la función de probabilidad binomial

$$p_{n,k} = \binom{n}{k} p^k (1 - p)^{n-k}, \quad (k = 0, 1, \dots, n).$$

◀

Para simplificar la notación vamos a suponer que X, Y son variables con valores i, j, donde $i, j \in \mathbb{Z}$. Entonces la ecuación (5.7) es

$$P(S = k) = \sum_{i+j=k} r_{i,j} = \sum_{i} r_{i,k-i}.$$

En particular, si X, Y son independientes,

$$P(S = k) = \sum_{i} p_i q_{k-i},$$

y si las variables X, Y son no-negativas tenemos

$$P(S = k) = \sum_{i=0}^{k} p_i q_{k-i}.$$

Ejemplos.

1. **Suma Binomial.** Si X tiene distribución binomial $b(n, p)$ e Y tiene distribución binomial $b(m, p)$, entonces la distribución de $S = X + Y$ es

$$\begin{aligned} P(S = k) &= \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i} \\ &= p^k q^{m+n-k} \sum_{i=0}^{k} \binom{n}{i} \binom{m}{k-i} \\ &= \binom{m+n}{k} p^k q^{m+n-k} \end{aligned}$$

de modo que $S = X + Y$ tiene distribución binomial $b(m + n, p)$.

2. **Suma Poisson.** Si X, Y son variables de Poisson independientes, con parámetros λ y μ respectivamente, entonces

$$\begin{aligned} P(S = k) &= \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k. \end{aligned}$$

de modo que S tiene distribución de Poisson con parámetro $\lambda + \mu$.

3. **Suma Geométrica.** Sean X, Y variables independientes con distribución geométrica de parámero p. Entonces para $k \geq 2$,

$$\begin{aligned} P(S = k) &= \sum_{i=1}^{k-1} q^{i-1} p q^{k-i-1} p \\ &= (k-1) q^{k-2} p^2, \end{aligned}$$

que es una distribución binomial negativa.

5.7. Sucesiones de Variables de Bernoulli

Vamos a considerar en esta sección sucesiones (X_n) de variables aleatorias de Bernoulli independientes, todas con la misma probabilidad de éxito p. Estas variables satisfacen la propiedad de que para cualquier colección de índices $i_1, i_2, \dots, i_k$ y cualquier vector $\mathbf{j} = (j_1, j_2, \dots, j_k)$ de componentes iguales a 0 ó 1, se tiene

$$P(X_{i_1} = j_1, X_{i_2} = j_2, \dots, X_{i_k} = j_k) = P(X_{i_1} = j_1) P(X_{i_2} = j_2) \cdots P(X_{i_k} = j_k).$$

Como todas las variables tienen la misma probabilidad de éxito, si en el vector $\mathbf{j}$ hay m componente iguales a 1 y $k - m$ iguales a 0, esta probabilidad es igual a

$$p^m (1 - p)^{k-m}.$$

Hemos estudiado algunas propiedades de estas sucesiones anteriormente. Por ejemplo, la distribución binomial corresponde al número de éxitos en una colección de n variables de este tipo: si $S_n = \sum_1^n X_i$ entonces S_n tiene distribución binomial $b(n, p)$. La distribución geométrica representa el tiempo de espera hasta el primer éxito mientras que la distribución binomial negativa representa el tiempo de espera hasta conseguir k éxitos. Veamos otras propiedades de estas sucesiones.

5.7.1. Intervalos entre Exitos

Llamemos T_k a la longitud del intervalo entre el $(k-1)$-ésimo y el k-ésimo éxito en la sucesión de Bernoulli. Vamos a incluir siempre al último éxito, pero no al anterior en nuestra contabilidad. Como ejemplo mostramos los valores de los T_k iniciales para un resultado particular:

$$\underbrace{0001}_{T_1}\underbrace{001}_{T_2}\underbrace{\ 1\ }_{T_3}\underbrace{00001}_{T_4}\underset{\dots}{\dots}$$

Ya conocemos la distribución de T_1, porque esta variable representa el tiempo de espera hasta el primer éxito, y tiene por lo tanto una distribución geométrica. Ahora bien, una vez que ocurre el primer éxito, la independencia implica que lo que ocurre después es como si iniciáramos el proceso de nuevo, y por lo tanto T_2 tiene la misma distribución que T_1 y además son variables independientes. Veamos esto en detalle.

Para i y j en $\mathbb{N}$ el evento $\{T_1 = i, T_2 = j\}$ ocurre si hay $(i-1)$ fracasos inicialmente, luego un éxito, luego $(j-1)$ fracasos más y finalmente un segundo éxito:

$$\underbrace{00\dots0}_{(i-1)}1\underbrace{00\dots0}_{(j-1)}1$$

y la probabilidad de que esto ocurra es

$$q^{i-1}pq^{j-1}p = P(T_1 = i)P(T_2 = j),$$

de modo que estas variables son independientes. De manera similar se puede demostrar que cualquier colección de variables $T_{i_1}, \dots, T_{i_k}$ son independientes.

Existe una relación fundamental entre estas variables T_k y las variables S_n. Supongamos que para cierta sucesión de ensayos de Bernoulli el k-ésimo éxito ha ocurrido antes de o en el n-ésimo experimento: $T_k \leq n$. Para que esto suceda, el número de éxitos en los primeros n ensayos debe ser mayor o igual a k: $S_n \geq k$. Recíprocamente, si $S_n \geq k$ entonces $T_k \leq n$.

Por otro lado, si $T_k = n$, el k-ésimo éxito ocurre en el experimento n. Esto quiere decir que en los primeros $n-1$ ensayos ocurrieron $k-1$ éxitos y el n-ésimo ensayo resultó en éxito: $S_{n-1} = k-1$ y $X_n = 1$. La relación recíproca también es cierta y es fácil de verificar. Resumimos estos resultados en la siguiente proposición.

Proposición 5.1 *Si $k \leq n$ tenemos*

a. $T_k \leq n$ si y sólo si $S_n \geq k$.

b. $T_k = n$ si y sólo si $S_{n-1} = k-1,\ X_n = 1$.

5.8. Paseo al Azar

Sea (X_n) una sucesión de variables de Bernoulli con probabilidad de éxito p. Definimos las variables $Y_n = 2X_n - 1$, que son independientes y toman valores 1 y -1 con probabilidades p y $1-p$. Si consideramos una serie de juegos en los que tenemos probabilidad p de ganar y $q = 1-p$ de perder y apostamos una unidad en cada uno, podemos representar el resultados de cada juego por una sucesión de variables de este tipo.

En esta situación la suma de las variables $Y_1 + Y_2 + \cdots + Y_n = S_n$ representa la evolución de la fortuna del jugador al cabo de n juegos, es decir, es la cantidad que el jugador gana o pierde luego de haber jugado n juegos. Si el capital inicial del jugador es A entonces su fortuna al cabo de n juegos es $S'_n = A + S_n$. La sucesión $(S_n)_{n\geq1}$ (o $(S'_n)_{n\geq1}$ si la fortuna inicial es distinta de 0) se conoce como el paseo al azar simple y en el caso $p = q = 1/2$ es el paseo simple simétrico. Podemos representarlo gráficamente como en la siguiente figura:

Es sencillo calcular la distribución de probabilidad de S_n.

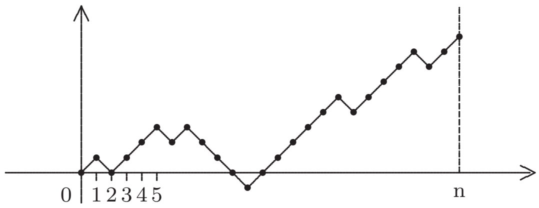


Figura 5.4

Proposición 5.2 *Para $0 \leq k \leq n$, $P(S_n = k) = 0$ si k y n no tienen la misma paridad. En caso contrario tenemos*

$$P(S_n = k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}.$$

Demostración. Sea s el número de $+1's$ en los n juegos y t el número de $-1's$. Claramente $s+t=n$. Para que $S_n = k$ es necesario que $s - t = k$. Resolviendo este sistema de ecuaciones tenemos

$$s = \frac{n+k}{2}, \qquad t = \frac{n-k}{2}$$

que son enteros si y sólo si n y k tienen la misma paridad. Si esto es cierto, tenemos $s = (n+k)/2$ éxitos en n ensayos de Bernoulli, que corresponde a una distribución binomial:

$$P(S_n = k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}.$$

■

En el capítulo 3 estudiamos un problema asociado al paseo al azar: el problema de la ruina del jugador. En el próximo capítulo analizaremos otro problema que es el de la duración promedio del juego.

5.9. Muestreo Secuencial

Ya hemos visto en capítulos anteriores el muestreo sin reposición, en el cual tenemos una población de n individuos que muestreamos sucesivamente pero sin reponer a la población los individuos que vamos observando. En esta situación los sucesivos resultados que obtenemos no son independientes entre sí, porque con cada selección el espacio muestral cambia. Para poblaciones grandes el efecto es menor pero con poblaciones pequeñas puede ser considerable.

Nos queremos enfocar en la dependencia entre los sucesivos resultados. Supongamos que tenemos una caja con r bola rojas y b bolas blancas con $r + b = N$. Seleccionamos una bola al azar de la caja, vemos su color y la colocamos aparte. Repetimos este procedimiento n veces y para $1 \leq i \leq n$ definimos las variables

$$X_i = \begin{cases} 1 & \text{si la } i\text{ésima bola es roja,} \\ 0 & \text{si la } i\text{ésima bola es blanca.} \end{cases}$$

Consideremos inicialmente el resultado de la primera extracción. Es inmediato que

$$p_0 = P(X_1 = 0) = \frac{b}{N}, \qquad p_1 = P(X_1 = 1) = \frac{r}{N}.$$

Veamos ahora la distribución de la segunda variable, usando la Ley de la Probabilidad Total,

$$\begin{aligned}
P(X_2 = 0) &= P(X_2 = 0|X_1 = 0)P(X_1 = 0) + P(X_2 = 0|X_1 = 1)P(X_1 = 1) \\
&= \frac{b-1}{N-1} \times \frac{b}{N} + \frac{b}{N-1} \times \frac{r}{N} \\
&= \frac{(b-1)b + br}{N(N-1)} \\
&= \frac{b(N-1)}{N(N-1)} = \frac{b}{N}
\end{aligned}$$

y de manera similar tenemos que

$$P(X_2 = 1) = \frac{r}{N},$$

de modo que las variables X_1 y X_2 tienen la misma distribución. Este hecho resulta inesperado a primera vista, pues la situación cambia una vez que hacemos la primera extracción, así que esperaríamos que X_2 tuviese una distribución distinta a la de X_1. Sin embargo hay que observar que no estamos tomando en cuenta el resultado de la primera extracción, es decir, estamos calculando la distribución de X_2 sin conocer el resultado de X_1, y en estas condiciones, a falta de mayor información, el resultado nos dice que la segunda extracción tiene la misma distribución de probabilidad que bajo las condiciones iniciales.

Es posible demostrar, aunque es un poco más trabajoso, que esto es cierto para cualquiera de las variables X_i, $1 \le i \le n$: Todas tienen la misma distribución.

Para ver que estas variables no son independientes, calculemos la distribución conjunta de X_1 y X_2.

$$P(X_1 = 0, X_2 = 0) = P(X_2 = 0|X_1 = 0)P(X_1 = 0) = \frac{b(b-1)}{N(N-1)}.$$

De manera similar tenemos

$$P(X_1 = 0, X_2 = 1) = \frac{br}{N(N-1)}, \quad P(X_1 = 1, X_2 = 0) = \frac{br}{N(N-1)}, \quad P(X_1 = 1, X_2 = 1) = \frac{r(r-1)}{N(N-1)},$$

y vemos que la función de probabilidad conjunta no es el producto de las funciones de distribución individuales de modo que las variables no son independientes.

Usando de nuevo la ley de la probabilidad total no es difícil ver que si $i_1, i_2, \dots, i_n$ toman valores 0 ó 1, y hay j unos (y $n-j$ ceros) con $0 \le j \le r$, $0 \le n-j \le b$, entonces

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \frac{\binom{N-n}{r-j}}{\binom{N}{n}}$$

5.10. Ejemplos.

1. Se extrae una muestra de tamaño dos con reposición de una bolsa que contiene dos bolas blancas, una negra y dos rojas. Definimos las variables aleatorias X_1 y X_2 de la siguiente manera: para $k = 1, 2$, $X_k = 1$ ó 0 según si la bola obtenida en la k-ésima extracción es blanca o no lo es.

   a. Describa la función de probabilidad conjunta de estas variables.

   b. Describa las funciones marginales de probabilidad.

   c. ¿Son independientes estas variables aleatorias?

   d. ¿Qué sucede si el muestreo se realiza sin reposición?

► a. Para el caso de muestreo con reposición tenemos

$$r_{00} = \frac{3}{5}\frac{3}{5} = \frac{9}{25}; \qquad r_{01} = \frac{3}{5}\frac{2}{5} = \frac{6}{25};$$
$$r_{10} = \frac{2}{5}\frac{3}{5} = \frac{6}{25}; \qquad r_{11} = \frac{2}{5}\frac{2}{5} = \frac{4}{25}.$$

b. Las funciones marginales de probabilidad son

$$p_0 = P(X_1 = 0) = \frac{3}{5}; \qquad p_1 = P(X_1 = 1) = \frac{2}{5};$$
$$q_0 = P(X_2 = 0) = \frac{3}{5}; \qquad q_1 = P(X_2 = 1) = \frac{2}{5}.$$

c. Es fácil verificar que para todo i, j se tiene

$$r_{ij} = p_i \, q_j$$

de modo que las variables X_1 y X_2 son independientes.

d. Si el muestreo es sin reposición, la función de probabilidad conjunta es

$$r_{00} = \frac{3}{5}\frac{2}{4} = \frac{6}{20}; \qquad r_{01} = \frac{3}{5}\frac{2}{4} = \frac{6}{20};$$
$$r_{10} = \frac{2}{5}\frac{3}{4} = \frac{6}{20}; \qquad r_{11} = \frac{2}{5}\frac{1}{4} = \frac{2}{20}.$$

Las funciones marginales de probabilidad son

$$p_0 = P(X_1 = 0) = \frac{3}{5}; \qquad p_1 = P(X_1 = 1) = \frac{2}{5};$$
$$q_0 = P(X_2 = 0) = \frac{3}{5}; \qquad q_1 = P(X_2 = 1) = \frac{2}{5}.$$

Las variables no son independientes en este caso ya que, por ejemplo, $r_{00} \neq p_0 \, q_0$. ◄

2. Supongamos que las variables (X, Y) pueden tomar los valores $(0, 0)$; $(1, 1)$; $(-1, 1)$; $(1, -1)$ y $(-1, -1)$ cada uno con probabilidad $1/5$. Determine si estas variables son independientes.

► La función de probabilidad conjunta está resumida en la siguiente tabla

		X		
		-1	0	1
	-1	$1/5$	0	$1/5$
Y	0	0	$1/5$	0
	1	$1/5$	0	$1/5$

Las funciones marginales de probabilidad son

$$P(X = -1) = p_{-1} = r_{-1,-1} + r_{-1,0} + r_{-1,1} = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

y similarmente

$$P(X = 0) = p_0 = \frac{1}{5}, \qquad P(X = 1) = p_1 = \frac{2}{5}.$$

Se verifica fácilmente que Y tiene la misma función de probabilidad. Ahora bien,

$$P(X = 0,\ Y = -1) = 0 \neq \frac{1}{5} \times \frac{2}{5} = P(X = 0)P(Y = -1)$$

y las variables no son independientes. ◄

3. Si X e Y son variables aleatorias independientes con la misma función de distribución F, ¿Cuál es la función de distribución $G(z)$ de la variable aleatoria $Z = \text{máx}(X, Y)$?

▶

$$\begin{aligned} G(z) &= P(Z \le z) = P(\text{máx}(X,Y) \le z) \\ &= P(X \le z,\ Y \le z) = P(X \le z)P(Y \le z) \\ &= F^2(z). \end{aligned}$$

◀

4. Sean X e Y variables aleatorias independientes con función de probabilidad uniforme en $\{1, 2, \dots, N\}$ (es decir, $P(X = i) = P(Y = i) = 1/N,\ i = 1, \dots, N$). Calcule la función de probabilidad de $X + Y$.

▶ Es evidente que para $j < 2$ ó $j > 2N$ tenemos

$$P(X + Y = j) = 0.$$

Si $2 \le j \le N$

$$P(X + Y = j) = \sum_{i=1}^{j-1} P(X = i,\ Y = j - i) = \sum_{i=1}^{j-1} \frac{1}{N^2} = \frac{j-1}{N^2}$$

mientras que para $N + 1 \le j \le 2N$, definiendo $i = j - N$ tenemos

$$\begin{aligned} P(X + Y = j) = P(X + Y = N + i) &= \sum_{k=i}^{N} P(X = k,\ Y = N + i - k) \\ = \frac{N - i + 1}{N^2} = \frac{2N - j + 1}{N^2}. \end{aligned}$$

Resumiendo

$$P(X + Y = j) = \begin{cases} (j-1)/N^2 & \text{para } 2 \le j \le N, \\ (2N - j + 1)/N^2 & \text{para } N + 1 \le j \le 2N, \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

◀

5. Si realizamos n ensayos de Bernoulli y U y V son el número de éxitos y fracasos respectivamente, sabemos que estas variables son dependientes. Supongamos que n es una variable de Poisson con parámetro λ y realizamos N ensayos de Bernoulli, de modo que el número de ensayos es aleatorio. Ahora obtenemos X éxitos e Y fracasos. Demuestre que X e Y son independientes.

▶ Tenemos

$$\begin{aligned} P(X = i, Y = j) &= P(X = i, Y = j, N = i + j) \\ &= P(X = i, Y = j | N = i + j)P(N = i + j) \\ &= \frac{(i+j)!}{i!j!} p^i q^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ &= \frac{(\lambda p)^i}{i!} e^{-\lambda p} \frac{(\lambda q)^j}{j!} e^{-\lambda q} \end{aligned}$$

Esta función de probabilidad se factoriza para todos los valores de X e Y, de modo que estas variables son independientes y tienen distribución de Poisson con parámetros λp y λq, respectivamente.

◀

Ejercicios

1. Dada la función de probabilidad conjunta definida por
$$p_{ij} = C(i+j) \tag{5.9}$$
en los puntos $(1,1)$; $(2,1)$; $(2,1)$ y $(3,1)$, donde C es una constante, determine en valor de C y obtenga la función de probabilidad marginal correspondiente a la primera variable.

2. Sean X, Y variables aleatorias con valores en $\{1, 2, \dots, n\}$ y con función de probabilidad conjunta dada por (5.9). Halle el valor de C y las distribuciones marginales.

3. La función $p_{i,j}$ está dada por $p_{i,j} = C\alpha^i\beta^j$ para $i, j \in \mathbb{N}$ y $0 < \alpha, \beta < 1$. Halle el valor de C para que $p_{i,j}$ sea una función de probabilidad.

4. ¿Es $p_{i,j} = (0.5)^{i+j}$ para $i, j \in \{0, 1, 2, \dots\}$ una función de probabilidad? Si la respuesta es positiva, calcule $P\{1 \leq i \leq 3, j \geq 2\}$.

5. La función $p_{i,j}$ está dada por
$$p_{i,j} = C\binom{10}{j}\Big(\frac{1}{2}\Big)^{10}.$$
Halle el valor de C y determine las funciones de probabilidad marginales.

6. Sea X una variable aleatoria de Bernoulli con probabilidad de éxito p y sean $Y = 1 - X$, $Z = XY$. Halle la distribución conjunta de (X, Y); (X, Z) y (Y, Z) ¿Es independiente alguna de estas parejas?

7. Sean X, Y variables aleatorias con función de probabilidad conjunta
$$r_{ij} = C\frac{i+j}{i!j!}\theta^{i+j},$$
para $i, j \geq 0$ donde $\theta > 0$ es una constante. Halle C, las distribuciones marginales de X e Y y $P(X+Y=k)$. ¿Son independientes estas variables aleatorias?

8. Sea $\Omega = \{\omega_1, \omega_2, \omega_3\}$ y P la distribución uniforme en Ω (todos los puntos tienen igual probabilidad). Definimos las variables X, Y y Z de la siguiente manera: $X(\omega_1) = Y(\omega_2) = Z(\omega_3) = 1$, $X(\omega_2) = Y(\omega_3) = Z(\omega_1) = 2$, $X(\omega_3) = Y(\omega_1) = Z(\omega_2) = 3$. Demuestre que estas tres variables tienen la misma función de probabilidad. Halle las funciones de probabilidad de $X+Y$, $Y+Z$ y $X+Z$.

9. Considere un grupo de cartas que consiste de J, Q, K y A de las cuatro pintas. Se extraen dos cartas del grupo sin reposición y llamamos X e Y al número de diamantes y corazones obtenidos, respectivamente. Obtenga la función de probabilidad conjunta y la función marginal correspondiente a X.

10. Una caja tiene 6 bolas numeradas del 1 al 6. Las bolas numeradas 1 y 2 son rojas mientras que las otras son blancas. Extraemos dos bolas al azar de la caja y sean X, Y las variables aleatorias que representan el número de bolas rojas y el número de bolas pares en la muestra, respectivamente. Halle la distribuciones de X e Y y su distribución conjunta. Determine si estas variables son independientes.

11. Una caja contiene ocho bolas numeradas del 1 al 8. Las primeras cuatro son rojas y las otras blancas. Seleccionamos dos bolas al azar de la caja y definimos las siguientes variables: X es el número de bolas blancas en la muestra, Y es el número de bolas pares y Z el número de bolas en la muestra cuyo número es menor que 6. Halle la distribución conjunta de las variables (X, Y); (X, Z); (Y, Z) y (X, Y, Z). Estudie la independencia de estas variables.

12. Sea X el número de aces en una mano de poker e Y el número de reinas. Halle la función de probabilidad conjunta para estas variables y sus funciones de probabilidad marginales. Determine si son independientes.

13. Lanzamos una moneda tres veces. Sea X el número de águilas en los dos primeros lanzamientos e Y el número de águilas en el tercer lanzamiento. Halle la distribución conjunta de X, Y, la distribución de $Z = X + Y$ y la de $W = X - Y$.

14. Sacamos cinco cartas con reposición de un paquete de 52 cartas. ¿Cuál es la probabilidad de obtener dos diamantes y un trébol? ¿Cuál es la probabilidad de obtener dos ases, dos reinas y un 10?

15. Un componente electrónico puede fallar de cuatro maneras distintas, y las probabilidades respectivas son $p_1 = 0.2; p_2 = 0.15; p_3 = 0.25; p_4 = 0.4$. Si examinamos 10 componentes, ¿Cuál es la probabilidad de que haya exactamente tres fallas de tipo 1, dos de tipo 2, dos de tipo 3 y y tres de tipo 4?

16. Las probabilidades de llenar una declaración de impuestos correctamente, con un error que favorezca al fisco, con un error que favorezca al declarante o con ambos tipos de errores son, respectivamente, 0.6; 0.2; 0.15 y 0.05. Calcule la probabilidad de que entre 10 declaraciones de impuestos 5 estén correctas, 3 tengan errores a favor del declarante, 1 tenga un error a favor del fisco y una tenga ambos tipos de errores.

17. A través de un estudio se ha determinado que al llegar a cierto cruce, 50 % de los vehículos continua de frente, 30 % da vuelta a la derecha y el resto da vuelta a la izquierda. Calcule la probabilidad de que de los siguientes cinco automóviles, uno de vuelta a la izquierda, dos a la derecha y los otros dos sigan de frente. Calcule la probabilidad de que entre los siguientes siete automóviles a lo sumo dos den vuelta a la izquierda.

18. Un taller mecánico hace afinaciones para vehículos de 4, 6 y 8 cilindros. Hay dos tipos de afinación en cada caso, según los puntos que se revisen y los cambios que se efectúen. El precio P es $100 \times k$, donde k es el número de cilindros, para el afinamiento normal (N) y $200 \times k$ para el afinamiento especial (E). La siguiente tabla muestra la función de probabilidad conjunta para estas variables.

		K		
		4	6	8
P	N	0.3	0.2	0.1
	E	0.05	0.05	0.1

Halle las funciones de probabilidad marginales para las variables K y P. Si Z es el costo total del afinamiento para un carro que va al taller, halle su función de probabilidad.

19. Considere dos variables aleatorias X e Y con distribución conjunta discreta definida por la siguiente tabla para la función de probabilidad conjunta, donde $h = 1/60$.

		X_1		
		0	1	2
	0	h	$2h$	$3h$
X_2	1	$2h$	$4h$	$6h$
	2	$3h$	$6h$	$9h$
	3	$4h$	$8h$	$12h$

Calcule

a. $P(X \leq 1,\ Y \leq 1)$ b. $P(X + Y \leq 1)$ c. $P(X + Y > 2)$ d. $P(X < 2Y)$

e. $P(X > 1)$ f. $P(X = Y)$ g. $P(X \geq Y | Y > 1)$ h. $P(X^2 + Y^2 \leq 1)$

20. Repita el ejercicio anterior para la siguiente función de probabilidad conjunta (de nuevo $h = 1/60$).

		X_1		
		0	1	2
	0	h	$6h$	$6h$
X_2	1	$2h$	$8h$	$9h$
	2	$3h$	$2h$	$12h$
	3	$4h$	$4h$	$3h$

21. Las funciones de probabilidad marginales de las variables aleatorias discretas X e Y están dadas en la siguiente tabla:

		X					
		1	2	3	4	5	
	1						5/14
	2						4/14
Y	3						3/14
	4						2/14
	5						1/14
		1/14	5/14	4/14	2/14	2/14	1

Para i, j entre 1 y 5 la probabilidad conjunta $P(X = i, Y = j)$ sólo puede tomar los valores 0 y 1/14. Determine la función de probabilidad conjunta para X e Y.

22. Si A es un conjunto medible hemos definido la función $\mathbf{1}_A(\omega)$ como la función que vale 1 si $\omega \in A$ y 0 si no. Demuestre que dos eventos A y B son independientes si y sólo si $\mathbf{1}_A$ y $\mathbf{1}_B$ son variables aleatorias independientes.

23. Demuestre que dos variables aleatorias X, Y son independientes si y sólo si para cualesquiera $x, y \in \mathbb{R}$,
$$P(X > x, Y > y) = P(X > x)P(Y > y).$$

24. Demuestre que dos variables aleatorias X, Y son independientes si y sólo si para cualesquiera $a, b, c, d \in \mathbb{R}$, con $a < b$ y $c < d$ se tiene
$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d).$$

25. Lanzamos un dado dos veces. Sea X el resultado del primer lanzamiento, Y el mayor de los resultados de los dos lanzamientos. Halle la distribución conjunta y las distribuciones marginales de estas variables. Determine si son independientes.

26. Lanzamos una moneda tres veces y definimos las siguientes variables aleatorias: X es el número de águilas, Y es la longitud de la mayor sucesión de águilas en la muestra. Por ejemplo $Y(A, S, A) = 1$, $Y(A, A, S) = 2$. Halle la distribución conjunta, las distribuciones marginales y determine si estas variables son independientes.

27. Lanzamos una moneda cuatro veces y definimos las siguientes variables aleatorias: X vale 1 si hay más águilas que soles y vale 0 si esto no es cierto. Por otro lado, Y representa la longitud de la mayor sucesión de águilas en la muestra. Hallar la distribución conjunta y las marginales. Determine si estas variables son independientes.

28. Una función de probabilidad conjunta está dada por $p_{0,0} = a$, $p_{0,1} = b$, $p_{1,0} = c$, $p_{1,1} = d$, donde necesariamente $a + b + c + d = 1$. Demuestre que una condición necesaria para que haya independencia es que $ad = bc$.

29. Sean X, Y variables aleatorias independientes e idénticamente distribuidas que toman valores $1, 2, 3$ con igual probabilidad. Definimos $Z = X - Y$, $W = X + Y$. Halle la distribución conjunta de estas variables y sus distribuciones marginales. ¿Son independientes?

30. Sean X, Y variables aleatorias con valores en $\{1, 2, \dots, n\}$ y función de probabilidad conjunta $p_{ij} = 1/n^2$. Halle las funciones de probabilidad marginales y determine si las variables son independientes. Calcule $P(X > Y)$ y $P(X = Y)$.

31. En un vivero se siembran n semillas. Cada una de ellas germina de manera independiente con probabilidad α. Las X plantas germinadas son transplantadas a macetas y sobreviven de manera independiente con probabilidad β. sea Y el número de plantas que sobreviven. Halle la función de distribución conjunta de X, Y y las marginales.

32. Consideremos un experimento que tiene resultados $\omega_1, \omega_2, \dots, \omega_8$ con probabilidades correspondientes $0.1; 0.1; 0.2; 0.2; 0.1; 0.1; 0.1; 0.1$. Sea X, Y y Z las variables aleatorias definidas por la siguiente tabla

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
X	1	2	1	2	1	2	1	2
Y	1	2	3	1	2	3	1	2
Z	1	2	3	4	1	2	3	4

Halle las distribuciones de probabilidad de X, Y y Z y las distribuciones conjuntas de (X, Y); (X, Z); (Y, Z) y (X, Y, Z).

33. Considere dos eventos A y B tales que $P(A) = 1/4$, $P(B|A) = 1/2$ y $P(A|B) = 1/4$. Definimos las variables X e Y por $X = \mathbf{1}_A$, $Y = \mathbf{1}_B$, donde $\mathbf{1}_E(x)$ vale 1 si $x \in E$ y vale 0 si $x \notin E$. Diga si las siguientes proposiciones son ciertas o falsas.

a. Las variables aleatorias X e Y son independientes.

b. $P(X^2 + Y^2 = 1) = 1/4$.

c. $P(XY = X^2Y^2) = 1$.

d. La variable aleatoria X tiene distribución uniforme en el intervalo $(0, 1)$.

e. Las variables X e Y tienen la misma distribución.

34. Sean X, Y variables aleatorias independientes, X con distribución geométrica de parámetro p e Y con distribución geométrica de parámetro r. Demuestre que $Z = \min\{X, Y\}$ tiene distribución geométrica de parámetro $p + r - pr$.

35. Sean X, Y variables aleatorias independientes, X con distribución geométrica de parámetro p e Y con distribución geométrica de parámetro r. Sean $U = \min\{X, Y\}, V = \text{máx}\{X, Y\}, W = V - U$. Halle la distribución conjunta de U y V y la de U y W. Demuestre que estas dos últimas variables son independientes.

36. Sean X, Y variables aleatorias independientes, ambas con distribución geométrica de parámetro p. Demuestre que

$$P(X = k | X + Y = n) = \frac{1}{n+1}$$

para $k = 0, 1, \dots, n$.

37. Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes discretas con distribución uniforme en el conjunto $\{1, 2, \dots, k\}$. Halle la función de distribución conjunta de la variables M_n y m_n definidas por

$$M_n = \text{máx}\{X_1, X_2, \dots, X_n\}; \qquad m_n = \min\{X_1, X_2, \dots, X_n\}$$

38. Sean X, Y variables aleatorias discretas con función de probabilidad conjunta
$$p_{ij} = \beta^{i+j+2}, \quad i, j = 0, 1, \dots$$
¿Para cuáles valores de β es esta una función de probabilidad? Halle las funciones de probabilidad marginales y determine si las variables son independientes.

39. Responda la pregunta anterior con la restricción $0 \le i < j < \infty$ sobre los valores posibles de X e Y.

40. Sea $X_1, \dots, X_n$ una colección de variables aleatorias con la propiedad de que para todo r, $1 \le r \le n-1$, la colección $\{X_1, \dots, X_r\}$ es independiente de X_{r+1}. Demuestre que las variable X_i, $1 \le i \le n$ son independientes.

41. Sean X, Y variables aleatorias discretas con función de probabilidad conjunta
$$p_{ij} = \frac{C}{(i+j-1)(i+j)(i+j+1)}, \qquad i, j \in \mathbb{N}.$$
Calcule C, halle las funciones de probabilidad marginales y determine si las variables son independientes.

42. Para las variables del ejercicio anterior halle la función de probabilidad de $U = X + Y$ y $V = X - Y$.

43. Sea X e Y variables aleatorias independientes, X con distribución uniforme en $\{1, 2, \dots, n\}$, Y con distribución uniforme en $\{1, 2, \dots, m\}$. Halle la distribución de probabilidad de $Z = X + Y$.

44. Sean X e Y variables aleatorias independientes con distribución de Poisson de parámetros λ y μ. Demuestre que la distribución condicional de X dado que $X+Y = n$ es binomial y halle sus parámetros.

45. Sean X e Y variables aleatorias independientes con distribución binomial, ambas de parámetros n y p y sea $Z = X + Y$. Demuestre que la distribución condicional de X dado que $Z = k$ es hipergeométrica.

46. Sea X, Y, Z variables aleatorias discretas con la probabilidad de que sus valores son distintos con probabilidad 1. Sea $a = P(X > Y, b = P(Y > Z), c = P(Z > X)$.
   a) Demuestre que $\min\{a, b, c\} \le 2/3$ y dé un ejemplo donde esta cota se alcance.
   b) Demuestre que si X, Y, Z son independientes e idénticamente distribuidas, entonces $a = b = c = 1/2$.

47. Sea X, Y variables aleatorias discretas. Demuestre que son independientes si y sólo si su función de probabilidad conjunta $P(X = i, Y = j) = r_{ij}$ se puede factorizar como el producto $s_i t_j$ de una funcíon de i por una función de j.

48. Sean $X_j, 1 \le j \le n$ variables aleatorias independientes simétricas respecto a 0, es decir, X_j y $-X_j$ tienen la misma distribución. Demuestre que para todo x, $P(S_n \ge x) = P(S_n \le -x)$, con $S_n = X_1 + X_2 + \cdots + X_n$.
¿Es cierta en general la conclusión si no suponemos independencia?

49. Considere un paseo al azar simétrico simple S con $S_0 = 0$. Sea $T = \min\{n \ge 1 : S_n = 0\}$ el instante del primer regreso al origen. Demuestre que
$$P(T = 2n) = \frac{1}{2n-1}\binom{2n}{n}2^{-2n}.$$

50. Considere un paseo al azar simétrico simple S con $S_0 = 0$. Definimos $U = \min\{0 \le j \le n : S_{2j} = S_{2n}\}$ el instante de la primera visita a la posición que ocupa en el instante $2n$. Demuestre que
$$P(U = 2k) = P(S_{2k} = 0)P(S_{2n-2k} = 0),$$
para $0 \le k \le n$.

Capítulo 6

Esperanza Matemática

6.1. Esperanza Matemática de Variables Aleatorias Discretas.

Recordemos que una variable aleatoria X es discreta, si existe una sucesión $(x_n)_{n\geq 1}$ de números reales tales que

$$\sum_{n=1}^{\infty} P(X = x_n) = 1.$$

Comenzaremos por definir la noción de valor esperado para variables aleatorias discretas.

Definición 6.1 Sea X una variable aleatoria discreta con la notación anterior, y llamemos $p_n = P(X = x_n)$, $n = 1, 2, \ldots$. Diremos que existe el *valor esperado*, la *media* o la *esperanza matemática* de X si la serie

$$\sum_{n=1}^{\infty} |x_n| p_n \tag{6.1}$$

es convergente. En ese caso, el valor esperado se denota $\mathrm{E}(X)$ y se define mediante la serie

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} x_n p_n \tag{6.2}$$

Ejemplo 6.1
Sea X el resultado de lanzar un dado, entonces X toma valores $\{1, 2, 3, 4, 5, 6\}$ con probabilidad uniforme en este conjunto. Por lo tanto

$$\mathrm{E}(X) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{21}{6} = 3.5$$

Observamos que en este caso el valor esperado no es un valor posible de la variable aleatoria.

6.1.1. Propiedades

La esperanza matemática de una variable discreta tiene las siguientes propiedades.

Propiedad 1. Si $X \geq 0$ y existe $\mathrm{E}(X)$, entonces $\mathrm{E}(X) \geq 0$.

Es obvio que si $X \geq 0$, los valores x_n que figuran en la suma (6.1) son no-negativos, y si dicha serie es convergente, la suma también será no-negativa.

Propiedad 2. Si X es una variable aleatoria acotada entonces existe $\mathrm{E}(X)$.

Decir que la variable aleatoria X es acotada es decir que existe una constante C tal que

$$|x_n| \leq C \quad \text{para todo } n,$$

y, por lo tanto,

$$\sum_{n=1}^{N} |x_n| p_n \leq C \sum_{n=1}^{N} p_n \leq C.$$

Es decir que las sumas parciales de la serie (6.1) resultan estar acotadas por la constante C. Ahora bien, recordemos que para una serie de términos no-negativos – es el caso de la serie (6.1) – es necesario y suficiente para que converja que sus sumas parciales estén acotadas. Como las de (6.1) lo están, esta serie es convergente y el valor esperado de X existe.

Observación 6.1 Una primera observación es que si una variable aleatoria toma sólamente un número finito de valores, entonces su esperanza matemática está bien definida, ya que (6.1) se reduce a una suma finita.

Por otra parte, no siempre existe el valor esperado de una variable aleatoria. Por ejemplo, consideremos la variable aleatoria X tal que

$$x_n = 2^n \qquad \text{y} \qquad p_n = P(X = 2^n) = \frac{1}{2^n}, \quad n \geq 1.$$

Es claro que

$$\sum_{n=1}^{\infty} p_n = 1.$$

Se tiene que $x_n p_n = 1$, para todo n y por lo tanto, la serie (6.1) es divergente, de modo que esta variable aleatoria no tiene esperanza matemática.

En algunos textos el lector encontrará que, en casos como el de este ejemplo, en el cual la variable aleatoria X es no-negativa y la serie (6.1) diverge, se conviene en definir $\mathrm{E}(X) = +\infty$ en lugar de hacer como hemos optado, esto es, decir que no existe la esperanza.

Se debe hacer especial atención, sin embargo, en que si la variable no es no-negativa (o, en general, de signo constante), no hay esa opción convencional. Por ejemplo, si modificamos el ejemplo considerado y tomamos la variable aleatoria Y tal que

$$P(Y = 2^n) = \frac{1}{2^{n+1}} \qquad P(Y = -2^n) = \frac{1}{2^{n+1}} \quad n \geq 1,$$

entonces, nuevamente (6.1) diverge y no existe $\mathrm{E}(Y)$.

Propiedad 3. Sea A un evento y $\mathbf{1}_A$ la variable aleatoria definida por:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

Entonces

$$\mathrm{E}(\mathbf{1}_A) = P(A).$$

Es claro que $\mathbf{1}_A$ es discreta, ya que toma solamente dos valores y

$$\mathrm{E}(\mathbf{1}_A) = 1 \times P(\mathbf{1}_A = 1) + 0 \times P(\mathbf{1}_A = 0) = P(A).$$

Propiedad 4. Sea $g : \mathbb{R} \to \mathbb{R}$ una función y consideremos la nueva variable aleatoria (discreta)

$$Y = g(X).$$

Entonces, si existe la esperanza matemática de Y, ésta es

$$\mathrm{E}(Y) = \sum_{n=1}^{\infty} g(x_n) p_n. \tag{6.3}$$

Demostración. Llamemos $\{y_m\}$ a los valores de Y, entonces

$$\mathrm{E}(Y) = \sum_m y_m P(Y = y_m). \tag{6.4}$$

Por otra parte el evento $\{Y = y_m\}$ es igual a

$$\{X = x_n \text{ para algún } x_n \text{ tal que } g(x_n) = y_m\},$$

puesto que decir que Y toma el valor y_m equivale a decir que X toma un valor cuya imagen por la función g es y_m. Por lo tanto

$$P(Y = y_m) = \sum_{\{n : g(x_n) = y_m\}} p_n$$

donde el conjunto de los valores de n sobre los que se suma es el conjunto de los valores tales que $g(x_n) = y_m$. Sustituyendo en (6.4)

$$\mathrm{E}(Y) = \sum_m y_m \sum_{\{n : g(x_n) = y_m\}} p_n = \sum_m \sum_{\{n : g(x_n) = y_m\}} g(x_n) p_n$$

y ahora, un instante de reflexión mostrará al lector que la suma doble que aparece en la última igualdad es, simplemente, la suma sobre todos los valores de n, ya que cada n aparece una y sólo una vez allí. En resumen

$$\mathrm{E}(Y) = \sum_n g(x_n) p_n$$

como afirmamos en el enunciado de la propiedad 4. ■

Observación 6.2 La ventaja que tiene la fórmula (6.3), es que nos permite calcular $\mathrm{E}(Y)$ sin necesidad de calcular previamente la función de probabilidad de la variable aleatoria Y, bastándonos con la función de probabilidad de X.

Propiedad 5. Si existe $\mathrm{E}(X)$ entonces $|\,\mathrm{E}(X)| \leq \mathrm{E}(|X|)$.

Demostración. De acuerdo a la propiedad 4 (tomando $g(x) = |x|$), se tiene

$$\mathrm{E}(|X|) = \sum_n |x_n| p_n \geq |\sum_n x_n p_n| = |\,\mathrm{E}(X)|$$

Observe que la desigualdad entre las sumas de las series, se deduce simplemente de que

$$\sum_{n=1}^{N} |x_n| p_n \geq |\sum_{n=1}^{N} x_n p_n|$$

en virtud de la desigualdad triangular entre números reales. Pasando al límite cuando $N \to \infty$, se obtiene la desigualdad análoga entre las sumas de las series.

Propiedad 6. Si λ es una constante y $\mathrm{E}(X)$ existe, entonces también existe $\mathrm{E}(\lambda X)$ y

$$\mathrm{E}(\lambda X) = \lambda\,\mathrm{E}(X).$$

La demostración es sencilla y queda a cargo del lector.

Propiedad 7. Si X e Y son variables aleatorias que tienen valor esperado, entonces también existe el valor esperado de $X+Y$ y se tiene

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y).$$

Demostración. Para demostrar esta propiedad utilizamos la notación anteriormente introducida:

$$r_{nm} = P(X = x_n, Y = y_m), \quad p_n = P(X = x_n), \quad q_m = P(Y = y_m),$$

para $n, m = 1, 2, \dots$ ($\{r_{nm}\}$ es la función de probabilidad conjunta de X e Y y $\{p_n\}, \{q_m\}$ las respectivas funciones de probabilidad marginales).

Sea $Z = X + Y$ y $\{z_k\}$ el conjunto de valores posibles de Z. Para ver que la variable aleatoria Z tiene valor esperado, tenemos que probar que

$$\sum_k |z_k| P(Z = z_k) < \infty.$$

Procediendo de manera parecida a lo que hicimos para demostrar la propiedad 4

$$\begin{aligned}\sum_k |z_k| P(Z = z_k) &= \sum_k |z_k| P(X + Y = z_k) \\ &= \sum_k |z_k| \sum_{\{(n,m): x_n + y_m = z_k\}} r_{nm}\end{aligned}$$

donde en la última suma hay que entender que la suma interior se efectúa sobre todas las parejas de índices (n, m) tales que $x_n + y_m = z_k$. La última expresión es:

$$\begin{aligned}\sum_k \sum_{\{(n,m): x_n + y_m = z_k\}} |x_n + y_m| r_{nm} &= \sum_{n,m} |x_n + y_m| r_{nm} \leq \sum_{n,m} (|x_n| + |y_m|) r_{nm} \\ &= \sum_n |x_n| \sum_m r_{nm} + \sum_m |y_m| \sum_n r_{nm} \\ &= \sum_n |x_n| p_n + \sum_m |y_m| q_m < \infty,\end{aligned}$$

ya que, por un lado,

$$\sum_m r_{nm} = p_n, \quad n = 1, 2, \dots; \qquad \sum_n r_{nm} = q_m, \quad m = 1, 2, \dots$$

y por otro, las dos series

$$\sum_n |x_n| p_n; \qquad \sum_m |y_m| q_m$$

son convergentes, dado que existen $\mathrm{E}(X)$ y $\mathrm{E}(Y)$.

Esto prueba que existe $\mathrm{E}(Z)$. Si ahora repetimos el cálculo anterior sin los valores absolutos, resulta

$$\mathrm{E}(Z) = \sum_k z_k P(Z = z_k) = \sum_n x_n p_n + \sum_m y_m q_m = \mathrm{E}(X) + \mathrm{E}(Y),$$

que es la propiedad que queríamos probar. ■

Propiedad 8. Si $X \leq Y$ y existen $\mathrm{E}(X)$ y $\mathrm{E}(Y)$, entonces $\mathrm{E}(X) \leq \mathrm{E}(Y)$.

Demostración. Para probar esta propiedad, recurrimos a las propiedades 1, 6 y 7. Tenemos

$$\mathrm{E}(Y) - \mathrm{E}(X) = \mathrm{E}(Y - X) \geq 0, \quad \text{ya que } Y - X \geq 0.$$

Propiedad 9. Si X e Y son variables aleatorias independientes con valor esperado, entonces existe $\mathrm{E}(XY)$ y

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y). \tag{6.5}$$

Demostración. Procedemos de manera análoga, nuevamente, a lo que hicimos para la propiedad 7, teniendo en cuenta además que la independencia de X e Y implica que

$$r_{nm} = P(X = x_n, Y = y_m) = P(X = x_n)P(Y = y_m) = p_n q_m,$$

y por lo tanto, si llamamos ahora $\{w_k\}$ a los valores de XY

$$\begin{aligned}
\sum_k |w_k| P(XY = w_k) &= \sum_k \sum_{\{(n,m): x_n y_m = w_k\}} |w_k| r_{nm} \\
&= \sum_k \sum_{\{(n,m): x_n y_m = w_k\}} |x_n y_m| p_n q_m \\
&= \sum_{n,m} |x_n||y_m| p_n q_m = \sum_n |x_n| p_n \sum_m |y_m| q_m < \infty
\end{aligned}$$

lo cual muestra que existe $\mathrm{E}(XY)$.

El mismo cálculo, quitando los valores absolutos, permite obtener la relación (6.5). ∎

6.2. Momentos de Orden Superior. Momentos Centrados.

Definición 6.2 Sea X una variable aleatoria discreta,

$$P(X = x_n) = p_n, \qquad \sum_n p_n = 1$$

y m un número positivo. Si

$$\sum_n |x_n|^m p_n < \infty,$$

se define el *momento de orden m de X* mediante

$$\mathrm{E}(X^m) = \sum_n x_n^m p_n.$$

(Ver la propiedad 4, con $g(x) = x^m$).

Se define también el *momento centrado de orden m de X* para $m \geq 2$ mediante

$$\mu_m = \mathrm{E}((X - \mathrm{E}(X))^m) = \sum_n (x_n - \mathrm{E}(X))^m p_n.$$

(Ver la propiedad 4, con $g(x) = (x - \mathrm{E}(X))^m$. En la definición se sobrentiende que existe la esperanza de la variable aleatoria X).

Se acostumbra denominar *varianza* ($\mathrm{Var}(X)$) al momento centrado de orden 2 – cuando existe – es decir

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) \tag{6.6}$$

y *desviación típica o estándar* de X a $(\mathrm{Var}(X))^{1/2}$.

También se define la *covarianza* de dos variables aleatorias mediante

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))], \tag{6.7}$$

siempre que la esperanza que figura en el segundo miembro esté bien definida, y el *coeficiente de correlación* de X e Y por

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \tag{6.8}$$

siempre que el segundo miembro tenga sentido.

Finalmente, se dice que X e Y no están correlacionadas, son *no-correlacionadas* o *incorreladas* si $\mathrm{Cov}(X,Y) = 0$.

6.2.1. Propiedades

Veamos a continuación algunas propiedades de estos momentos.

Propiedad 1. Si a es una constante y X una variable aleatoria que tiene varianza, entonces

$$\mathrm{Var}(X) = \mathrm{Var}(X + a)$$

y

$$\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X).$$

La demostración resulta de aplicar la definición y queda como ejercicio para el lector.

Propiedad 2. $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

Demostración. Observar simplemente que

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[(X - \mathrm{E}(X))^2] = \mathrm{E}[X^2 - 2\,\mathrm{E}(X)X + (\mathrm{E}(X))^2]\\ &= \mathrm{E}(X^2) - 2\,\mathrm{E}(X)\,\mathrm{E}(X) + (\mathrm{E}(X))^2\\ &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2.\end{aligned}$$

Propiedad 3. (Desigualdad de Cauchy-Schwarz)

$$|\,\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

Demostración. Para demostrar esta desigualdad, consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida por

$$f(x) = \mathrm{E}[(U - xV)^2]\ ,$$

donde $U = X - \mathrm{E}(X)$ y $V = Y - \mathrm{E}(Y)$. Usando las propiedades de la esperanza matemática

$$f(x) = \mathrm{E}[U^2 - 2xUV + x^2V^2] = \mathrm{E}(U^2) - 2x\,\mathrm{E}(UV) + x^2\,\mathrm{E}(V^2)$$

de modo que f es un polinomio de segundo grado, y como

$$f(x) \geq 0 \quad \text{para todo } x \in \mathbb{R}$$

porque $(U - xV)^2 \geq 0$ para todo $x \in \mathbb{R}$, se deduce que el discriminante de este polinomio es menor o igual que cero (ya que, de no ser así, habría valores de $x \in \mathbb{R}$ donde $f(x)$ sería negativa). Es decir que

$$(-2\,\mathrm{E}(UV))^2 - 4\,\mathrm{E}(V^2)\,\mathrm{E}(U^2) \leq 0 \Rightarrow (\mathrm{E}(UV))^2 \leq \mathrm{E}(U^2)\,\mathrm{E}(V^2).$$

Tomando en cuenta quienes son U y V y las definiciones de varianza y covarianza, resulta la desigualdad que queríamos probar. ■

Propiedad 4. Sean X e Y dos variables aleatorias tales que está definido el coeficiente de correlación. Entonces

$$-1 \leq \rho(X, Y) \leq 1 \tag{6.9}$$

Demostración. Es una consecuencia sencilla de la propiedad anterior.

Propiedad 5. Si $\mathrm{Var}(X) = 0$ entonces existe una constante c tal que

$$P(X = c) = 1.$$

(Esta propiedad de que $P(X = c) = 1$ se enuncia a veces diciendo que X es una variable aleatoria "*casi seguramente*" igual a la constante c).

Demostración. Para probar esta propiedad, recordemos que X es una variable aleatoria discreta. (El resultado también será cierto en general, como ocurre con casi todas las propiedades que hemos estudiado, pero eso lo veremos más adelante).

Entonces la variable aleatoria $Y = (X - \mathrm{E}(X))^2$ también es discreta y toma valores mayores o iguales a cero, digamos $\{y_m\}$. Nuestra hipótesis es que

$$\mathrm{E}(Y) = 0.$$

Sea $q_m = P(Y = y_m)$. Si hubiera algún $y_{m_0} > 0$ tal que $p_{m_0} > 0$, se tendría que

$$\mathrm{E}(Y) = \sum_m y_m p_m \geq y_{m_0} p_{m_0} > 0$$

contradiciendo la hipótesis. Por lo tanto no hay un tal y_{m_0}, y si $y_m > 0$ necesariamente $p_m = 0$. En consecuencia

$$P(Y > 0) = \sum_{\{m: y_m > 0\}} p_m = 0,$$

de donde se deduce que

$$P(Y = 0) = P(Y \geq 0) - P(Y > 0) = 1 - 0 = 1.$$

Pero $\{Y = 0\}$ es el mismo evento que $\{X = \mathrm{E}(X)\}$, de modo que

$$P(X = \mathrm{E}(X)) = 1,$$

y la propiedad se cumple, tomando $c = \mathrm{E}(X)$. ■

Observación 6.3 Es obvio que el recíproco de esta propiedad es cierto, ya que si

$$P(X = c) = 1 \quad \text{entonces} \quad \mathrm{E}(X) = c$$

y

$$\mathrm{Var}(X) = \mathrm{E}((X - c)^2) = 0,$$

puesto que

$$P(X - c = 0) = P(X = c) = 1.$$

Propiedad 6. Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes, cada una de las cuales tiene varianza, entonces la suma también tiene varianza y

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n). \tag{6.10}$$

Demostración. Basta probar la igualdad (6.10) para $n = 2$. Para $n > 2$ se sigue por inducción completa, y queda como ejercicio.

Sea entonces $n = 2$ y pongamos $Y_1 = X_1 - \mathrm{E}(X_1)$, $Y_2 = X_2 - \mathrm{E}(X_2)$. Como X_1, X_2 son variables aleatorias independientes, también lo son Y_1, Y_2 y, en consecuencia, por la propiedad 9 de la sección 6.1.1

$$\mathrm{E}(Y_1 Y_2) = \mathrm{E}(Y_1)\,\mathrm{E}(Y_2) = 0$$

Se deduce que

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= \mathrm{E}[(X_1 + X_2 - \mathrm{E}(X_1 + X_2))^2] = \mathrm{E}[(Y_1 + Y_2)^2] \\
&= \mathrm{E}[Y_1^2 + Y_1^2 + 2Y_1Y_2] = \mathrm{E}(Y_1^2) + \mathrm{E}(Y_2^2) + 2\,\mathrm{E}(Y_1Y_2) \\
&= \mathrm{E}(Y_1^2) + \mathrm{E}(Y_2^2) = \mathrm{E}[(X_1 - \mathrm{E}(X_1))^2] + \mathrm{E}[(Y_2 - \mathrm{E}(Y_2))^2] \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)
\end{aligned}$$

que es lo que queríamos probar.

Propiedad 7. Sea X una variable aleatoria y supongamos que existe su momento de orden m. Entonces también existe su momento de orden k, para cualquier k, que satisfaga $0 \leq k < m$.

Demostración. Para probar que existe el momento de orden k hay que ver que

$$\sum_n |x_n|^k p_n < \infty, \quad \text{donde } p_n = P(X = x_n),\ \sum_n p_n = 1.$$

Pero tenemos la desigualdad, válida para todo número $x \in \mathbb{R}$,

$$|x|^k \leq 1 + |x|^m. \tag{6.11}$$

En efecto, si $|x| \leq 1$ la desigualdad es cierta ya que, en este caso, el primer miembro es menor o igual que 1 y el segundo es mayor o igual que 1. Por otra parte, si $|x| > 1$, es claro que

$$|x|^k < |x|^m < 1 + |x|^m$$

y también se verifica (6.11). En resumen,

$$\sum_n |x_n|^k p_n \leq \sum_n (1 + |x_n|^m) p_n = 1 + \sum_n |x_n|^m p_n < \infty.$$

■

Propiedad 8. Sean X e Y variables aleatorias con covarianza $\mathrm{Cov}(X, Y)$, entonces

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

Propiedad 9. Sean X e Y variables aleatorias con covarianza $\mathrm{Cov}(X, Y)$, entonces

$$\mathrm{Cov}(X + Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Las demostraciones quedan como ejercicio.

6.3. Ejemplos y Aplicaciones.

1. *Distribución Binomial.* Recordemos que la distribución binomial es la distribución de la variable S_n que representa el número de veces que ocurre un cierto evento A en n observaciones independientes del mismo. Por ejemplo, el número de veces que extraemos un objeto defectuoso de una cierta población de objetos fabricados en una línea de producción, cuando la extracción se hace al azar y con reposición.

   Definimos las variables aleatorias

$$X_i = \begin{cases} 1 & \text{si en la } i\text{-ésima observación ocurre } A, \\ 0 & \text{si en la } i\text{-ésima observación no ocurre } A. \end{cases}$$

   que admitimos que son independientes en nuestro modelo. Entonces

$$S_n = X_1 + X_2 + \cdots + X_n.$$

   Si denotamos $p = P(A)$, entonces

$$P(X_i = 1) = p \qquad i = 1, 2, \dots$$

   y

$$\mathrm{E}(X_i) = 1 \times P(X_i = 1) + 0 \times P(X_i = 0) = p$$

   Se deduce que

$$\mathrm{E}(S_n) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = np.$$

   Además, en virtud de que $X_1, X_2, \dots, X_n$ son variables aleatorias independientes

$$\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

   y

$$\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - (\mathrm{E}(X_i))^2 = p - p^2 = p(1-p),$$

   ya que $X_i^2 = X_i$ puesto que $X_i = 0$ ó 1. Reemplazando en la igualdad anterior, resulta

$$\mathrm{Var}(S_n) = np(1-p)$$

   También podemos calcular $\mathrm{E}(S_n)$ y $\mathrm{Var}(S_n)$ directamente, ya que conocemos la función de probabilidad de S_n, que es binomial $b(n,p)$, es decir que

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

   Por lo tanto, aplicando la definición de esperanza matemática en forma directa,

$$\begin{aligned} \mathrm{E}(S_n) &= \sum_{k=0}^{n} k P(S_n = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np. \end{aligned}$$

Observamos que la última suma es igual a 1 ya que no es otra cosa que la suma de todas las probabilidades ($j = 0, 1, \dots, n-1$) correspondientes a una distribución binomial con $(n-1)$ observaciones. O, también, es el desarrollo de Newton del binomio $(p+1-p)^{n-1} = 1$.

Con un cálculo parecido también reencontramos el resultado que hemos obtenido para $\mathrm{Var}(S_n)$; observemos primero que

$$\mathrm{Var}(S_n) = \mathrm{E}(S_n^2) - (\mathrm{E}(S_n))^2 = \mathrm{E}(S_n(S_n-1)) + \mathrm{E}(S_n) - (\mathrm{E}(S_n))^2$$

y entonces

$$\begin{aligned}
\mathrm{E}(S_n(S_n-1)) &= \sum_{k=0}^{n} k(k-1)P(S_n = k) \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2}(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j(1-p)^{n-2-j} = n(n-1)p^2.
\end{aligned}$$

La última suma nuevamente vale 1, por un argumento análogo al ya empleado. Reemplazando en la igualdad anterior

$$\mathrm{Var}(S_n) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

2. *Distribución de Poisson.* Recordemos que la variable aleatoria X tiene distribución de Poisson con parámetro $\lambda > 0$ si

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

Entonces

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda,
\end{aligned}$$

donde hemos empleado el desarrollo de Taylor de la función exponencial $e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$. Por lo tanto

$$\mathrm{E}(X) = \lambda.$$

Además

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\
&= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2
\end{aligned}$$

y en consecuencia

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}$$

Es decir
$$\mathrm{Var}(X) = \lambda.$$

3. *Distribución Geométrica.* Sea X una variable aleatoria con función de probabilidad
$$P(X = n) = p(1-p)^{n-1} \qquad n = 1, 2, \dots$$
con $0 < p < 1$. Entonces
$$\mathrm{E}(X) = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} np(1-p)^{n-1}. \tag{6.12}$$
Para calcular esta suma, recordemos que si
$$\sum_{n=0}^{\infty} c_n z^n$$
es una serie de potencias y R es su radio de convergencia (lo cual significa que si $|z| < R$ la serie converge y que, cuando $R < \infty$, si $|z| > R$ la serie no converge), si denotamos por $f(z)$ la suma de la serie para $|z| < R$:
$$f(z) = \sum_{n=0}^{\infty} c_n z^n \qquad (|z| < R),$$
entonces f es derivable y
$$f'(z) = \sum_{n=0}^{\infty} n c_n z^{n-1} \qquad (|z| < R).$$
Lo cual significa que, para $|z| < R$, podemos calcular la derivada de la función $f(z)$ como si fuera una suma finita, es decir, sumando las derivadas de cada uno de los términos.

Apliquemos este resultado a la serie de potencias
$$\sum_{n=0}^{\infty} z^n.$$
Primero, se verifica que el radio de convergencia de esta serie es igual a 1. Esto es directo: si $|z| > 1$ entonces $|z|^n \to \infty$, es decir que el término general no tiende a cero y la serie no converge, y si $|z| < 1$, la suma parcial de orden N es
$$\sum_{n=0}^{N} z^n = 1 + z + z^2 + \cdots + z^N = \frac{1 - z^{N+1}}{1-z} \overset{N\to\infty}{\longrightarrow} \frac{1}{1-z}$$
es decir que la serie converge cuando $|z| < 1$ y su suma es $\frac{1}{1-z}$, o sea
$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \qquad |z| < 1.$$
Derivando término a término como hemos indicado anteriormente, se tiene
$$\frac{1}{(1-z)^2} = \sum_{n=0}^{\infty} n z^{n-1} = \sum_{n=1}^{\infty} n z^{n-1} \qquad |z| < 1.$$
Volviendo ahora al problema que estábamos considerando, reemplazando z por $1-p$ en la ecuación (6.12) resulta
$$\mathrm{E}(X) = p \sum_{n=1}^{\infty} n(1-p)^{n-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Para calcular la varianza de esta distribución comenzamos por la relación

$$\mathrm{Var}(X) = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2, \tag{6.13}$$

y calculamos

$$\begin{aligned}\mathrm{E}(X(X-1)) &= \sum_{n=1}^{\infty} n(n-1)P(X=n) = \sum_{n=1}^{\infty} n(n-1)p(1-p)^{n-1} \\ &= p(1-p)\sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2}.\end{aligned}$$

Ahora bien, para calcular $f''(z)$ procedemos como antes, derivando término a término la serie que define $f'(z)$. Es decir, que para $|z| < 1$,

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \Rightarrow \frac{1}{(1-z)^2} = \sum_{n=1}^{\infty} nz^{n-1} \Rightarrow \frac{2}{(1-z)^3} = \sum_{n=2}^{\infty} n(n-1)z^{n-2}.$$

En consecuencia, reemplazando z por $1-p$, resulta

$$\mathrm{E}(X(X-1)) = p(1-p)\frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}$$

y volviendo a (6.13)

$$\mathrm{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

4. **Muestreo de poblaciones finitas.** Se considera una población dividida en r grupos dos a dos disjuntos (llamados "estratos") cuyas proporciones respectivas con relación al total de los miembros de la población son $q_1, q_2, \dots, q_r$. Se extrae una muestra de tamaño n de la población, al azar y con reposición.

   a. Calcular el valor esperado del número de estratos no representados en la muestra.

   b. Hacer el cálculo efectivo de dicho valor esperado, en los siguientes casos numéricos, suponiendo que todos los estratos son del mismo tamaño:

   i) $r = n = 5$.

   ii) $r = n = 100$.

► a. Consideremos las variables aleatorias $X_1, X_2, \dots, X_r$ definidas por

$$X_i = \begin{cases} 1 & \text{si el estrato } i \text{ no está representado en la muestra,} \\ 0 & \text{si el estrato } i \text{ está representado en la muestra.} \end{cases}$$

Entonces, el número de estratos no representados en la muestra es, evidentemente,

$$X_1 + X_2 + \cdots + X_r$$

y su esperanza matemática

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = \sum_{i=1}^{r} \mathrm{E}(X_i) = \sum_{i=1}^{r} P(X_i = 1)$$

ya que

$$\mathrm{E}(X_i) = 1 \times P(X_i = 1) + 0 \times P(X = 0).$$

Ahora bien, ¿cuánto vale $P(X_i = 1)$? El evento $\{X_i = 1\}$ sucede cuando el i-ésimo estrato no está representado en la muestra, es decir, que ninguno de los n individuos seleccionados pertenece al estrato i. En otros términos

$$\begin{aligned}\{X_i = 1\} =&\{1^{er} \text{ individuo no pertenece al estrato } i\}\\ &\cap \{2^{o} \text{ individuo no pertenece al estrato } i\}\\ &\cdots \cap \{n\text{-ésimo individuo no pertenece al estrato } i\}.\end{aligned}$$

Como el muestreo es con reposición, los n eventos que figuran en esta intersección son independientes y, en consecuencia, la probabilidad de su intersección es igual al producto de sus probabilidades:

$$P(X_i = 1) = \prod_{j=1}^{n} P(j\text{-ésimo individuo no pertenece al estrato } i).$$

Además,

$$P(j\text{-ésimo individuo no pertenece al estrato } i) = 1 - P(j\text{-ésimo individuo pertenece al estrato } i)$$

y esta última probabilidad es q_i, es decir, la proporción de individuos que contiene el estrato i con relación al total. Reemplazando obtenemos

$$P(X_i = 1) = (1 - q_i)^n$$

y

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = \sum_{i=1}^{r} (1 - q_i)^n.$$

b. Que todos los estratos son de igual tamaño implica que

$$q_1 = q_2 = \cdots = q_r$$

y como

$$\sum_{i=1}^{r} q_i = 1,$$

se deduce que

$$q_i = \frac{1}{r} \qquad (i = 1, 2, \ldots, r).$$

La esperanza matemática del número de estratos no representados en la muestra resulta ser

$$r\left(1 - \frac{1}{r}\right)^n.$$

Si $r = n = 5$, esto es $5\left(\frac{4}{5}\right)^5 \approx 1.64$, mientras que si $r = n = 100$ el resultado es $100\left(1 - \frac{1}{100}\right)^{100} \approx 36.6$. Es útil ver cómo encontramos un valor aproximado para esta expresión. Usando la fórmula de Taylor para $\log(1 - x)$

$$\log(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots \qquad (|x| < 1)$$

y entonces, si $|x| < 1$

$$(1 - x)^{\frac{1}{x}} = e^{\frac{1}{x}\log(1-x)} = e^{\frac{1}{x}(-x-\frac{x^2}{2}-\frac{x^3}{3}-\ldots)} = e^{-1-\frac{x}{2}-\frac{x^2}{3}-\ldots}$$

que es aproximadamente igual a e^{-1} cuando x es pequeño. Con $x = 1/100$ se tiene que la esperanza matemática es aproximadamente $100e^{-1}$, o sea que esperamos encontrar una proporción $1/e$ de estratos *no* representados en una muestra de $n = r$ individuos. Observe que esto es cierto si $n = r$ es suficientemente grande. ◀

5. Se lanzan dos dados simétricos independientes. Sean X_1, X_2 los resultados obtenidos en cada uno e $Y = \text{máx}\{X_1, X_2\}$.

   a. Hallar la probabilidad conjunta de X_1 e Y.

   b. Calcular $\text{E}(X_1)$, $\text{E}(Y)$, $\text{Var}(X_1)$, $\text{Var}(Y)$ y $\text{Cov}(X_1, Y)$.

► a. Usemos la notación $r_{ij} = P(X_1 = i, Y = j)$ para $i, j = 1, 2, 3, 4, 5, 6$. Es claro que si $i > j$ entonces $r_{ij} = 0$ porque $Y \geq X_1$. Si $i < j$, entonces

$$r_{ij} = P(X_1 = i, X_2 = j) = P(X_1 = i)P(X_2 = j) = \left(\frac{1}{6}\right)^2,$$

mientras que para $i = j$ tenemos

$$r_{ii} = P(X_1 = i, X_2 \leq i) = P(X_1 = i)P(X_2 \leq i) = \frac{1}{6}\frac{i}{6}.$$

Resumiendo

$$\begin{aligned} i < j &\Longrightarrow r_{ij} = \frac{1}{36}, \\ i = j &\Longrightarrow r_{ij} = \frac{i}{36}, \\ i > j &\Longrightarrow r_{ij} = 0. \end{aligned}$$

b.

$$\begin{aligned} \text{E}(X_1) &= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6}, \\ \text{Var}(X_1) &= \text{E}(X_1^2) - (\text{E}(X_1))^2 \\ &= \frac{1}{6}(1 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{21}{6}\right)^2 = \frac{35}{12}. \end{aligned}$$

Para calcular $\text{E}(Y)$ y $\text{Var}(Y)$, observemos que

$$P(Y = i) = \frac{2i - 1}{36} \qquad (i = 1, 2, 3, 4, 5, 6)$$

lo cual puede calcularse directamente o usando la parte (a) de este ejercicio. Por lo tanto

$$\begin{aligned} \text{E}(Y) &= 1\frac{1}{36} + 2\frac{3}{36} + 3\frac{5}{36} + 4\frac{7}{36} + 5\frac{9}{36} + 6\frac{11}{36} = \frac{161}{36} \\ \text{Var}(Y) &= \text{E}(Y^2) - (\text{E}(Y))^2 \\ &= 1\frac{1}{36} + 2^2\frac{3}{36} + 3^2\frac{5}{36} + 4^2\frac{7}{36} + 5^2\frac{9}{36} + 6^2\frac{11}{36} - \left(\frac{1}{36}\right)^2 \\ &= \frac{7365}{(36)^2}. \end{aligned}$$

Finalmente

$$\text{Cov}(X_1, Y) = \text{E}[(X_1 - \text{E}(X_1))(Y - \text{E}(Y))] = \text{E}(X_1 Y) - \text{E}(X_1)\,\text{E}(Y).$$

$\mathrm{E}(X_1Y)$ se calcula utilizando la probabilidad conjunta que encontramos en la parte a.

$$\begin{aligned}
\mathrm{E}(X_1Y) &= \sum_{i,j=1}^{6} ijr_{ij} = \sum_{i<j} ij\frac{1}{36} + \sum_{i=1}^{6} i^2\frac{i}{36} \\
&= \frac{1}{36}\left[\sum_{i=1}^{5} i \sum_{j=i+1}^{6} j + \sum_{i=1}^{6} i^3\right] \\
&= \frac{1}{36}\left[\sum_{i=1}^{5} \frac{(6-i)(6+i+1)}{2} i + \sum_{i=1}^{6} i^3\right] = \frac{154}{9}
\end{aligned}$$

lo que implica que $\mathrm{Cov}(X_1, Y) = \frac{35}{8}$. ◄

6. Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes y supongamos que

$$\mathrm{E}(X_i) = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2, \quad (i = 1, \dots, n).$$

Definimos la media muestral correspondiente a esta muestra por

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

y la varianza muestral por

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Es importante no confundirlas con la media μ y varianza σ^2 de cada una de las observaciones. Estas son constantes, mientras que las primeras son variables aleatorias.

Mostrar que

$$\mathrm{E}(\overline{X}) = \mu; \qquad \mathrm{E}(s^2) = \sigma^2.$$

► Comenzamos con

$$\mathrm{E}(\overline{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \mu.$$

Por otro lado

$$\mathrm{E}(s^2) = \mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}((X_i - \mu + \mu - \overline{X})^2). \tag{6.14}$$

Calculemos el valor de cada sumando:

$$\mathrm{E}((X_i - \mu)^2) - 2\,\mathrm{E}((X_i - \mu)(\overline{X} - \mu)) + \mathrm{E}((\overline{X} - \mu)^2) \tag{6.15}$$

y ahora cada uno de estos términos:

$\mathrm{E}((X_i-\mu)^2)$ es la varianza de X_i, es decir, σ^2.

$$\begin{aligned}
\mathrm{E}((X_i-\mu)(\overline{X}-\mu)) &= \mathrm{E}((X_i-\mu)(\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu))) \\
&= \mathrm{E}((X_i-\mu)\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)) \\
&= \frac{1}{n}\,\mathrm{E}(\sum_{j=1}^{n}(X_i-\mu)(X_j-\mu)) \\
&= \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}((X_i-\mu)(X_j-\mu)).
\end{aligned}$$

En esta suma, si $i \neq j$, como $X_i,\ X_j$ son variables aleatorias independientes, tenemos

$$\mathrm{E}((X_i-\mu)(X_j-\mu)) = \mathrm{E}(X_i-\mu)\,\mathrm{E}(X_j-\mu) = 0,$$

y por lo tanto

$$\mathrm{E}((X_i-\mu)(\overline{X}-\mu)) = \frac{1}{n}\,\mathrm{E}((X_i-\mu)^2) = \frac{\sigma^2}{n}.$$

Veamos a continuación el tercer término de (6.15):
$\mathrm{E}((\overline{X}-\mu)^2)$ es la varianza de $\overline{X}$, ya que

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{n\mu}{n} = \mu.$$

Pero, por otra parte,

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

ya que $X_1,\dots,X_n$ son variables aleatorias independientes.
En consecuencia, resulta que la suma en (6.15) es igual a

$$\sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} = \sigma^2\frac{n-1}{n},$$

y reemplazando en la igualdad (6.14)

$$\mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) = \frac{1}{n-1}n\sigma^2\frac{n-1}{n} = \sigma^2.$$

◀

Observación 6.4 Es fácil verificar que todo el cálculo realizado en este ejercicio, y por lo tanto la conclusión, permanece válido si las variables aleatorias $X_1,\dots,X_n$ no están correlacionadas, en lugar de exigirse que sean independientes.

7. Dar un ejemplo de dos variables aleatorias X e Y que no sean independientes, pero que, sin embargo,

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y),$$

lo cual implica que no están correlacionadas.

► Sea X una variable aleatoria que toma los valores $1, 0$ y -1, cada uno con probabilidad $1/3$, y definamos $Y = X^2$.

Es obvio que X e Y no son variables aleatorias independientes, ya que si conocemos el valor de X, también conocemos el valor de Y. Más precisamente, por ejemplo,

$$P(X = 1, Y = 1) = P(X = 1),$$

debido a que el evento $\{X = 1\}$ está contenido en el evento $\{Y = 1\}$. En consecuencia

$$P(X = 1, Y = 1) = \frac{1}{3} \neq P(X = 1)P(Y = 1) = \frac{1}{3} \times \frac{2}{3}$$

ya que

$$P(Y = 1) = P(X = 1) + P(X = -1) = \frac{2}{3}.$$

Sin embargo, X e Y cumplen que

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y).$$

Calculemos ambos miembros y veamos que dan lo mismo. El segundo es evidentemente igual a cero, ya que

$$\mathrm{E}(X) = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0.$$

En cuanto al primero, obervamos que $XY = X^3 = X$, ya que como X vale $1, 0$ ó -1, X^3 es siempre igual a X. En conclusión

$$\mathrm{E}(XY) = \mathrm{E}(X^3) = \mathrm{E}(X) = 0.$$

◄

Observación 6.5 En el enunciado se señala que si $E(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$, entonces X e Y son variables aleatorias no-correlacionadas. Esto se deduce de la Propiedad 9 de la sección 6.1.1.

6.4. Esperanza Matemática de Variables Aleatorias con Densidad.

Definición 6.3 Sea X una variable aleatoria, F_X su función de distribución y supongamos que ésta posee una densidad f_X. El *valor esperado* o *esperanza matemática* de X existe y es finito si y sólo si la integral

$$\int_{-\infty}^{\infty} |x| f_X(x) dx \tag{6.16}$$

es finita, y en este caso $\mathrm{E}(X)$ se define mediante la fórmula

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{6.17}$$

Veamos a continuación algunos ejemplos de aplicación de la fórmula (6.17).

Ejemplos.

1. *Distribución Uniforme*. Supongamos que X tiene distribución uniforme en el intervalo (a, b), $a < b$. Sabemos que ello implica que X tiene densidad

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b, \\ 0 & \text{si } x \leq a \text{ ó } x \geq b. \end{cases}$$

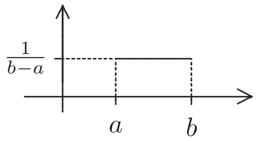


Figura 6.1

Tenemos

$$\begin{aligned}\mathrm{E}(X) &= \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_a^b x \frac{1}{b-a}\, dx \\ &= \frac{1}{b-a}\frac{1}{2}(b^2 - a^2) = \frac{b+a}{2}\end{aligned}$$

1. *Distribución Exponencial.* Si X tiene distribución exponencial con parámetro λ, $(\lambda > 0)$, su densidad es

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Entonces

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_0^{+\infty} x\, \lambda e^{-\lambda x}\, dx.$$

Integrando por partes

$$\begin{aligned}\mathrm{E}(X) &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda x} x \Big|_0^{+\infty} - \int_0^{+\infty} -\frac{1}{\lambda} e^{-\lambda x}\, dx \right] \\ &= \int_0^{+\infty} e^{-\lambda x}\, dx = \frac{1}{\lambda}\end{aligned}$$

es decir $\mathrm{E}(X) = 1/\lambda$.

2. *Distribución Normal.* Si X tiene distribución normal con parámetros (μ, σ^2), su densidad es

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

Se verifica fácilmente que

$$\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx < \infty,$$

y efectuando el cambio de variables $u = (x - \mu)/\sigma$, resulta

$$\begin{aligned}\mathrm{E}(X) &= \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\} dx \\ &= \int_{-\infty}^{+\infty} (\sigma u + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-u^2/2} du + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.\end{aligned}$$

La integral en el primer término vale cero, dado que la función $u\, e^{-u^2/2}$ es impar y tiene integral finita. En cuanto a la integral en el segundo término, vale 1, ya que no es sino la integral sobre toda la recta de la densidad normal con parámetros $(0, 1)$. Resumiendo $\mathrm{E}(X) = \mu$.

6.5. Cambio de Variables. Momentos de Orden Superior.

Sea $f(x)$ la densidad de la variable aleatoria X y

$$g : \mathbb{R} \to \mathbb{R}$$

una función con alguna regularidad. Consideremos la variable aleatoria

$$Y = g(X).$$

Entonces, el valor esperado de Y se puede calcular mediante la fórmula

$$\mathrm{E}(Y) = \int_{\mathbb{R}} g(x)f(x)dx \tag{6.18}$$

siempre que la integral de la derecha sea absolutamente convergente, es decir que

$$\int_{\mathbb{R}} |g(x)|f(x)dx < \infty$$

Un caso particular merece una mención aparte. Si $g : \mathbb{R} \to \mathbb{R}$ es la función

$$g(x) = x^k,$$

entonces

$$\mathrm{E}(g(X)) = \mathrm{E}(X^k)$$

se denomina el *momento de orden k de la variable aleatoria* X, que es finito si y sólo si

$$\mathrm{E}(|X|^k) < \infty.$$

Si X tiene densidad f_X y si

$$\int_{-\infty}^{+\infty} |x|^k f_X(x)dx < \infty,$$

entonces la fórmula (6.18) se traduce en

$$\mathrm{E}(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x)\, dx$$

Al igual que en el caso discreto definimos el *momento centrado de orden m* de X para $m \geq 2$ mediante

$$\mu_m = \mathrm{E}((X - \mathrm{E}(X))^m) = \int_{-\infty}^{+\infty} (x - \mathrm{E}(X))^m f_X(x)\, dx,$$

suponiendo que la integral exista.

Las definiciones de *varianza, desviación típica, covarianza* y *correlación* son iguales a las del caso discreto.

Observación 6.6 Las propiedades que demostramos para los momentos de orden superior en la sección 6.2.1 para el caso discreto también son ciertas para el caso continuo y aun con mayor generalidad. Para algunas propiedades las demostraciones no requieren modificación, como en el caso de las propiedades 1, 2 y 6, pero en otros es necesario dar demostraciones distintas, que no incluiremos.

Ejemplos.

1. *Distribución Uniforme.* Ya vimos que si X tiene distribución uniforme en (a,b), entonces

$$\mathrm{E}(X) = \frac{a+b}{2}.$$

Calculemos

$$\begin{aligned}
\mathrm{Var}(X) = \mathrm{E}[(X-\mathrm{E}(X))^2] &= \int_{-\infty}^{+\infty}\left(x-\frac{a+b}{2}\right)^2 f_X(x)dx \\
&= \frac{1}{b-a}\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx = \frac{1}{b-a}\frac{1}{3}\left(x-\frac{a+b}{2}\right)^3\Bigg|_{x=a}^{x=b} \\
&= \frac{1}{3(b-a)}\left[\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right] = \frac{(b-a)^2}{12}
\end{aligned}$$

2. *Distribución Exponencial.* Calculemos los momentos de orden k ($k \geq 1$ entero) de una variable aleatoria X con distribución exponencial con parámetro λ ($\lambda > 0$).

Recurriendo a la forma de la densidad respectiva, e integrando por partes

$$\mathrm{E}(X^k) = \int_0^{+\infty} x^k \lambda e^{-\lambda x} dx = -e^{-\lambda x}x^k\Big|_0^{\infty} + k\int_0^{\infty} x^{k-1}e^{-\lambda x}dx$$

de donde obtenemos

$$\mathrm{E}(X^k) = \frac{k}{\lambda}\,\mathrm{E}(X^{k-1}).$$

Procediendo de manera inductiva se deduce de aquí que

$$\mathrm{E}(X^k) = \frac{k}{\lambda}\frac{k-1}{\lambda}\cdots\frac{1}{\lambda}\,\mathrm{E}(X^0) = \frac{k!}{\lambda^k}$$

3. *Distribución Normal.* Sea X una variable aleatoria con distribución normal de parámetros $(0,1)$, es decir que su densidad es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \qquad \text{para } x \in \mathbb{R}.$$

Calculemos el momento de orden k ($k \geq 1$ entero) de X. Es claro que estos momentos son finitos, ya que una acotación sencilla (que dejamos como ejercicio) muestra que

$$\int_{-\infty}^{+\infty} |x|^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx < \infty.$$

Sabemos que

$$\mathrm{E}(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}dx. \tag{6.19}$$

Si k es impar, como el integrando resulta ser una función impar, la integral vale cero y $\mathrm{E}(X^k) = 0$.

Si k es un número par, digamos $k = 2p$, p entero positivo, integrando por partes en el segundo miembro de (6.19) se obtiene

$$\begin{aligned}
\mathrm{E}(X^{2p}) &= \frac{1}{\sqrt{2\pi}}\left[-x^{2p-2}e^{-x^2/2}\Big|_{-\infty}^{+\infty} + (2p-1)\int_{-\infty}^{+\infty} x^{2p-2}e^{-x^2/2}dx\right] \\
&= (2p-1)\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\, x^{2p-2}e^{-x^2/2}dx,
\end{aligned}$$

y por lo tanto

$$\mathrm{E}(X^{2p}) = (2p-1)\,\mathrm{E}(X^{2p-2}).$$

Aplicando ahora un procedimiento inductivo, se tiene

$$\mathrm{E}(X^{2p}) = (2p-1)(2p-3)\cdots 3\,\mathrm{E}(X) = \frac{(2p)!}{2^p\,p!} \tag{6.20}$$

Una variante sencilla es el cálculo del momento de orden k de la variable aleatoria $X-\mu$, donde X tiene distribución normal con parámetros (μ, σ^2).

En este caso

$$\mathrm{E}((X-\mu)^k) = \int_{-\infty}^{+\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}dx.$$

Haciendo el cambio de variables $t = (x-\mu)/\sigma$, resulta $dx = \sigma dt$ y

$$\mathrm{E}((X-\mu)^k) = \int_{-\infty}^{+\infty} (\sigma t)^k \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2} \sigma dt = \sigma^k \int_{-\infty}^{+\infty} t^k \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \sigma dt$$

y la integral de la derecha es la que venimos de calcular. Por lo tanto

$$k \text{ impar } \Rightarrow \mathrm{E}((X-\mu)^k) = 0$$
$$k = 2p \text{ par } \Rightarrow \mathrm{E}((X-\mu)^{2p}) = \sigma^{2p}\frac{(2p)!}{2^p\,p!}$$

En especial, si $k=2$ tenemos $\mathrm{Var}(X) = \mathrm{E}((X-\mu)^2) = \sigma^2$.

6.6. Dos Fórmulas para el Cálculo de Valores Esperados.

En primer lugar, sea X una variable aleatoria discreta, cuyos valores son enteros positivos:

$$P(X=n) = p_n \quad n = 0, 1, \dots, \qquad \sum_{n=0}^{\infty} p_n = 1.$$

Denotamos

$$q_n = P(X > n) = \sum_{k=n+1}^{\infty} p_k, \qquad n = 0, 1, \dots$$

entonces

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} q_n, \tag{6.21}$$

donde se debe entender que si $\mathrm{E}(X) = +\infty$ la serie diverge.

Para probar (6.21) observamos que

$$\begin{aligned}\sum_{n=1}^{N} np_n &= (p_1 + \cdots + p_N) + (p_2 + \cdots + p_N) + \cdots + (p_{N-1} + p_N) + p_N \\ &= \sum_{n=0}^{N-1} q_n - Nq_N \qquad (6.22)\end{aligned}$$

Esto ya muestra que si $\mathrm{E}(X) = +\infty$, caso en el cual el primer miembro tiende a $+\infty$ cuando $N \to \infty$, entonces

$$\sum_{n=0}^{\infty} q_n = +\infty,$$

ya que

$$\sum_{n=1}^{N} np_n \leq \sum_{n=0}^{N-1} q_n.$$

Si, por el contrario, $\mathrm{E}(X) < \infty$, es decir, si la serie

$$\sum_n np_n$$

es convergente, debe cumplirse que

$$Nq_N \to 0 \quad \text{cuando } N \to \infty,$$

porque

$$0 \leq Nq_N = N \sum_{n=N+1}^{\infty} p_n \leq \sum_{n=N+1}^{\infty} np_n \to 0 \quad \text{cuando } N \to \infty,$$

dado que el último miembro de esta cadena de desigualdades es la "cola"de una serie convergente. Pero entonces, pasando al límite en (6.22) cuando $N \to \infty$, se obtiene (6.21).

La segunda fórmula que veremos, de la cual en realidad (6.21) es un caso particular, es la siguiente: Sea X una variable aleatoria no-negativa y F_X su función de distribución. Entonces

$$\mathrm{E}(X) = \int_0^{+\infty} (1 - F_X(x))\, dx. \tag{6.23}$$

Hay que entender que en la fórmula (6.23),

$$\int_0^{+\infty} (1 - F_X(x))\, dx$$

se define como integral impropia, es decir,

$$\int_0^{+\infty} (1 - F_X(x))\, dx = \lim_{A\to\infty} \int_0^{A} (1 - F_X(x))\, dx,$$

y también que si $\mathrm{E}(X) = +\infty$, la integral en el segundo miembro de (6.36) es divergente, y en este sentido, también vale la igualdad. El lector interesado puede encontrar la demostración de (6.23) en el apéndice de este capítulo.

6.7. Ejemplos.

1. *Distribución Geométrica*. Sea X una variable aleatoria con distribución geométrica de parámetro p, $0 < p < 1$, es decir que
$$P(X = n) = p(1-p)^{n-1} \qquad n = 1, 2, \dots$$
Calcular $E(X)$ mediante aplicación de la fórmula (6.34).

► Sabemos que X es el tiempo que tarda en ocurrir el primer éxito en una sucesión de pruebas de Bernoulli con parámetro p. Por lo tanto, el suceso $\{X > n\}$ corresponde a que en las primeras n experiencias, el evento que estamos observando no ocurre, y esto implica que

$$q_n = P(X > n) = (1-p)^n$$

y entonces

$$\mathrm{E}(X) = \sum_{n=0}^{\infty}(1-p)^n = \frac{1}{1-(1-p)} = \frac{1}{p}$$

donde hemos usado la suma de una serie geométrica de razon $1-p$. ◄

2. *Distribución Normal.* Sea X una variable aleatoria con distribución normal de parámetros $(0, \sigma^2)$. Calcular $\mathrm{E}(|X|)$ y $\mathrm{Var}(|X|)$.

► Hagamos la transformación $Y = X/\sigma$, entonces Y es normal típica y

$$\mathrm{E}(|X|) = \sigma\,\mathrm{E}(|Y|) \qquad \mathrm{Var}(|X|) = \sigma^2\,\mathrm{Var}(|Y|)$$

y nos basta considerar el caso de Y. Tenemos

$$\begin{aligned}\mathrm{E}(|Y|) &= \int_{-\infty}^{+\infty} |x|\, f_Y(x)\,dx = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x\, e^{-x^2/2} dx \\ &= \sqrt{\frac{2}{\pi}} \left(-e^{-x^2/2} \right|_0^{+\infty} = \sqrt{\frac{2}{\pi}}.\end{aligned}$$

Por otro lado

$$\mathrm{Var}(|Y|) = \mathrm{E}(|Y|^2) - (\mathrm{E}(|Y|))^2, \qquad \mathrm{E}(|Y|^2) = \mathrm{Var}(Y) = 1.$$

(ya hemos visto que si una variable aleatoria tiene distribución normal de parámetros (μ, σ^2), entonces su varianza es σ^2. Por lo tanto la varianza de Y es igual a 1). Reemplazando

$$\mathrm{Var}(|Y|) = 1 - \frac{2}{\pi},$$

y volviendo al principio

$$\mathrm{E}(|X|) = \sigma\sqrt{\frac{2}{\pi}}, \qquad \mathrm{Var}(|Y|) = \sigma^2(1 - \frac{2}{\pi}).$$

◄

3. Calcular el valor esperado y la varianza de la variable aleatoria

$$Z = X_1^2 + X_2^2 + \cdots + X_n^2,$$

donde $X_1, \dots, X_n$ son variables aleatorias independientes cada una de las cuales tiene distribución normal de parámetros $(0, 1)$.

Nota. La distribución de Z se denomina distribución ji-cuadrado con n grados de libertad *y se denota χ_n^2.*

► El cálculo del valor esperado es sencillo:

$$\mathrm{E}(Z) = \mathrm{E}(X_1^2 + X_2^2 + \cdots + X_n^2) = \mathrm{E}(X_1^2) + \mathrm{E}(X_2^2) + \cdots + \mathrm{E}(X_n^2)$$

y como $\mathrm{E}(X_i^2)$ es la varianza de X_i, es claro que $\mathrm{E}(X_i) = 1$ para todo $i = 1, \dots, n$. Por lo tanto

$$\mathrm{E}(Z) = n.$$

En cuanto a la varianza, como $X_1, \dots, X_n$ son variables aleatorias independientes, se deduce que también $X_1^2, \dots, X_n^2$ son variables independientes, y por lo tanto

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \mathrm{Var}(X_i^2).$$

Como todas las X_i tienen igual distribución de probabilidad, todos los términos de esta suma son iguales. Además

$$\mathrm{Var}(X_1^2) = \mathrm{E}(X_1^4) - (\mathrm{E}(X_1^2))^2 = \frac{4!}{2^2 2!} - 1 = 2$$

donde hemos aplicado la fórmula (6.20) para calcular $\mathrm{E}(X_1^4)$. Resumiendo

$$\mathrm{Var}(Z) = 2n.$$

◀

4. Sean $X_1, X_2, \dots, X_n$ n observaciones independientes con distribución uniforme en el intervalo $(0, 1)$. Ordenando estas variables en forma creciente, obtenemos los llamados *estadísticos de orden* de las observaciones $X_1, X_2, \dots, X_n$:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

a. Calcular $\mathrm{E}(X_{(n)})$ y $\mathrm{Var}(X_{(n)})$.

b. Calcular, en general, $\mathrm{E}(X_{(k)})$ y $\mathrm{Var}(X_{(k)})$.

▶ a. La distribución de $X_{(n)}$ está dada por

$$P(X_{(n)} \leq x) = x^n \quad \text{si } 0 \leq x \leq 1,$$

ya que

$$\{X_{(n)} \in [0, x]\} = \cap_{i=1}^{n} \{X_i \leq x\}$$

y los eventos que aparecen en esta intersección son independientes. Por lo tanto,

$$F_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^n & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

y $X_{(n)}$ tiene densidad

$$f_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x \notin (0, 1) \\ n x^{n-1} & \text{si } 0 < x < 1 \end{cases}$$

y calculamos

$$\begin{aligned} \mathrm{E}(X_{(n)}) &= \int_{-\infty}^{+\infty} x f_{X_{(n)}}(x) dx = \int_0^1 x n x^{n-1} \, dx = \frac{n}{n+1} \\ \mathrm{E}(X_{(n)}^2) &= \int_0^1 x^2 n x^{n-1} \, dx = \frac{n}{n+2} \\ \mathrm{Var}(X_{(n)}) &= \mathrm{E}(X_{(n)}^2) - (\mathrm{E}(X_{(n)}))^2 \\ &= \frac{n}{n+2} - \left(\frac{n}{n+1} \right)^2 = \frac{n}{(n+2)(n+1)^2} \end{aligned}$$

b. En general, la densidad de $X_{(k)}$ está dada por

$$f_{X_{(k)}}(x) = \begin{cases} n\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k} & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

Entonces

$$\begin{aligned} \mathrm{E}(X_{(k)}) &= \int_0^1 xn\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k}\,dx, \\ \mathrm{E}(X_{(k)}^2) &= \int_0^1 x^2n\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k}\,dx. \end{aligned}$$

Si introducimos la notación

$$B(k,l) = \int_0^1 x^k(1-x)^l dx \quad \text{donde } k,l = 0,1,2\dots$$

entonces $\mathrm{E}(X_{(k)})$ y $\mathrm{E}(X_{(n)}^2)$ se escriben como

$$\begin{aligned} \mathrm{E}(X_{(k)}) &= n\binom{n-1}{k-1}B(k,n-k) \\ \mathrm{E}(X_{(n)}^2) &= n\binom{n-1}{k-1}B(k+1,n-k) \end{aligned}$$

y nuestro problema se reduce a calcular $B(k,l)$, lo cual hacemos a continuación.

Primero observemos que

$$B(k,l) = B(l,k)$$

haciendo el cambio de variables $y = 1 - x$ en la integral (6.39). Calculamos $B(k,l)$ integrando por partes. Sea $l \geq 1$:

$$\begin{aligned} B(k,l) &= \int_0^1 x^k(1-x)^l dx \\ &= \frac{1}{k+1}x^{k+1}(1-x)^l\Big|_0^1 + \frac{l}{k+1}\int_0^1 x^{k+1}(1-x)^{l-1}dx \\ &= \frac{l}{k+1}B(k+1,l-1). \end{aligned}$$

Esta fórmula de recurrencia, aplicada reiteradamente nos da

$$\begin{aligned} B(k,l) = \frac{l}{k+1}B(k+1,l-1) &= \frac{l(l-1)}{(k+1)(k+2)}B(k+2,l-2) \\ &= \frac{l(l-1)\cdots 1}{(k+1)\cdots(k+l)}B(k+l,0) \\ &= \frac{l!\,k!}{(k+l)!}\frac{1}{k+l+1} \end{aligned}$$

ya que $B(k+l,0) = 1/(k+l+1)$. Resumiendo,

$$B(k,l) = \frac{l!\,k!}{(k+l+1)!}$$

y reemplazando en las fórmulas para $\mathrm{E}(X_{(k)})$ y $\mathrm{E}(X_{(n)}^2)$ obtenemos

$$\begin{aligned}
\mathrm{E}(X_{(k)}) &= \frac{n!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}\\
\mathrm{E}(X_{(n)}^2) &= \frac{n!}{(k-1)!(n-k)!}\frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)}\\
\mathrm{Var}((X_{(k)}) &= \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} = \frac{k(n-k+1)}{(n+2)(n+1)^2}
\end{aligned}$$

◄

5. Sea X una variable aleatoria con distribución de Cauchy de densidad:

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}.$$

Demostrar que no existe $\mathrm{E}(X)$ pero que, sin embargo

$$\lim_{a\to\infty}\int_{-a}^{a} x\, f(x)\, dx = 0.$$

► Para ver que la esperanza matemática de X no existe, mostraremos que la integral

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx$$

es divergente. Efectivamente

$$\begin{aligned}
\frac{1}{\pi}\int_{-a}^{a}\frac{|x|}{1+x^2}\,dx &= \frac{2}{\pi}\int_{0}^{a}\frac{x}{1+x^2}\,dx\\
&= \frac{2}{\pi}\left(\frac{1}{2}\log(1+x^2)\right|_0^a = \frac{1}{\pi}\log(1+a^2)\\
&\to +\infty \quad \text{cuando } a\to\infty
\end{aligned}$$

Sin embargo, es obvio que

$$\frac{1}{\pi}\int_{-a}^{a}\frac{x}{1+x^2}\,dx = 0$$

ya que el integrando es una función impar, y por lo tanto,

$$\lim_{a\to\infty}\frac{1}{\pi}\int_{-a}^{a}\frac{x}{1+x^2}\,dx = 0$$

◄

6. Sean X e Y dos variables aleatorias, $\mathrm{Var}(X) \neq 0$, $\mathrm{Var}(Y) \neq 0$ y sea ρ su coeficiente de correlación

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Probar que si $\rho = 1$ ó $\rho = -1$, entonces existen dos constantes m y n tales que

$$P(Y = mX + n) = 1.$$

► Recordemos como probamos que $-1 \le \rho \le 1$ (Propiedad 4, sección 6.1.1). Consideramos la función $f(x) = \mathrm{E}((V - xU)^2)$, donde $U = X - \mathrm{E}(X)$ y $V = Y - \mathrm{E}(Y)$. Desarrollando el cuadrado

$$\begin{aligned} f(x) &= \mathrm{E}(V^2) - 2x\,\mathrm{E}(UV) + x^2\,\mathrm{E}(U^2) \\ &= \mathrm{Var}(Y) - 2x\,\mathrm{Cov}(X,Y) + x^2\,\mathrm{Var}(X) \end{aligned}$$

El discriminante de este polinomio de segundo grado es positivo (observar que el grado del polinomio es 2 ya que hemos supuesto que $\mathrm{Var}(X) \neq 0$). Por otra parte, este polinomio tiene discriminante

$$4(\mathrm{Cov}(X,Y))^2 - 4\,\mathrm{Var}(X)\,\mathrm{Var}(Y) = 0$$

ya que $\rho = 1$ ó $\rho = -1$. Esto quiere decir que ese polinomio tiene una raíz real doble, que llamaremos m. O sea que $f(m) = 0$, es decir

$$\mathrm{E}((V - mU)^2) = 0.$$

Pero $(V - mU)^2$ es una variable aleatoria no-negativa, cuya esperanza es nula y por lo tanto (ver propiedad 5, sección 6.2.1 y la observación 6.6 de la sección 6.5)

$$P((V - mU)^2 = 0) = 1 \Rightarrow P(V - mU = 0) = 1$$

ya que $\{V - mU = 0\}$ y $\{(V - mU)^2 = 0\}$ son el mismo evento. Reemplazando V por $Y - \mathrm{E}(Y)$ y U por $X - \mathrm{E}(X)$, resulta

$$P(Y - \mathrm{E}(Y) - m(X - \mathrm{E}(X)) = 0) = 1$$

y eligiendo $n = \mathrm{E}(Y) - m\,\mathrm{E}(X)$, se obtiene el resultado. ◄

Nota. El número ρ, que sabemos que está comprendido entre -1 y 1, es una medida de la dependencia lineal entre las variables aleatorias X e Y. Lo que dice este ejercicio es que si $\rho = 1$ ó $\rho = -1$, entonces la probabilidad de que el par aleatorio (X,Y) caiga en la recta de ecuación $y = mx + n$ es igual a 1, o, lo que es lo mismo, que la probabilidad de que caiga fuera de dicha recta es igual a cero.

Cuando $\rho = 1$ ó $\rho = -1$, la pendiente m de la recta en cuestión es la raíz doble de la ecuación $f(x) = 0$, es decir que

$$m = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

De modo que, si $\rho = 1$, $\mathrm{Cov}(X,Y) > 0$ y la pendiente es positiva y si $\rho = -1$, $\mathrm{Cov}(X,Y) < 0$ y m resulta negativo.

7. Un pasajero llega al terminal de autobuses en el instante T, que es una variable aleatoria con distribución uniforme entre las 11 y las 12 horas. De acuerdo con el horario, está previsto que del terminal partan un autobús a las 11 y otro a las 12, pero éstos salen con retardos aleatorios X e Y respectivamente, teniendo cada una de estas dos variables aleatorias distribución uniforme en el intervalo $(0, 1/2$ hora$)$.

   Si ambos autobuses le sirven al pasajero y las variables aleatorias T, X, Y son independientes, ¿cuál es el valor esperado del tiempo que el pasajero permanecerá en el terminal?

► Llamemos Z al tiempo que debe esperar el pasajero. Sea A el evento $\{11 \le T \le 11:30\}$, es decir "el pasajero llega entre las 11 y las 11 y 30" y B el evento $\{T - 11 \le X\}$, es decir, "llega antes de que el primer autobus salga". Entonces

– Si ocurren A y B, el tiempo de espera es $Z = X - (T - 11)$ (ver figura 6.2).

– Si ocurre A y no ocurre B, $Z = 12 - T + Y$ (ver figura 6.3).

– Si no ocurre A entonces $Z = 12 - T + Y$(ver figura 6.4).

En resumen,

$$\begin{aligned} &\text{en } A \cap B, && Z = X - (T - 11), \\ &\text{en } (A \cap B)^c, && Z = 12 - T + Y, \end{aligned}$$

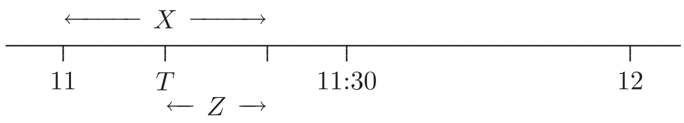


Figura 6.2

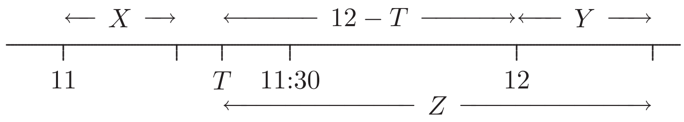


Figura 6.3

lo que es lo mismo que

$$Z = (X - (T - 11))\mathbf{1}_{A\cap B} + (12 - T + Y)\mathbf{1}_{(A\cap B)^c}$$

y si observamos que

$$\mathbf{1}_{(A\cap B)^c} = 1 - \mathbf{1}_{A\cap B}$$

obtenemos que

$$\begin{aligned} Z &= (X - (T - 11))\mathbf{1}_{A\cap B} + (12 - T + Y)(1 - \mathbf{1}_{(A\cap B)}) \\ &= (12 - T + Y) + (X - T + 11 - 12 + T - Y)\mathbf{1}_{A\cap B} \\ &= (12 - T + Y) + (X - Y - 1)\mathbf{1}_{A\cap B}. \end{aligned}$$

Por lo tanto

$$\mathrm{E}(Z) = 12 - \mathrm{E}(T) + \mathrm{E}(Y) + \mathrm{E}((X - Y - 1)\mathbf{1}_{A\cap B}).$$

Es claro que

$$\mathrm{E}(T) = 11.5 \qquad \mathrm{E}(Y) = 0.25.$$

Para calcular el último término, observamos que la densidad conjunta de las variables aleatorias T, X e Y es

$$f_{T,X,Y}(t, x, y) = \begin{cases} 4 & \text{si } 11 \le t \le 12;\ 0 \le x \le 0.5;\ 0 \le y \le 0.5 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Esto resulta inmediatamente de que T, X e Y son independientes y de que tienen densidad uniforme, la primera en $(11,\ 12)$ y en $(0,\ 0.5)$ las otras dos.

Por lo tanto, teniendo en cuenta la definición de las eventos A y B,

$$\mathrm{E}((X - Y - 1)\mathbf{1}_{A\cap B}) = \int_{11}^{11{,}5} dt \int_{t-11}^{0{,}5} dx \int_0^{0{,}5} (x - y - 1)4\, dy = -\frac{11}{48}$$

Reemplazando se obtiene

$$\mathrm{E}(Z) = \frac{25}{48} \text{ horas.}$$

◄

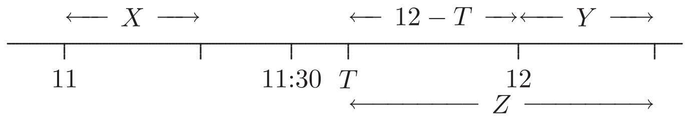


Figura 6.4

Ejercicios

1. Un jugador lanza dos monedas. Si ambas caen águila, gana \$ 10, si sólo hay un águila, gana \$ 2, mientras que si no hay ninguna águila pierde \$ 12. Determine el valor esperado de la ganancia para un lanzamiento de las dos monedas.

2. En una lotería se venden 1,000,000 de boletos de \$10 cada uno. Hay un primer premio de \$3,000,000, 10 premios de \$200,000, 100 premios de \$2,000, 1,000 premios de \$100 y 10,000 boletos reciben un reembolso del costo del boleto. Halle el valor esperado de la ganancia neta por boleto.

3. Extraemos al azar cartas de un juego de barajas con reposición hasta obtener un as y sea X el número de cartas extraídas. Halle $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

4. Lanzamos una moneda repetidamente hasta obtener dos águilas o dos soles, lo que ocurra primero. Sea X el númeo de lanzamientos de la moneda. Halle el valor esperado y la varianza de X.

5. Sea X una variable aleatoria con distribución de Poisson de parámetro λ. Halle el valor esperado de la variable $Y = e^X$.

6. Sean X, Y variables aleatorias cada una de las cuales toma únicamente dos valores distintos. Demuestre que X e Y son independientes si y sólo si $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

7. Lanzamos dos dados, si el lanzamiento es un *doble* (dos caras iguales) los dados se vuelven a lanzar y así hasta que las dos caras que se obtienen sean distintas. Sea X el número de lanzamientos necesarios para que esto ocurra. Halle el valor esperado y la varianza de X. Halle también el valor esperado y la varianza de 2^X

8. En una bolsa hay cinco boletos que corresponden a un premio de \$1,000 y cuatro premios de \$20. Cinco personas sacan, sucesivamente, un boleto al azar y sea X_i el premio que le corresponde a la i-ésima persona en sacar un boleto. Calcule $\mathrm{E}(X_1)$ y $\mathrm{Var}(X_1)$, (b) $\mathrm{E}(X_3)$ y $\mathrm{Var}(X_5)$, (c) $\mathrm{E}(X_1 + \cdots + X_5)$ y $\mathrm{Var}(X_1 + \cdots + X_5)$.

9. Una caja contiene 5 bolas numeradas del 1 al 5. Se extraen 200 bolas con reposición y sea X_i el número en la i-ésima bola extraída, Y la suma de los 200 números obtenidos y Z el promedio de los 200 números. Halle (a) $\mathrm{E}(X)$ y $\mathrm{Var}(X)$, (b) $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$, (c) $\mathrm{E}(Z)$ y $\mathrm{Var}(Z)$.

10. Lanzamos un par de dados, sea X_1 el número que aparece en el primero y X_2 el del segundo. Definimos $Y = X_1 + X_2$, $Z = XY$. Halle (a) $\mathrm{E}(X_1)$, (b) $\mathrm{E}(X_2^2)$, (c) $\mathrm{E}(Y)$, (d) $\mathrm{E}(Z)$, (e) $\mathrm{E}(YZ)$, (f) $\mathrm{E}(Y^2)$, (g) $\mathrm{E}(Z^2)$.

11. En un concurso hay cinco cajas idénticas cerradas. En una de ellas hay \$100,000, otra contiene \$10,000, una tercera \$1,000, la cuarta \$100 y la última tiene un signo de ALTO. El concursante escoge una caja y gana el contenido. Este proceso se repite hasta que salga el signo d ALTO, momento en el cual el concurso termina. Halle el valor esperado de la cantidad que gana el concursante.

12. Una máquina produce objetos que son defectuosos con probabilidad 0.01 y cuando esto ocurre, la máquina se detiene y es ajustada. Halle el valor promedio del número de objetos buenos producidos entre dos objetos defectuosos.

13. Calcule media y varianza para las distribuciones de los ejercicios 1, 2, 4, 5, 7, 8, 9, 10, 11, 18 y 19 del Capítulo 4.

14. Sea X una variable aleatoria con $\mathrm{E}(X) = 2,\ \mathrm{Var}(X) = 1,\ \mathrm{E}(X^4) = 34$. Calcular media y varianza para las siguientes variables aleatorias: $U = 2X + 1,\ V = X^2,\ Z = -X^2 + 2$.

15. Las variables X, Y son independientes y $\mathrm{E}(X) = -3,\ \mathrm{E}(X^4) = 100,\ \mathrm{E}(Y) = 4,\ \mathrm{E}(Y^4) = 500,\ \mathrm{Var}(X) = 0.5,\ \mathrm{Var}(Y) = 2$. Calcular la media y varianza para las variables $U = 3X - 2Y$ y $V = X^2 - Y^2$.

16. Suponga que X, Y tienen igual media e igual varianza y además son independientes. Demuestre que
$$\mathrm{E}((X - Y)^2) = 2\,\mathrm{Var}(X).$$

17. Suponga que X, Y tienen igual varianza. Demuestre que
$$\mathrm{E}((X + Y)(X - Y)) = \mathrm{E}(X + Y)\,\mathrm{E}(X - Y).$$

18. Suponga que X, Y son independientes y ambas tienen media 3 y varianza 1. Halle la media y varianza de $X + Y$ y XY.

19. Demuestre que
$$\mathrm{Var}(X + Y) + \mathrm{Var}(X - Y) = 2\,\mathrm{Var}(X) + 2\,\mathrm{Var}(Y).$$

20. Sea X una variable con valor esperado finito $\mathrm{E}(X)$. Demuestre que
$$(\mathrm{E}\,X)^2 \le \mathrm{E}(X^2)$$

21. Sean $X_1, \dots, X_n$ variables aleatorias independientes idénticamente distribuidas con media μ y varianza σ^2. Definimos la media muestral por $\overline{X} = n^{-1}\sum_{k=1}^{n} X_k$. Demuestre que
$$\mathrm{Cov}(\overline{X}, X_k - \overline{X}) = 0, \quad 1 \le k \le n.$$

22. Se considera el siguiente juego de azar entre dos jugadores. El primero elige al azar un punto X en el intervalo $(0, 2)$ mientras que el segundo elige un punto Y en el intervalo $(1, 3)$, también con distribución uniforme. Suponemos que X e Y son variables aleatorias independientes. Entonces

    – Si $X < Y$, el primer jugador paga $a(Y - X)$ unidades al segundo.

    – Si $X \ge Y$, el segundo jugador paga $b(X - Y)$ unidades al primero,

    donde a y b son constantes positivas.

    a. Hallar la relación b/a para que el juego sea equitativo (esto significa que la esperanza matemática de la ganancia de cada jugador es igual a cero).

    b. Con la relación b/a calculada en la parte anterior, calcular la varianza de la ganancia del primer jugador.

23. Sea X, Y variables aleatorias independientes, ambas con distribución de Bernoulli con probabilidad de éxito 1/2. Demuestre que $X + Y$ y $|X + Y|$ son dependientes pero no están correlacionadas.

24. Sea X una variable aleatoria con distribución geométrica de parámetro p y M un número entero positivo. Calcular la esperanza matemática de la variable aleatoria $Y = \min\{X, M\}$.

25. (a) Se lanza una moneda perfecta repetidamente. Sea ν la variable aleatoria que indica el número de veces seguidas que ocurre lo mismo que en el primer lanzamiento. (Por ejemplo, si A es águila y S es sol, con la sucesión de resultados $AAASSAS\dots$ se tiene $\nu = 3$ mientras que con la sucesión $SSSSSASAS\dots$ se tiene $\nu = 5$). Calcular $\mathrm{E}(\nu)$ y $\mathrm{Var}(\nu)$.

    (b) Rehacer el cálculo de la parte a suponiendo que, en lugar de una moneda perfecta, se dispone de una moneda tal que la probabilidad de águila en un lanzamiento es igual a p. ¿Qué ocurre si $p = 0$ ó $p = 1$?

26. (a) Una ciudad está dividida en cuatro partes, con número respectivo de habitantes H_1, H_2, H_3 y H_4. Supongamos que el promedio $m = (H_1 + H_2 + H_3 + H_4)/4$ de habitantes de las partes, es de 1000 personas y sea

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i - m)^2.$$

Se eligen al azar en un muestreo sin reposición dos de las cuatro partes y se cuenta el número de habitantes obtenidos en cada una de ellas, que denominamos X_1 y X_2. Mostrar que

$$\mathrm{E}(X_1 + X_2) = 2 \times 1000$$
$$\mathrm{Var}(X_1 + X_2) = \frac{4}{3}\sigma^2$$

(b) Generalizar al caso en el cual, en lugar de cuatro, la ciudad está dividida en n partes, y en lugar de seleccionar al azar dos de ellas, se seleccionan r. Se obtiene

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = mr,$$

donde m es el promedio del número de habitantes de las partes, y

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_r) = \frac{r(n-r)}{n-1}\sigma^2.$$

27. Se escriben n cartas y sus respectivos sobres, y se ensobran las cartas al azar, de modo que la probabilidad de cualquiera de las posibles permutaciones de las cartas, es la misma. Calcular la esperanza y la varianza del número de cartas que se ensobran correctamente.

Sugerencia. Escribir Y como $Y = \sum_{i=1}^{n} X_i$, donde

$$X_i = \begin{cases} 1 & \text{si la i-ésima carta va en el i-ésimo sobre} \\ 0 & \text{en caso contrario.} \end{cases}$$

28. En una bolsa hay n tarjetas numeradas de 1 a n. Se extraen las tarjetas sucesivamente con reposición. (a) ¿Cuál es el valor esperado del número de extracciones hasta repetir el primer número? (b) ¿Cuál es el valor esperado del número de extracciones hasta que ocurra la primera repetición?

29. Se considera un conjunto de n personas. Calcular el valor esperado del número de días del año en que cumplen años exactamente k de ellos. Suponga que el año tiene 365 días, que todas las distribuciones posibles son igualmente probables y que $n \geq k$.

30. Una persona con n llaves trata de abrir una puerta probando las llaves sucesiva e independientemente. Calcule la esperanza y la varianza del número ν de intentos requeridos hasta encontrar la llave correcta, suponiendo que ésta es una sola, en las dos situaciones siguientes.

(a) Si la selección de las llaves es con reposición, es decir, que una llave inservible no es quitada del lote, una vez probada. (b) Si la selección es sin reposición.

31. Un sistema permite establecer comunicaciones de acuerdo al diagrama 6.1. Cada bloque rectangular demora una unidad de tiempo y T es la variable aleatoria que indica el tiempo que se demora en establecer una comunicación buena.

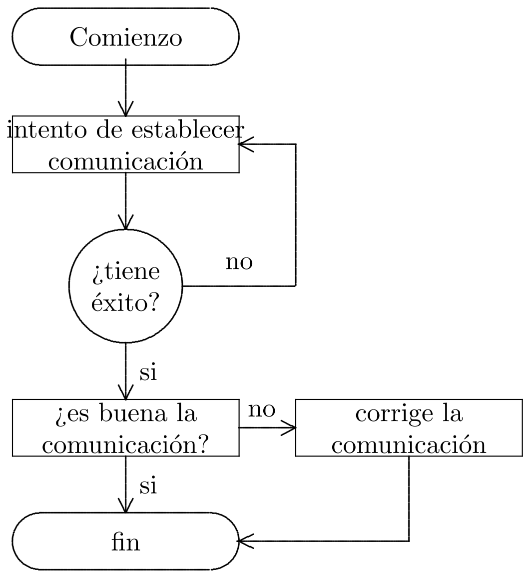


Diagrama 6.1

Se supone que los intentos para establecer comunicación son independientes y que la probabilidad de éxito en cada uno es p. Una vez establecida, la comunicación puede ser buena o mala, y la probabilidad de que sea buena es b. Si es mala, en una operación más se corrige y deviene buena. Calcular la función de probabilidad de T y su esperanza matemática.

32. Se lanza un dado n veces. Sea X_i el número de veces que se obtiene el resultado i. Calcular la covarianza de X_1 y X_6.

33. En un estanque en el que originalmente hay n peces, se pesca sucesivamente por medio de redes, que retiran $n_1,\ n_2, \dots$, donde n_k denota el número (aleatorio) de peces extraídos la k-ésima vez. Suponiendo que la probabilidad de que cada pez sea capturado es p, calcular la esperanza matemática del número de peces extraídos en la n-ésima ocasión.

34. Una bolsa contiene bolas numeradas de 1 a N. Sea X el mayor número obtenido en n extracciones al azar y con reposición, donde n es un número fijo.

    (a) Hallar la distribución de probabilidad de X.

    (b) Calcular $\mathrm{E}(X)$ y probar que si N es grande, $\mathrm{E}(X)$ es aproximadamente igual a $N/(n+1)$.

    (c) Hallar $\mathrm{E}(X^2)$ y deducir una expresión asintótica para $\mathrm{Var}(X)$ cuando $N \to \infty$.

35. Calcular $\mathrm{E}(X^3)$ donde X tiene distribución binomial con parámetros (n, p).

36. Sea X, Y variables aleatorias discretas con función de probabilidad

$$p_{ij} = \frac{1}{1 + 2n(n+1)}, \quad \text{para } |i - j| \le n, |i + j| \le n.$$

    Demuestre que X e Y son independientes y que $\mathrm{Cov}(X, Y) = 0$.

37. Sea X, Y variables independientes e idénticamente distribuidas con distribución normal $\mathcal{N}(\mu, \sigma^2)$. Halle media y varianza de la variable $Z = \max(X, Y)$. (Ayuda: si x, y son números reales demuestre que $2 \max(x, y) = |x - y| + x + y$).

38. Sea X una variable con distribución $\mathcal{N}(\mu, \sigma^2)$. Halle media y varianza de $|X - c|$ cuando (a) c es una constante dada, (b) $\sigma = \mu = c = 1$, (c) $\sigma = \mu = 1, c = 2$.

39. Calcule media y varianza para las distribuciones de los ejercicios 18, 19, 21, 22, 24, 25, 26, 29 y 30 del Capítulo 4.

40. En el ejercicio 31 del capítulo 4 vimos que de acuerdo a la ley de Maxwell, la velocidad v de una molécula de gas de masa m a temperatura absoluta T es una variable aleatoria con densidad

$$f_v(x) = \frac{4}{\sqrt{\pi}} \frac{1}{\alpha^3} x^2 e^{-x^2/k^2}$$

para $x > 0$ con $\alpha = (2kT/m)^{1/2}$ y k es la constante de Boltzman. Halle media y varianza de (a) la velocidad v de la molécula, (b) la energía cinética $E = mv^2/2$ de la molécula.

41. Sea X una variable aleatoria con distribución uniforme en $(-\frac{\pi}{2}, \frac{\pi}{2})$. Calcular $\mathrm{E}(Y)$ cuando (a) $Y = \mathrm{sen} X$. (b) $Y = \cos X$ (c) $Y = 3X^2 + 2$ (d) $Y = 1/|X|^\alpha$ En el último caso, ¿para cuáles valores de α se tiene que $\mathrm{E}(Y) < \infty$?

42. Calcular la esperanza y la varianza de las distribuciones cuyas densidades se indican a continuación:

$$(a)\ f(x) = \begin{cases} 2x & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases} \qquad (b)\ f(x) = \begin{cases} |x| & \text{si } x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

43. Sea X una variable con distribución exponencial. (a) Halle $P(\mathrm{sen}\, X > 1/2)$, (b) $\mathrm{E}(X^n)$ para $n \geq 1$.

44. La duración T de cierto tipo de llamadas telefónicas satisface la relación

$$P(T > t) = ae^{-\lambda t} + (1 - a)e^{-\mu t}, \quad t \geq 0,$$

donde $0 \leq a \leq 1, \lambda > 0, \mu > 0$ son constantes. halle media y varianza para T.

45. El tiempo de vida de cierto componente electrónico tiene distribución exponencial con parámetro $\lambda = 0.01$. Calcule el tiempo de vida promedio. Si el aparato ha durado 50 horas, ¿cuál es el valor esperado del tiempo de vida que le queda?

46. Escogemos n puntos al azar en $[0, 1]$ con distribución uniforme. Sea $m_n = \min\{X_1, \dots, X_n\}$, $M_n = \text{máx}\{X_1, \dots, X_n\}$, $R_n = M_n - m_n$. Halle el valor esperado de estas tres variables aleatorias.

47. La proporción de tiempo que una máquina trabaja durante una semana de 40 horas es una variable aleatoria con densidad $f(x) = 2x$ para $0 \leq x \leq 1$. Halle $\mathrm{E}(X)$ y $\mathrm{Var}(X)$. La ganancia semanal Y para esta máquina está dada por $Y = 200X - 60$; determine $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$.

48. Sea X una variable aleatoria con densidad $f(x) = Cx^{a-1}e^{-x^a}$, $x \geq 0$. Halle el valor de C y calcule $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

49. Sea X una variable aleatoria con densidad $f(x) = 6x(1 - x)$ para $0 < x < 1$. Compruebe que f es una densidad y obtenga media y varianza para esta distribución.

50. Para ciertas muestras de minerales la proporción de impurezas por muestra X es una variable aleatoria con densidad dada por $f(x) = 1.5x^2 + x$ para $0 \leq x \leq 1$. El valor en pesos de cada muestra es $Y = 5 - 0.5X$. Encuentre el valor esperado y la varianza de Y

51. Sea X una variable aleatoria con densidad $f(x) = |\,\mathrm{sen}(x)|/4$ para $0 \leq x \leq 2\pi$. Calcule $\mathrm{E}(X)$.

52. La proporción de tiempo por día en que todas las cajas registradoras están ocupadas a la salida de cierto supermercado es una variable aleatoria X con una densidad $f(x) = Cx^2(1 - x)^4$ para $0 \leq x \leq 1$. Determine el valor de C y $\mathrm{E}(X)$.

53. **Distribución de Pareto** La distribución de Pareto con parámetros r y A, $r > 0,\ A > 0$, es aquella que tiene la densidad

$$f(x) = \begin{cases} \frac{rA^r}{x^{r+1}} & \text{si } x \geq A, \\ 0 & \text{si } x < A. \end{cases}$$

(a) Calcular la esperanza y la varianza de esta distribución de probabilidad.

(b) Calcular y graficar la función

$$Q(y) = F(\mu + y\sigma) - F(\mu - y\sigma)$$

para $y \geq 0$ donde μ es la esperanza y σ^2 la varianza halladas en a, y F es la función de distribución de probabilidad de Pareto, en el caso $A = 1,\ r = 3$.

54. **Distribución Beta** Sea X una variable aleatoria con densidad

$$f(x) = Cx^{a-1}(1-x)^{b-1}, \quad 0 \leq x \leq 1.$$

Halle el valor de C y calcule $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

55. En un proceso de manufactura se ensamblan cuantro componente sucesivamente de longitudes X_1, X_2, X_3 y X_4, respectivamente con $\mathrm{E}(X_1) = 20,\ \mathrm{E}(X_2) = 30,\ \mathrm{E}(X_3) = 40,\ \mathrm{E}(X_4) = 60$ y $\mathrm{Var}(X_j) = 4$ para $J = 1, 2, 3, 4$. Halle la media y varianza para la longitud total $L = X_1 + X_2 + X_3 + X_4$ (a) si las longitudes de los componentes no están correlacionadas, (b) si $\rho(X_j, X_k) = 0.2$ para $1 \leq j < k \leq 4$.

56. **Distribución de Rayleigh** Una variable tiene distribución de Rayleigh si su densidad es

$$f(x) = \frac{2x}{\theta} e^{-x^2/\theta}$$

para $x > 0$. Calcule $\mathrm{E}(X)$, $\mathrm{Var}(X)$ y obtenga la densidad de $Y = X^2$.

57. **Distribución de Laplace** Una variable X tiene distribución de Laplace si su densidad está dada por

$$f(x) = \frac{1}{2\beta} \exp\Big(-\frac{|x-\alpha|}{\beta}\Big), \quad x \in \mathbb{R},$$

para ciertas constantes α y $\beta > 0$. Halle media y varianza para una variable con esta distribución.

58. El tiempo de vida de ciertos focos tiene distribución de Rayleigh con función de distribución

$$F(x) = 1 - \exp\Big(\frac{x^2}{300}\Big), \quad \text{para } x > 0.$$

Hallar la densidad de la variable aleatoria $Y = X^2$ y su valor esperado.

59. Sea X una variable aleatoria con tercer momento finito $\mathrm{E}(|X|^3) < \infty$). Definimos el coeficiente de asimetría de X, $\alpha(X)$ por

$$\alpha(X) = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3}$$

donde $\mu = \mathrm{E}(X)$ y $\sigma^2 = \mathrm{Var}(X)$.

(a) Demuestre que para cualquier variable X se tiene que

$$\alpha(X) = \frac{\mathrm{E}(X^3) - 3\mu\,\mathrm{E}(X^2) + 2\mu^3}{\sigma^3}$$

Halle $\alpha(X)$ si X tiene distribución

(b) Bernoulli con parámetro p.

(c) Poisson con parámetro λ.

(d) Geométrica con parámetro p.

(e) Demuestre que si $X_1, \dots, X_n$ son variables aleatorias independientes e idénticamente distribuidas entonces

$$\alpha(X_1 + \cdots + X_n) = \frac{\alpha(X_1)}{\sqrt{n}}.$$

(f) Halle $\alpha(X)$ si X tiene distribución binomial.

60. Sea $\{p_n\}$ la distribución de probabilidad de la variable aleatoria X, concentrada en los enteros no-negativos, es decir que

$$P(X = n) = p_n, \qquad n = 0, 1, 2 \dots \qquad \sum_{n=0}^{\infty} p_n = 1$$

Se define la función generatriz de la distribución definida por $\{p_n\}$ mediante

$$g(z) = \sum_{n=0}^{\infty} p_n z^n = \mathrm{E}(z^X). \tag{6.24}$$

(a) Probar que la serie (6.40) converge, por lo menos, cuando $|z| \leq 1$, es decir que su radio de convergencia es mayor o igual que 1.

(b) Observar que la función $g(z)$ determina unívocamente las probabilidades $\{p_n\}$.

(c) Calcular la función generatriz de la distribución binomial con parámetros n y p.

(d) Calcular la función generatriz de la distribución geométrica.

(e) Calcular la función generatriz de la distribución de Poisson de parámetro λ.

(f) Probar que

$$\mathrm{E}(X) = g'(1), \qquad \mathrm{E}(X(X-1)) = g''(1), \dots$$
$$\mathrm{E}(X(X-1)\dots(X-k+1)) = g^{(k)}(1).$$

bajo la hipótesis de que las esperanzas que allí figuran son finitas. ¿Qué ocurre si las esperanzas son infinitas?

(g) Sea g la función generatriz de la variable aleatoria X y h la función generatriz de la variable aleatoria Y. Probar que si X e Y son independientes, entonces $g(z)h(z)$ es la función generatriz de $X + Y$.

(h) Utilizando funciones generatrices, mostrar que $X_1, X_2, \dots, X_n$ son variables aleatorias independientes, cada una con distribución de Poisson, con parámetros respectivos $\lambda_1, \lambda_2, \dots, \lambda_n$, entonces $X_1 + X_2 + \cdots + X_n$ también tiene distribución de Poisson y su parámetro es $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

(i) Hallar la distribución de probabilidad de $X_1 + X_2 + \cdots + X_n$ donde las X_i, $i = 1, \dots, m$ son variables aleatorias independientes, cada una con distribución binomial, y los parámetros respectivos son (n_i, p), $i = 1, \dots, m$.

Apéndice

6.7.1. Definición del Valor Esperado en el Caso General.

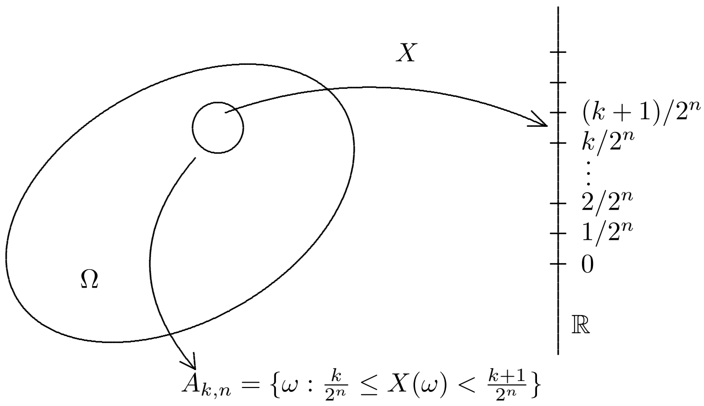
Hasta el momento hemos definido y calculado el *valor esperado* o *esperanza matemática* de una variable aleatoria, solamente en el caso en el que la distribución de probabilidad de la misma es discreta. A continuación nos proponemos hacer lo mismo en el caso general, es decir, para variables no necesariamente discretas. El estudio del tema requiere una extensión bastante mayor de la que estamos en condiciones de asignarle aquí. Aún así, el lector que no esté especialmente interesado, en una primera instancia, en contar con una fundamentación sobre el tema, puede pasar directamente a la sección 6.5 en la cual se estudia el caso de las variables aleatorias que tienen densidad, y adoptar la fórmula (6.21) como una definición. Así mismo, en una primera instancia, puede saltar las demostraciones de las propiedades básicas de la esperanza matemática en el caso general, que están marcadas con una estrella $\star$.

Como criterio general, la manera de definir la esperanza matemática y sus propiedades, consiste en aproximar una variable aleatoria cualquiera por variables discretas, y utilizar la definición y propiedades de la esperanza para estas últimas, que ya conocemos de nuestro estudio anterior. Este proceso de aproximación tiene algunas dificultades, que aparecen en las secciones mencionadas, marcadas también con una estrella $\star$.

Sea X una variable aleatoria no negativa. Para cada número natural n, $n = 1, 2, \dots$ definimos la variable aleatoria aproximante X_n mediante

$$X_n(\omega) = \frac{k}{2^n} \quad \text{si} \quad \frac{k}{2^n} \le X(\omega) < \frac{k+1}{n} \quad (k = 0, 1, 2, \dots) \tag{6.25}$$

Para entender esta definición procedemos de la siguiente forma: dividimos la semirecta $[0, \infty)$ de los números reales no negativos, en intervalos de logitud $1/2^n$, a partir de 0.



Para cada intervalo

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

consideramos la preimagen del intervalo por la variable aleatoria X, recordando que X es una función de Ω en $\mathbb{R}$. Dicha preimagen es

$$A_{k,n} = \left\{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\right\},$$

es decir, el conjunto de puntos de Ω cuya imagen por X está en

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right).$$

Entonces, definimos X_n en $A_{k,n}$ igual a $k/2^n$. Por lo tanto, una manera de escribir X_n es

$$X_n(\omega) = \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{A_{k,n}}(\omega) = \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{\{\omega: \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\}}(\omega)$$

donde, como es habitual, la función indicatriz $\mathbf{1}_{A_{k,n}}(\omega)$ está definida por

$$\mathbf{1}_{A_{k,n}}(\omega) = \begin{cases} 1 & \text{si } \omega \in A_{k,n}, \\ 0 & \text{si } \omega \notin A_{k,n}. \end{cases}$$

Las siguientes observaciones son consecuencia de la definición de X_n:

1. Para cada n, X_n es una variable aleatoria discreta y no negativa. Esto es claro ya que X_n sólo puede tomar los valores
$$0, \frac{1}{2^n}, \frac{2}{2^n}, \frac{3}{2^n}, \dots, \frac{k}{2^n}, \dots$$
que es un conjunto numerable de números mayores o iguales que cero.

2. Para cada $\omega \in \Omega$, $X_n(\omega)$ es una sucesión creciente (en sentido amplio), de números reales.

   En efecto, observamos que si aumentamos n en una unidad, es decir, en lugar de n ponemos $n+1$, lo que hacemos es subdividir cada uno de los intervalos
$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$$
en dos partes de igual longitud:
$$I_1 = \left[\frac{2k+1}{2^{n+1}}, \frac{k+1}{2^n}\right) \qquad I_2 = \left[\frac{k}{2^n}, \frac{2k+1}{2^{n+1}}\right).$$
Si tenemos que $X_n(\omega) = k/2^n$, lo que quiere decir que $\omega \in A_{k,n}$, o aún, de manera equivalente, que
$$\frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n},$$
tenemos dos alternativas posibles (ver figura 6.2):

   En primer lugar, que $X(\omega) \in I_1$, lo que implica que
$$X_{n+1}(\omega) = \frac{2k+1}{2^{n+1}} > \frac{k}{2^n} = X_n(\omega).$$
En segundo lugar, que $X(\omega) \in I_2$, lo que implica que
$$X_{n+1}(\omega) = \frac{k}{2^n} = X_n(\omega).$$

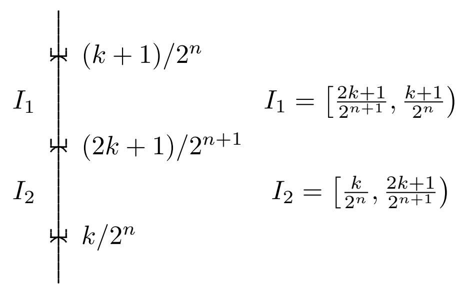


Figura 6.2

En cualquiera de los dos casos se verifica que $X_{n+1}(\omega) \geq X_n(\omega)$.

3. Se tiene
$$0 \leq X(\omega) - X_n(\omega) < \frac{1}{2^n} \qquad \text{para todo } \omega \in \Omega. \tag{6.26}$$
De aquí se deduce que
$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \text{para todo } \omega \in \Omega.$$

4. Puesto que X_n es una variable aleatoria discreta y no-negativa, está definida su esperanza matemática
$$\begin{aligned} \mathrm{E}(X_n) &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(X_n = \frac{k}{2^n}) = \sum_{k=0}^{\infty} \frac{k}{2^n} P(A_{k,n}) \\ &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}). \end{aligned} \tag{6.27}$$
Observamos que, cuando definimos la esperanza matemática de una variable aleatoria discreta, dijimos que ésta existe cuando la serie que aparece en la fórmula (6.18) es convergente. En caso contrario, y para simplificar la exposición que sigue, vamos a convenir en que $\mathrm{E}(X_n)$ es $+\infty$. Esta convención es útil a los fines de la presentación del tema.

Estamos ahora en condiciones de definir, de manera general, la esperanza matemática de una variable aleatoria. En una primera instancia consideremos una variable no-negativa X, y la sucesión aproximante X_n definida en (6.25). Tenemos dos alternativas posibles:

- O bien, para algún valor de n, $\mathrm{E}(X_n) = +\infty$, en cuyo caso definimos $\mathrm{E}(X) = +\infty$.
- O bien, para todo $n = 1, 2, \ldots$, $\mathrm{E}(X_n) < \infty$. En este caso, la sucesión de números reales no-negativos $\mathrm{E}(X_n)$ es monótona creciente, ya que, de acuerdo a la propiedad 8 de la sección 6.1, del hecho de que $X_n \leq X_{n+1}$ se deduce que $\mathrm{E}(X_n) \leq \mathrm{E}(X_{n+1})$. Pero, una sucesión creciente de números reales tiene límite, finito o infinito. Definimos entonces
$$\mathrm{E}(X) = \lim_{n\to\infty} \mathrm{E}(X_n) \tag{6.28}$$
y esto asigna un valor a la esperanza matemática $\mathrm{E}(X)$, para variables aleatorias no-negativas.

Consideremos ahora una variable aleatoria cualquiera X. Definimos otras dos variables X^+ y X^-, denominadas respectivamente, la parte positiva y la parte negativa de X, de la siguiente forma:
$$X^+(\omega) = \begin{cases} X(\omega) & \text{si } X(\omega) \geq 0, \\ 0 & \text{si } X(\omega) < 0. \end{cases}$$
$$X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geq 0, \\ -X(\omega) & \text{si } X(\omega) < 0. \end{cases}$$
Se verifica fácilmente que
$$\begin{aligned} X(\omega) &= X^+(\omega) - X^-(\omega) \\ |X(\omega)| &= X^+(\omega) + X^-(\omega). \end{aligned}$$

En efecto, si $X(\omega) \geq 0$, de acuerdo a las definiciones

$$X(\omega) = X^+(\omega) \qquad \text{y} \qquad |X(\omega)| = X^+(\omega),$$

mientras que si $X(\omega) < 0$,

$$X(\omega) = -X^-(\omega) \qquad \text{y} \qquad |X(\omega)| = X^-(\omega),$$

y en ambos casos se satisfacen ambas igualdades.

Ahora bien, X^+ y X^- son ambas no-negativas y podemos definir $\mathrm{E}(X^+)$ y $\mathrm{E}(X^-)$ utilizando (6.19). Basándonos en esto, definimos $\mathrm{E}(X)$ de la siguiente manera:

1. Si $\mathrm{E}(X^+) < \infty$ y $\mathrm{E}(X^-) < \infty$ definimos $\mathrm{E}(X) = E(X^+) - E(X^-)$.

2. Si $\mathrm{E}(X^+) = +\infty$ y $\mathrm{E}(X^-) < \infty$ definimos $\mathrm{E}(X) = +\infty$.

3. Si $\mathrm{E}(X^+) < \infty$ y $\mathrm{E}(X^-) = +\infty$ definimos $\mathrm{E}(X) = -\infty$.

4. Si $\mathrm{E}(X^+) = +\infty$ y $\mathrm{E}(X^-) = +\infty$ diremos que el valor esperado de la variable aleatoria X no está definido.

Lo primero que verificaremos es que nuestra nueva definición coincide con la anterior (sección 6.1) en el caso de variables aleatorias discretas. Para ello, basta con hacerlo para variables aleatorias que son no-negativas ya que, en caso contrario, podemos considerar por separado X^+ y X^-, que si lo son, y luego sumar.

Sea entonces X una variable aleatoria discreta no-negativa, es decir que

$$P(X = x_k) = p_k, \qquad x_k \geq 0, \quad k = 1, 2, \dots \qquad \sum_{k=1}^{\infty} p_k = 1.$$

Vamos a aplicarle a esta variable aleatoria X nuestra nueva definición. Supongamos primero que

$$\sum_{k=1}^{\infty} x_k p_k < \infty.$$

Para cada n, $n = 1, 2, \dots$ definimos la variable aleatoria aproximante X_n correspondiente a X de acuerdo a (6.16). Sabemos, por la desigualdad (6.17), que

$$0 \leq X - X_n < \frac{1}{2^n}$$

y puesto que X y X_n son variables aleatorias discretas, aplicando la propiedad 8 de la sección 6.1 para valores esperados de variables aleatorias discretas, se tiene

$$0 \leq \sum_{k=1}^{\infty} x_k p_k - \mathrm{E}(X_n) \leq \frac{1}{2^n}$$

Haciendo tender n a ∞, el tercer miembro tiende a cero y $\mathrm{E}(X_n)$ converge a $\mathrm{E}(X)$ (de acuerdo a nuestra nueva definición). Por lo tanto

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} x_k p_k,$$

o sea que las definiciones vieja y nueva de la esperanza matemática de X coinciden.

Si

$$\sum_{k=1}^{\infty} x_k p_k$$

es una serie divergente, dejamos como ejercicio para el lector probar que $E(X) = +\infty$, de acuerdo a la nueva definición. (Sugerencia: muestre que, en este caso, $E(X_n) = +\infty$ para cualquier valor de n).

6.7.2. Esperanza Matemática de Variables Aleatorias con Densidad.

Teorema 6.1 *Sea X una variable aleatoria, F_X su función de distribución y supongamos que ésta posee una densidad f_X. El valor esperado de X existe y es finito si y sólo si la integral*

$$\int_{-\infty}^{\infty} |x| f_X(x) dx \tag{6.29}$$

es finita, y en este caso $\mathrm{E}(X)$ *se calcula mediante la fórmula*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{6.30}$$

La validez de la fórmula anterior es completamente general cuando la integral en (6.16) es convergente. A continuación procedemos a probar la fórmula (6.17) cuando la función f_X es bastante regular como para que las integrales que aquí aparecen tengan sentido para nosotros, es decir, que puedan ser manejadas con las herramientas usuales del Cálculo. Por ejemplo, ese es el caso si f_X es una función continua, salvo a lo sumo en un número finito de puntos. El lector no especialmente interesado puede saltarse la demostración de (6.21) en una primera lectura. De todos modos, la aplicación de (6.21) es directa, como veremos en los ejemplos, cuando f_X es una función conocida.

Demostración del teorema 6.1

Consideremos X^+ y X^-, la parte positiva y la parte negativa de la variable aleatoria X y sus variables aleatorias aproximantes X_n^+ y X_n^-. Tenemos

$$\begin{aligned}
\mathrm{E}(X^+) &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X^+ < \frac{k+1}{2^n}) = \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) \\
&= \sum_{k=0}^{\infty} \frac{k}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx = \sum_{k=0}^{\infty} \left[\int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} x f_X(x)\, dx + \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx \right]
\end{aligned}$$

Observe el lector que hemos tenido en cuenta que, dado que f_X es la densidad de X,

$$P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) = \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx.$$

Por lo tanto,

$$\mathrm{E}(X^+) - \int_0^{+\infty} x\, f_X(x)\, dx = \sum_{k=0}^{\infty} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx$$

y se deduce que

$$\begin{aligned}
\left| \mathrm{E}(X^+) - \int_0^{+\infty} x\, f_X(x)\, dx \right| &\leq \sum_{k=0}^{\infty} \frac{1}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx \\
&= \frac{1}{2^n} \int_0^{+\infty} f_X(x)\, dx \leq \frac{1}{2^n}
\end{aligned}$$

que converge a cero cuando n tiende a $+\infty$.

En el razonamiento anterior hemos utilizado la desigualdad triangular, la acotación

$$\left| \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx \right| \leq \frac{1}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx,$$

y el hecho de que

$$\int_0^{+\infty} f_X(x)\,dx \leq \int_{-\infty}^{+\infty} f_X(x)\,dx = 1.$$

Se concluye entonces que

$$E(X^+) = \lim_{n\to\infty} \mathrm{E}(X_n^+) = \int_0^{+\infty} x\,f_X(x)\,dx \tag{6.31}$$

Dejamos como ejercicio la verificación de este resultado si $\mathrm{E}(X^+) = +\infty$.

De manera enteramente similar a lo que hicimos para probar (6.31) se demuestra que

$$E(X^-) = \lim_{n\to\infty} \mathrm{E}(X_n^-) = \int_0^{+\infty} (-x)\,f_X(x)\,dx \tag{6.32}$$

De (6.31) y (6.32) resulta inmediatamente (6.17). Observamos además, que $\mathrm{E}(X^+)$ y $\mathrm{E}(X^-)$ son ambas finitas, y por lo tanto también lo es $\mathrm{E}(|X|)$, si y sólo si las dos integrales en (6.31) y (6.32) son finitas, es decir, si y sólo si

$$\int_{-\infty}^{+\infty} |x| f_X(x)\,dx < \infty$$

■

6.8. Cambio de Variables. Momentos de Orden Superior.

Sea $f(x_1,\dots,x_n)$ la densidad conjunta de las variables aleatorias $X_1,\dots,X_n$ y

$$g : \mathbb{R}^n \to \mathbb{R}$$

una función de n variables que tiene alguna regularidad. Consideremos la variable aleatoria

$$Y = g(X_1, X_2, \dots, X_n).$$

Entonces, el valor esperado de Y se puede calcular mediante la fórmula

$$\mathrm{E}(Y) = \int \cdots \int_{\mathbb{R}^n} g(x_1,\dots,x_n) f(x_1,\dots,x_n) dx_1 \cdots dx_n \tag{6.33}$$

siempre que la integral de la derecha sea absolutamente convergente, es decir que

$$\int \cdots \int_{\mathbb{R}^n} |g(x_1,\dots,x_n)| f(x_1,\dots,x_n) dx_1 \cdots dx_n < \infty$$

La integral en la fórmula (6.24) la indicaremos, a los efectos de abreviar la notación, por

$$\int_{\mathbb{R}^n} g(x) f(x) dx,$$

donde debe entenderse que x representa el vector $(x_1, x_2, \dots, x_n)$.

A continuación demostraremos la fórmula (6.33). Al igual que en algunas situaciones anteriores, el lector puede saltar esta demostración en una primera lectura y pasar directamente a los ejemplos de aplicación de la fórmula (6.33).

Demostración de la fórmula (6.24)

Basta considerar el caso en que $g \geq 0$, ya que si esto no ocurre, separamos el estudio para g^+ y para g^-, y luego sumamos, donde definimos

$$g^+(x) = \begin{cases} g(x) & \text{si } g(x) \geq 0 \\ 0 & \text{si } g(x) < 0. \end{cases} \qquad g^-(x) = \begin{cases} 0 & \text{si } g(x) \geq 0 \\ -g(x) & \text{si } g(x) < 0. \end{cases}$$

Sea entonces $g(x) \geq 0$ para todo $x \in \mathbb{R}^n$. Definimos (analogamente a lo hecho para variables aleatorias), para $m = 1, 2, \dots$

$$g_m(x) = \frac{k}{2^m} \quad \text{si } \frac{k}{2^m} \leq g(x) < \frac{k+1}{2^m} \quad k = 0, 1, \dots$$

Es claro que

$$g(x) - \frac{1}{2^m} \leq g_m(x) \leq g(x) \quad \text{para todo } x \in \mathbb{R}^n \tag{6.34}$$

y por lo tanto

$$g(X(\omega)) - \frac{1}{2^m} \leq g_m(X(\omega)) \leq g(X(\omega)),$$

donde

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

es nuestro vector aleatorio, cuyas coordenadas son las variables aleatorias dadas.

Tomando valores esperados (ver las propiedades en la sección siguiente), resulta

$$\mathrm{E}(g(X)) - \frac{1}{2^m} \leq \mathrm{E}(g_m(X)) \leq E(g(X)). \tag{6.35}$$

Ahora bien,

$$g_m(X) = g_m(X_1, X_2, \dots, X_n)$$

es una variable aleatoria discreta, que puede tomar sólo valores de la forma $k/2^m$ para $k = 0, 1, 2, \dots$, y por lo tanto su valor esperado se calcula directamente

$$\begin{aligned}
\mathrm{E}(g_m(X)) &= \sum_{k=0}^{\infty} \frac{k}{2^m} P(g_m(X) = \frac{k}{2^m}) \\
&= \sum_{k=0}^{\infty} \frac{k}{2^m} P(\frac{k}{2^m} \leq g_m(X) < \frac{k+1}{2^m}) \\
&= \sum_{k=0}^{\infty} \frac{k}{2^m} P\left(X \in g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)\right) \qquad (6.36)
\end{aligned}$$

donde hemos indicado, como es usual

$$g^{-1}(I) = \{x : x \in \mathbb{R}^n, g(x) \in I\}.$$

En consecuencia, (6.27) no es sino

$$\begin{aligned}
\mathrm{E}(g_m(X)) &= \sum_{k=0}^{\infty} \frac{k}{2^m} \int_{g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)} f(x)dx \\
&= \sum_{k=0}^{\infty} \int_{g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)} g_m(x)\, f(x)dx \\
&= \int_{\mathbb{R}^n} g_m(x) f(x)dx \qquad (6.37)
\end{aligned}$$

En virtud de (6.35), cuando $m \to \infty$, el primer miembro de (6.37) tiende a $E(g(X))$. En cuanto al último miembro, tenemos, haciendo intervenir (6.34),

$$\int_{\mathbb{R}^n} g(x)f(x)dx - \frac{1}{2^m} \leq \int_{\mathbb{R}^n} g_m(x)f(x)dx \leq \int_{\mathbb{R}^n} g(x)f(x)dx$$

y por lo tanto, cuando $m \to \infty$, resulta que el último miembro de (6.37) tiende a

$$\int_{\mathbb{R}^n} g(x)f(x)dx.$$

Resumiendo:

$$E(g(X)) = \int_{\mathbb{R}^n} g(x)f(x)dx,$$

que es justamente (6.33). ∎

Observación 6.7 Si el lector interesado observa con cuidado la demostración que hemos hecho de la fórmula (6.33), verá que algunas afirmaciones debieran merecer una mayor atención. Por lo pronto, en (6.35)

$$\left\{\omega : X(\omega) \in g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)\right\},$$

tiene que ser un evento en el espacio de probabilidad subyacente. Para ello, se requiere que la función $g : \mathbb{R}^n \to \mathbb{R}$ sea lo que se denomina una función "boreliana", lo cual significa que la preimagen por g de un intervalo en $\mathbb{R}$, es un conjunto de Borel en $\mathbb{R}^n$. Por cierto, todas las funciones corrientes que aparecen en el cálculo, son funciones borelianas, de modo que, en este contexto, esto no implica ninguna limitación efectiva.

En segundo término, tiene que estar bien definida la integral

$$\int_{g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)} f(x)dx$$

que figura en la fórmula (6.37).

En tercer término, la última igualdad que aparece en (6.37), dice que, si denotamos por

$$B_{k,m} = g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right) = \left\{x : x \in \mathbb{R}^n, g(x) \in \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right\}$$

entonces

$$\sum_{k=0}^{\infty} \int_{B_{k,m}} h(x)\,dx = \int_{\mathbb{R}^n} h(x)\,dx,$$

donde $h(x) = g_m(x)f(x)$, es una cierta función no-negativa.

Todas estas afirmaciones son correctas, pero requieren un mayor desarrollo, y la tercera de ellas debe ser probada. El lector puede consultar, si desea comprender con la profundidad adecuada estos temas que están más allá del nivel de esta exposición, los libros de M. Loéve, D. Billingsley y R. M. Dudley incluidos en la bibliografía.

6.9. Propiedades del Valor Esperado en el Caso General.

Tal como hemos mencionado anteriormente, con la definición general el valor esperado sigue teniendo todas las propiedades enunciadas y probadas en la sección 6.1 para el caso discreto. La mayor parte de ellas son de demostración muy sencilla, y el lector puede probarlas por si mismo, sin mayor dificultad. Las únicas dos que requieren una mayor atención son las propiedades 7 y 9, a saber:

Propiedad 7. Si X e Y son variables aleatorias que tienen valor esperado, entonces también existe el valor esperado de $X + Y$ y se tiene

$$\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y). \tag{6.38}$$

Propiedad 9. Si X e Y son variables aleatorias independientes con valor esperado, entonces existe $\mathrm{E}(XY)$ y

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y). \tag{6.39}$$

El lector que no esté especialmente interesado puede saltar las demostraciones de estas dos propiedades, en una primera lectura.

Demostración de la propiedad 7

Son obvias las desigualdades

$$(X+Y)^+ \leq |X+Y| \leq |X| + |Y|$$
$$(X+Y)^- \leq |X+Y| \leq |X| + |Y|$$

y como por hipótesis se cumplen

$$\mathrm{E}(|X|) < \infty, \qquad \mathrm{E}(|Y|) < \infty,$$

se deducen

$$\mathrm{E}((X+Y)^+) < \infty, \qquad \mathrm{E}((X+Y)^-) < \infty \Rightarrow E(X+Y) < \infty.$$

Probaremos la igualdad (6.38) cuando X e Y son no-negativas; cuando esto no ocurre, descomponemos Ω en los conjuntos donde $X,\ Y$ y $X+Y$ tienen signo constante, y aplicamos el resultado que vamos a demostrar. Los detalles quedan a cargo del lector.

Supongamos entonces, $X \geq 0,\ Y \geq 0$. Por lo tanto, también $X+Y \geq 0$ y usamos la notación de la sección 6.4, llamando

$$X_n,\ Y_n,\ (X+Y)_n$$

las variables aleatorias aproximantes de $X,\ Y,\ X+Y$ respectivamente. Sabemos que

$$X_n \leq X < X_n + \frac{1}{2^n}$$
$$Y_n \leq Y < Y_n + \frac{1}{2^n}$$
$$(X+Y)_n \leq X+Y < (X+Y)_n + \frac{1}{2^n}$$

Se deduce que

$$\begin{aligned}\frac{-2}{2^n} &= X - \frac{1}{2^n} + Y - \frac{1}{2^n} - (X+Y) < X_n + Y_n - (X+Y)_n \\ &< X + Y - (X+Y) - \frac{1}{2^n} = \frac{1}{2^n}\end{aligned}$$

es decir,

$$\frac{-2}{2^n} < X_n + Y_n - (X+Y)_n < \frac{1}{2^n} \tag{6.40}$$

Como $X_n,\ Y_n,\ (X+Y)_n$ son variables aleatorias discretas, sabemos que (6.33) implica que

$$\frac{-2}{2^n} < \mathrm{E}(X_n) + \mathrm{E}(Y_n) - \mathrm{E}((X+Y)_n) < \frac{1}{2^n}$$

Si ahora hacemos tender n a infinito, en virtud de la definición de la esperanza matemática en el caso general, se tiene que

$$\mathrm{E}(X_n) \to \mathrm{E}(X), \qquad \mathrm{E}(Y_n) \to \mathrm{E}(Y), \qquad \mathrm{E}((X+Y)_n) \to \mathrm{E}(X+Y)$$

y por lo tanto,

$$0 \leq \mathrm{E}(X) + \mathrm{E}(Y) - \mathrm{E}(X+Y) \leq 0,$$

lo cual prueba (6.38). ■

Para demostrar la propiedad 9 podemos seguir un método similar de aproximación por variables discretas, haciendo valer nuevamente el hecho de que ya sabemos que la propiedad es válida para el caso de las variables aleatorias discretas.

6.9.1. Demostración de (6.23)

Para demostrar (6.23), apliquemos la definición del valor esperado. Sea X_n la variable aleatoria aproximante de X definida en la sección 6.4. Tenemos

$$\begin{aligned} \mathrm{E}(X) &= \lim_{n\to\infty} \mathrm{E}(X_n) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \le X < \frac{k+1}{2^n}) \\ &= \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} \left[P(X \ge \frac{k}{2^n}) - P(X \ge \frac{k+1}{2^n}) \right]. \end{aligned}$$

Introducimos la notación

$$G(x) = 1 - F_X(x^-) = P(X \ge x)$$

y reemplazando, resulta

$$\mathrm{E}(X) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} \left[G(\frac{k}{2^n}) - G(\frac{k+1}{2^n}) \right].$$

Mas aún, si ponemos

$$p_k = G(\frac{k}{2^n}) - G(\frac{k+1}{2^n}), \qquad q_k = \sum_{j=k+1}^{\infty} p_j G(\frac{k+1}{2^n}).$$

entonces podemos aplicar la fórmula (6.21), de la forma siguiente:

$$\begin{aligned} \mathrm{E}(X_n) &= \sum_{k=0}^{\infty} \frac{k}{2^n} p_k = \frac{1}{2^n} \sum_{k=0}^{\infty} k p_k = \frac{1}{2^n} \sum_{k=0}^{\infty} q_k \\ &= \sum_{k=0}^{\infty} \frac{1}{2^n} G(\frac{k+1}{2^n}) \le \int_0^{+\infty} G(x)\, dx. \end{aligned} \tag{6.41}$$

Para comprender la última desigualdad, hay que tener en cuenta que la función G es decreciente (en sentido amplio), ya que F_X es creciente (en sentido amplio), y que estamos comparando una suma de Riemann calculada evaluando en el extremo derecho de cada intervalo, con la integral (ver figura 6.4).

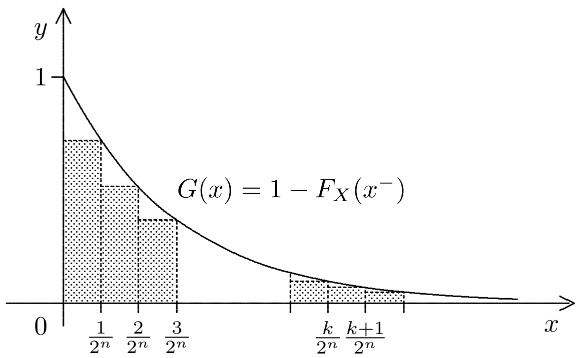


Figura 6.4

De (6.41) ya resulta que si $\mathrm{E}(X) = +\infty$, como el primer miembro tiende a $+\infty$ cuando $n \to \infty$, entonces

$$\int_0^{+\infty} G(x)\, dx \quad \text{diverge.}$$

Consideremos entonces el caso en el cual $\mathrm{E}(X) < \infty$. Pasando al límite cuando $n \to \infty$ en (6.41), se obtiene

$$\mathrm{E}(X) \leq \int_0^{+\infty} G(x)\,dx. \tag{6.42}$$

Por otro lado, si N es un número natural positivo cualquiera, la integral en el intervalo $[0, N]$ es el límite de las sumas de Riemann, cuando la partición se afina, y entonces

$$\begin{aligned}\int_0^N G(x)\,dx &= \lim_{n\to\infty} \sum_{k=0}^{N2^n-1} \frac{1}{2^n} G(\frac{k+1}{2^n}) \leq \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{1}{2^n} G(\frac{k+1}{2^n}) \\ &= \lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).\end{aligned}$$

Haciendo $N \to +\infty$, resulta

$$\int_0^{+\infty} G(x)\,dx \leq \mathrm{E}(X),$$

que conjuntamente con (6.42), implica (6.23). ■

Observación 6.8 En la fórmula (6.23), si en lugar de $1 - F_X(x^-)$ ponemos en el integrando la función $1 - F_X(x)$, el resultado es el mismo, es decir

$$\mathrm{E}(X) = \int_0^{+\infty} (1 - F_X(x))\,dx. \tag{6.43}$$

El lector puede verificar esto por su cuenta. También es posible ver que si F_X no tiene saltos entonces (6.23) y (6.43) son la misma cosa, pero si hay saltos, ambos integrandos no coinciden en todo punto.

Índice alfabético

Bibliografía

[1] K. Baclawski. *Introduction to Probability with R.* Chapman & Hall/CRC, Boca Raton, 2008.

[2] P. Brémaud. *Introduction aux Probabilités.* Springer, Berlin, 1984.

[3] K. L. Chung. *Elementary Probability Theory with Stochastic Processes.* Springer, New York, 3rd. edition, 1979.

[4] J. P. D'angelo y D. B. West. *Mathematical Thinking. Problem Solving and Proofs.* Prentice-Hall, Upper Saddle River, 1997.

[5] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 1. J. Wiley & Sons, New York, 1968.

[6] M. A. García Alvarez. *Introducción a la Teoría de la Probabilidad. Primer Curso.* Fondo de Cultura Económica, Mexico, 2005.

[7] B. Gnedenko y A. Khinchin. *An Elementary Introduction to the Theory of Probability.* Dover Publications, Inc., New York, 1962.

[8] H. Gordon. *Discrete Probability.* Springer, New York, 1997.

[9] G. Grimmett y D. Stirzaker. *One Thousand Exercises in Probability.* Oxford University Press, Oxford, 2001.

[10] G. Grimmett y D. Welsh. *Probability. An Introduction.* Oxford University Press, Oxford, 1986.

[11] R. Isaac. *The Pleasures of Probability.* Springer, New York, 1995.

[12] S. M. Ross. *Simulation.* Academic Press, San Diego, 3rd. edition, 2002.

[13] Y. A. Rozanov. *Probability Theory: A Concise Course.* Dover Publications, Inc., New York, 1969.

[14] D. Stirzaker. *Elementary Probability.* Cambridge. Cambridge Univ. Press, 1994.

[15] D. Stirzaker. *Probability and Random Variables. A Beginner's Guide.* Cambridge. Cambridge Univ. Press, 1999.