

Capítulo 1

Conceptos Básicos 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 con exactitud. Aún cuando la historia de la teoría de probabilidades tiene ya varios siglos, y muchos autores consideran que se inició con la correspondencia entre Blaise Pascal y Pierre de Fermat sobre juegos de azar en el siglo XVII, se puede decir que no fue hasta el siglo XX cuando esta teoría alcanzó un desarrollo notable.

Uno de los principales problemas por los cuales este desarrollo no ocurrió antes fue la ausencia de una axiomatización adecuada de las probabilidades, que le diese una base sólida y le permitiese desarrollarse al igual que otras ramas de la Matemática. En 1933, A. N. Kolmogorov propone una axiomatización usando las ideas de la Teoría de Medida, desarrollada a principios del siglo XX por H. Lebesgue. Esta axiomatización propone modelar los experimentos que tienen comportamiento aleatorio usando un espacio de medida. Aún cuando el desarrollo pleno de estas ideas está más allá del alcance del material que presentamos, vamos a considerar este enfoque axiomático en las próximas secciones, presentando numerosas aplicaciones de estas ideas.

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 la letra griega Ω el espacio muestral, y mediante ω los eventos elementales (o puntos de Ω).

Ejemplos 1.1

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 el 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 puede superar 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 más 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 al azar y se conecta cada uno de ellos hasta que se daña, observando en cada caso el tiempo de vida. 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 atmosférica y la 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 están 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 ejemplo 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 ejemplo 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{F} de eventos.

Definición 1.1 Si al realizar un experimento obtenemos como resultado el evento elemental ω , decimos que el evento $A \subset \Omega$ ha ocurrido si $\omega \in A$

Veamos qué condiciones debe cumplir la familia de eventos \mathcal{F} . En primer lugar

a. $\boxed{\Omega \in \mathcal{F}}$

es decir que al realizar el experimento el resultado es un elemento de Ω . A Ω lo llamaremos *evento cierto*.

b. $\boxed{A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}}$

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

Es decir, si A es un evento, pediremos que “no ocurre A ” también sea un evento.

Finalmente, la familia \mathcal{F} 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. $\boxed{A_n \in \mathcal{F} (n = 1, 2, \dots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}}$

Definición 1.2 Una familia \mathcal{F} 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 que las familias de eventos son σ -álgebras. Las siguientes propiedades son consecuencias inmediatas de la definición:

1. El conjunto vacío, \emptyset , es un evento, ya que $\emptyset = \Omega^c$.
2. $A_1, A_2, \dots, A_k \in \mathcal{F} \Rightarrow \bigcup_{n=1}^k A_n \in \mathcal{F}$. Basta considerar $A_{n+1} = A_{n+2} = \dots = \emptyset$ y aplicar la propiedad anterior y c.
3. $A_1, A_2, \dots, A_n, \dots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$. 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 1.2

1. 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 $\mathcal{P}(\Omega)$.

2. *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,

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

3. *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 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 1.3 Sean Ω un espacio muestral y \mathcal{F} una familia de eventos de Ω , es decir, una σ -álgebra de subconjuntos de Ω . Estamos interesados en asignar a cada evento $A \in \mathcal{F}$ un número real $P(A)$, que llamaremos la probabilidad de A , de modo tal que se cumplan las siguientes condiciones:

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

La probabilidad de un evento cualquiera es un número real no negativo.

2. $P(\Omega) = 1$

El evento cierto tiene probabilidad igual a 1.

3. Si $A_n \in \mathcal{F}$ 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

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

Definición 1.4 Una terna (Ω, \mathcal{F}, P) , formada por un espacio muestral Ω , una familia \mathcal{F} 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.

$$(1) A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

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

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

$$(2) P(\emptyset) = 0.$$

En efecto, consideremos $A_1 = \Omega$ y $A_2 = \emptyset$ Usando la propiedad anterior $P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$ y en consecuencia $P(\emptyset) = 0$.

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

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 $A_2 = A_1 + (A_2 \cap A_1^c)$ (ver figura 1.1) resulta $P(A_2) = P(A_1) + P(A_2 \cap A_1^c)$

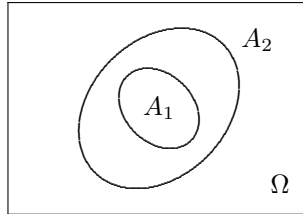


Figura 1.1: Diagrama de Venn para un conjunto incluido en otro

y en consecuencia $P(A_1) \leq P(A_2)$ ya que $P(A_2 \cap A_1^c) \geq 0$.

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

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

$$(6) A_1 \subset A_2 \subset \dots \subset A_n \subset \dots \Rightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Sean

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

entonces los B_i son disjuntos dos a dos y $A_n = \bigcup_{i=1}^n B_i$ (ver figura 1.2). Además

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

y entonces

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

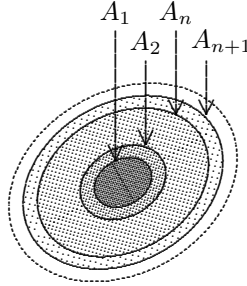


Figura 1.2: Diagrama de Venn para una sucesión creciente de conjuntos.

$$(7) A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \Rightarrow P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Como la sucesión $\{A_n^c\}$ es creciente, usando (6) obtenemos

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \\ &= 1 - \lim_{n \rightarrow \infty} P(A_n^c) = 1 - \lim_{n \rightarrow \infty} (1 - P(A_n)) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

$$(8) P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

En efecto, considerando que (ver figura 1.3)

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

después de aplicar la propiedad (2) a ambas igualdades y restar resulta la proposición (8).

$$(9) P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Para ver esto apliquemos (8) 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

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

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

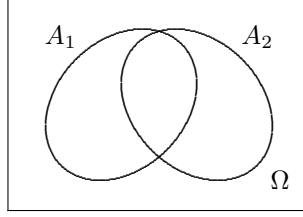


Figura 1.3: Diagrama de Venn para dos conjuntos con intersección no vacía.

$$(10) \quad P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i<j=2}^n P(A_i \cap A_j) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n).$$

Para demostrar esta proposición se procede por inducción completa en n siguiendo la misma idea que en la anterior, que corresponde al caso $n = 3$. Para $n = 2$ es la propiedad (8). Dejamos esto como ejercicio.

Observación 1.1 Consideremos la propiedad (10). Si en lugar de usar todos los términos en el lado derecho de la identidad usamos solo algunos, obtenemos una aproximación a la probabilidad $P(\bigcup_{i=1}^n A_i)$. Si la aproximación es por exceso o por defecto depende del signo del primer sumando no tomado en cuenta. Si éste es de signo negativo, la aproximación es por exceso mientras que si es positivo la aproximación es por defecto. Así,

$$\sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad (1.1)$$

y más generalmente, para cualquier a tal que $2a + 1 \leq n$,

$$\sum_{k=1}^{2a} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{k=1}^{2a+1} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \quad (1.2)$$

Estas se conocen como las desigualdades de Bonferroni. La primera de estas desigualdades, que corresponde al lado derecho de la ecuación (1.1), es conocida como la propiedad de subaditividad de las probabilidades, que es válida para cualquier colección de conjuntos A_1, \dots, A_n :

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

1.5. Casos Particulares

1.5.1. Probabilidades en Espacios Finitos.

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

$$\begin{cases} p_i \geq 0, & \text{para todo } i, \\ \sum_{i=1}^m p_i = 1. \end{cases}$$

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

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

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, 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 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. Las técnicas de teoría combinatoria elemental, que no revisaremos en estas notas, resultan útiles en estos casos.

Ejemplos 1.3

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.

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

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

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

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 7 es $1/6$.

1.5.2. Probabilidades en Espacios Numerables.

Un caso similar al desarrollado en la sección anterior se presenta cuando Ω es un conjunto infinito numerable:

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

donde los números p_i verifican

$$\begin{cases} p_i \geq 0, & \text{para todo } i, \\ \sum_{i=1}^{\infty} p_i = 1. \end{cases}$$

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.

Ejemplo 1.4

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 que 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:

$$1 + r + r^2 + r^3 + \dots = \frac{1}{1-r} \quad (1.3)$$

y multiplicando ambos lados por r obtenemos

$$r + r^2 + r^3 + r^4 + \dots = \frac{r}{1-r} \quad (1.4)$$

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 ‘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 ‘Aguila se obtiene por primera vez en un número par de lanzamientos’. Tenemos que $A = \{2, 4, 6, \dots\}$ y

$$P(A) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots = \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \dots$$

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. Probabilidades en Espacios No Numerables.

El caso de los espacios no numerables es más complicado y no lo vamos a poder considerar en todo detalle, pues carecemos de las herramientas matemáticas necesarias. Sin embargo, vamos a analizar un ejemplo sencillo que permite presentar las ideas y resaltar las dificultades que se presentan en este caso.

Consideremos $\Omega = [0, 1]$ y sea el experimento que consiste en escoger al azar un número en el intervalo $[0, 1]$. No podemos, en este caso, utilizar el procedimiento que usamos para los espacios numerables ya que en el caso continuo, una probabilidad uniforme no puede asignarle probabilidad positiva a un conjunto de un solo elemento, es decir, si $x \in [0, 1]$, necesariamente $P(\{x\}) = 0$ (ver ejercicio ?? en la sección ??).

En consecuencia, nos enfocamos en la probabilidad de que el número seleccionado al azar caiga dentro de un intervalo de valores y por uniformidad, la probabilidad de escoger un número en el intervalo $[c, d] \subset [0, 1]$ debe depender únicamente de la longitud del intervalo y no de su ubicación, es decir, debe ser invariante bajo traslaciones. No es difícil probar que una tal probabilidad P debe verificar

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

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

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

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

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

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

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

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 $k \rightarrow \infty$ resulta

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

Por la observación que hicimos al inicio de esta sección, tenemos que, en este caso,

$$P([0, x)) = P([0, x]) = P((0, x]) = P((0, x)).$$

Como consecuencia de estos resultados vemos que la probabilidad de cada intervalo es su longitud:

$$P([c, d]) = d - c, \quad \text{para todo } [c, d] \subset [0, 1]. \quad (1.6)$$

¿Qué ocurre si tenemos un subconjunto cualquiera $A \subset [0, 1]$ y queremos calcular $P(A)$? Lo que estamos planteando es cómo extender la noción de 'longitud' a cualquier subconjunto de $[0, 1]$. Desafortunadamente, no es posible definir una medida de probabilidad sobre todos los subconjuntos de $[0, 1]$ que satisfaga la propiedad (1.6). La demostración de este hecho está fuera de los objetivos de este curso, pero esto implica que hay conjuntos que no son 'medibles', es decir, a los cuales no podemos asignarles una probabilidad.

Por lo tanto, es necesario restringirse a una clase más pequeña \mathcal{B} de subconjuntos de $[0, 1]$, que sea una σ -álgebra, es decir, que satisfaga las condiciones de la definición 1.2 y contenga a los subintervalos de $[0, 1]$. Una posibilidad es usar la clase de los conjuntos borelianos en $[0, 1]$, que es la menor σ -álgebra generada por los subintervalos de $[0, 1]$. En la próxima sección nos detendremos en este punto. Sin embargo, es importante observar que en este caso hay otras σ -álgebras que pueden considerarse.

1.6. σ -Algebras

Vamos a considerar en mayor detalle el problema que presentamos en la sección anterior. Sea Ω un espacio muestral y consideremos el conjunto de partes de Ω , $\mathcal{P}(\Omega)$, que contiene todos los subconjuntos de Ω y que, como hemos visto anteriormente, es una σ -álgebra. De hecho, es la mayor σ -álgebra de subconjuntos de Ω en el sentido de que cualquier otra σ -álgebra \mathcal{F} está contenida en el conjunto de partes: $\mathcal{F} \subset \mathcal{P}(\Omega)$.

En la próxima proposición mostramos que la intersección de cualquier colección de σ -álgebras es también una σ -álgebra.

Proposición 1.1 *Sea T un conjunto de índices cualquiera y sea $\{\mathcal{F}_t, t \in T\}$ una colección de σ -álgebras de subconjuntos del espacio muestral Ω . Sea \mathcal{G} la intersección de esta colección:*

$$\mathcal{G} = \bigcap_{t \in T} \mathcal{F}_t$$

Entonces \mathcal{G} es una σ -álgebra.

Demostración. Tenemos que demostrar que \mathcal{G} satisface las tres propiedades de la definición 1.2. En primer lugar, $\Omega \in \mathcal{F}_t$ para todo t , pues \mathcal{F}_t es una σ -álgebra, y en consecuencia

$$\Omega \in \bigcap_{t \in T} \mathcal{F}_t = \mathcal{G}.$$

En segundo lugar consideremos $A \in \mathcal{G}$, entonces para todo $t \in T$ se tiene que $A \in \mathcal{F}_t$ y por la propiedad b de la definición 1.2 se tiene que $A^c \in \mathcal{F}_t$ para todo t . En consecuencia

$$A \in \mathcal{G} \Rightarrow A^c \in \mathcal{G}.$$

Finalmente, si $A_n, n \geq 1$ es una colección de eventos en \mathcal{G} , un argumento similar muestra que $\bigcup_{n \geq 1} A_n \in \mathcal{F}_t$ para todo $t \in T$. En consecuencia $\bigcup_{n \geq 1} A_n$ va a estar en la intersección de todas estas σ -álgebras, es decir,

$$\bigcup_{n \geq 1} A_n \in \bigcap_{t \in T} \mathcal{F}_t = \mathcal{G}.$$

Con este resultado podemos introducir la siguiente definición. ■

Corolario 1.1 Sea \mathcal{C} una colección de subconjuntos de Ω . Existe una σ -álgebra, que denotamos por $\sigma(\mathcal{C})$, que contiene a \mathcal{C} y es la menor entre todas las σ -álgebras que contienen a \mathcal{C} en el siguiente sentido: si $\mathcal{C} \subset \mathcal{F}$ entonces $\sigma(\mathcal{C}) \subset \mathcal{F}$.

Demostración. Consideremos la colección de todas las σ -álgebras de subconjuntos de Ω que contienen a \mathcal{C} . Esta clase no es vacía pues ya vimos que $\mathcal{P}(\Omega)$ está en ella. $\sigma(\mathcal{C})$ es la intersección de todas las σ -álgebras en esta colección. La proposición anterior demuestra que $\sigma(\mathcal{C})$ es una σ -álgebra y al ser la intersección de todas las que contienen a \mathcal{C} , contiene a \mathcal{C} y es la menor de todas. ■

Definición 1.5 La σ -álgebra $\sigma(\mathcal{C})$ se conoce como la σ -álgebra generada por \mathcal{C} .

Veamos algunos ejemplos.

Ejemplo 1.5

Sea $\Omega = \{1, 2, 3, 4, 5, 6\}$ y $\mathcal{C} = \{\{3\}\}$. Es fácil verificar que la σ -álgebra generada por \mathcal{C} está dada por

$$\sigma(\mathcal{C}) = \{\Omega, \emptyset, \{3\}, \{1, 2, 4, 5, 6\}\}.$$

En efecto, toda σ -álgebra debe incluir como elementos a Ω y \emptyset , por las dos primeras condiciones de la definición 1.2, también tiene que contener a \mathcal{C} , es decir, debe tener como elemento al conjunto $\{3\}$ y por la segunda propiedad de la definición también debe contener a su complemento, $\{1, 2, 4, 5, 6\}$. Solo falta verificar que esta colección de conjuntos satisface las condiciones de una σ -álgebra, lo cual es sencillo.

Ejemplo 1.6

En el mismo contexto del ejemplo anterior, si tenemos $\mathcal{C} = \{\{3\}, \{4\}\}$ entonces

$$\sigma(\mathcal{C}) = \{\Omega, \emptyset, \{3\}, \{4\}, \{3, 4\}, \{1, 2, 4, 5, 6\}, \{1, 2, 3, 5, 6\}, \{1, 2, 5, 6\}\}.$$

Un argumento similar al del ejemplo anterior muestra que ésta es la menor σ -álgebra que contiene a \mathcal{C} .

Ejemplo 1.7 (La σ -álgebra de Borel)

Esta σ -álgebra tiene una importancia central en la teoría de la medida y se define como la menor σ -álgebra que contiene a los subintervalos de $[0, 1]$. Usualmente se denota por \mathcal{B} o $\mathcal{B}[0, 1]$ si queremos resaltar que solo estamos considerando subconjuntos de $[0, 1]$.

Hay varias definiciones alternativas de esta σ -álgebra. Por ejemplo, también se puede definir como la menor σ -álgebra que contiene a los intervalos abiertos (a, b) , $0 \leq a < b \leq 1$ o la menor σ -álgebra que contiene a los intervalos cerrados $[a, b]$, $0 \leq a \leq b \leq 1$. También es la menor σ -álgebra que contiene a los intervalos de la forma $(a, b]$ o a los de la forma $[a, b)$. No es difícil demostrar que todas estas definiciones son equivalentes.

También es posible definir la σ -álgebra de Borel como la σ -álgebra generada por los subconjuntos abiertos de $[0, 1]$, y se puede demostrar que esta definición es equivalente a cualquiera de las anteriores. Esta última definición tiene la ventaja de que podemos usarla en cualquier espacio que tenga una topología, por ejemplo, en cualquier espacio métrico.

Las definiciones anteriores pueden extenderse sin ninguna dificultad a los subconjuntos de la recta real \mathbb{R} .

1.7. Ejemplos y Aplicaciones.

Ejemplo 1.8 (Muestreo con Reposición.)

Retomemos el ejemplo 1.2.2 sobre el muestreo con reposición, donde

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

y sea $\mathcal{F} = \mathcal{P}(\Omega)$. 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

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

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

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

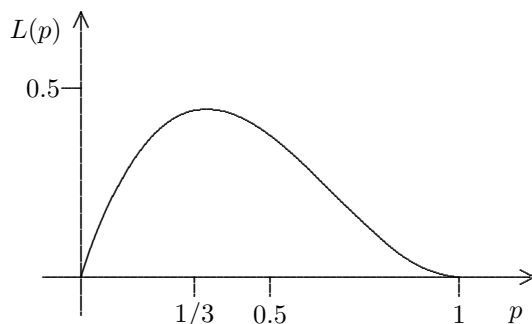


Figura 1.4: Gráfica de la función $3p(1-p)^2$ para $0 \leq p \leq 1$.

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

El gráfico de la función $L(p)$ se muestra en la figura 1.4, 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.

Ejemplo 1.9 (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.

Determinemos ahora la probabilidad del evento A : “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

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

y su probabilidad es

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

1.8. Probabilidad Condicional

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

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

Veamos ahora el siguiente problema: Si sabemos 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 1.6 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y sea $B \in \mathcal{F}$ tal que $P(B) > 0$. Definimos una nueva función $P(\cdot|B)$ de \mathcal{F} en \mathbb{R} mediante

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

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

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

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

3. Sean A_1, A_2, \dots conjuntos disjuntos en \mathcal{F} , entonces

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

$P(\cdot|B)$ 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 1.10

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 algún 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

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

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 , $1 \leq k \leq 6$?
 - b) Si el primer lanzamiento resultó en 3, ¿cuál es la probabilidad de que el segundo sea k , $1 \leq k \leq 6$?
 - 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 y Y el del segundo. Sabemos que $P(X = k) = P(Y = k) = 1/6$, $1 \leq k \leq 6$.
- 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 $2 \leq k \leq 6$ 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

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

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 seleccionada al azar 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

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

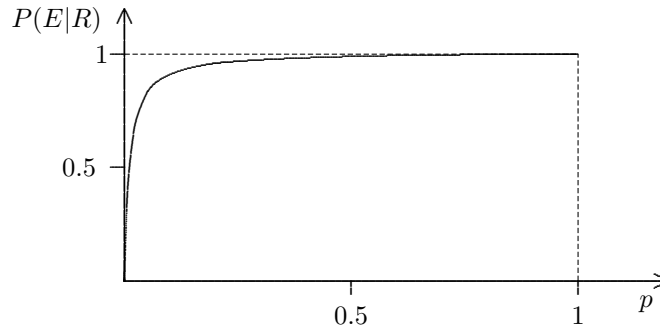


Figura 1.5: Gráfica de $P(E|R)$ como función de p .

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 1.5) 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 $P(E|R) \simeq 0.083$.

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. **El Tiempo de Vida.** 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

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

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

$$\varphi(s + h) = \varphi(s)\varphi(h) \tag{1.7}$$

para cualquier par de números positivos s y h . Esta ecuación funcional se conoce como la *ecuación de Hamel*. Si φ no es idénticamente nula veamos que

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

- De (1.7), 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

$$\varphi(1/n) = (\varphi(1))^{1/n}$$

poniendo ahora $s = 1/m$, se tiene

$$\varphi(n/m) = (\varphi(1/m))^n = (\varphi(1))^{n/m},$$

es decir que

$$\varphi(t) = (\varphi(1))^t \quad \text{si } t \in \mathbb{Q}, t \geq 0.$$

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

$$t_k < t < t'_k, \quad t_k \rightarrow t, \quad t'_k \rightarrow t,$$

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 $k \rightarrow \infty$ 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.

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

$$b) P(T \geq 2000 | T \geq 1000) = \frac{P((T \geq 2000) \cap (T \geq 1000))}{P(T \geq 1000)} = \frac{P(T \geq 2000)}{P(T \geq 1000)} = \frac{e^{-2}}{e^{-1}} = e^{-1} \approx 0.367879.$$

▲

1.9. Resultados Básicos sobre Probabilidad Condicional.

Estudiaremos en esta sección algunas propiedades sencillas pero fundamentales de la probabilidad condicional.

Proposición 1.2 *Para cualquier colección finita de eventos A_1, \dots, A_n se tiene*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

siempre que $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$.

Demostración. Como

$$P(A_1) \geq P(A_1 \cap A_2) \geq \dots \geq P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$$

todas las probabilidades condicionales que aparecen están bien definidas. Si escribimos explícitamente el segundo miembro de la igualdad obtenemos

$$P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})}$$

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 1.3 (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.

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(A \cap (\cup_i B_i)) \\ &= P(\cup_i (A \cap B_i)) = \sum_i P(A \cap B_i) \\ &= \sum_i P(B_i)P(A|B_i). \end{aligned}$$

■

Ejemplos 1.11

1. 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

$$\begin{aligned} P(A) &= \frac{(m-1)m}{(n-1)n} + \frac{m}{(n-1)} \frac{(n-m)}{n} \\ &= \frac{m}{(n-1)n} (m-1 + n-m) = \frac{m}{n}. \end{aligned}$$

▲

2. 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 la probabilidad de un cliente le dé propina es 0.5, 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 $p = 1/2$. Las probabilidades condicionales son:

$$P(A|B_1) = 0, \quad P(A|B_2) = \left(\frac{1}{2}\right)^2,$$

$$\begin{aligned} P(A|B_3) &= P(2 \text{ éxitos en 3 intentos}) + P(3 \text{ éxitos en 3 intentos}) \\ &= \binom{3}{2} \frac{1}{2^3} + \binom{3}{3} \frac{1}{2^3} = \frac{1}{2}, \end{aligned}$$

$$P(A|B_4) = \left[\binom{4}{2} + \binom{4}{3} + \binom{4}{4} \right] \frac{1}{2^4},$$

$$P(A|B_5) = \left[\binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} \right] \frac{1}{2^5},$$

y finalmente obtenemos $P(A) = \frac{9}{20}$. ▲

3. ¿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:

$$\begin{aligned} P(E_1) &= 1 \\ P(E_2|E_1) &= 5/6 \\ P(E_3|E_1 \cap E_2) &= 4/6 \\ &\vdots \\ P(E_6|E_1 \cap E_2 \cap \dots \cap E_5) &= 1/6 \end{aligned}$$

y usando la proposición 3.1

$$P(E_1 \cap E_2 \cap \dots \cap E_6) = 1 \times \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} = \frac{6!}{6^6}.$$

4. **El juego de dados.** El juego de dados que se conoce en inglés como *craps* tiene las siguientes reglas. El jugador lanza un par de dados, si el resultado es 7 u 11, gana. Si es 2, 3 ó 12, pierde. Si sale cualquier otro número, ese número se convierte en su *punto* y a partir de ese momento el jugado continua lanzando hasta obtener el punto, en cuyo caso gana, o 7, en cuyo caso pierde. ¿Cuál es la probabilidad de ganar en este juego?

► Llamemos A_j a obtener j en el primer lanzamiento, con $j = 2, 3, \dots, 12$ y G indica el evento de ganar el juego. Por la ley de probabilidad total tenemos

$$P(G) = \sum_{j=2}^{12} P(G|A_j)P(A_j) \tag{1.8}$$

Ya sabemos que si $j = 2, 3$ ó 12 , el jugador pierde, por lo que los términos correspondientes en la suma valen 0. Si $j = 7$ u 11 , el jugador gana, por lo que $P(G|A_j) = 1$ en estos dos casos. Para el resto de los resultados tenemos que calcular la probabilidad condicional $P(G|A_j)$, es decir, la probabilidad de ganar si el punto es j , para $j = 4, 5, 6, 8, 9, 10$.

En ejemplo 1.10.1 resolvimos un problema similar. tenemos que la probabilidad de obtener j antes que 7 si el punto es j , que es lo mismo que $P(G|A_j)$, es

$$P(G|A_j) = P(\{j\}|\{j, 7\}) = \frac{P(\{j\})}{P(\{j, 7\})}.$$

Usando esta fórmula obtenemos

$$P(G|A_4) = P(G|A_{10}) = \frac{1}{3}; \quad P(G|A_5) = P(G|A_9) = \frac{2}{5}; \quad P(G|A_6) = P(G|A_8) = \frac{5}{11};$$

Regresando a la ecuación (1.8) obtenemos

$$\begin{aligned} P(G) &= \sum_{j=4}^{11} P(G|A_j)P(A_j) \\ &= \frac{1}{3} \times \frac{3}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{5}{11} \times \frac{5}{36} + \frac{6}{36} + \frac{5}{11} \times \frac{5}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{1}{3} \times \frac{3}{36} + \frac{2}{36} = 0.49293 \end{aligned}$$

1.10. El Teorema de Bayes

El último resultado fundamental sobre probabilidad condicional que incluimos se conoce como el Teorema de Bayes y es útil en situaciones en las cuales queremos invertir la probabilidad condicional: se conocen probabilidades de la forma $P(A|B_i)$ y $P(B_i)$ y se desea determinar $P(B_i|A)$.

Proposición 1.4 (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:

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} \quad (1.9)$$

y

$$P(A \cap B_i) = P(A|B_i)P(B_i). \quad (1.10)$$

Por la proposición 3.2 tenemos

$$P(A) = \sum_j P(A|B_j)P(B_j). \quad (1.11)$$

Usando (1.10) y (1.11) en (1.9) obtenemos el resultado. ■

Ejemplos 1.12

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,

$$\begin{aligned} P(T) &= \frac{10}{100} = 0.1; & P(T^c) &= \frac{90}{100} = 0.9; \\ P(C|T) &= 0.75; & P(C|T^c) &= 0.50. \end{aligned}$$

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 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:

$$\begin{aligned} P(A) &= 0.01, & P(B) &= 0.005, & P(C) &= 0.02 \\ P(H|A) &= 0.9, & P(H|B) &= 0.95, & P(H|C) &= 0.75 \end{aligned}$$

Usando el Teorema de Bayes tenemos

$$\begin{aligned} P(A|H) &= \frac{P(H|A)P(A)}{P(H|A)P(A) + P(H|B)P(B) + P(H|C)P(C)} \\ &= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.95 \times 0.005 + 0.75 \times 0.02} = \frac{9}{28.75} = 0.313. \end{aligned}$$

▲

3. 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

$$P(C) = 0.8; \quad P(A|C^c) = 0.25; \quad P(A|C) = 1.$$

Usando el Teorema de Bayes

$$\begin{aligned} P(C|A) &= \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)} \\ &= \frac{0.8}{0.8 + (0.25)(0.2)} = \frac{0.8}{0.85} = 0.941. \end{aligned}$$

▲

1.11. Eventos Independientes

En los ejemplos que hemos considerado en relación al concepto de probabilidad condicional, hemos visto que en ciertos casos, cuando saber que ha ocurrido el 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:

$$P(A \cap B) = P(A)P(B). \tag{1.12}$$

Observamos además que (1.12) es válida aún cuando A o B sean sucesos de probabilidad nula. Usaremos esta relación como definición.

Definición 1.7 Dos eventos A y B son independientes si vale (1.12).

Es fácil verificar que Ω y \emptyset son independientes de cualquier evento.

Proposición 1.5 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 1.13

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 1.8 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:

$$P\left(\bigcap_{j=1}^n A_{i_j}\right) = \prod_{j=1}^n P(A_{i_j}). \quad (1.13)$$

En este caso decimos también que \mathcal{C} es una familia de eventos independientes.

Observemos que en la definición *solo* 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 (1.13) para las parejas de eventos. En efecto, sea el experimento de lanzar dos dados y $\mathcal{C} = \{A_1, A_2, A_3\}$ con

A_1 : “se obtiene un 6 en el primer dado”.

A_2 : “se obtiene un 1 en el segundo dado”.

A_3 : “la suma de los dos dados es 7”.

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

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{36} \neq \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = P(A_1)P(A_2)P(A_3),$$

de modo que los tres eventos no son independientes.

Ejemplos 1.14

1. 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 $P(A) = P(B) = 0.4$. ▲

2. ¿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

$$\binom{8}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

▲

3. 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

$$P(A) = \frac{4}{6}, \quad P(B) = \frac{4}{6} \quad \text{y} \quad P(A \cap B) = \left(\frac{4}{6}\right)^2,$$

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 1.11.1 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. ▲

4. 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

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

▲

5. *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 , ($r \leq n$) 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

$$\binom{n}{r} p^r (1 - p)^{n-r}, \quad 0 \leq r \leq n,$$

es decir, la distribución binomial que habíamos encontrado anteriormente. ▲

6. 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 $\mathcal{C} = \{A_i, i \geq 1\}$. \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 1.5). Entonces

$$\begin{aligned} p_n &= P(\text{el primero de tipo } A \text{ es el } n\text{-ésimo}) \\ &= P(A_1^c \cap \dots \cap A_{n-1}^c \cap A_n) = (1-p)^{n-1}p, \end{aligned}$$

que se conoce como la distribución geométrica.

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”.

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 1.6).

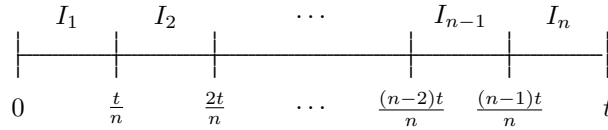


Figura 1.6: División del intervalo $[0, t]$ en n subintervalos iguales.

Si A_k es el evento “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, \dots, A_n son independientes, es decir que el hecho de que no ocurran desperfectos en los intervalos I_1, \dots, 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 $\frac{o(h)}{h} \rightarrow 0$ ($h \rightarrow 0$), resulta que

$$\begin{aligned} P(T > t) &= P(\bigcap_{k=1}^n A_k) = \prod_{k=1}^n P(A_k) \\ &= \left(1 - \frac{\lambda t}{n} + o\left(\frac{t}{n}\right)\right)^n \\ &\rightarrow e^{-\lambda t} \quad (n \rightarrow \infty) \end{aligned}$$

y volvemos a encontrar la distribución exponencial. ▲

7. 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

$$P(T > x) = e^{-\lambda x} \quad \text{para cierto } \lambda.$$

¿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 medimos 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 $L(\lambda) \rightarrow 0$ cuando $\lambda \rightarrow 0^+$ y $L(\lambda) \rightarrow 0$ cuando $\lambda \rightarrow \infty$, 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 $t_1 = t_2 = \dots = t_n = 0$ y esto sólo puede suceder con probabilidad 0. ▲

1.12. Variables Aleatorias

A la realización de un experimento aleatorio le hemos asociado un modelo matemático representado por un espacio de probabilidad (Ω, \mathcal{F}, P) , donde Ω es el conjunto de resultados posibles del experimento, \mathcal{F} es la colección de eventos, de acuerdo a las definiciones de la sección 1.2 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 1.3.

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

$$y = \varphi(x_1, x_2, \dots, x_k) \tag{1.14}$$

donde $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$ 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:

$$x_i = \bar{x}_i + \delta_i \quad (i = 1, \dots, k)$$

donde $\delta_1, \delta_2, \dots, \delta_k$ son los errores de redondeo.

Si en lugar de (1.14) 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*.

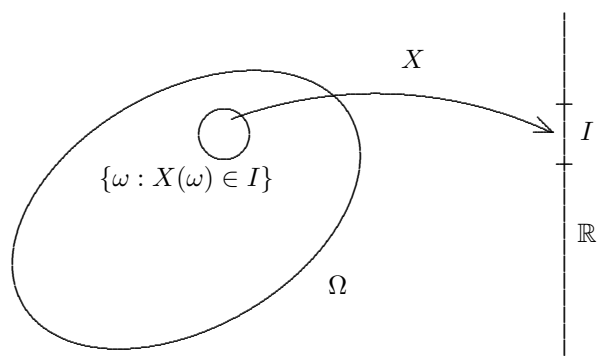


Figura 1.7: Definición de variable aleatoria.

Con frecuencia va a resultar de interés poder calcular la probabilidad de que una variable aleatoria X tome valores en un intervalo I (ver figura 1.7), es decir, nos interesa calcular 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} .

Definición 1.9 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Una función

$$X : \Omega \rightarrow \mathbb{R}$$

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}).

Observamos que la definición 1.9 no requiere realmente que X esté definida en un espacio de probabilidad, basta con que esté definida en un espacio Ω que tenga asociada una σ -álgebra \mathcal{F} de subconjuntos de Ω . En general, si $g : Y \rightarrow \mathbb{R}$ y \mathcal{F} es una σ -álgebra de subconjuntos de Y , decimos que g es *medible* si para todo intervalo $I \subset \mathbb{R}$, se tiene que

$$\{\omega : X(\omega) \in I\} \in \mathcal{F}.$$

Para definir el concepto de variable aleatoria vectorial, introducimos primero el concepto de intervalo en \mathbb{R}^m .

Definición 1.10 Un intervalo en \mathbb{R}^m es un conjunto de la forma:

$$\{(x_1, x_2, \dots, x_m) : x_i \in I_1, x_2 \in I_2, \dots, x_m \in I_m\}$$

donde I_1, I_2, \dots, 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 1.11 Una función

$$Y : \Omega \rightarrow \mathbb{R}^m$$

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

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

y la función $X(\omega) = \omega$. X no es una variable aleatoria, porque, por ejemplo,

$$X^{-1}([0, 3/2]) = \{1\} \notin \mathcal{F}.$$

Ejemplos 1.15

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 $X : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$\{\omega : X(\omega) \in I\} = \begin{cases} \Omega, & \text{si } 0, 1 \in I \\ A, & \text{si } 0 \notin I, 1 \in I \\ A^c, & \text{si } 0 \in I, 1 \notin I \\ \emptyset, & \text{si } 0 \notin I, 1 \notin I \end{cases}$$

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 (Ω, \mathcal{F}, P) que sólo toma los valores 1 y 0, entonces X es la variable indicadora del evento

$$A = \{\omega : X(\omega) = 1\}.$$

La medibilidad de una función $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ depende de la σ -álgebra \mathcal{F} . Una función puede ser medible respecto a una σ -álgebra \mathcal{F}_1 y no respecto a otra \mathcal{F}_2 . Sin embargo, está claro a partir de la definición, que si X es medible respecto a \mathcal{F}_1 y $\mathcal{F}_1 \subset \mathcal{F}_2$, entonces X también es medible respecto a \mathcal{F}_2 . La menor σ -álgebra respecto a la cual una variable aleatoria X es medible se conoce como la σ -álgebra generada por X , y se denota por $\sigma(X)$. Esta σ -álgebra no es otra cosa que la intersección de todas las σ -álgebras respecto a las cuales X es medible.

Por ejemplo, si X sólo tiene una cantidad numerable de valores posibles x_1, x_2, \dots , los conjuntos

$$A_i = \{\omega : X(\omega) = x_i\}, \quad i = 1, 2, \dots$$

forman una partición numerable de Ω , es decir,

$$\Omega = \bigcup_{i=1}^{\infty} A_i, \quad \text{y} \quad A_i \cap A_j = \emptyset \quad \text{si } i \neq j.$$

En este caso \mathcal{F} está compuesta por los conjuntos \emptyset, Ω y por todos los conjuntos que sean unión de algunos de los A_i .

1.13. Operaciones con Variables Aleatorias.

Esta sección está destinada a probar que si se efectúan 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 1.1 Sea $X : \Omega \rightarrow \mathbb{R}^m$ una función. X es una variable aleatoria si y sólo si la preimagen de cualquier conjunto de Borel es un evento, es decir

$$X^{-1}(B) \in \mathcal{F} \quad \text{para todo } B \in \mathcal{B} \quad (1.15)$$

donde \mathcal{B} denota la familia de conjuntos de Borel.

Demostración. Puesto que los intervalos son conjuntos de Borel, si se cumple (1.15) 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

$$\bigcup_{n=1}^{\infty} D_n \in \mathcal{F}$$

ya que

$$\begin{aligned} X^{-1}(\bigcup_{n=1}^{\infty} D_n) &= \{\omega : X(\omega) \in \bigcup_{n=1}^{\infty} D_n\} \\ &= \bigcup_{n=1}^{\infty} \{\omega : X(\omega) \in D_n\} \\ &= \bigcup_{n=1}^{\infty} X^{-1}(D_n) \in \mathcal{F}. \end{aligned}$$

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

$$\Omega \xrightarrow{X} \mathbb{R}^m \xrightarrow{g} \mathbb{R}$$

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}^m . Definimos la *función compuesta* $Y : \Omega \rightarrow \mathbb{R}$ mediante

$$Y(\omega) = g(X(\omega)).$$

Lema 1.2 Con las notaciones anteriores, si X es una variable aleatoria y g es una función medible, Y también es una variable aleatoria.

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} (ver fig. 1.8). 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 función medible, $g^{-1}(I) \in \mathcal{B}^m$, y como X es una variable aleatoria, usando el lema 1.1 obtenemos

$$X^{-1}(g^{-1}(I)) \in \mathcal{F}.$$

■

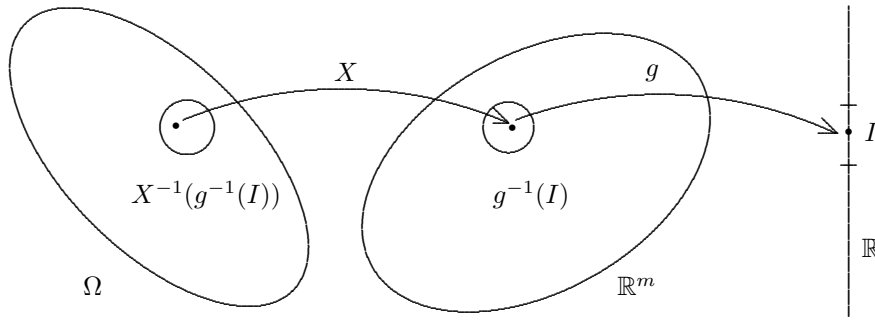


Figura 1.8: Composición de funciones.

Observación 1.2 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

$$X_i : \Omega \rightarrow \mathbb{R}$$

y por lo tanto también podemos escribir

$$Y(\omega) = g(X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

Lema 1.3 Las siguientes condiciones son equivalentes

1. $\{\omega : X(\omega) \in I\} \in \mathcal{F}$ para todo intervalo $I \subset \mathbb{R}$.
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

$$\{X < c\}^c = \{X \geq c\} \quad \text{y} \quad \{X > c\}^c = \{X \leq c\}$$

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

$$\{\omega : X(\omega) \in I\} = \{\omega : a \leq X(\omega) < b\} = \{\omega : a \leq X(\omega)\} \cap \{\omega : X(\omega) < b\}.$$

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 1.6 *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

$$X_1 < r \quad \text{y} \quad r < c - X_2$$

por lo tanto, si \mathbb{Q} es el conjunto de los números 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 1.3 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 1.12 Decimos que $g : \mathbb{R} \rightarrow \mathbb{R}$ 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. *estrictamente decreciente*). Decimos que g es *monótona* cuando es creciente o decreciente.

Proposición 1.7 Sea $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función monótona. Entonces $Y : \Omega \rightarrow \mathbb{R}$ 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 1.2 es suficiente demostrar que g es una función medible, 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 1.3 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:

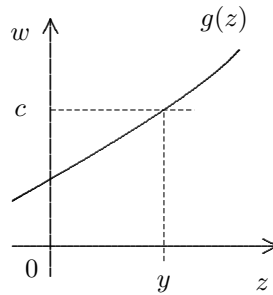


Figura 1.9: Gráfica de la función g estrictamente creciente y continua.

- (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 1.9) 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.

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 1.10).

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 concluye la demostración. ■

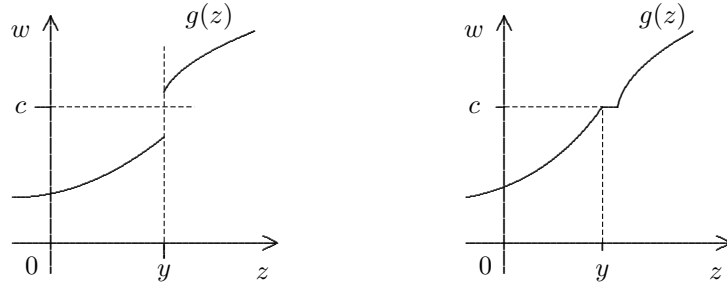


Figura 1.10: Gráfica de una función creciente pero no continua (izq.) y de una función creciente pero no estrictamente creciente (der.)

1.14. Función de Distribución.

Definición 1.13 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}$ 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 1.8 Si F es una función de distribución, satisface las siguientes propiedades:

1. F es creciente (en sentido amplio).
2. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.
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, $x_n \rightarrow -\infty$. 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

$$\lim_{n \rightarrow -\infty} F(x_n) = \lim_{n \rightarrow -\infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\emptyset) = 0.$$

Esto prueba que $\lim_{n \rightarrow -\infty} F(x_n) = 0$. Del mismo modo, si $\{x_n\}$ es una sucesión creciente y $x_n \rightarrow \infty$, 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

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\Omega) = 1.$$

Esto prueba que $\lim_{n \rightarrow \infty} F(x_n) = 1$.

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

$$\lim_{n \rightarrow \infty} F(x_n) = F(a).$$

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 $F : \mathbb{R} \rightarrow [0, 1]$ 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} igual a 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 1.9 *Sea X una variable aleatoria con función de distribución F . Entonces*

$$P(X \in (a, b]) = F(b) - F(a), \tag{1.16}$$

$$P(X \in [a, b]) = F(b) - F(a^-), \tag{1.17}$$

$$P(X \in (a, b)) = F(b^-) - F(a), \tag{1.18}$$

$$P(X \in [a, b)) = F(b^-) - F(a^-). \tag{1.19}$$

Demostración. Veamos la demostración de (1.17), las demás son similares:

$$P(X \in [a, b]) = P(X \leq b) - P(X < a) = F(b) - F(a^-)$$

■

Si ponemos ahora $a = b = x$ en (1.17) obtenemos que

$$P(X = x) = F(x) - F(x^-) \tag{1.20}$$

de modo que la función de distribución es continua en x si y sólo si $P(X = x) = 0$.

1.15. Variables Aleatorias Discretas.

Definición 1.14 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 $X : \Omega \rightarrow \mathbb{R}$ 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 sólo toma valores en el conjunto $\{0, 1, 2, \dots, n\}$.

Para ver cómo 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

$$F_X(x) = P(X \leq x) = \sum_{n: x_n \leq x} p_n$$

donde la suma se extiende a aquellos valores de n para los cuales $x_n \leq x$. La figura 1.11 (a) representa una gráfica típica de la función de distribución de una variable aleatoria discreta, mientras que la figura 1.11 (b) representa la función de probabilidad correspondiente.

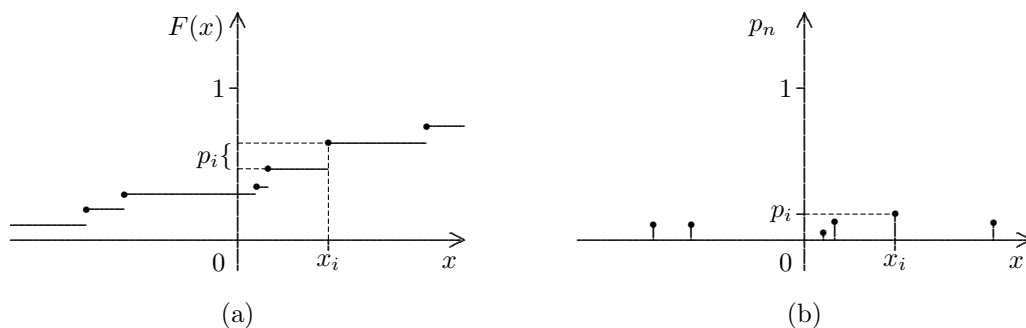


Figura 1.11: (a) Función de distribución y (b) función de probabilidad de una variable aleatoria discreta.

Ejemplos 1.16

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:

$$\begin{array}{l} x_i : 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \\ p_i : \frac{1}{36} \quad \frac{2}{36} \quad \frac{3}{36} \quad \frac{4}{36} \quad \frac{5}{36} \quad \frac{6}{36} \quad \frac{5}{36} \quad \frac{4}{36} \quad \frac{3}{36} \quad \frac{2}{36} \quad \frac{1}{36} \end{array}$$

El gráfico correspondiente a esta función de probabilidad es el siguiente:

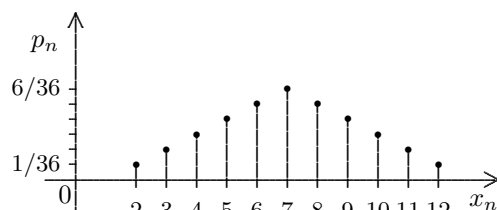


Figura 1.12: Función de probabilidad para la suma de dos dados.

La función de distribución de esta variable se representa en la figura 1.13.

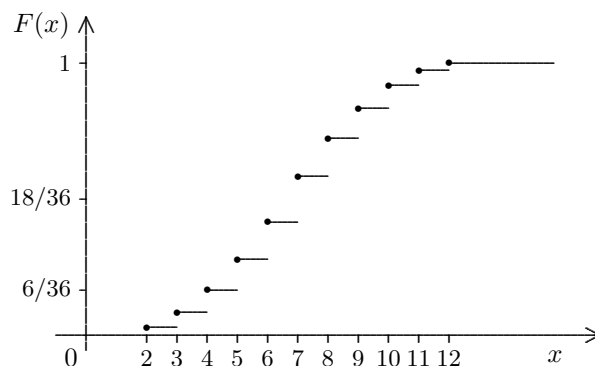


Figura 1.13: Función de distribución para la suma de dos dados.

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 $X : \Omega \rightarrow \mathbb{R}$ que estamos considerando está definida por

$$X(\omega_1, \omega_2) = \max\{\omega_1, \omega_2\}$$

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 :

$$\begin{array}{l} x_i : 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ p_i : \frac{1}{36} \quad \frac{3}{36} \quad \frac{5}{36} \quad \frac{7}{36} \quad \frac{9}{36} \quad \frac{11}{36} \end{array}$$

y su representación gráfica se presenta en la figura 1.14.

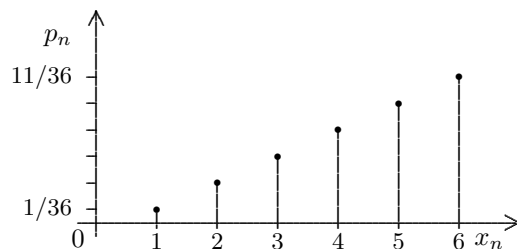


Figura 1.14: Muestreo con reposición.

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 $\{2, 3, 4, 5, 6\}$.

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

$$\begin{array}{l} x_i : 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ p_i : \frac{2}{30} \quad \frac{4}{30} \quad \frac{6}{30} \quad \frac{8}{30} \quad \frac{10}{30} \end{array}$$

y su representación gráfica es la figura 1.15

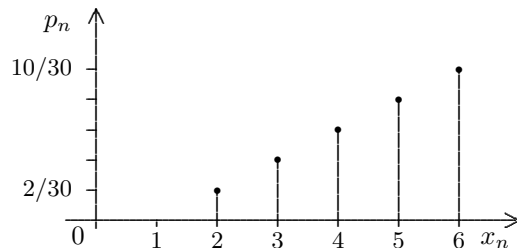


Figura 1.15: Muestreo sin reposición

▲

- Se lanza al aire una moneda repetidamente y se observa en cuál de los lanzamientos la moneda cae aguilá por primera vez. Hallar la función de probabilidad de esta variable.

► Si A denota aguilá y S sol, cada evento elemental es una sucesión infinita de estos símbolos:

$$\omega = (A, A, A, A, S, A, A, S, S, S, \dots)$$

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:

$$P(X = 1) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{2} \frac{1}{2} = \left(\frac{1}{2}\right)^2,$$

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 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

$$p_n = P(X = n) = \frac{1}{10}, \quad (n = 0, 1, \dots, 9)$$

La gráfica de la función de probabilidad se presenta en la figura 1.16.

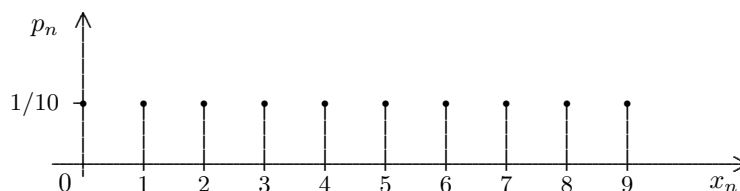


Figura 1.16: Función de probabilidad para el tercer dígito en el desarrollo decimal de un número seleccionado al azar en el intervalo $[0, 1]$.

1.16. 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, anteriormente consideramos el tiempo de vida útil T de una maquina y obtuvimos, bajo ciertas condiciones específicas, que esta variable T satisface

$$P(T > x) = e^{-\lambda x}, \quad \text{para } x > 0 \text{ y algún } \lambda > 0.$$

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

$$P(T = x) = 0 \tag{1.21}$$

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:

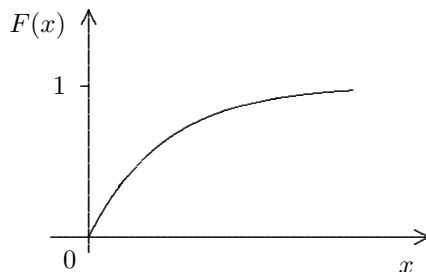


Figura 1.17: Función de distribución exponencial.

En general diremos que una variable aleatoria X es *continua* si su función de distribución lo es. Por (1.20) esto equivale a pedir

$$P(X = x) = 0, \quad \text{para todo } x \in \mathbb{R}.$$

Ejemplo 1.17

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

$$C = \frac{1}{|D|} \quad \text{y} \quad P(\omega \in A) = \frac{|A|}{|D|}.$$

En el razonamiento anterior hay un punto fundamental que hemos pasado por alto y es el concepto de área de un subconjunto de D . 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

$$\{\omega \in D : \text{la primera coordenada de } \omega \text{ es racional}\}.$$

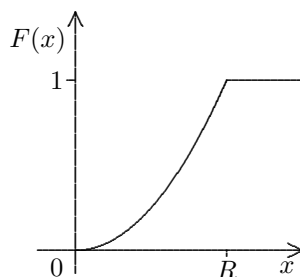


Figura 1.18: Gráfica de la función de distribución.

¿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

$$F(x) = P(X \leq x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}, \quad 0 \leq x \leq R.$$

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 la figura 1.18. ▲

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 1.19 está en esta clase.

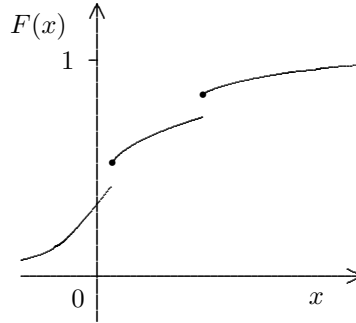


Figura 1.19: Ejemplo de función de distribución que no es discreta ni continua.

1.17. Densidades.

Definición 1.15 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

$$F(x) = \int_{-\infty}^x f(t) dt, \quad \text{para todo } x \in \mathbb{R}. \quad (1.22)$$

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 1.8 resulta que

$$\int_{-\infty}^{\infty} f(t) dt = 1. \quad (1.23)$$

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

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(t) dt.$$

Geoméricamente, 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 (ver figura 1.20).

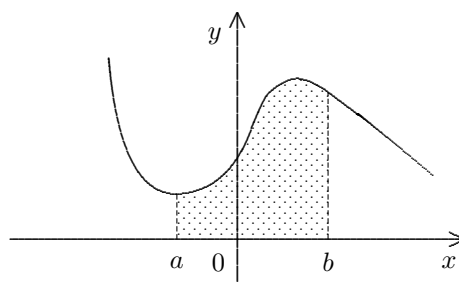


Figura 1.20: Densidad de una variable aleatoria continua.

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 (1.23), la función F definida por (1.22) es una función de distribución, es decir, que satisface las condiciones 1, 2 y 3 de la proposición 1.8. 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 $n \rightarrow \infty$ 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 (1.23).

1.17.1. La Distribución Uniforme.

Definición 1.16 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

$$K(b - a) = 1 \quad \text{de donde} \quad K = \frac{1}{b - a}.$$

En consecuencia la distribución de probabilidad de X es la siguiente

$$F_X(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } x > b. \end{cases}$$

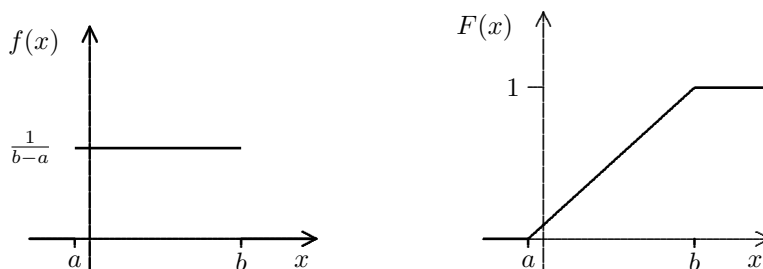


Figura 1.21: Densidad (izq.) y función de distribución (der) para la distribución uniforme en el intervalo $[a, b]$.

Esta distribución tiene como densidad la función f_X definida por

$$f_X(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{1}{b-a}, & \text{si } a \leq x \leq b, \\ 0, & \text{si } x > b, \end{cases}$$

ya que se verifica inmediatamente que

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{para todo } x \in \mathbb{R}.$$

En la figura 1.21 representamos las gráficas de estas funciones. Usaremos la notación $X \sim \mathcal{U}[a, b]$

Ejemplo 1.18

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

$$P(X \in B) = \frac{2}{15}.$$

▲

1.17.2. La Distribución Exponencial.

Definición 1.17 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

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - e^{-\lambda x}, & \text{si } x \geq 0. \end{cases}$$

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

$$P(X > a + b) = P(X > a)P(X > b). \quad (1.24)$$

La verificación de esta propiedad es inmediata a partir de la definición. Una forma equivalente de escribir esta propiedad es

$$P(X > a + b | X > a) = P(X > b), \quad a \geq 0, \quad b \geq 0, \quad (1.25)$$

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 1.10.4 vimos que si (1.25), o equivalentemente (1.24), 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 (1.24) o (1.25) caracteriza a la distribución exponencial.

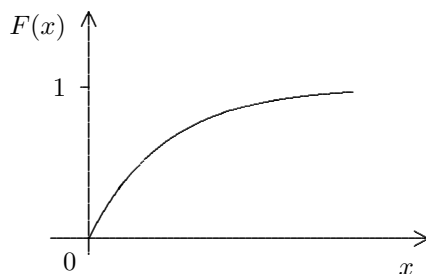


Figura 1.22: Función de distribución exponencial.

Ejemplos 1.19

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 . ▲

1.17.3. 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 $z = (y - \mu)/\sqrt{2}\sigma$ resulta

$$\begin{aligned} \int_{-\infty}^{\infty} n(y; \mu, \sigma^2) dy &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2} \sigma \sqrt{2} dz \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz. \end{aligned} \tag{1.26}$$

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 1.23).

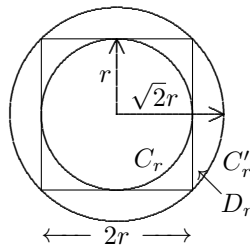


Figura 1.23

Dado que el integrando común en las integrales siguientes no es negativo, tenemos que

$$\iint_{C_r} e^{-(u^2+v^2)} du dv \leq \iint_{D_r} e^{-(u^2+v^2)} du dv \leq \iint_{C'_r} e^{-(u^2+v^2)} du dv \tag{1.27}$$

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 (1.27). Pasando a coordenadas polares ρ , θ por medio de la transformación $u = \rho \cos \theta$, $v = \rho \sin \theta$, obtenemos

$$\begin{aligned} \iint_{C_r} e^{-(u^2+v^2)} du dv &= \int_0^{2\pi} d\theta \int_0^r e^{-\rho^2} \rho d\rho \\ &= 2\pi \left[-\frac{1}{2} e^{-\rho^2} \right]_0^r \\ &= 2\pi \left[\frac{1}{2} (1 - e^{-r^2}) \right] \\ &= \pi(1 - e^{-r^2}). \end{aligned}$$

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 (1.27)

$$\pi(1 - e^{-r^2}) \leq \left(\int_{-r}^r e^{-u^2} du \right)^2 \leq \pi(1 - e^{-2r^2}),$$

haciendo $r \rightarrow \infty$

$$\pi \leq \left(\int_{-\infty}^{\infty} e^{-u^2} du \right)^2 \leq \pi$$

y por lo tanto

$$\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}.$$

Sustituyendo en (1.26) 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

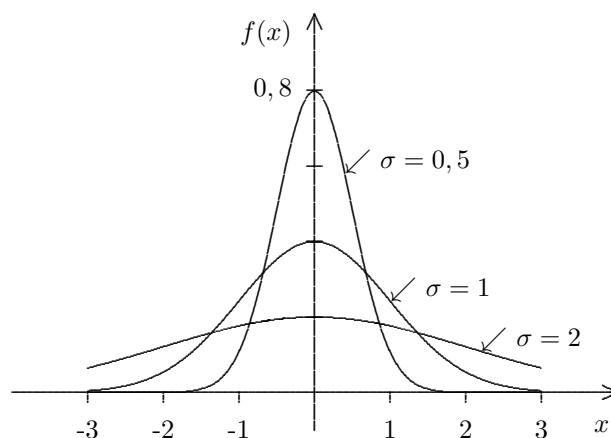
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

En la figura 1.24 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 explícita y debe ser calculada numéricamente. Es posible calcular los valores de esta función en \mathbf{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 basta conocer los valores de $\Phi(x)$ para $x \geq 0$.

Figura 1.24: Densidad Gaussiana para tres valores de σ .

Sea X una variable aleatoria con densidad ϕ y consideremos $Y = \mu + \sigma X$, donde $\sigma > 0$. Es posible demostrar que Y tiene como densidad a la función

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = n(y; \mu, \sigma^2) \quad (1.28)$$

y este hecho nos permite calcular cualquier distribución normal a partir de Φ ya que

$$P(Y \leq y) = P(\mu + \sigma X \leq y) = P\left(X \leq \frac{y-\mu}{\sigma}\right) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Por lo tanto, si Y tiene una distribución con densidad $n(y; \mu, \sigma^2)$ y $a \leq b$ entonces

$$P(a \leq Y \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Ejemplo 1.20

Supongamos que Y tiene una distribución normal con parámetros $\mu = 0.5$ y $\sigma = 4$ y queremos calcular la probabilidad de que $-0.5 \leq Y \leq 2.4$:

$$P(-0.5 \leq Y \leq 2.4) = \Phi\left(\frac{1}{2}\right) - \Phi\left(\frac{-1}{4}\right) = 0.6915 - 0.4013 = 0.2902.$$

1.18. Simulación de Variables Aleatorias.

Los generadores de números aleatorios simulan valores de la distribución $\mathcal{U}[0, 1]$. En esta sección estudiaremos métodos que nos permiten generar valores de variables aleatorias con otras distribuciones F . Un método general para hacer esto se conoce como el método de transformada inversa, que explicamos a continuación.

Sea X una variable aleatoria con función de distribución continua y $g : \mathbb{R} \rightarrow (a, b)$ 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 1.10 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

$$F_Y(y) = F_X(g^{-1}(y)). \quad (1.29)$$

Demostración. Como g es estrictamente creciente los eventos $\{X \leq g^{-1}(y)\}$ y $\{g(X) \leq y\}$ son iguales. Por lo tanto,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

■

Si g es estrictamente decreciente entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Corolario 1.2 Sea F una función de distribución estrictamente 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

$$F_Z(z) = F_U(F(z)) = F(z) \quad (1.30)$$

de modo que Z tiene función de distribución F .

■

Observación 1.3 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:

$$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$$

En consecuencia, para cualquier función de distribución F , la variable aleatoria $Z = F^{\leftarrow}(U)$ tiene función de distribución F .

El Corolario 1.2 y la Observación 1.3 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.

1.18.1. 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 :

$$[0, p_1); \quad [p_1, p_1 + p_2); \quad [p_1 + p_2, p_1 + p_2 + p_3); \quad \dots \quad \left[\sum_{j < n} p_j, 1\right].$$

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

$$P(X = x_i) = p_i, \quad \text{para } 1 \leq i \leq n.$$

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 probabilidades 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 las variables el valor x_k si

$$\frac{k-1}{n} < U \leq \frac{k}{n}$$

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 .

1.18.2. 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 1.2. 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 sección 1.17.2: 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:

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2.$$

1.18.3. 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 `rdist`, 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	<code>rbinom(n, size, prob)</code>
Poisson	<code>rpois(n, lambda)</code>
Geométrica	<code>rgeom(n, prob)</code>
Hipergeométrica	<code>rhyper(nn, m, n, k)</code>
Binomial Negativa	<code>rnbinom(n, size, prob)</code>
Multinomial	<code>rmultinom(n, size, prob)</code>
Uniforme	<code>runif(n, min=0, max=1)</code>
Exponencial	<code>rexp(n, rate=1)</code>
Gaussiana	<code>rnorm(n, mean=0, sd=1)</code>
Gamma	<code>rgamma(n, shape, scale=1)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
Beta	<code>rbeta(n, shape1, shape2)</code>
t	<code>rt(n, df)</code>
Fisher	<code>rf(n, df1, df2)</code>
χ^2	<code>rchisq(n, df)</code>
Logística	<code>rlogis(n, location=0, scale=1)</code>
Lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>

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.

1.19. Distribución Conjunta de Dos Variables Aleatorias.

Sean X e Y dos variables aleatorias sobre un espacio de probabilidad común (Ω, \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 1.25 (a).

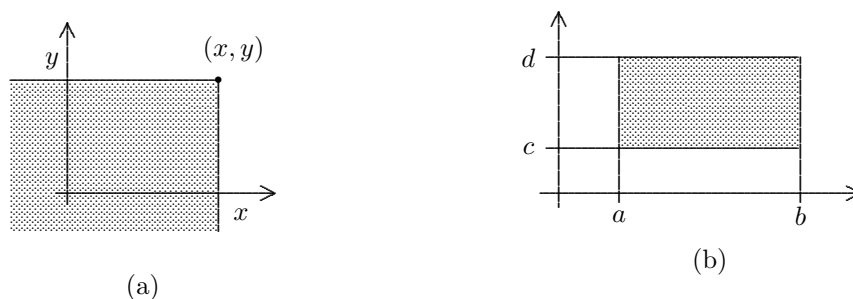


Figura 1.25: Función de distribución en el plano.

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 1.25 (b))

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) &= P(X \leq b, Y \leq d) - P(X \leq b, Y \leq c) \\ &\quad - P(X \leq a, Y \leq d) + P(X \leq a, Y \leq c) \\ &= F(b, d) - F(b, c) - F(a, d) + F(a, c). \end{aligned} \quad (1.31)$$

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

$$\{\omega : X(\omega) \leq x\} \subset \{\omega : X(\omega) \leq x'\}$$

y por lo tanto

$$\begin{aligned} F(x, y) &= P(\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\}) \\ &\leq P(\{\omega : X(\omega) \leq x'\} \cap \{\omega : Y(\omega) \leq y\}) \\ &= F(x', y). \end{aligned}$$

2. $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0$; $\lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1$.

y como la función es creciente en ambas variables se deduce que, para cualesquiera x, y ,

$$0 \leq F(x, y) \leq 1.$$

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

$$F(x, y) = \begin{cases} 0 & \text{si } x + y < 0 \\ 1 & \text{si } x + y \geq 0 \end{cases}$$

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 1.26).

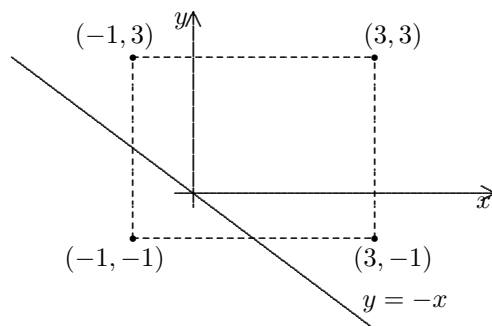


Figura 1.26

La función es creciente, continua por la derecha y satisface la propiedad 2. Sin embargo, si aplicamos la fórmula (1.31) 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

$$P(-1 < X \leq 3, -1 < Y \leq 3) = F(3, 3) - F(3, -1) - F(-1, 3) + F(-1, -1) = -1$$

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 (1.31) no sea negativo para ninguna colección de números $a < b$, $c < d$.

Teorema 1.1 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 $a < b$, $c < d$,

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0.$$

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

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x, Y < \infty) \\ &= \lim_{y \rightarrow \infty} P(X \leq x, Y \leq y) \\ &= \lim_{y \rightarrow \infty} F(x, y) \end{aligned}$$

y de manera similar, para cualquier $y \in \mathbb{R}$

$$F_Y(y) = \lim_{x \rightarrow \infty} F(x, y).$$

Las funciones F_X y F_Y se conocen como las *funciones de distribución marginales* de X e Y , respectivamente.

1.20. Variables Aleatorias Independientes.

Se dice que las variables X e Y son *independientes* si cualesquiera que sean los intervalos $(a, b]$ y $(c, d]$, se verifica que los eventos

$$\{X \in (a, b]\} \quad \text{y} \quad \{Y \in (c, d]\}$$

son independientes, es decir que

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d). \quad (1.32)$$

En términos menos precisos, de acuerdo a lo que hemos visto sobre independencia de eventos en el Capítulo 3, esta relación dice que el saber que el valor de X está en cierto intervalo no arroja información alguna sobre la probabilidad de que Y esté en otro intervalo.

Es fácil ver que la condición (1.32) es equivalente a la condición

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \text{para todo } x, y \in \mathbb{R}. \quad (1.33)$$

En efecto, si se cumple (1.32) basta poner $b = x$, $d = y$ y hacer tender $a \rightarrow -\infty$, $c \rightarrow -\infty$, para obtener (1.33). Recíprocamente, si se cumple (1.33), poniendo $F_{X,Y} = F$ tenemos

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \\ &= F_X(b)F_Y(d) - F_X(a)F_Y(d) - F_X(b)F_Y(c) + F_X(a)F_Y(c) \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c)) \\ &= P(a < X \leq b)P(c < Y \leq d) \end{aligned}$$

o sea que (1.33) implica (1.32), cualesquiera sean a , b , c y d .

Las relaciones (1.32) y (1.33) 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 (1.32), y semirectas cerradas a la derecha, en el caso de (1.33). Es posible probar, aunque no lo haremos en este texto, que (1.33), o equivalentemente (1.32), 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 .

Si tenemos una colección de variables aleatorias $\{X_s, s \in S\}$, donde S es cualquier conjunto de índices, diremos que son independientes si para cualquier subconjunto finito de índices $K \subset S$ y cualesquiera intervalos $I_k = (a_k, b_k]$, $k \in K$, se tiene que

$$P(\cap_{k \in K} \{X_k \in I_k\}) = \prod_{k \in K} P(X_k \in I_k).$$

Ejemplo 1.21

Si tenemos tres variables aleatorias X, Y y Z , para verificar que son independientes es necesario verificar que son independientes a pares y además que se satisface

$$P(X \in I_1, Y \in I_2, Z \in I_2) = P(X \in I_1)P(Y \in I_2)P(Z \in I_2)$$

Para ver que independencia a pares no es suficiente para garantizar que las variables sean independientes tenemos el siguiente ejemplo. Consideramos el lanzamiento de dos monedas. X vale 1 si la primera moneda es Aguila y 0 si es Sol. Y se define de manera similar para la segunda moneda mientras que Z vale 1 si exactamente una de las dos monedas es Aguila, y vale 0 en otro caso. Vemos que el vector (X, Y, Z) toma los valores $(0, 0, 0); (0, 1, 1); (1, 0, 1); (1, 1, 0)$ cada uno con probabilidad $1/4$.

Es fácil ver que las funciones de probabilidad marginales de las tres variables son iguales:

$$P(X = 0) = P(Y = 0) = P(Z = 0) = \frac{1}{2}; \quad P(X = 1) = P(Y = 1) = P(Z = 1) = \frac{1}{2}$$

y también es sencillo ver que son independientes dos a dos. Veamos que Y y Z son independientes. Para esto calculamos la función de probabilidad conjunta:

$$\begin{aligned} P(Y = 0, Z = 0) &= P(0, 0, 0) = \frac{1}{4}; & P(Y = 0, Z = 1) &= P(1, 0, 1) = \frac{1}{4}; \\ P(Y = 1, Z = 0) &= P(1, 1, 0) = \frac{1}{4}; & P(Y = 1, Z = 1) &= P(0, 1, 1) = \frac{1}{4}. \end{aligned}$$

Sin embargo las tres variables no son independientes pues la función de probabilidad conjunta de (X, Y, Z) no es el producto de las marginales. Por ejemplo, $P(X = 1, Y = 1, Z = 1) = 0$ pero $P(X = 1)P(Y = 1)P(Z = 1) = 1/8$.

1.21. 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

$$\sum_{n=1}^{\infty} P(X = x_n) = 1.$$

Comenzamos por definir la noción de valor esperado para variables aleatorias discretas.

Definición 1.18 Sea X una variable aleatoria discreta con la notación anterior y sea $p_n = P(X = x_n)$, $n = 1, 2, \dots$. Diremos que existe el *valor esperado*, la *media* o la *esperanza matemática* de X si la serie

$$\sum_{n=1}^{\infty} |x_n| p_n \tag{1.34}$$

es convergente. En este caso, el valor esperado se denota $E(X)$ y se define mediante la serie

$$E(X) = \sum_{n=1}^{\infty} x_n p_n. \tag{1.35}$$

Ejemplo 1.22

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

$$E(X) = \sum_{i=1}^6 i \frac{1}{6} = \frac{21}{6} = 3.5$$

Observamos que en este caso el valor esperado no es un valor posible de la variable aleatoria.

Observación 1.4 Una primera observación es que si una variable aleatoria toma sólo un número finito de valores, entonces su esperanza matemática está bien definida, ya que (1.34) 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

$$x_n = 2^n \quad \text{y} \quad p_n = P(X = 2^n) = \frac{1}{2^n}, \quad n \geq 1.$$

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 (1.34) 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 (1.34) diverge, se conviene en definir $E(X) = +\infty$ 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

$$P(Y = 2^n) = \frac{1}{2^{n+1}} \quad P(Y = -2^n) = \frac{1}{2^{n+1}} \quad n \geq 1,$$

entonces, nuevamente (1.34) diverge y no existe $E(Y)$.

1.21.1. Propiedades

La esperanza matemática de una variable discreta tiene las siguientes propiedades.

Propiedad 1. Si $X \geq 0$ y existe $E(X)$, entonces $E(X) \geq 0$.

Es obvio que si $X \geq 0$, los valores x_n que figuran en la suma (1.35) 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 $E(X)$.

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 (1.34) resultan estar acotadas por la constante C . Ahora bien, recordemos que para una serie de términos no-negativos – es el caso de la serie (1.34) – es necesario y suficiente para que converja que sus sumas parciales estén acotadas. Como las de (1.34) lo están, esta serie es convergente y el valor esperado de X existe.

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

$$E(\mathbf{1}_A) = P(A).$$

Es claro que $\mathbf{1}_A$ es discreta, ya que toma solamente dos valores y

$$E(\mathbf{1}_A) = 1 \times P(\mathbf{1}_A = 1) + 0 \times P(\mathbf{1}_A = 0) = P(A).$$

Propiedad 4. Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función y consideremos la nueva variable aleatoria (discreta)

$$Y = g(X).$$

Entonces, si existe la esperanza matemática de Y , ésta es

$$E(Y) = \sum_{n=1}^{\infty} g(x_n) p_n. \quad (1.36)$$

Demostración. Llamemos $\{y_m\}$ a los valores de Y , entonces

$$E(Y) = \sum_m y_m P(Y = y_m). \quad (1.37)$$

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 (1.37)

$$E(Y) = \sum_m y_m \sum_{\{n:g(x_n)=y_m\}} p_n = \sum_m \sum_{\{n:g(x_n)=y_m\}} g(x_n)p_n$$

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

$$E(Y) = \sum_n g(x_n)p_n$$

como afirmamos en el enunciado de la propiedad 4. ■

Observación 1.5 La ventaja que tiene la fórmula (6.3), es que nos permite calcular $E(Y)$ 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 $E(X)$ entonces $|E(X)| \leq E(|X|)$.

Demostración. De acuerdo a la propiedad 4 (tomando $g(x) = |x|$), se tiene

$$E(|X|) = \sum_n |x_n|p_n \geq \left| \sum_n x_n p_n \right| = |E(X)|.$$

Observe que la desigualdad entre las sumas de las series, se deduce simplemente de que

$$\sum_{n=1}^N |x_n|p_n \geq \left| \sum_{n=1}^N x_n p_n \right|$$

en virtud de la desigualdad triangular entre números reales. Pasando al límite cuando $N \rightarrow \infty$, se obtiene la desigualdad análoga entre las sumas de las series.

Propiedad 6. Si λ es una constante y $E(X)$ existe, entonces también existe $E(\lambda X)$ y

$$E(\lambda X) = \lambda E(X).$$

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

$$E(X + Y) = E(X) + E(Y).$$

Demostración. Para demostrar esta propiedad utilizamos la notación:

$$r_{nm} = P(X = x_n, Y = y_m), \quad p_n = P(X = x_n), \quad q_m = P(Y = y_m),$$

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,

$$\sum_m r_{nm} = p_n, \quad n = 1, 2, \dots; \quad \sum_n r_{nm} = q_m, \quad m = 1, 2, \dots$$

y por otro, las dos series

$$\sum_n |x_n| p_n; \quad \sum_m |y_m| q_m$$

son convergentes, dado que existen $E(X)$ y $E(Y)$.

Esto prueba que existe $E(Z)$. Si ahora repetimos el cálculo anterior sin los valores absolutos, resulta

$$E(Z) = \sum_k z_k P(Z = z_k) = \sum_n x_n p_n + \sum_m y_m q_m = E(X) + E(Y),$$

que es la propiedad que queríamos probar. ■

Propiedad 8. Si $X \leq Y$ y existen $E(X)$ y $E(Y)$, entonces $E(X) \leq E(Y)$.

Demostración. Para probar esta propiedad, recurrimos a las propiedades 1, 6 y 7. Tenemos

$$E(Y) - E(X) = E(Y - X) \geq 0, \quad \text{ya que } Y - X \geq 0.$$

Propiedad 9. Si X e Y son variables aleatorias independientes con valor esperado, entonces existe $E(XY)$ y

$$E(XY) = E(X) E(Y). \quad (1.38)$$

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 $E(XY)$.

El mismo cálculo, quitando los valores absolutos, permite obtener la relación (1.38). ■

1.22. Momentos de Orden Superior. Momentos Centrados.

Definición 1.19 Sea X una variable aleatoria discreta,

$$P(X = x_n) = p_n, \quad \sum_n p_n = 1$$

y m un número positivo. Si

$$\sum_n |x_n|^m p_n < \infty,$$

se define el *momento de orden m de X* mediante

$$E(X^m) = \sum_n x_n^m p_n.$$

(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

$$\mu_m = E((X - E(X))^m) = \sum_n (x_n - E(X))^m p_n.$$

(Ver la propiedad 4, con $g(x) = (x - E(X))^m$. En la definición se sobrentiende que existe la esperanza de la variable aleatoria X).

Se acostumbra denominar *varianza* ($\text{Var}(X)$) al momento centrado de orden 2 – cuando existe – es decir

$$\text{Var}(X) = E((X - E(X))^2) \tag{1.39}$$

y *desviación típica o estándar* de X a $(\text{Var}(X))^{1/2}$.

También se define la *covarianza* de dos variables aleatorias mediante

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))], \tag{1.40}$$

siempre que la esperanza que figura en el segundo miembro esté bien definida, y el *coeficiente de correlación* de X e Y por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}, \tag{1.41}$$

siempre que el segundo miembro tenga sentido.

Finalmente, se dice que X e Y no están correlacionadas, son *no-correlacionadas* o *incorreladas* si $\text{Cov}(X, Y) = 0$.

1.22.1. 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

$$\text{Var}(X) = \text{Var}(X + a)$$

y

$$\text{Var}(aX) = a^2 \text{Var}(X).$$

La demostración resulta de aplicar la definición y queda como ejercicio para el lector.

Propiedad 2. $\text{Var}(X) = E(X^2) - (E(X))^2$.

Demostración. Observar simplemente que

$$\begin{aligned}\text{Var}(X) &= E[(X - E(X))^2] = E[X^2 - 2E(X)X + (E(X))^2] \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2.\end{aligned}$$

Propiedad 3. (Desigualdad de Cauchy-Schwarz)

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}.$$

Demostración. Para demostrar esta desigualdad, consideremos la función $f: \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$f(x) = E[(U - xV)^2],$$

donde $U = X - E(X)$ y $V = Y - E(Y)$. Usando las propiedades de la esperanza matemática

$$f(x) = E[U^2 - 2xUV + x^2V^2] = E(U^2) - 2xE(UV) + x^2E(V^2)$$

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

$$(-2E(UV))^2 - 4E(V^2)E(U^2) \leq 0 \Rightarrow (E(UV))^2 \leq E(U^2)E(V^2).$$

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

$$-1 \leq \rho(X, Y) \leq 1 \tag{1.42}$$

Demostración. Es una consecuencia sencilla de la propiedad anterior.

Propiedad 5. Si $\text{Var}(X) = 0$ 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 todas las propiedades que hemos estudiado, pero eso lo veremos más adelante).

Entonces la variable aleatoria $Y = (X - E(X))^2$ también es discreta y toma valores mayores o iguales a cero, digamos $\{y_m\}$. Nuestra hipótesis es que

$$E(Y) = 0.$$

Sea $p_m = P(Y = y_m)$. Si hubiera algún $y_{m_0} > 0$ tal que $p_{m_0} > 0$, se tendría que

$$E(Y) = \sum_m y_m p_m \geq y_{m_0} p_{m_0} > 0$$

contradiendo 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 $\{X = E(X)\}$, de modo que

$$P(X = E(X)) = 1,$$

y la propiedad se cumple, tomando $c = E(X)$. ■

Observación 1.6 Es obvio que el recíproco de esta propiedad es cierto, ya que si

$$P(X = c) = 1 \quad \text{entonces} \quad E(X) = c$$

y

$$\text{Var}(X) = E((X - c)^2) = 0,$$

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

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n). \quad (1.43)$$

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 $Y_1 = X_1 - E(X_1)$, $Y_2 = X_2 - E(X_2)$. 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 1.21.1

$$E(Y_1 Y_2) = E(Y_1) E(Y_2) = 0$$

Se deduce que

$$\begin{aligned} \text{Var}(X_1 + X_2) &= E[(X_1 + X_2 - E(X_1 + X_2))^2] = E[(Y_1 + Y_2)^2] \\ &= E[Y_1^2 + Y_2^2 + 2Y_1 Y_2] = E(Y_1^2) + E(Y_2^2) + 2E(Y_1 Y_2) \\ &= E(Y_1^2) + E(Y_2^2) = E[(X_1 - E(X_1))^2] + E[(X_2 - E(X_2))^2] \\ &= \text{Var}(X_1) + \text{Var}(X_2) \end{aligned}$$

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

$$\sum_n |x_n|^k p_n < \infty, \quad \text{donde } p_n = P(X = x_n), \quad \sum_n p_n = 1.$$

Pero tenemos la desigualdad, válida para todo número $x \in \mathbb{R}$,

$$|x|^k \leq 1 + |x|^m. \quad (1.44)$$

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 (1.44). 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 $\text{Cov}(X, Y)$, entonces

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y).$$

Propiedad 9. Sean X e Y variables aleatorias con covarianza $\text{Cov}(X, Y)$, entonces

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Las demostraciones quedan como ejercicio.

1.23. Ejemplos y Aplicaciones.

1. **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,

$$X_1 + X_2 + \dots + X_r$$

y su esperanza matemática

$$E(X_1 + X_2 + \dots + X_r) = \sum_{i=1}^r E(X_i) = \sum_{i=1}^r P(X_i = 1)$$

ya que

$$E(X_i) = 1 \times P(X_i = 1) + 0 \times P(X_i = 0).$$

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

$$\begin{aligned} \{X_i = 1\} = & \{1^{\text{er}} \text{ individuo no pertenece al estrato } i\} \\ & \cap \{2^{\text{o}} \text{ individuo no pertenece al estrato } i\} \\ & \cdots \cap \{n\text{-ésimo individuo no pertenece al estrato } i\}. \end{aligned}$$

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

$$E(X_1 + X_2 + \cdots + X_r) = \sum_{i=1}^r (1 - q_i)^n.$$

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

$$q_i = \frac{1}{r} \quad (i = 1, 2, \dots, r).$$

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 $\log(1 - x)$

$$\log(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots \quad (|x| < 1)$$

y entonces, si $|x| < 1$

$$(1 - x)^{\frac{1}{x}} = e^{\frac{1}{x} \log(1-x)} = e^{\frac{1}{x}(-x - \frac{x^2}{2} - \frac{x^3}{3} - \dots)} = e^{-1 - \frac{x}{2} - \frac{x^2}{3} - \dots}$$

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. ▲

2. Se lanzan dos dados simétricos independientes. Sean X_1, X_2 los resultados obtenidos en cada uno e $Y = \max\{X_1, X_2\}$.
- Hallar la probabilidad conjunta de X_1 e Y .
 - Calcular $E(X_1)$, $E(Y)$, $\text{Var}(X_1)$, $\text{Var}(Y)$ 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

$$r_{ij} = \begin{cases} \frac{1}{36}, & i < j, \\ \frac{i}{36}, & i = j, \\ 0, & i > j. \end{cases}$$

b.

$$\begin{aligned} E(X_1) &= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6}, \\ \text{Var}(X_1) &= E(X_1^2) - (E(X_1))^2 \\ &= \frac{1}{6}(1 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{21}{6}\right)^2 = \frac{35}{12}. \end{aligned}$$

Para calcular $E(Y)$ y $\text{Var}(Y)$, observemos que

$$P(Y = i) = \frac{2i - 1}{36} \quad (i = 1, 2, 3, 4, 5, 6)$$

lo cual puede calcularse directamente o usando la parte (a) de este ejercicio. Por lo tanto

$$\begin{aligned} E(Y) &= 1 \frac{1}{36} + 2 \frac{3}{36} + 3 \frac{5}{36} + 4 \frac{7}{36} + 5 \frac{9}{36} + 6 \frac{11}{36} = \frac{161}{36} \\ \text{Var}(Y) &= E(Y^2) - (E(Y))^2 \\ &= 1 \frac{1}{36} + 2^2 \frac{3}{36} + 3^2 \frac{5}{36} + 4^2 \frac{7}{36} + 5^2 \frac{9}{36} + 6^2 \frac{11}{36} - \left(\frac{161}{36}\right)^2 \\ &= \frac{2555}{(36)^2}. \end{aligned}$$

Finalmente

$$\text{Cov}(X_1, Y) = E[(X_1 - E(X_1))(Y - E(Y))] = E(X_1 Y) - E(X_1) E(Y).$$

$E(X_1Y)$ se calcula utilizando la probabilidad conjunta que encontramos en la parte *a*.

$$\begin{aligned} E(X_1Y) &= \sum_{i,j=1}^6 ijr_{ij} = \sum_{i<j} ij \frac{1}{36} + \sum_{i=1}^6 i^2 \frac{i}{36} \\ &= \frac{1}{36} \left[\sum_{i=1}^5 i \sum_{j=i+1}^6 j + \sum_{i=1}^6 i^3 \right] \\ &= \frac{1}{36} \left[\sum_{i=1}^5 \frac{(6-i)(6+i+1)}{2} i + \sum_{i=1}^6 i^3 \right] = \frac{154}{9} \end{aligned}$$

lo que implica que $\text{Cov}(X_1, Y) = \frac{35}{24}$. ▲

3. Sean X_1, X_2, \dots, X_n variables aleatorias independientes y supongamos que

$$E(X_i) = \mu, \quad \text{Var}(X_i) = \sigma^2, \quad (i = 1, \dots, n).$$

Definimos la media muestral correspondiente a esta muestra por

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

y la varianza muestral por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

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

$$E(\bar{X}) = \mu; \quad E(s^2) = \sigma^2.$$

► Comenzamos con

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu.$$

Por otro lado

$$E(s^2) = E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} \sum_{i=1}^n E((X_i - \mu + \mu - \bar{X})^2). \quad (1.45)$$

Calculemos el valor de cada sumando:

$$E((X_i - \mu)^2) - 2E((X_i - \mu)(\bar{X} - \mu)) + E((\bar{X} - \mu)^2) \quad (1.46)$$

y ahora cada uno de estos términos:

$E((X_i - \mu)^2)$ es la varianza de X_i , es decir, σ^2 .

$$\begin{aligned} E((X_i - \mu)(\bar{X} - \mu)) &= E((X_i - \mu)\left(\frac{1}{n} \sum_{j=1}^n (X_j - \mu)\right)) \\ &= E((X_i - \mu) \frac{1}{n} \sum_{j=1}^n (X_j - \mu)) \\ &= \frac{1}{n} E\left(\sum_{j=1}^n (X_i - \mu)(X_j - \mu)\right) \\ &= \frac{1}{n} \sum_{j=1}^n E((X_i - \mu)(X_j - \mu)). \end{aligned}$$

En esta suma, si $i \neq j$, como X_i, X_j son variables aleatorias independientes, tenemos

$$E((X_i - \mu)(X_j - \mu)) = E(X_i - \mu) E(X_j - \mu) = 0,$$

y por lo tanto

$$E((X_i - \mu)(\bar{X} - \mu)) = \frac{1}{n} E((X_i - \mu)^2) = \frac{\sigma^2}{n}.$$

Veamos a continuación el tercer término de (6.15):

$E((\bar{X} - \mu)^2)$ es la varianza de \bar{X} , ya que

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{n\mu}{n} = \mu.$$

Pero, por otra parte,

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

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)

$$E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} n\sigma^2 \frac{n-1}{n} = \sigma^2.$$

▲

Observación 1.7 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.

4. Dar un ejemplo de dos variables aleatorias X e Y que no sean independientes, pero que, sin embargo,

$$E(XY) = E(X) E(Y),$$

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

$$E(XY) = E(X)E(Y).$$

Calculemos ambos miembros y veamos que dan lo mismo. El segundo es evidentemente igual a cero, ya que

$$E(X) = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0.$$

En cuanto al primero, observamos que $XY = X^3 = X$, ya que como X vale $1, 0$ ó -1 , X^3 es siempre igual a X . En conclusión

$$E(XY) = E(X^3) = E(X) = 0.$$

▲

Observación 1.8 En el enunciado se señala que si $E(XY) = E(X)E(Y)$, entonces X e Y son variables aleatorias no-correlacionadas. Esto se deduce de la Propiedad 9 de la sección 1.21.1.

1.24. Esperanza Matemática de Variables con Densidad.

Definición 1.20 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

$$\int_{-\infty}^{\infty} |x|f_X(x)dx \tag{1.47}$$

es finita, y en este caso $E(X)$ se define mediante la fórmula

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx. \tag{1.48}$$

Veamos a continuación algunos ejemplos de aplicación de la fórmula (1.48).

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

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b, \\ 0 & \text{si } x \leq a \text{ ó } x \geq b. \end{cases}$$

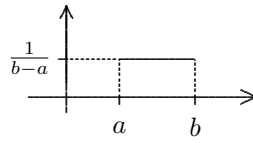


Figura 1.27: Densidad de la distribución uniforme

Tenemos

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \frac{1}{2} (b^2 - a^2) = \frac{b+a}{2} \end{aligned}$$

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

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx.$$

Integrando por partes

$$\begin{aligned} E(X) &= [-e^{-\lambda x} x]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \end{aligned}$$

es decir $E(X) = 1/\lambda$.

2. *Distribución Normal.* Si X tiene distribución normal con parámetros (μ, σ^2) , su densidad es

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

Se verifica fácilmente que

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty,$$

y efectuando el cambio de variables $u = (x - \mu)/\sigma$, resulta

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right\} dx \\ &= \int_{-\infty}^{+\infty} (\sigma u + \mu) \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-u^2/2} du + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du. \end{aligned}$$

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 $E(X) = \mu$.

1.25. Cambio de Variables. Momentos de Orden Superior.

Sea $f(x)$ la densidad de la variable aleatoria X y

$$g : \mathbb{R} \rightarrow \mathbb{R}$$

una función medible. Consideremos la variable aleatoria

$$Y = g(X).$$

Entonces, el valor esperado de Y se puede calcular mediante la fórmula

$$E(Y) = \int_{\mathbb{R}} g(x)f(x)dx \quad (1.49)$$

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 $g : \mathbb{R} \rightarrow \mathbb{R}$ es la función

$$g(x) = x^k,$$

entonces

$$E(g(X)) = E(X^k)$$

se denomina el *momento de orden k de la variable aleatoria X* , que es finito si y sólo si

$$E(|X|^k) < \infty.$$

Si X tiene densidad f_X y si

$$\int_{-\infty}^{+\infty} |x|^k f_X(x)dx < \infty,$$

entonces la fórmula (1.49) se traduce en

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx.$$

Al igual que en el caso discreto definimos el *momento centrado de orden m* de X para $m \geq 2$ mediante

$$\mu_m = E((X - E(X))^m) = \int_{-\infty}^{+\infty} (x - E(X))^m f_X(x) dx,$$

suponiendo que la integral exista.

Las definiciones de *varianza*, *desviación típica*, *covarianza* y *correlación* son iguales a las del caso discreto.

Observación 1.9 Las propiedades que demostramos para los momentos de orden superior en la sección 1.22.1 para el caso discreto también son ciertas para el caso continuo y aún 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

$$E(X) = \frac{a+b}{2}.$$

Calculemos

$$\begin{aligned} \text{Var}(X) &= E[(X - E(X))^2] = \int_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 f_X(x) dx \\ &= \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{1}{b-a} \frac{1}{3} \left(x - \frac{a+b}{2}\right)^3 \Big|_{x=a}^{x=b} \\ &= \frac{1}{3(b-a)} \left[\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3 \right] = \frac{(b-a)^2}{12} \end{aligned}$$

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

$$E(X^k) = \int_0^{+\infty} x^k \lambda e^{-\lambda x} dx = -e^{-\lambda x} x^k \Big|_0^{\infty} + k \int_0^{\infty} x^{k-1} e^{-\lambda x} dx$$

de donde obtenemos

$$E(X^k) = \frac{k}{\lambda} E(X^{k-1}).$$

Procediendo de manera inductiva se deduce de aquí que

$$E(X^k) = \frac{k}{\lambda} \frac{k-1}{\lambda} \cdots \frac{1}{\lambda} E(X^0) = \frac{k!}{\lambda^k}$$

3. *Distribución Normal.* Sea X una variable aleatoria con distribución normal de parámetros $(0, 1)$, es decir que su densidad es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{para } x \in \mathbb{R}.$$

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

$$E(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (1.50)$$

Si k es impar, como el integrando resulta ser una función impar, la integral vale cero y $E(X^k) = 0$.

Si k es un número par, digamos $k = 2p$, p entero positivo, integrando por partes en el segundo miembro de (1.50) se obtiene

$$\begin{aligned} E(X^{2p}) &= \frac{1}{\sqrt{2\pi}} \left[-x^{2p-1} e^{-x^2/2} \Big|_{-\infty}^{+\infty} + (2p-1) \int_{-\infty}^{+\infty} x^{2p-2} e^{-x^2/2} dx \right] \\ &= (2p-1) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^{2p-2} e^{-x^2/2} dx, \end{aligned}$$

y por lo tanto

$$E(X^{2p}) = (2p - 1)E(X^{2p-2}).$$

Aplicando ahora un procedimiento inductivo, se tiene

$$E(X^{2p}) = (2p - 1)(2p - 3) \cdots 3 E(X) = \frac{(2p)!}{2^p p!} \quad (1.51)$$

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

$$E((X - \mu)^k) = \int_{-\infty}^{+\infty} (x - \mu)^k \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} dx.$$

Haciendo el cambio de variables $t = (x - \mu)/\sigma$, resulta $dx = \sigma dt$ y

$$E((X - \mu)^k) = \int_{-\infty}^{+\infty} (\sigma t)^k \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2} \sigma dt = \sigma^k \int_{-\infty}^{+\infty} t^k \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \sigma dt$$

y la integral de la derecha es la que venimos de calcular. Por lo tanto

$$\begin{aligned} k \text{ impar} &\Rightarrow E((X - \mu)^k) = 0 \\ k = 2p \text{ par} &\Rightarrow E((X - \mu)^{2p}) = \sigma^{2p} \frac{(2p)!}{2^p p!} \end{aligned}$$

En especial, si $k = 2$ tenemos $\text{Var}(X) = E((X - \mu)^2) = \sigma^2$.