

Estadística para Informática

Versión 4.0

Félix Rincón de Rojas

Departamento de Matemática Aplicada a las Tecnologías de la Información y las
Comunicaciones
Escuela Técnica Superior de Ingeniería de Sistemas Informáticos
Universidad Politécnica de Madrid

Espacios probabilísticos

1.1 Fenómenos aleatorios

Como se ha dicho en el prólogo, la estadística inferencial observa lo que ocurre en una muestra y extiende los resultados a toda la población. Por este motivo:

- Trabaja en condiciones de *incertidumbre*, pues no dispone de toda la información.
- Comete un error, que se mide en términos de *probabilidad*, como corresponde a un *fenómeno aleatorio*.

Vamos a formalizar este concepto.

Definición 1.1.1 (Fenómenos deterministas y aleatorios)

Un fenómeno es *determinista* cuando, repetido en las mismas condiciones, produce siempre el mismo resultado. En caso contrario, se llama *aleatorio*. ◀

Ejemplo 1.1.2 (Dado de quinelas)

Un dado de quinelas tiene tres caras marcadas con 1, dos caras marcadas con \times y una cara marcada con 2. Se deja caer varias veces desde la misma altura y se observa:

- El tiempo que tarda en llegar al suelo. Siempre es el mismo, luego este fenómeno es determinista.
- El resultado obtenido. Varía en cada lanzamiento, luego este fenómeno es aleatorio.

En el segundo experimento, el conjunto de resultados posibles es $E := \{1, \times, 2\}$, que tiene 3 elementos. Algunos sucesos relevantes son estos dos:

$$A : \text{“el equipo local gana”} = \{1\} \quad \text{y} \quad B : \text{“el equipo visitante gana”} = \{2\}.$$

Otros sucesos, relacionados con estos, son:

- “el equipo local no gana” = $\{x, 2\} = \bar{A}$.
- “el equipo local no pierde” = $\{1, x\} = \bar{B} = \text{“no gana el visitante”}$.
- “algún equipo gana” = $\{1, 2\} = A \cup B$.

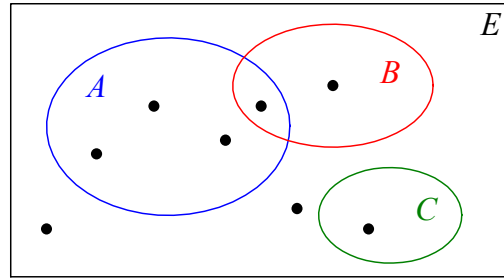


Figura 1.1: Espacios muestral y de sucesos.

- “los dos equipos ganan” = $\emptyset = A \cap B$.
- “algún equipo no gana” = $\{1, \times, 2\} = \overline{A} \cup \overline{B} = E$.
- “algún equipo gana, pero no es el local” = $\{2\} = (A \cup B) - A = B$.

Todos estos sucesos se obtienen a partir de A y B por medio de operaciones conjuntistas. ◀

La siguiente definición generaliza lo observado en el ejemplo anterior.

Definición 1.1.3 (Espacio muestral y sucesos)

Dado un fenómeno aleatorio, se definen (figura 1.1):

- *Espacio muestral* es el conjunto de posibles resultados del experimento. Se representa $E := \{a, b, c, \dots\}$.
- *Sucesos* son los subconjuntos de E . Se representan $A, B, C, \dots \subseteq E$. En particular:
 - *Suceso seguro* es el espacio muestral, E .
 - *Suceso imposible* es el conjunto vacío, \emptyset .
 - *Sucesos elementales* son los subconjuntos unitarios de E . Se representan $\{a\}, \{b\}, \{c\}, \dots$
 - *Suceso contrario* del suceso A es su complementario, \overline{A} .
 - *Sucesos incompatibles* son dos sucesos A y B *disjuntos* ($A \cap B = \emptyset$).
- *Espacio de sucesos* es el conjunto de todos los sucesos¹. Se representa S . ◀

Ejemplo 1.1.4 (Diversos espacios muestrales)

Los siguientes fenómenos aleatorios tienen asociados los espacios muestrales que se indican:

1. Lanzar una moneda. Representando cara por c y cruz por $+$, $E := \{c, +\}$, que es finito.
2. Lanzar una moneda tres veces.

$E := \{ccc, cc+, c+c, c++, +cc, +c+, ++c, +++\}$, que es finito.

¹Con esta definición, $S = \mathcal{P}(E)$, cuyo cardinal es $\text{card}(\mathcal{P}(E)) = 2^{\text{card}(E)}$. En ciertos casos, este conjunto es demasiado grande, por lo que se adopta una definición más restrictiva de espacio de sucesos. Puede verse en Fernández-Abascal [8, página 28, Definición 2.3].

3. Contar el número de trabajos pendientes en una cola de impresión en un periodo fijo de tiempo. $E := \{0, 1, 2, 3, \dots\} = \mathbb{N}$, que es numerable.
4. Medir el tiempo transcurrido entre dos caídas consecutivas del sistema operativo de un ordenador. $E := [0, \infty)$, que es continuo. ◀

En el ejemplo anterior se observa que un espacio muestral puede ser:

- *Discreto*, es decir, finito o numerable.
- *Continuo*, es decir, unión de intervalos de \mathbb{R} .

1.2 Probabilidad

Al lanzar un dado de quinielas, parece *más probable* el suceso A : “obtener 1” que el suceso B : “obtener 2”. Pero, ¿cuánto de probable es cada uno? La probabilidad debe medir la incertidumbre de un fenómeno aleatorio, asignando un número a cada suceso. Esto se puede hacer de dos maneras:

- Enfoque axiomático. El dado tiene 6 caras, que se pueden considerar equiprobables. De ellas, 3 son favorables a A y 1 es favorable a B . Por tanto, $p(A) := 3/6 = 1/2$ y $p(B) := 1/6$.
- Enfoque experimental. Se lanza el dado muchas veces. A la larga, las frecuencias relativas de ambos sucesos se estabilizan y sus valores son, aproximadamente, $f_A \simeq 1/2$ y $f_B \simeq 1/6$. Por tanto, $p(A) := 1/2$ y $p(B) := 1/6$.

En este tema se define la probabilidad formalmente y se extraen sus propiedades. El enfoque es axiomático, pero apoyado en el método experimental: si al lanzar un dado de quinielas 600 veces el 2 solo aparece 10 veces, habrá que pensar que el dado está cargado y asignar $p(B) = 10/600 = 1/60$ en lugar de $1/6$.

La probabilidad de un suceso se puede interpretar como su frecuencia relativa cuando el experimento aleatorio se repite muchas veces. Por tanto, un modelo axiomático de probabilidad debe ajustarse a las propiedades de las frecuencias relativas.

Ejemplo 1.2.1 (Moneda y $S = \mathcal{P}(E)$)

Al lanzar una moneda, el espacio muestral es $E := \{c, +\}$ y el espacio de sucesos es $S := \{\emptyset, \{c\}, \{+\}, E\} = \mathcal{P}(E)$. ¿Qué probabilidad se debe asignar a cada suceso? Tres posibilidades son:

S	p_1	p_2	p_3
\emptyset	0	0	0
$\{c\}$	1/2	2/3	2/3
$\{+\}$	1/2	1/2	1/3
E	1	1	1

Con la función p_1 cara y cruz tienen probabilidad $1/2$. Esto concuerda con el enfoque axiomático, luego p_1 es una probabilidad válida. Sin embargo, si en muchos lanzamientos de la

moneda ha salido cara el doble de veces que cruz, se debe sospechar que dicha moneda no está equilibrada. En tal caso, es razonable asignar a cara probabilidad $2/3$, como hace p_2 . Sin embargo, ya no es razonable asignar a cruz probabilidad $1/2$, porque entonces

$$p_2(\{c\}) + p_2(\{+\}) = \frac{2}{3} + \frac{1}{2} = \frac{7}{6} > 1.$$

Esto contraviene la propiedad de que las frecuencias relativas suman 1, luego p_2 no es válida como probabilidad. Lo que se debe hacer es asignar a cruz probabilidad $1/3$, como hace p_3 , que también es una probabilidad válida. ◀

Puesto que la probabilidad es una medida de la incertidumbre, su asignación inicial depende de la información disponible. Dos observadores pueden asignar probabilidades muy diferentes al mismo experimento, ambas correctas desde sus respectivos puntos de vista. Sin embargo, una probabilidad no puede ser cualquier cosa y para garantizar la coherencia están los axiomas.

Definición 1.2.2 (Espacio probabilístico)

Sean E un espacio muestral y S un espacio de sucesos. Una *probabilidad* sobre S es una función $p : S \rightarrow \mathbb{R}$ que satisface los tres *axiomas de Kolmogorov*:²

- A1. Es *positiva*, es decir, la probabilidad de cualquier suceso es mayor o igual que 0: $p(A) \geq 0$ para todo $A \in S$.
- A2. Está *normalizada*, es decir, la probabilidad del suceso seguro es 1: $p(E) = 1$.
- A3. Es *aditiva*, es decir, si dos sucesos son incompatibles, entonces la probabilidad de su unión es la suma de sus probabilidades:³

$$A, B \in S, A \cap B = \emptyset \Rightarrow p(A \cup B) = p(A) + p(B).$$

La terna (E, S, p) se llama *espacio probabilístico* (figura 1.2). ◀

Ejemplo 1.2.3 (Dado de quinielas)

Se lanza un dado de quinielas y se observa el resultado. El espacio muestral es $E := \{1, \times, 2\}$

2



Andrey Nikolaevich Kolmogorov (Tambov, Rusia, 1903 – Moscú, URSS, 1987).

Matemático conocido por sus trabajos en probabilidad, topología, lógica intuicionista, turbulencia, mecánica clásica y complejidad computacional.

Kolmogorov estudió y enseñó en la Universidad Estatal de Moscú, donde llegó a ser catedrático y decano.

En 1933, Kolmogorov publicó el libro *Foundations of the Theory of Probability*, donde establece los axiomas de la probabilidad.

³Cuando el espacio muestral E es infinito, el axioma A3 debe extenderse a uniones numerables de sucesos disjuntos, es decir, dada una sucesión de sucesos incompatibles dos a dos, la probabilidad de su unión es la suma de sus probabilidades:

$$\left. \begin{array}{l} A_1, A_2, \dots, A_n, \dots \in S \\ A_i \cap A_j = \emptyset \text{ cuando } i \neq j \end{array} \right\} \Rightarrow p\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} p(A_n).$$

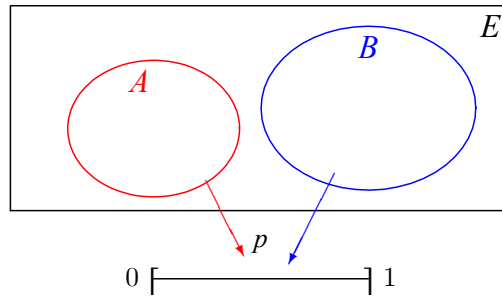


Figura 1.2: Espacio probabilístico.

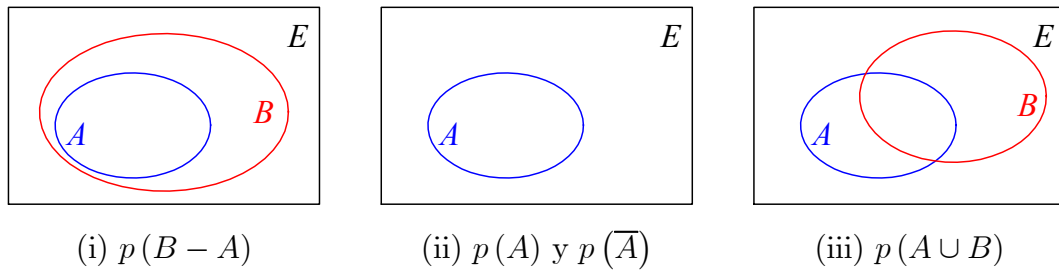


Figura 1.3: Propiedades de la probabilidad.

y el espacio de sucesos es $S := \mathcal{P}(E)$, que tiene $2^3 = 8$ elementos. Es razonable definir las probabilidades de los sucesos elementales como:

$$p(\{1\}) = 1/2, \quad p(\{\times\}) = 1/3 \quad \text{y} \quad p(\{2\}) = 1/6,$$

comprobando que suman $1/2 + 1/3 + 1/6 = 1$. Esta probabilidad se extiende a los demás sucesos aplicando el axioma 3. Por ejemplo:

$$A : \text{“el equipo local no pierde”} = \{1, \times\} = \{1\} \cup \{\times\} \Rightarrow p(A) = p(\{1\}) + p(\{\times\}) = 1/2 + 1/3 = 5/6.$$

En resumen:

S	\emptyset	$\{1\}$	$\{\times\}$	$\{2\}$	$\{1, \times\}$	$\{1, 2\}$	$\{\times, 2\}$	E
p	0	1/2	1/3	1/6	5/6	2/3	1/2	1

A partir de los axiomas de probabilidad se obtienen varias propiedades, que serán útiles para calcular la probabilidad de sucesos complejos.

Teorema 1.2.4 (Propiedades de la probabilidad)

Sea (E, S, p) un espacio probabilístico. Para todo $A, B \in S$ se verifican:

(i) La probabilidad es creciente: si $A \subseteq B$, entonces:

$$p(A) \leq p(B) \quad \text{y} \quad p(B - A) = p(B) - p(A).$$

(ii) $p(A) \leq 1$.

(iii) Regla del complementario: $p(\bar{A}) = 1 - p(A)$.

(iv) $p(\emptyset) = 0$.

(v) Regla de la unión: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

Demostración: Apartado (i)

Si $A \subseteq B$, entonces A y $B - A$ particionan B (figura 1.3.i). Por el axioma de aditividad, se verifica:

$$\left. \begin{array}{l} B = A \cup (B - A) \\ A \cap (B - A) = \emptyset \end{array} \right\} \Rightarrow p(B) = p(A) + p(B - A) \Rightarrow p(B - A) = p(B) - p(A).$$

Por el axioma de positividad, $p(B - A) \geq 0$. Por tanto, $p(A) \leq p(B)$.

Apartados (ii) y (iii)

Aplicando la propiedad (i) a $A \subseteq E$ (figura 1.3.ii), se verifica:

$$p(A) \leq p(E) = 1 \quad \text{y} \quad p(\overline{A}) = p(E - A) = p(E) - p(A) = 1 - p(A).$$

Apartado (iv)

Aplicando la propiedad (iii), se verifica que $p(\emptyset) = p(\overline{E}) = 1 - p(E) = 1 - 1 = 0$.

Apartado (v)

Por una parte, A y $B - A$ particionan $A \cup B$ (figura 1.3.iii), luego:

$$\left. \begin{array}{l} A \cup B = A \cup (B - A) \\ A \cap (B - A) = \emptyset \end{array} \right\} \Rightarrow p(A \cup B) = p(A) + p(B - A).$$

Por otro lado, $A \cap B$ y $B - A$ particionan B (figura 1.3.iii), luego:

$$\left. \begin{array}{l} B = (A \cap B) \cup (B - A) \\ (A \cap B) \cap (B - A) = \emptyset \end{array} \right\} \Rightarrow p(B) = p(A \cap B) + p(B - A).$$

Despejando $p(B - A)$ en la segunda igualdad y sustituyendo en la primera, resulta que:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B). \quad \blacktriangleleft$$

Ejemplo 1.2.5 (Dado y regla de la unión)

Se lanza un dado y se toman como espacio muestral $E := \{1, 2, 3, 4, 5, 6\}$ y como espacio de sucesos $S := \mathcal{P}(E)$, que tiene $2^6 = 64$ elementos. Se puede suponer que los sucesos elementales son equiprobables, es decir:

$$p(\{a\}) = 1/6 \text{ para todo } a \in E.$$

La probabilidad de cualquier otro suceso se calcula aplicando el axioma de aditividad. Por ejemplo:

$$\begin{aligned} A : \text{“obtener un número par”} &= \{2, 4, 6\} = \{2\} \cup \{4\} \cup \{6\}, \\ B : \text{“obtener un número mayor que 2”} &= \{3, 4, 5, 6\} = \{3\} \cup \{4\} \cup \{5\} \cup \{6\}, \\ p(A) &= p(\{2\}) + p(\{4\}) + p(\{6\}) = 3 \cdot 1/6 = 1/2, \\ p(B) &= p(\{3\}) + p(\{4\}) + p(\{5\}) + p(\{6\}) = 4 \cdot 1/6 = 2/3. \end{aligned}$$

El suceso “obtener par y mayor que 2” coincide con $A \cap B$. Su probabilidad vale:

$$p(A \cap B) = p(\{4, 6\}) = 2 \cdot 1/6 = 1/3.$$

El suceso “obtener par o mayor que 2” coincide con $A \cup B$. Aplicando la regla de la unión, su probabilidad vale:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{1}{2} + \frac{2}{3} - \frac{1}{3} = \frac{3 + 4 - 2}{6} = \frac{5}{6}. \quad \blacktriangleleft$$

1.3 Espacios muestrales discretos

Si E es un espacio muestral discreto, entonces se puede tomar $S = \mathcal{P}(E)$ y cada suceso es unión finita o numerable de sucesos elementales. Para definir una probabilidad sobre S , basta hacerlo sobre cada suceso elemental y comprobar que su suma vale 1. La probabilidad de cualquier otro suceso se calcula aplicando el axioma 3.

Teorema 1.3.1 (Regla de Laplace)

Sea E un espacio muestral equiprobable. Para todo $A \in S = \mathcal{P}(E)$ se verifica:

$$p(A) = \frac{\text{card}(A)}{\text{card}(E)} = \frac{\text{número de casos favorables a } A}{\text{número de casos posibles}}.$$

Esta igualdad se llama regla de Laplace⁴.

Demostración: Si E es equiprobable, entonces todos los sucesos elementales tienen la misma probabilidad. Por tanto, E es finito y para todo $a \in E$ se verifica:

$$p(\{a\}) = \frac{1}{\text{card}(E)}.$$

Por otra parte, todo subconjunto de E es unión finita de sucesos elementales:

$$A = \{a_1, a_2, \dots, a_k\} = \{a_1\} \cup \{a_2\} \cup \dots \cup \{a_k\},$$

luego $A \in S$ y $S = \mathcal{P}(E)$. Además, los sucesos elementales son disjuntos dos a dos, luego:

$$p(A) = p(\{a_1\}) + \dots + p(\{a_k\}) = k \frac{1}{\text{card}(E)} = \frac{\text{card}(A)}{\text{card}(E)}. \quad \blacktriangleleft$$

4



Pierre-Simon de Laplace (Beaumont-en-Auge, Francia, 1749 – París, Francia, 1827).

Laplace ocupó puestos relevantes en la *Académie des Sciences* y la *École Normale* francesas.

Laplace es conocido por sus trabajos sobre ecuaciones en diferencias y diferenciales, probabilidad, mecánica celeste y teoría del calor. Su *Théorie Analytique des Probabilités* contiene una definición de probabilidad y sus aplicaciones en temas tan diversos como: errores de observaciones, determinación de las masas de Júpiter, Saturno y Urano, métodos de triangulación en topografía y problemas de geodesia, particularmente la determinación del meridiano de Francia.

Ejemplo 1.3.2 (Reglas de Laplace y de la unión)

De una baraja española de 40 cartas, se extrae una al azar. Se trata de calcular la probabilidad de obtener oro o figura. Para ello, se consideran los sucesos:

$$A : \text{“obtener oro”} \quad \text{y} \quad B : \text{“obtener figura”}.$$

Se trata de calcular $p(A \cup B)$. El espacio muestral consta de 40 resultados equiprobables. Aplicando la regla de Laplace, se verifica:

$$p(A) = \frac{10}{40} = \frac{1}{4}, \quad p(B) = \frac{12}{40} = \frac{3}{10}, \quad p(A \cap B) = \frac{3}{40}.$$

Aplicando ahora la regla de la unión, se obtiene:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{10}{40} + \frac{12}{40} - \frac{3}{40} = \frac{19}{40}. \quad \blacktriangleleft$$

Ejemplo 1.3.3 (Falacia con la regla de Laplace)

Se lanzan dos monedas y se considera el espacio muestral:

$$E = \{\text{dos caras, una cara y una cruz, dos cruces}\}.$$

Se considera el suceso:

$$A : \text{“obtener al menos una cara”} = \{\text{dos caras, una cara y una cruz}\}.$$

Según la regla de Laplace, parece que la probabilidad de A vale:

$$p(A) = \frac{\text{card}(A)}{\text{card}(E)} = \frac{2}{3}.$$

Sin embargo, las probabilidades de los sucesos elementales son:

E	dos caras	una cara y una cruz	dos cruces
p	1/4	2/4 = 1/2	1/4

Por tanto, dichos sucesos no son equiprobables y la regla de Laplace no es aplicable. De hecho:

$$p(A) = p(\text{dos caras}) + p(\text{una cara y una cruz}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$

Por el contrario, el espacio muestral $E_2 = \{cc, c+, +c, ++\}$ es equiprobable, luego:

$$p(A) = p(\{cc, c+, +c\}) = \frac{\text{card}(A)}{\text{card}(E_2)} = \frac{3}{4},$$

porque ahora sí es aplicable la regla de Laplace. ◀

Ejemplo 1.3.4 (Espacio de sucesos numerable)

Se lanza una moneda hasta que sale cruz. El espacio muestral es:

$$E := \{+, c+, cc+, ccc+, \dots, \overbrace{cc \dots c}^n+, \dots\}.$$

Este conjunto es numerable. El espacio de sucesos puede ser $S := \mathcal{P}(E)$, que también es numerable y contiene sucesos tanto finitos como infinitos. Por ejemplo:

A : “obtener menos de 3 caras” = $\{+, c+, cc+\}$ es finito,

B : “obtener un número par de caras” = $\{+, cc+, cccc+, \dots\}$ es infinito.

Como el espacio muestral es numerable, basta asignar probabilidad a los sucesos elementales, que ahora no pueden ser equiprobables:

$$\begin{array}{c|cccccc} E & + & c+ & cc+ & \dots & \overbrace{cc\dots c}^n+ & \dots \\ \hline p & 1/2 & 1/4 & 1/8 & \dots & 1/2^{n+1} & \dots \end{array}$$

Como comprobación, estas cantidades forman una progresión geométrica de razón $1/2$, luego su suma vale:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^{n+1}} + \dots = \frac{1/2}{1 - 1/2} = 1.$$

Algunos ejemplos de probabilidades son:

$$p(A) = p(\{+\}) + p(\{c+\}) + p(\{cc+\}) = 1/2 + 1/4 + 1/8 = 7/8,$$

$$p(B) = 1/2 + 1/8 + 1/32 + \dots = \frac{1/2}{1 - 1/4} = \frac{1/2}{3/4} = 2/3.$$

En el segundo caso, se suma una progresión geométrica de razón $1/4$. ◀

1.4 Probabilidad condicionada

La probabilidad es una medida de la incertidumbre y por tanto depende de la información sobre el fenómeno aleatorio disponible en cada momento.

Ejemplo 1.4.1 (Sucesos condicionados)

Si se lanza una moneda 2 veces, el espacio muestral es $E := \{cc, c+, +c, ++\}$. Se considera el suceso:

$$A : \text{“obtener dos resultados iguales”} = \{cc, ++\}.$$

y vamos a calcular su probabilidad en varias situaciones:

- Antes de lanzar la moneda, los 4 elementos de E son equiprobables (figura 1.4.i), luego $p(A) = 2/4 = 1/2$ por la regla de Laplace.
- Si la moneda ya ha sido lanzada las 2 veces, pero se ignora el resultado, la incertidumbre sobre A es la misma. Por tanto, $p(A)$ sigue valiendo $1/2$.
- Por el contrario, si se sabe que al menos una de las 2 tiradas ha salido cruz, entonces la incertidumbre cambia (figura 1.4.ii), pues ha ocurrido el suceso:

$$B : \text{“otener al menos una cruz”} = \{c+, +c, ++\}.$$

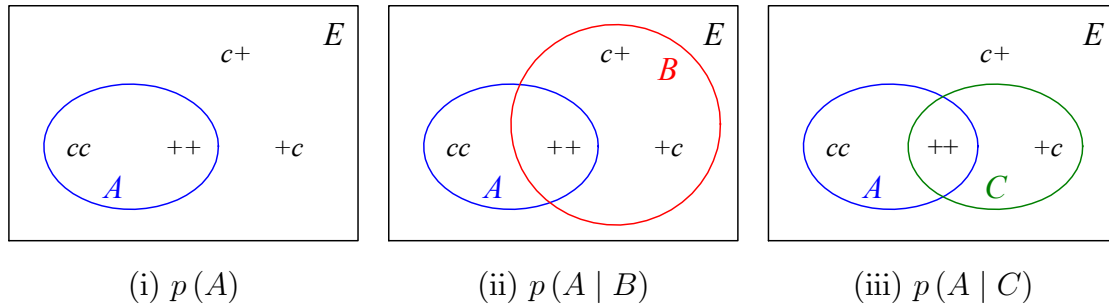
(i) $p(A)$ (ii) $p(A | B)$ (iii) $p(A | C)$

Figura 1.4: Probabilidades condicionadas.

De los 3 resultados que hay en B solo 1 es favorable a A . Por tanto, la nueva probabilidad de A vale:

$$\frac{\text{card}(A \cap B)}{\text{card}(B)} = \frac{1}{3}.$$

Esta probabilidad se calcula sabiendo que ha ocurrido B . Para indicarlo, se escribe $p(A | B)$, que se lee probabilidad de A condicionada a B . En este caso,

$$p(A | B) = \frac{1}{3} = \frac{\text{card}(A \cap B) / \text{card}(E)}{\text{card}(B) / \text{card}(E)} = \frac{p(A \cap B)}{p(B)}.$$

Es decir, la probabilidad de A condicionada a B es la proporción entre la probabilidad de la parte de B favorable a A (es decir, $A \cap B$) y la probabilidad de B .

- De modo análogo, si se sabe que en la primera tirada ha salido cruz, entonces ha ocurrido el suceso:

$$C : \text{“obtener cruz en la primera tirada”} = \{+c, ++\}.$$

La probabilidad de A condicionada a C vale (figura 1.4.iii):

$$p(A | C) = \frac{\text{card}(A \cap C)}{\text{card}(C)} = \frac{1}{2} = \frac{\text{card}(A \cap C) / \text{card}(E)}{\text{card}(C) / \text{card}(E)} = \frac{p(A \cap C)}{p(C)}.$$

Se observa que $p(A | B) \neq p(A)$, pero $p(A | C) = p(A)$. Más adelante se analizará el motivo. ◀

La probabilidad condicionada mide la incertidumbre cuando se dispone de información parcial sobre el resultado de un fenómeno aleatorio. El ejemplo anterior motiva el siguiente resultado.

Teorema 1.4.2 (Probabilidad condicionada)

Sea (E, S, p) un espacio probabilístico. Sea $B \in S$ un suceso de probabilidad no nula. La función $p(\cdot | B) : S \rightarrow \mathbb{R}$ tal que para todo $A \in S$ vale

$$p(A | B) = \frac{p(A \cap B)}{p(B)}$$

es otra probabilidad sobre S , llamada probabilidad condicionada a B .

Demostración: Se trata de verificar los tres axiomas de Kolmogorov:

1. $p(A | B) = p(A \cap B) / p(B) \geq 0$ para todo $A \in S$, pues $p(A \cap B) \geq 0$ y $p(B) > 0$.
2. $p(E | B) = p(E \cap B) / p(B) = p(B) / p(B) = 1$.
3. Si $A_1, A_2 \in S$ con $A_1 \cap A_2 = \emptyset$, entonces:

$$\begin{aligned} p(A_1 \cup A_2 | B) &= \frac{p((A_1 \cup A_2) \cap B)}{p(B)} = \frac{p((A_1 \cap B) \cup (A_2 \cap B))}{p(B)} = \\ &= \frac{p(A_1 \cap B) + p(A_2 \cap B)}{p(B)} = p(A_1 | B) + p(A_2 | B), \end{aligned}$$

ya que $(A_1 \cap B) \cap (A_2 \cap B) \subseteq A_1 \cap A_2 = \emptyset$. ◀

La *regla de la intersección* calcula la probabilidad de una intersección usando probabilidad condicionada:

$$p(A \cap B) = p(B) p(A | B) = p(A) p(B | A).$$

Ejemplo 1.4.3 (Regla de la intersección)

Se saca una carta al azar de una baraja española y se considera el suceso “obtener oro y figura”. Dicho suceso es $A \cap B$, donde:

A : “obtener oro” y B : “obtener figura”.

La probabilidad de $A \cap B$ se puede calcular de dos maneras:

- Por la regla de Laplace. El espacio muestral consta de 40 elementos equiprobables y hay 3 casos favorables. Por tanto, $p(A \cap B) = 3/40$.
- Por la regla de la intersección:

$$\begin{aligned} p(A \cap B) &= p(A) p(B | A) = \frac{10}{40} \cdot \frac{3}{10} = \frac{3}{40} \\ &= p(B) p(A | B) = \frac{12}{40} \cdot \frac{3}{12} = \frac{3}{40}. \end{aligned} \quad \blacktriangleleft$$

La regla de la intersección se generaliza a más de dos sucesos.

Corolario 1.4.4 (Regla de la intersección)

Sea (E, S, p) es un espacio probabilístico. Si $A_1, A_2, \dots, A_n \in S$ son sucesos tales que

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

entonces:

$$\begin{aligned} p(A_1 \cap A_2 \cap \dots \cap A_n) &= p(A_1) p(A_2 | A_1) p(A_3 | A_1 \cap A_2) p(A_4 | A_1 \cap A_2 \cap A_3) \dots \\ &\dots p(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned}$$

Demostración: Teniendo en cuenta que:

$$p(A_1) \geq p(A_1 \cap A_2) \geq p(A_1 \cap A_2 \cap A_3) \geq \dots \geq p(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0,$$

todas las probabilidades condicionadas del segundo miembro de la igualdad tienen sentido. Desarrollando sus valores, se obtiene:

$$p(A_1) \frac{p(A_2 \cap A_1)}{p(A_1)} \frac{p(A_3 \cap A_1 \cap A_2)}{p(A_1 \cap A_2)} \frac{p(A_4 \cap A_1 \cap A_2 \cap A_3)}{p(A_1 \cap A_2 \cap A_3)} \cdots \frac{p(A_n \cap A_1 \cap A_2 \cap \cdots \cap A_{n-1})}{p(A_1 \cap A_2 \cap \cdots \cap A_{n-1})}.$$

Simplificando esta expresión, se obtiene el primer miembro de la igualdad. ◀

Ejemplo 1.4.5 (Paradoja del cumpleaños)

Se consideran los cumpleaños de n personas. Se trata de calcular la probabilidad del suceso:

A_n : “en un grupo de n personas, al menos dos cumpleaños coinciden”.

Si $n > 365$, al menos dos cumpleaños coinciden, luego $p(A_n) = 1$. Si $n \leq 365$, se pueden suponer los cumpleaños equiprobables y aplicar las reglas del complementario y de la intersección:

$$p(A_n) = 1 - p(\overline{A_n}) = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - (n - 1)}{365} = 1 - \frac{365!}{365^n (365 - n)!}.$$

Esta probabilidad tiende a 1 rápidamente con n , como muestra la siguiente tabla:

n	2	10	20	30	40	50	57
$p(A_n)$	0.0027	0.12	0.41	0.71	0.89	0.97	0.99

1.5 Independencia de sucesos

Dados dos sucesos, A y B , puede que la probabilidad de A se mantenga cuando ocurre B (es decir, $p(A | B) = p(A)$) o que se altere.

Ejemplo 1.5.1 (Sucesos dependientes e independientes)

Se lanza una moneda 2 veces y se consideran los sucesos:

A : “obtener dos resultados iguales”,

B : “otener al menos una cruz”,

C : “obtener cruz en la primera tirada”.

Se ha visto en el ejemplo 1.4.1 que:

$$p(A | B) \neq p(A) \quad \text{y} \quad p(A | C) = p(A).$$

Por tanto, la probabilidad de A se altera cuando ocurre B , pero se mantiene cuando ocurre C . Se dice que A y B son *sucesos dependientes*, mientras que A y C son *sucesos independientes*. ◀

Observemos que:

$$p(A | B) = p(A) \Leftrightarrow \frac{p(A \cap B)}{p(B)} = p(A) \Rightarrow p(A \cap B) = p(A) p(B).$$

La última igualdad no exige que $p(B) > 0$ y es la que se emplea en la definición.

Definición 1.5.2 (Sucesos independientes)

Sea (E, S, p) un espacio probabilístico. Dos sucesos $A, B \in S$ son *independientes* si y solo si $p(A \cap B) = p(A) p(B)$. ◀

Si $p(A) > 0$, entonces la independencia equivale a $p(B | A) = p(B)$. Si $p(B) > 0$, entonces la independencia equivale a $p(A | B) = p(A)$.

Ejemplo 1.5.3 (Dado)

Si se lanza un dado, el espacio muestral es $E = \{1, 2, 3, 4, 5, 6\}$, que consta de 6 sucesos equiprobables. Se consideran los sucesos:

$$\begin{aligned} A : \text{“obtener un número par”} &= \{2, 4, 6\}, & p(A) &= 3/6 = 1/2, \\ B : \text{“otener un múltiplo de 3”} &= \{3, 6\}, & p(B) &= 2/6 = 1/3, \\ C : \text{“obtener un número menor que 5”} &= \{1, 2, 3, 4\}, & p(C) &= 4/6 = 2/3. \end{aligned}$$

La independencia de A y B se puede analizar de tres maneras:

$$\begin{aligned} p(A | B) &= \frac{p(A \cap B)}{p(B)} = \frac{1/6}{2/6} = \frac{1}{2} = p(A) \Rightarrow A \text{ y } B \text{ son independientes,} \\ p(B | A) &= \frac{p(B \cap A)}{p(A)} = \frac{1/6}{3/6} = \frac{1}{3} = p(B) \Rightarrow A \text{ y } B \text{ son independientes,} \\ p(A \cap B) &= \frac{1}{6} = p(A) p(B) \Rightarrow A \text{ y } B \text{ son independientes.} \end{aligned}$$

Por otro lado A y C verifican que:

$$p(A \cap C) = \frac{2}{6} = \frac{1}{3} = p(A) p(C) \Rightarrow A \text{ y } C \text{ son independientes.}$$

Finalmente, B y C verifican que:

$$p(B \cap C) = \frac{1}{6} \neq p(B) p(C) \Rightarrow B \text{ y } C \text{ son dependientes.} \quad \blacktriangleleft$$

Ejemplo 1.5.4 (Moneda y dado)

Se lanzan una moneda y un dado. El espacio muestral,

$$E = \{c1, c2, c3, c4, c5, c6, +1, +2, +3, +4, +5, +6\},$$

consta de 12 sucesos equiprobables. Se trata de hallar la probabilidad del suceso:

$$A : \text{“obtener cruz y múltiplo de 3”} = \{+3, +6\}.$$

Aplicando la regla de Laplace,

$$p(A) = \frac{2}{12} = \frac{1}{6}.$$

Como alternativa, se puede observar que A es la intersección de los sucesos:

$$B : \text{“obtener cruz”} \quad \text{y} \quad C : \text{“obtener múltiplo de 3”}.$$

Dichos sucesos son independientes, luego:

$$p(A) = p(B \cap C) = p(B) p(C) = \frac{1}{2} \frac{2}{6} = \frac{1}{6}.$$

Las probabilidades de B y C se calculan por la regla de Laplace. ◀

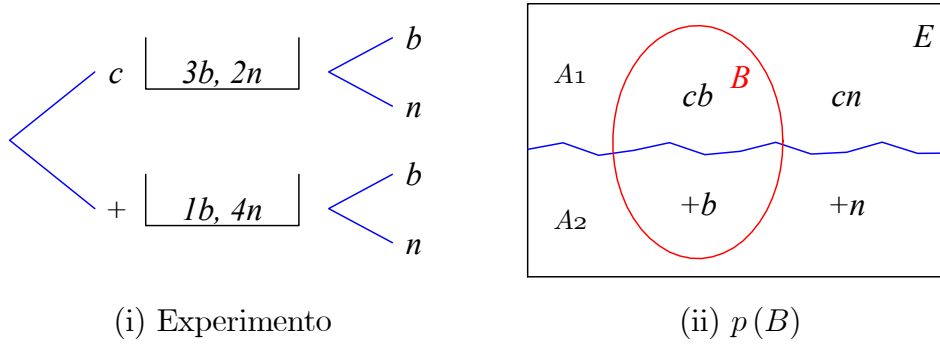


Figura 1.5: Moneda y urnas.

Cuando un fenómeno aleatorio consta de varias fases, cada una de las cuales no influye en las siguientes, se habla de *pruebas compuestas independientes*, como las del ejemplo anterior.

Teorema 1.5.5 (Propiedades de la independencia de sucesos)

Sea (E, S, p) un espacio probabilístico. Si $A, B \in S$ son sucesos independientes, entonces también son independientes las parejas de sucesos:

$$A \text{ y } \bar{B}, \quad \bar{A} \text{ y } B \quad \text{y} \quad \bar{A} \text{ y } \bar{B}.$$

Demostración: Veamos el primer caso. Puesto que $A \cap B \subseteq A$, se verifica que:

$$\begin{aligned} p(A \cap \bar{B}) &= p(A - A \cap B) = p(A) - p(A \cap B) = p(A) - p(A)p(B) = \\ &= p(A)(1 - p(B)) = p(A)p(\bar{B}). \end{aligned}$$

Por tanto, A y \bar{B} son independientes. ◀

1.6 Reglas de la probabilidad total y de Bayes

A veces, la probabilidad de un suceso se calcula con más facilidad si se obtiene primero su probabilidad condicionada a la ocurrencia de otros sucesos auxiliares que particionan el espacio muestral, tal como sucede en el siguiente ejemplo.

Ejemplo 1.6.1 (Moneda y urnas)

Se lanza una moneda. Si sale cara, se extrae una bola de una urna, que contiene 3 bolas blancas y 2 negras. Si sale cruz, la bola se extrae de otra urna, que contiene 1 bola blanca y 4 negras (figura 1.5.i). Se trata de hallar la probabilidad del suceso:

B : “extraer bola blanca”.

El espacio muestral asociado a este experimento es (figura 1.5 ii):

$$E := \{cb, cn, +b, +n\},$$

Sin embargo, los 4 sucesos elementales no son equiprobables. Por ejemplo, la proporción de bolas blancas en la primera urna ($3/5$) es mayor que en la segunda ($1/5$); por tanto, $p(cb)$ debe ser mayor que $p(+b)$.

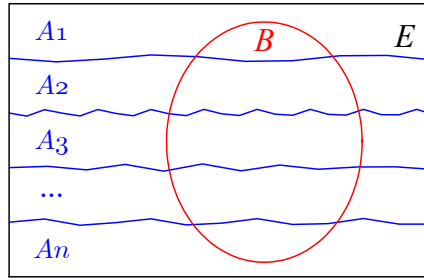


Figura 1.6: Regla de la probabilidad total.

Para calcular $p(B)$, se consideran los sucesos auxiliares:

$$A_1 : \text{“obtener cara”} \quad \text{y} \quad A_2 : \text{“obtener cruz”}.$$

Por la regla de Laplace, se verifica:

$$p(A_1) = p(A_2) = 1/2, \quad p(B | A_1) = 3/5 \quad \text{y} \quad p(B | A_2) = 1/5.$$

Además, A_1 y A_2 *particionan* el espacio muestral, es decir, $A_1 \cup A_2 = E$ y $A_1 \cap A_2 = \emptyset$ (figura 1.5.ii). Por tanto, aplicando el axioma de aditividad:

$$\left. \begin{array}{l} B = (A_1 \cap B) \cup (A_2 \cap B) \\ (A_1 \cap B) \cap (A_2 \cap B) = \emptyset \end{array} \right\} \Rightarrow p(B) = p(A_1 \cap B) + p(A_2 \cap B).$$

Aplicando ahora la regla de la intersección, se deduce que:

$$p(B) = p(A_1) p(B | A_1) + p(A_2) p(B | A_2) = \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{5} = \frac{2}{5}.$$

Por otra parte, si la bola extraída es blanca, es más probable que haya salido cara. Ambas probabilidades valen:

$$\begin{aligned} p(A_1 | B) &= \frac{p(A_1 \cap B)}{p(B)} = \frac{p(A_1) p(B | A_1)}{p(B)} = \frac{1/2 \cdot 3/5}{2/5} = 3/4, \\ p(A_2 | B) &= \frac{p(A_2) p(B | A_2)}{p(B)} = \frac{1/2 \cdot 1/5}{2/5} = 1/4 = p(\overline{A_1} | B) = 1 - p(A_1 | B). \quad \blacktriangleleft \end{aligned}$$

Generalizando lo observado en el ejemplo anterior, se obtiene el siguiente resultado.

Teorema 1.6.2 (Reglas de la probabilidad total y de Bayes)

Sean (E, S, p) un espacio probabilístico y $A_1, A_2, \dots, A_n, \dots \in S$ una sucesión de sucesos de probabilidad no nula que *particionan* E (figura 1.6). Para todo suceso $B \in S$ se verifican:

(i) Regla de la probabilidad total:

$$p(B) = p(A_1) p(B | A_1) + \dots + p(A_n) p(B | A_n) + \dots = \sum_{n=1}^{\infty} p(A_n) p(B | A_n).$$

(ii) Regla de Bayes: si $p(B) > 0$, entonces para todo $i = 1, 2, \dots$

$$p(A_i | B) = \frac{p(A_i) p(B | A_i)}{\sum_{n=1}^{\infty} p(A_n) p(B | A_n)}.$$

Demostración: Apartado (i)

Puesto que $A_1, A_2, \dots, A_n, \dots$ particionan E , se verifica que:

$$B = E \cap B = \left(\bigcup_{n=1}^{\infty} A_n \right) \cap B = \bigcup_{n=1}^{\infty} (A_n \cap B).$$

Además, si $i \neq j$, entonces:

$$(A_i \cap B) \cap (A_j \cap B) = (A_i \cap A_j) \cap B = \emptyset \cap B = \emptyset.$$

En consecuencia, aplicando las reglas de la unión y de la intersección,

$$p(B) = \sum_{n=1}^{\infty} p(A_n \cap B) = \sum_{n=1}^{\infty} p(A_n) p(B | A_n).$$

Apartado (ii)

Puesto que $p(B) > 0$, se verifica que:

$$p(A_i | B) = \frac{p(A_i \cap B)}{p(B)} = \frac{p(A_i) p(B | A_i)}{\sum_{n=1}^{\infty} p(A_n) p(B | A_n)},$$

donde se ha aplicado la regla de la intersección en el numerador y la regla de la probabilidad total en el denominador. ◀

Las reglas de la probabilidad total y de Bayes⁵ son útiles cuando el fenómeno aleatorio consta de dos etapas tales que:

- Se puede particionar el espacio muestral en sucesos $A_1, A_2, \dots, A_n, \dots$ correspondientes a la primera etapa y calcular sus probabilidades.
- El suceso B corresponde a la segunda etapa y es sencillo calcular, $p(B | A_n)$ para todo $n = 1, 2, \dots$

En tal caso, la regla de la probabilidad total proporciona $p(B)$ y la de Bayes $p(A_n | B)$.

5



Thomas Bayes (Londres, Inglaterra, 1702 – Tunbridge Wells, Kent, Inglaterra, 1761).

Bayes fue un clérigo que estudió lógica y teología en la universidad de Edinburgh, aunque llegó a ser miembro de la Royal Society.

Como matemático, su principal obra es *Essay towards solving a problem in the doctrine of chances*, publicada póstumamente. En ella, Bayes establece la regla que lleva su nombre. Además, analiza la distribución binomial, investiga la probabilidad condicionada y se ocupa de la evaluación de la función beta incompleta.

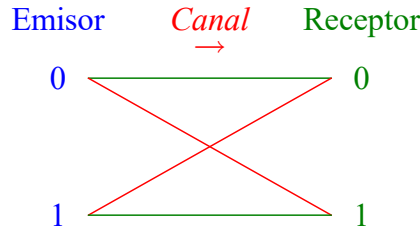


Figura 1.7: Canal binario con ruido y sin pérdidas.

Ejemplo 1.6.3 (Canal binario con ruido y sin pérdidas)

Un *canal binario con ruido* y sin pérdidas transporta bits (figura 1.7). A causa del ruido, cuando se emite un 0, el 10 % de las veces se recibe un 1, y cuando se emite un 1, el 5 % de las veces se recibe un 0. Además, se sabe que el 40 % de los bits emitidos son 0. La información sobre el canal se puede expresar mediante los sucesos siguientes:

$$\begin{aligned}
 A_0 &: \text{“emitir un 0”}, & p(A_0) &= 0.4, \\
 A_1 &: \text{“emitir un 1”}, & p(A_1) &= p(\overline{A_0}) = 0.6, \\
 B_0 &: \text{“recibir un 0”}, & p(B_0 | A_1) &= 0.05, \\
 B_1 &: \text{“recibir un 1”}, & p(B_1 | A_0) &= 0.1.
 \end{aligned}$$

¿Cuál es la probabilidad de recibir un 0? Usando la regla de la probabilidad total y la regla del complementario, se verifica:

$$\begin{aligned}
 p(B_0) &= p(A_0) p(B_0 | A_0) + p(A_1) p(B_0 | A_1) = \\
 &= p(A_0) (1 - p(B_1 | A_0)) + (1 - p(A_0)) p(B_0 | A_1) = \\
 &= 0.4 \cdot 0.9 + 0.6 \cdot 0.05 = 0.39.
 \end{aligned}$$

La probabilidad de recibir un 1 se puede hallar por el mismo método o bien usando la regla del complementario a partir de $p(B_0)$:

$$\begin{aligned}
 p(B_1) &= p(A_0) p(B_1 | A_0) + p(A_1) p(B_1 | A_1) = \\
 &= 0.4 \cdot 0.1 + 0.6 \cdot 0.95 = 0.61 = 1 - p(B_0).
 \end{aligned}$$

Si se ha recibido un 0, ¿qué probabilidad hay de que se haya emitido un 0? Este valor se obtiene con la regla de Bayes:

$$p(A_0 | B_0) = \frac{p(A_0 \cap B_0)}{p(B_0)} = \frac{p(A_0) p(B_0 | A_0)}{p(B_0)} = \frac{0.4 \cdot 0.9}{0.39} \simeq 0.923077.$$

Del mismo modo se calcula la probabilidad de haber emitido un 1 cuando se recibe un 1:

$$p(A_1 | B_1) = \frac{p(A_1 \cap B_1)}{p(B_1)} = \frac{p(A_1) p(B_1 | A_1)}{p(B_1)} = \frac{0.6 \cdot 0.95}{0.61} \simeq 0.934426.$$

Finalmente, ¿cuál es la probabilidad de que un bit emitido se reciba correctamente? Este suceso se puede expresar como unión de sucesos incompatibles:

$$C : \text{“un bit emitido se recibe correctamente”} = (A_0 \cap B_0) \cup (A_1 \cap B_1).$$

Por tanto, usando la aditividad, la regla de la intersección y la regla del complementario, se verifica:

$$\begin{aligned} p(C) &= p(A_0 \cap B_0) + p(A_1 \cap B_1) = p(A_0) p(B_0 | A_0) + p(A_1) p(B_1 | A_1) = \\ &= 0.4 \cdot 0.9 + 0.6 \cdot 0.95 = 0.93. \end{aligned}$$

Cuando $p(B_0 | A_0) = p(B_1 | A_1) = p$, se dice que el canal es *simétrico*. En tal caso, la probabilidad de que un bit emitido se reciba correctamente vale:

$$p(C) = p(A_0) p(B_0 | A_0) + p(A_1) p(B_1 | A_1) = p(A_0) p + (1 - p(A_0)) p = p \quad \blacktriangleleft$$

y esto cualquiera que sea la proporción de ceros que se emiten.

Variables aleatorias discretas

2.1 Variables aleatorias discretas

Al estudiar un fenómeno aleatorio, a menudo interesa analizar alguna característica numérica del experimento más que los propios sucesos. Por ejemplo, al lanzar una moneda 3 veces, puede interesar el número de caras obtenidas, pero no la secuencia en que aparecen. Es decir, interesa estudiar la característica:

X : “nº de caras obtenidas al lanzar una moneda 3 veces”.

Al controlar la calidad en cierta fabrica de chips, puede interesar:

X : “nº de transistores defectuosos en un chip tomado al azar”,

Y : “duración de un chip tomado al azar (años)”.

Al analizar la complejidad de un algoritmo de ordenación, en función de los datos de entrada, puede interesar:

X : “nº de veces que se ejecuta el bucle principal”,

Y : “nº de operaciones aritméticas realizadas”.

Al estudiar las características de un grupo de personas, puede interesar:

X : “edad de una persona tomada al azar (años)”,

Y : “estatura de una persona tomada al azar (cm)”,

Z : “peso de una persona tomada al azar (kg)”.

Todas las expresiones anteriores son reglas que asignan un valor numérico a cada resultado de un fenómeno aleatorio, es decir, a cada elemento del espacio muestral E . Una característica numérica de este tipo se denomina *variable aleatoria*. Vamos a estudiar con detalle una de estas variables aleatorias.

Ejemplo 2.1.1 (Número de caras con 3 monedas)

Lanzar una moneda 3 veces es un fenómeno aleatorio, cuyo espacio muestral asociado es:

$$E := \{ccc, cc+, c+c, c++, +cc, +c+, ++c, +++\}.$$

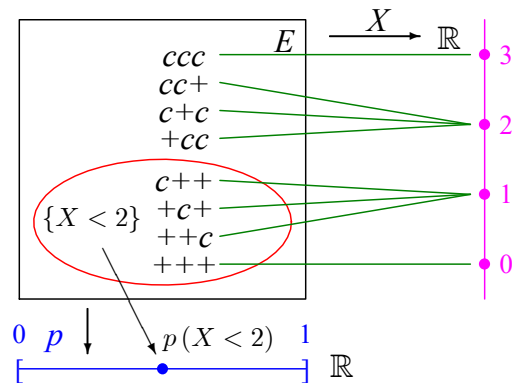


Figura 2.1: Variable aleatoria n° de caras.

Estos 8 resultados posibles son equiprobables, con probabilidad $1/8$. El espacio de sucesos es $S = \mathcal{P}(E)$ y por tanto (E, S, p) es un espacio probabilístico. Sobre E se define la característica:

X : “nº de caras obtenidas al lanzar una moneda 3 veces”.

La característica X es una función¹ $X : E \rightarrow \mathbb{R}$, que se llama *variable aleatoria*. Su imagen es $\text{Im}(X) = \{0, 1, 2, 3\}$, que es un conjunto finito. La figura 2.1 ilustra esta situación.

Muchos sucesos asociados a este fenómeno aleatorio se pueden expresar en términos de los valores de X . Por ejemplo:

$$\begin{aligned} \text{“obtener 2 caras”} &= \{cc+, c+c, +cc\} = \{X = 2\}, \\ \text{“obtener menos de 2 caras”} &= \{c++, +c+, ++c, +\} = \{X < 2\}, \\ \text{“obtener entre 2 y 4 caras”} &= \{ccc, cc+, c+c, +cc\} = \{2 \leq X \leq 4\}, \\ \text{“obtener al menos 1 cara y menos de 2 caras”} &= \{c++, +c+, ++c\} = \{1 \leq X < 2\}, \\ \text{“obtener más de 2 caras y a lo sumo 4 caras”} &= \{ccc\} = \{2 < X \leq 4\}. \end{aligned}$$

Las probabilidades de dichos sucesos valen:

$$\begin{aligned} p(X = 2) &= 3/8, & p(X < 2) &= 4/8 = 1/2, & p(2 \leq X \leq 4) &= 4/8 = 1/2, \\ p(1 \leq X < 2) &= 3/8, & p(2 < X \leq 4) &= 1/8. \end{aligned}$$

Para cada $x \in \mathbb{R}$, interesa conocer $p(X = x)$, que vale 0 si x no pertenece a $\text{Im}(X)$. Los demás casos aparecen en la figura 2.2.i, que representa la llamada *función de masa* de X . Se trata de la función:

$$f : \mathbb{R} \rightarrow \mathbb{R} \text{ tal que } f(x) := p(X = x).$$

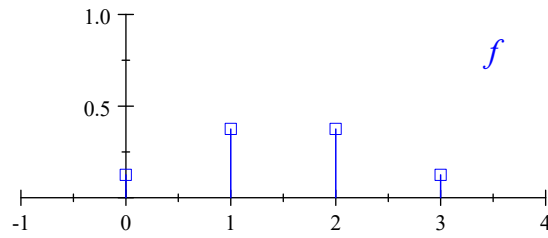
La suma de las probabilidades de la tabla o de las alturas de las barras es 1.

La *función de distribución* de X se define de modo similar:

$$F : \mathbb{R} \rightarrow \mathbb{R} \text{ tal que } F(x) := p(X \leq x).$$

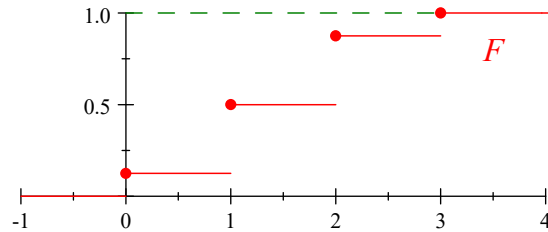
¹En matemáticas, es usual designar las funciones con letras minúsculas, como f , g o h . Sin embargo, en estadística se acostumbra a utilizar letras mayúsculas, como X , Y o Z .

$\text{Im}(X)$	$p(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8



(i) Función de masa.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$



(ii) Función de distribución.

Figura 2.2: Probabilidades de la variable aleatoria n° de caras.

La figura 2.2.ii muestra sus expresiones analítica y gráfica. Se observa que F es una función escalonada², cuyas discontinuidades están en $\text{Im}(X)$. La altura de cada peldaño es la probabilidad de que X tome el valor correspondiente.

Tanto la función de masa como la función de distribución permiten calcular probabilidades relativas a X . Por ejemplo:

$$\begin{aligned} p(1 < X < 3) &= \begin{cases} f(2) & = \frac{3}{8}, \\ F(3^-) - F(1^+) = 7/8 - 1/2 & = \frac{3}{8}, \end{cases} \\ p(1 \leq X < 3) &= \begin{cases} f(1) + f(2) = 3/8 + 3/8 & = \frac{3}{4}, \\ F(3^-) - F(1^-) = 7/8 - 1/8 & = \frac{3}{4}, \end{cases} \\ p(1 < X \leq 3) &= \begin{cases} f(2) + f(3) = 3/8 + 1/8 & = \frac{1}{2}, \\ F(3^+) - F(1^+) = 1 - 1/2 & = \frac{1}{2}, \end{cases} \\ p(1 \leq X \leq 3) &= \begin{cases} f(1) + f(2) + f(3) = 3/8 + 3/8 + 1/8 & = \frac{7}{8}, \\ F(3^+) - F(1^-) = 1 - 1/8 & = \frac{7}{8}. \end{cases} \end{aligned}$$

Las notaciones $F(a^-)$ y $F(a^+)$ representan los límites de la función F en el punto a por la izquierda y por la derecha, respectivamente.

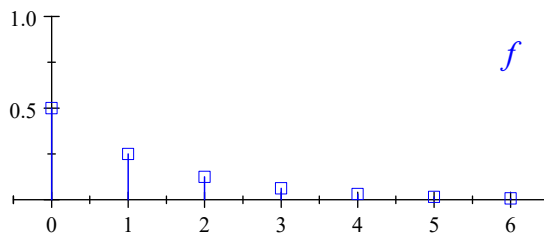
La relación entre f y F es directa: las barras de f están situadas en los puntos de discontinuidad de F y sus alturas son los saltos de F en dichos puntos. ◀

También se puede definir variables aleatorias sobre espacios muestrales numerables, como en el ejemplo siguiente.

Ejemplo 2.1.2 (Variable aleatoria numerable)

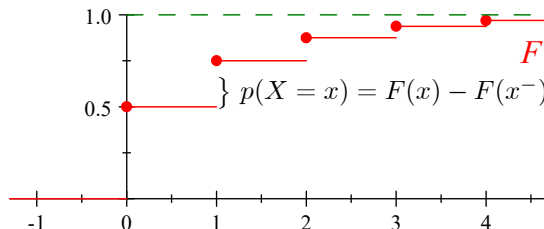
²Es decir, constante en una cantidad finita o numerable de intervalos que particionan \mathbb{R} .

$\text{Im}(X)$	$p(X = x)$
0	$1/2$
\vdots	\vdots
n	$1/2^{n+1}$
\vdots	\vdots



(i) Función de masa.

$$F(x) = \begin{cases} 0 & x < 0 \\ \vdots & \vdots \\ 1 - 1/2^{n+1} & n \leq x < n + 1 \\ \vdots & \vdots \end{cases}$$



(ii) Función de distribución.

Figura 2.3: Variable aleatoria numerable.

Lanzar una moneda hasta que sale cruz es un fenómeno aleatorio, cuyo espacio muestral es:

$$E := \{+, c+, cc+, ccc+, \dots, cc \dots c+, \dots\}.$$

El espacio de sucesos es $S := \mathcal{P}(E)$. Se considera la variable aleatoria:

X : “número de caras obtenidas antes de salir cruz”.

Su imagen es $\text{Im}(X) = \{0, 1, 2, 3, \dots\} = \mathbb{N}$, un conjunto numerable. Para construir la función de masa de X , hay que calcular $p(X = n)$ para todo $n \in \mathbb{N}$:

$$p(X = n) = p\left(\overbrace{cc \dots c}^n +\right) = 1/2^{n+1}.$$

Se trata de la función f representada en la figura 2.3.i. En este caso, la tabla tiene infinitas filas y la gráfica tiene infinitas barras, pero de nuevo la suma de las probabilidades, es decir, de las alturas de las barras, es 1.

Para obtener la función de distribución de X , se necesita calcular, para cada $n \in \mathbb{N}$, la suma de una progresión geométrica de razón $1/2$:

$$p(X \leq n) = \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{n+1}} = \frac{1/2 - (1/2^{n+1})(1/2)}{1 - 1/2} = 1 - \frac{1}{2^{n+1}}.$$

Las expresiones analítica y gráfica de F aparecen en la figura 2.3.ii. De nuevo, F es escalonada, sus discontinuidades están en $\text{Im}(X)$ y la altura de cada peldaño es la probabilidad de que X tome el valor correspondiente.

Tanto la función de masa como la función de distribución permiten calcular probabilidades relativas a X . Por ejemplo:

$$p(1 < X \leq 3) = \begin{cases} f(2) + f(3) = 1/2^3 + 1/2^4 \\ F(3) - F(1) = (1 - 1/2^4) - (1 - 1/2^2) \end{cases} = \frac{3}{16},$$

$$p(1 \leq X \leq 3) = \begin{cases} f(1) + f(2) + f(3) = 1/2^2 + 1/2^3 + 1/2^4 \\ F(3) - F(1^-) = (1 - 1/2^4) - (1 - 1/2) \end{cases} = \frac{7}{16}.$$



En los ejemplos 2.1.1 y 2.1.2, las funciones de masa tienen una cantidad finita o numerable de barras, cuyas alturas suman 1. Esta propiedad caracteriza a las variables aleatorias discretas, que se definen a continuación.

Definición 2.1.3 (Variable aleatoria discreta, función de masa y función de distribución)

Sea (E, S, p) un espacio probabilístico.

- Una *variable aleatoria discreta* es una función³ $X : E \rightarrow \mathbb{R}$ tal que existe un conjunto finito o numerable de puntos, $\{x_1, x_2, \dots, x_i, \dots\} \subset \mathbb{R}$, verificando $\sum_{i=1}^{\infty} p(X = x_i) = 1$.
- La *función de masa* de X asigna a cada número real x la probabilidad de que X tome el valor x , es decir: $f : \mathbb{R} \rightarrow \mathbb{R}$ es tal que $f(x) := p(X = x)$.
- La *función de distribución* de X asigna a cada número real x la probabilidad de que X tome valores menores o iguales que x , es decir: $F : \mathbb{R} \rightarrow \mathbb{R}$ es tal que $F(x) := p(X \leq x)$.

Cuando se necesita especificar la variable, se nota f_X en lugar de f y F_X en lugar de F . ◀

Toda función de distribución tiene ciertas propiedades que la caracterizan.

Teorema 2.1.4 (Propiedades de la función de distribución de una variable aleatoria discreta)

Sea X una variable aleatoria discreta. Su función de distribución, F , verifica:

- (i) Es creciente: $a \leq b \Rightarrow F(a) \leq F(b)$ y además $p(a < X \leq b) = F(b) - F(a)$.
- (ii) Empieza en 0 y termina en 1, es decir, $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.
- (iii) Es continua por la derecha, es decir, $\lim_{x \rightarrow a^+} F(x) = F(a^+) = F(a)$ para todo $a \in \mathbb{R}$.
- (iv) Es escalonada.

Demostración: Apartado (i)

Si $a, b \in \mathbb{R}$ con $a \leq b$, entonces $\{X \leq a\} \subseteq \{X \leq b\}$ y $\{X \leq b\} - \{X \leq a\} = \{a < X \leq b\}$. Por tanto:

$$\begin{aligned} p(X \leq a) &\leq p(X \leq b) \Rightarrow F(a) \leq F(b), \\ p(a < X \leq b) &= p(X \leq b) - p(X \leq a) = F(b) - F(a). \end{aligned}$$

Apartados (ii) y (iii)

Ver Rohatgi [15, página 45, Theorem 3]. ◀

La función de distribución permite calcular la probabilidad de cualquier intervalo de \mathbb{R} , incluso reducido a un punto.

³Si $S = \mathcal{P}(E)$, está garantizado que los conjuntos de la forma $\{X \in I\}$, con I intervalo de \mathbb{R} , son sucesos. En otro caso, se impone a X la condición de que $\{X \leq a\} \in S$ para todo $a \in \mathbb{R}$. De aquí se deduce que $\{X \in I\} \in S$ para todo intervalo I .

En la práctica, dicho requisito se cumple en todos los casos que se van a manejar en este texto, por lo que no se comprobará.

Teorema 2.1.5 (Cálculo de probabilidades mediante la función de distribución)

Sean X una variable aleatoria discreta con función de distribución F y $a \in \mathbb{R}$ un número real.

Se verifica:

$$(i) \quad p(X < a) = \lim_{x \rightarrow a^-} F(x) = F(a^-).$$

(ii) Si I es un intervalo de \mathbb{R} , entonces la probabilidad $p(X \in I)$ es el salto de F en I .

(iii) $p(X = a) = F(a) - F(a^-)$ y vale 0 si y solo si F es continua en a .

Demostración: Apartado (i)

Ver Rohatgi [15, página 51, Theorem 4].

Apartado (ii)

Si el intervalo I tiene un extremo infinito y el otro finito, entonces:

$$p(X \leq a) = F(a) \quad (\text{definición de función de distribución}),$$

$$p(X < a) = F(a^-) \quad (\text{apartado i}),$$

$$p(X \geq a) = 1 - p(\overline{X \geq a}) = 1 - p(X < a) = 1 - F(a^-) \quad (\text{regla del complementario}),$$

$$p(X > a) = 1 - p(\overline{X > a}) = 1 - p(X \leq a) = 1 - F(a) \quad (\text{regla del complementario}).$$

Si el intervalo I tiene extremos a y b , con $a, b \in \mathbb{R}$ y $a \leq b$, entonces:

$$p(a \leq X \leq b) = p(\{X \leq b\} - \{X < a\}) = p(X \leq b) - p(X < a) = F(b) - F(a^-),$$

$$p(a \leq X < b) = p(\{X < b\} - \{X < a\}) = p(X < b) - p(X < a) = F(b^-) - F(a^-),$$

$$p(a < X \leq b) = p(\{X \leq b\} - \{X \leq a\}) = p(X \leq b) - p(X \leq a) = F(b) - F(a),$$

$$p(a < X < b) = p(\{X < b\} - \{X \leq a\}) = p(X < b) - p(X \leq a) = F(b^-) - F(a).$$

En todos estos casos se ha aplicado el álgebra de sucesos.

Apartado (iii)

En el caso particular de que $b = a$, se verifica:

$$p(X = a) = p(a \leq X \leq a) = F(a) - F(a^-).$$

Esta expresión vale 0 si y solo si $F(a) = F(a^-)$, es decir, si y solo si F es continua en a . ◀

Nota 2.1.6 (Identificación de variables aleatorias discretas)

En general, una variable aleatoria discreta verifica lo siguiente:

- La variable *cuenta* eventos.
- Su imagen es finita o numerable.
- Su función de distribución es escalonada. ◀

2.2 Medidas de una variable aleatoria discreta

La función de masa de una variable aleatoria discreta proporciona toda la información relativa a dicha variable. En particular, su gráfico puede dar una buena perspectiva de la variable en

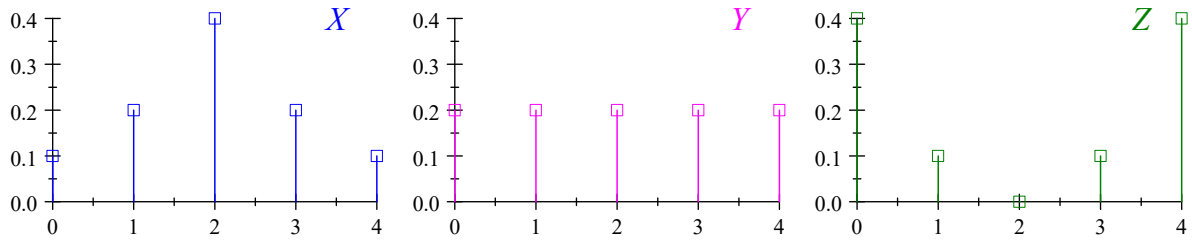


Figura 2.4: X , Y y Z tienen igual media, pero distinta dispersión.

cuestión. Sin embargo, a menudo interesa resumir toda la información en unos pocos números. Estos valores, llamados *medidas de la variable aleatoria*, tienen dos objetivos:

- Resumir las propiedades esenciales de la distribución de probabilidad, proporcionando una idea rápida de las características cuantitativas de la variable.
- Permitir la comparación de variables.

Las *medidas de posición* dan una idea de los valores de la variable alrededor de los cuales se agrupa su distribución. Dicho de otro modo, hallan el *centro* de la distribución. Ahora se define la media y en el apéndice A se estudiarán otras medidas de posición.

Definición 2.2.1 (Media de una variable aleatoria discreta)

Sea X una variable aleatoria discreta. La *media* o *esperanza* de X es la suma⁴ de cada valor de X multiplicado por su probabilidad:

$$\mu(X) = E(X) := \sum_{x \in \text{Im}(X)} x p(X = x).$$

La media de X se expresa en las mismas unidades que X . ◀

Ejemplo 2.2.2 (Variables con igual media, pero distinta dispersión)

La figura 2.4 muestra las funciones de masa de tres variables aleatorias discretas: X , Y y Z . La media de cada una vale:

$$\begin{aligned} \mu(X) &= 0 \cdot 0.1 + 1 \cdot 0.2 + 2 \cdot 0.4 + 3 \cdot 0.2 + 4 \cdot 0.1 = 2, \\ \mu(Y) &= 0 \cdot 0.2 + 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.2 = 2, \\ \mu(Z) &= 0 \cdot 0.4 + 1 \cdot 0.1 + 2 \cdot 0 + 3 \cdot 0.1 + 4 \cdot 0.4 = 2. \end{aligned}$$

Es decir, las tres tienen media 2. Sin embargo, es evidente que los valores están más agrupados en torno a dicha media en el caso de X que en el de Y y más en el de Y que en el de Z . Se necesita una medida que describa este comportamiento. ◀

Un caso extremo de separación respecto de la media es el siguiente: si una persona se come un pollo y otra persona ayuna, en promedio han comido medio pollo cada una. Obviamente, la media no describe bien esta situación: los datos se apartan tanto de ella que la inhabilitan como representante de la información. En general, las medidas de posición no resumen completamente la información contenida en una variable. Por ello, se definen las *medidas de dispersión*, que

⁴Para que exista la media, se exige que la serie $\sum_{i=1}^{\infty} |x_i| p(X = x_i)$ converja.

indican cuánto se apartan los valores de la variable de sus promedios y permiten valorar la representatividad de éstos. Cuanto más grande es una medida de dispersión, mayor es dicha separación.

Definición 2.2.3 (Medidas de dispersión de una variable aleatoria discreta)

Sea X una variable aleatoria discreta.

- La *varianza* de X es la media de los cuadrados de las desviaciones respecto de $\mu(X)$:

$$\sigma^2(X) = V(X) := \mu((X - \mu(X))^2).$$

La varianza de X se expresa en el cuadrado de las unidades de X .

- La *desviación típica* de X es la raíz cuadrada de su varianza:

$$\sigma(X) := \sqrt{\sigma^2(X)} = \sqrt{V(X)}.$$

La desviación típica de X tiene las mismas unidades que X . Junto con la varianza, mide la dispersión absoluta de la variable.

- El *coeficiente de variación* de X es el cociente entre su desviación típica y el valor absoluto de su media:

$$CV(X) := \frac{\sigma(X)}{|\mu(X)|} \quad \text{si } \mu(X) \neq 0.$$

El coeficiente de variación es adimensional. Mide la dispersión relativa de la variable y permite comparar la dispersión de dos variables. ◀

Veamos cómo se calculan todas estas medidas en un caso concreto.

Ejemplo 2.2.4 (Número de caras con 3 monedas)

En el ejemplo 2.1.1 se ha estudiado la variable aleatoria:

X : “nº de caras obtenidas al lanzar una moneda 3 veces”.

A partir de la función de masa, la media de X vale:

$$\mu(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Esto es lógico, porque la distribución es simétrica respecto al punto $3/2$. La varianza de X es la media de $(X - \mu(X))^2$:

$$\begin{aligned} \sigma^2(X) &= \left(0 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} = \\ &= \frac{9}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{9}{4} \cdot \frac{1}{8} = \frac{3}{4}. \end{aligned}$$

La varianza de X se puede calcular también como $\mu(X^2) - \mu(X)^2$ (ver el teorema 2.2.5). Como comprobación:

$$\mu(X^2) - \mu(X)^2 = \left(0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8}\right) - \left(\frac{3}{2}\right)^2 = \frac{3}{4} = \sigma^2(X).$$

La desviación típica y el coeficiente de variación de X valen, respectivamente:

$$\sigma(X) = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2} \simeq 0.866\,025, \quad \text{CV}(X) = \frac{\sqrt{3}/2}{3/2} = \frac{\sqrt{3}}{3} \simeq 57.735\%.$$

El coeficiente de variación es grande, luego X es bastante dispersa. ◀

Si X es una variable aleatoria y a es un número real, entonces es posible sumar a a cada valor de X o multiplicar por a cada valor de X . De este modo se obtienen las variables aleatorias $X + a$ y aX . El siguiente resultado muestra, entre otras cosas, el comportamiento de la media y de la varianza respecto de las *operaciones algebraicas* con una variable aleatoria.

Teorema 2.2.5 (Propiedades de la media y de la varianza)

Sean X una variable aleatoria discreta y $a, b \in \mathbb{R}$ dos números reales.

- (i) Si la función de masa de X es simétrica respecto de la recta $x = a$, entonces $\mu(X) = a$.
- (ii) Media de una combinación afín: $\mu(aX + b) = a\mu(X) + b$.
- (iii) Positividad de la varianza: $\sigma^2(X) \geq 0$.
- (iv) La varianza es igual a la media del cuadrado menos el cuadrado de la media: $\sigma^2(X) = \mu(X^2) - \mu(X)^2$.
- (v) Varianza de una combinación afín: $\sigma^2(aX + b) = a^2\sigma^2(X)$.

Demostración: Apartado (ii)

Si X es una variable aleatoria discreta, entonces $Y = aX + b$ también es una variable aleatoria discreta. La media de Y vale:

$$\begin{aligned} \mu(Y) &= \sum_{y \in \text{Im}(Y)} yp(Y = y) = \sum_{x \in \text{Im}(X)} (ax + b)p(ax + b = ax + b) = \\ &= \sum_{x \in \text{Im}(X)} (ax + b)p(X = x) = a \sum_{x \in \text{Im}(X)} xp(X = x) + b \sum_{x \in \text{Im}(X)} p(X = x) = \\ &= a\mu(X) + b1 = a\mu(X) + b. \end{aligned}$$

Apartado (iii)

La varianza de X es la media de la variable aleatoria $(X - \mu(X))^2$. Esta variable solo toma valores positivos, luego su media es positiva.

Apartado (iv)

Desarrollando el cuadrado y usando la linealidad de la media, se verifica:

$$\begin{aligned} \sigma^2(X) &= \mu((X - \mu(X))^2) = \mu(X^2 - 2X\mu(X) + \mu(X)^2) = \\ &= \mu(X^2) - 2\mu(X)\mu(X) + \mu(X)^2 = \mu(X^2) - \mu(X)^2. \end{aligned}$$

Apartado (v)

Usando la linealidad de la media, se verifica:

$$\begin{aligned}\sigma^2(aX + b) &= \mu(((aX + b) - \mu(aX + b))^2) = \mu((aX + b - a\mu(X) - b)^2) = \\ &= \mu((a(X - \mu(X)))^2) = \mu(a^2(X - \mu(X))^2) = \\ &= a^2(\mu(X - \mu(X))^2) = a^2\sigma^2(X).\end{aligned}$$

2.3 Independencia de variables aleatorias

Intuitivamente, dos variables aleatorias X e Y son independientes cuando cualquier suceso asociado solo a los valores de X es independiente de cualquier suceso asociado solo a los valores de Y . La formalización matemática es la siguiente.

Definición 2.3.1 (Independencia de variables aleatorias)

Dos variables aleatorias X e Y son *independientes* si y solo si están definidas sobre el mismo espacio probabilístico y para todo $x, y \in \mathbb{R}$ se verifica:

$$p(\{X \leq x\} \cap \{Y \leq y\}) = p(X \leq x)p(Y \leq y).$$

La definición anterior se generaliza a n variables aleatorias X_1, \dots, X_n .

Ejemplo 2.3.2 (Variables aleatorias discretas independientes y dependientes)

Una moneda tiene sus caras marcadas con 0 y 1, respectivamente. Se lanza la moneda dos veces, con lo cual se generan dos bits aleatoriamente, y se consideran las variables aleatorias:

$$\begin{aligned}X &: \text{“número obtenido en el primer lanzamiento”}, & \text{Im}(X) &= \{0, 1\}, \\ Y &: \text{“número obtenido en el segundo lanzamiento”}, & \text{Im}(Y) &= \{0, 1\}, \\ Z &: \text{“producto de ambos números”} = X \cdot Y, & \text{Im}(Z) &= \{0, 1\}.\end{aligned}$$

El espacio muestral asociado a este fenómeno aleatorio consta de 4 sucesos equiprobables:

$$E = \{00, 01, 10, 11\}.$$

Es intuitivo que las variables X e Y no dependen una de otra. Para demostrarlo, se calcula $p(\{X \leq x\} \cap \{Y \leq y\})$ y se comprueba que siempre coincide con $p(X \leq x)p(Y \leq y)$. Los 4 casos a considerar son:

$$\begin{aligned}p(\{X \leq 0\} \cap \{Y \leq 0\}) &= p(00) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = p(X \leq 0)p(Y \leq 0), \\ p(\{X \leq 0\} \cap \{Y \leq 1\}) &= p(00, 01) = \frac{2}{4} = \frac{1}{2} \cdot 1 = p(X \leq 0)p(Y \leq 1), \\ p(\{X \leq 1\} \cap \{Y \leq 0\}) &= p(00, 10) = \frac{2}{4} = 1 \cdot \frac{1}{2} = p(X \leq 1)p(Y \leq 0), \\ p(\{X \leq 1\} \cap \{Y \leq 1\}) &= p(00, 01, 10, 11) = \frac{4}{4} = 1 \cdot 1 = p(X \leq 1)p(Y \leq 1).\end{aligned}$$

También es obvio que Z depende tanto de X como de Y . Para probar que X y Z son dependientes, basta hallar un caso donde $p(\{X \leq x\} \cap \{Z \leq z\})$ no coincida con $p(X \leq x)p(Z \leq z)$, por ejemplo:

$$p(\{X \leq 0\} \cap \{Z \leq 0\}) = p(00, 01) = \frac{2}{4} \neq \frac{1}{2} \cdot \frac{3}{4} = p(X \leq 0) p(Z \leq 0).$$

Se ha usado que $p(Z \leq 0) = p(00, 01, 10) = 3/4$. ◀

Si X e Y son variables aleatorias definidas sobre el mismo espacio probabilístico, entonces $X + Y$ también lo es. El siguiente resultado muestra el comportamiento de la media y de la varianza respecto de la suma de variables aleatorias.

Teorema 2.3.3 (Propiedades de la media y de la varianza en el caso de dos variables)

Sean X e Y dos variables aleatorias discretas definidas sobre el mismo espacio probabilístico.

(i) *Media de una suma:* $\mu(X + Y) = \mu(X) + \mu(Y)$.

(ii) *Varianza de una suma:* si X e Y son independientes, entonces $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$.

Demostración: Ver Fernández-Abascal [8, página 290, Propiedad P1] y [8, página 291, Propiedad P2]. ◀

Ejemplo 2.3.4 (Propiedades de la media y de la varianza)

Sean X , Y y $Z := X \cdot Y$ las variables aleatorias del ejemplo 2.3.2, cuyas funciones de masa respectivas son:

Im(X)	$p(X = x)$	Im(Y)	$p(Y = y)$	Im(Z)	$p(Z = z)$
0	1/2	0	1/2	0	$p(00, 01, 10) = 3/4$
1	1/2	1	1/2	1	$p(11) = 1/4$

Las medias de estas tres variables valen, respectivamente:

$$\mu(X) = \mu(Y) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2} \quad \text{y} \quad \mu(Z) = 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = \frac{1}{4}.$$

Consideremos otras dos variables aleatorias, obtenidas por combinación lineal de X , Y y Z :

$$T = 2X - 3Y + 1 \quad \text{y} \quad W = 2X - 3Z + 1.$$

Vamos a comparar las medias de T y de W con las de X , Y y Z . De acuerdo con el teorema 2.2.5, se verifica:

$$\begin{aligned} \mu(T) &= \mu(2X - 3Y + 1) = 2\mu(X) - 3\mu(Y) + 1 = 2 \cdot \frac{1}{2} - 3 \cdot \frac{1}{2} + 1 = \frac{1}{2}, \\ \mu(W) &= \mu(2X - 3Z + 1) = 2\mu(X) - 3\mu(Z) + 1 = 2 \cdot \frac{1}{2} - 3 \cdot \frac{1}{4} + 1 = \frac{5}{4}. \end{aligned}$$

Para comprobarlo, construimos las funciones de masa de T y de W :

X	Y	$2X - 3Y + 1$	$p(T = t)$	X	Z	$2X - 3Z + 1$	$p(W = w)$
0	0	1	$p(00) = 1/4$	0	0	1	$p(00, 01) = 2/4 = 1/2$
0	1	-2	$p(01) = 1/4$	1	0	3	$p(10) = 1/4$
1	0	3	$p(10) = 1/4$	1	1	0	$p(11) = 1/4$

En el caso de W , se ha tenido en cuenta que si X vale 0, entonces Z no puede valer 1. Usando la definición de media, se obtienen los valores esperados:

$$\begin{aligned}\mu(T) &= -2 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{1}{2}, \\ \mu(W) &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = \frac{5}{4}.\end{aligned}$$

También vamos a comparar las varianzas de T y de W con las de X , Y y Z , que valen, respectivamente:

$$\begin{aligned}\sigma^2(X) = \sigma^2(Y) &= \mu(X^2) - \mu(X)^2 = \left(0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2}\right) - \left(\frac{1}{2}\right)^2 = \frac{1}{4}, \\ \sigma^2(Z) &= \mu(Z^2) - \mu(Z)^2 = \left(0^2 \cdot \frac{3}{4} + 1^2 \cdot \frac{1}{4}\right) - \left(\frac{1}{4}\right)^2 = \frac{3}{16}.\end{aligned}$$

Como T es combinación lineal de X y de Y , que son independientes, se puede aplicar el teorema anterior, combinado con 2.2.5, y resulta:

$$\begin{aligned}\sigma^2(T) &= \sigma^2(2X - 3Y + 1) = \sigma^2(2X + (-3Y)) = \sigma^2(2X) + \sigma^2(-3Y) = \\ &= 2^2 \sigma^2(X) + (-3)^2 \sigma^2(Y) = 4 \cdot \frac{1}{4} + 9 \cdot \frac{1}{4} = \frac{13}{4}.\end{aligned}$$

Si se calcula directamente, se obtiene el mismo resultado:

$$\sigma^2(T) = \mu(T^2) - \mu(T)^2 = \left((-2)^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4}\right) - \left(\frac{1}{2}\right)^2 = \frac{13}{4}.$$

El caso de la varianza de W es diferente. Si se intenta hacer lo mismo que con T , resulta:

$$\sigma^2(W) = \sigma^2(2X - 3Z + 1) = 2^2 \sigma^2(X) + (-3)^2 \sigma^2(Z) = 4 \cdot \frac{1}{4} + 9 \cdot \frac{3}{16} = \frac{43}{16}.$$

Pero el valor verdadero es:

$$\sigma^2(W) = \mu(W^2) - \mu(W)^2 = \left(0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 3^2 \cdot \frac{1}{4}\right) - \left(\frac{5}{4}\right)^2 = \frac{19}{16}.$$

Los dos resultados no coinciden porque W es combinación lineal de X y de Z , que son dependientes. ◀

2.4 Distribución binomial

Ciertas variables aleatorias están relacionadas con un experimento que se repite un número fijo de veces. Por ejemplo:

- X : “nº de caras obtenidas al lanzar una moneda 3 veces”,
- Y : “nº de bits transmitidos correctamente en un paquete de 32 bits”,
- Z : “nº de tornillos defectuosos en una caja de 100 tornillos”.

En todos estos casos la variable cuenta el número de veces que ocurre cierto suceso al repetir el experimento. Estas variables tienen una definición similar y vamos a ver que comparten ciertas propiedades.

Ejemplo 2.4.1 (Pruebas de Bernoulli)

Una urna contiene bolas blancas en proporción p . Supongamos que se realiza un número fijo de extracciones con reemplazamiento, digamos k . En cada extracción se puede dar el suceso “obtener bola blanca”, que por convenio se denomina *éxito*, o su contrario, llamado *fracaso*.

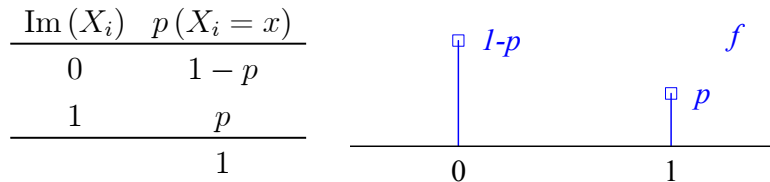
Consideremos las variables aleatorias:

- X : “nº de bolas blancas obtenidas en k extracciones”,
- X_i : “nº de bolas blancas obtenidas en la extracción i ”, $i = 1, 2, \dots, k$.

Es evidente que $X = X_1 + X_2 + \dots + X_k$ y por eso empezamos por estudiar las variables X_i . Cada variable X_i es *booleana*, porque vale 1 si en la extracción número i sale una bola blanca (éxito) y 0 en otro caso (fracaso). Además, como las extracciones se hacen con reemplazamiento, se verifican dos propiedades:

- Cada bola extraída es blanca o no independientemente de las demás, es decir, *las variables X_i son independientes*.
- La probabilidad de obtener bola blanca vale p para todas las extracciones, es decir, *la probabilidad de éxito es constante*.

La función de masa de cada variable X_i es:



En consecuencia, la media de cada variable X_i vale:

$$\mu(X_i) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Además, $X_i^2 = X_i$, luego $\mu(X_i^2) = \mu(X_i) = p$ y la varianza de X_i vale:

$$\sigma^2(X_i) = \mu(X_i^2) - \mu(X_i)^2 = p - p^2 = p(1 - p).$$

En cuanto a la variable aleatoria X , se puede decir que X cuenta éxitos en un número fijo de pruebas (k) independientes y con probabilidad de éxito constante (p). La imagen de la variable X es $\text{Im}(X) = \{0, 1, \dots, k\}$, luego X es discreta. Para construir la función de masa de X hay que calcular $p(X = x)$ cuando $x \in \{0, 1, \dots, k\}$. El suceso $\{X = x\}$ se obtiene cuando x de las k bolas extraídas son blancas y el resto no. Esto equivale a decir que x de las variables X_1, \dots, X_k valen 1 y las restantes $k - x$ valen 0. Una posibilidad es que las x variables que valen 1 sean las primeras, es decir:

$$\overbrace{\{X_1 = 1\} \cap \dots \cap \{X_x = 1\}}^x \cap \overbrace{\{X_{x+1} = 0\} \cap \dots \cap \{X_k = 0\}}^{k-x}.$$

Este suceso es intersección de sucesos independientes, luego su probabilidad vale $p^x (1 - p)^{k-x}$. Ahora bien, existen $\binom{k}{x}$ formas de elegir las x variables que valen 1 entre las k . Por tanto, la probabilidad de obtener x bolas blancas en k extracciones vale:

$$p(X = x) = \binom{k}{x} p^x (1 - p)^{k-x} \quad \text{si } x \in \{0, 1, \dots, k\}.$$

Esta expresión representa el número de formas de obtener x éxitos en k extracciones, multiplicado por la probabilidad de éxito elevada al número de éxitos y multiplicado por la probabilidad de fracaso elevada al número de fracasos. Por otro lado, la media de X vale:

$$\mu(X) = \mu(X_1 + \dots + X_k) = \mu(X_1) + \dots + \mu(X_k) = p + \dots + p = kp.$$

Esta expresión es el producto del número de extracciones por la probabilidad de éxito. En cuanto a la varianza de X , teniendo en cuenta que las variables X_i son independientes, se verifica:

$$\begin{aligned} \sigma^2(X) &= \sigma^2(X_1 + \dots + X_k) = \sigma^2(X_1) + \dots + \sigma^2(X_k) = \\ &= p(1 - p) + \dots + p(1 - p) = kp(1 - p). \end{aligned}$$

Esta expresión es el producto del número de extracciones por la probabilidad de éxito y por la probabilidad de fracaso. ◀

Generalizando el ejemplo anterior, se obtiene la distribución binomial.

Definición 2.4.2 (Distribución binomial)

Una *prueba de Bernoulli*⁵ es un fenómeno aleatorio con dos resultados posibles, llamados *éxito* y *fracaso*, que tienen probabilidades respectivas p y $1 - p$.

Si se realizan $k \in \mathbb{N} - \{0\}$ pruebas de Bernoulli independientes y con probabilidad de éxito constante, p , entonces se dice que la variable aleatoria

X : “número de éxitos en k pruebas de Bernoulli”

sigue una *distribución binomial de parámetros k y p* . Se nota $X \sim B(k, p)$. ◀

La distribución binomial modela variables aleatorias como:

X : “nº de caras obtenidas al lanzar una moneda 3 veces” $\sim B(3, 1/2)$,

Y : “nº de bits transmitidos correctamente en un paquete de 32 bits” $\sim B(32, p)$,

Z : “nº de tornillos defectuosos en una caja de 100 tornillos” $\sim B(100, p)$,

T : “nº de bolas blancas en k extracciones con reemplazamiento” $\sim B(k, p)$.

5



Jakob (Jacques) Bernoulli (Basel, Suiza, 1654 – Basel, Suiza, 1705).

Matemático y miembro de una familia de eminentes científicos.

Estudió filosofía y teología, porque su padre quería que fuese clérigo. Sin embargo, terminó dedicado a las matemáticas y a la astronomía. Ocupó la cátedra de matemáticas de la universidad de Basel desde 1687 hasta su muerte.

Son importantes sus contribuciones en álgebra, cálculo infinitesimal, cálculo en variaciones, mecánica, series y probabilidad. Su trabajo más conocido es *Ars Conjectandi* (el Arte de la Conjetura), publicado póstumamente. Esta obra contiene el conocimiento de la época sobre probabilidad y combinatoria. Además, incluye la aplicación de la probabilidad a los juegos de azar y la *ley de los grandes números*. El concepto de *prueba de Bernoulli* se deriva de este trabajo, así como los *números de Bernoulli*.

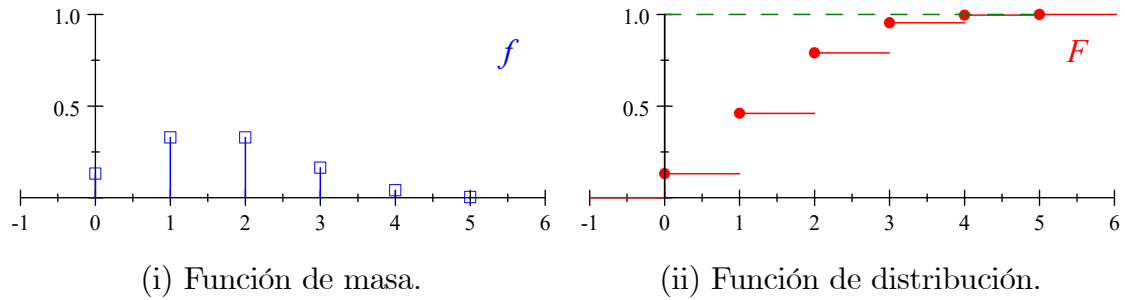


Figura 2.5: Distribución binomial $B(k, p)$.

En este contexto el “éxito” es subjetivo. Por ejemplo, para Y el éxito es “transmitir un bit correctamente”, mientras que para Z el éxito es “observar un tornillo defectuoso”.

Teorema 2.4.3 (Propiedades de la distribución binomial)

Sea $X \sim B(k, p)$.

(i) Función de masa (figura 2.5.i):

$$p(X = x) = \begin{cases} \binom{k}{x} p^x (1-p)^{k-x} & x \in \text{Im}(X) = \{0, 1, 2, \dots, k\} \\ 0 & \text{en otro caso} \end{cases}$$

(ii) Función de distribución (figura 2.5.ii):

$$F(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^x \binom{k}{i} p^i (1-p)^{k-i} & 0 \leq x < k \\ 1 & k \leq x \end{cases}$$

Esta expresión es poco manejable, por lo que se halla tabulada.

(iii) Media: $\mu(X) = kp$ (producto del número de intentos por la probabilidad de éxito).

(iv) Varianza: $\sigma^2(X) = kp(1-p)$ (producto del número de intentos por la probabilidad de éxito y por la probabilidad de fracaso).

(v) Reproductividad respecto del primer parámetro: si $X \sim B(k_X, p)$ e $Y \sim B(k_Y, p)$ son independientes, entonces $X + Y \sim B(k_X + k_Y, p)$.

Demostración: Apartado (i)

La suma de estas probabilidades es el desarrollo de un binomio:

$$\sum_{x=0}^k \binom{k}{x} p^x (1-p)^{k-x} = (p + (1-p))^k = 1.$$

De aquí proviene el nombre que recibe la distribución.

Apartado (v)

Tanto X como Y se pueden expresar como suma de variables booleanas independientes. Por tanto, $X + Y$ también:

$$X + Y = (X_1 + \cdots + X_{k_X}) + (Y_1 + \cdots + Y_{k_Y}),$$

donde $X_1, \dots, X_{k_X}, Y_1, \dots, Y_{k_Y} \sim B(1, p)$. Por tanto, $X + Y \sim B(k_X + k_Y, p)$. ◀

Ejemplo 2.4.4 (Transmisión de paquetes de bits)

En un canal binario simétrico (ejemplo 1.6.3), el 98 % de los bits se transmiten correctamente. Se considera la variable aleatoria:

$$X : \text{“número de bits alterados en un paquete de 32”} \sim B(32, 0.02).$$

Esta variable cuenta éxitos en un número fijo de pruebas (32) independientes y con probabilidad de éxito constante ($1 - 0.98$). La probabilidad de que el paquete tenga más de 1 bit alterado vale:

$$\begin{aligned} p(X > 1) &= 1 - p(X \leq 1) = 1 - p(X = 0) - p(X = 1) = \\ &= 1 - \binom{32}{0} 0.02^0 \cdot 0.98^{32} - \binom{32}{1} 0.02^1 \cdot 0.98^{31} \simeq 0.133989. \end{aligned}$$

El número esperado de bits alterados en el paquete es la media de X , que vale:

$$\mu(X) = 32 \cdot 0.02 = 0.64.$$

La varianza del número anterior vale:

$$\sigma^2(X) = 32 \cdot 0.02 \cdot 0.98 = 0.6272.$$

Si se transmite un paquete de 32 bits y otro de 16 bits, el número total de bits alterados es una variable aleatoria con distribución $B(48, 0.02)$. ◀

2.5 Distribución geométrica

Consideremos las siguientes variables aleatorias:

X : “número de cruces obtenidas al lanzar una moneda antes de obtener cara”,

Y : “número de bits alterados antes de recibir el primero correcto”,

Z : “número de tornillos correctos observados antes del primero defectuoso”,

T : “número de bolas extraídas con reemplazamiento antes de la primera blanca”.

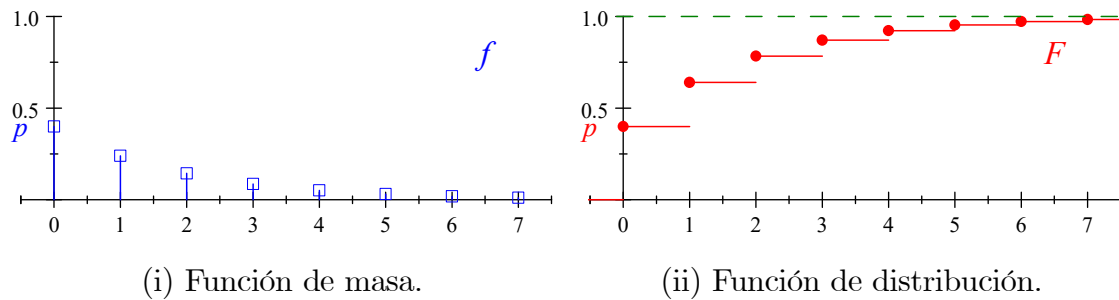
Todas ellas cuentan fracasos antes del primer éxito en pruebas de Bernoulli. Esto caracteriza a la distribución geométrica.

Definición 2.5.1 (Distribución geométrica)

Se realizan pruebas de Bernoulli independientes y con probabilidad de éxito constante, $0 < p \leq 1$. La variable aleatoria:

$$X : \text{“número de fracasos antes del primer éxito”}$$

sigue una *distribución geométrica de parámetro p* . Se nota $X \sim G(p)$. ◀

Figura 2.6: Distribución geométrica $G(p)$.**Teorema 2.5.2** (Propiedades de la distribución geométrica)Sea $X \sim G(p)$.

(i) Función de masa (figura 2.6.i): $p(X = x) = \begin{cases} (1-p)^x p & x \in \text{Im}(X) = \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$

(ii) Función de distribución (figura 2.6.ii):

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - (1-p)^{n+1} & n \leq x < n+1, \quad n \in \mathbb{N} \end{cases}$$

(iii) Media: $\mu(X) = \frac{1-p}{p} = \frac{1}{p} - 1$.

(iv) Varianza: $\sigma^2(X) = \frac{1-p}{p^2}$.

(v) Falta de memoria: si $n, x \in \mathbb{N}$, entonces:

$$p(X = n+x \mid X \geq n) = p(X = x).$$

Es decir, la probabilidad de obtener $n+x$ fracasos, sabiendo que ya se han producido n fracasos, coincide con la probabilidad de obtener x fracasos.

Demostración: Apartado (i)

El suceso $\{X = x\}$ significa obtener x fracasos seguidos de un éxito. Usando la independencia, para todo $x \in \mathbb{N}$ se verifica que:

$$p(X = x) = (1-p)^x p.$$

Estas probabilidades forman una progresión geométrica de razón $0 \leq 1-p < 1$, cuya suma es el primer término partido por 1 menos la razón:

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

De aquí proviene el nombre de la distribución.

Apartado (ii)

Si $n \in \mathbb{N}$ y $n \leq x < n + 1$, entonces:

$$F(x) = p(X \leq x) = 1 - p(X > x) = 1 - \sum_{i=n+1}^{\infty} (1-p)^i p = 1 - \frac{(1-p)^{n+1} p}{1 - (1-p)} = 1 - (1-p)^{n+1}.$$

Se ha usado la regla del complementario y se ha sumado una progresión geométrica de razón $1-p < 1$.

Apartado (v)

Se trata de calcular:

$$\begin{aligned} p(X = n+x | X \geq n) &= \frac{p(\{X = n+x\} \cap \{X \geq n\})}{p(X \geq n)} = \frac{p(X = n+x)}{p(X \geq n)} = \\ &= \frac{(1-p)^{n+x} p}{(1-p)^n} = (1-p)^x p = p(X = x). \end{aligned} \quad \blacktriangleleft$$

Ejemplo 2.5.3 (Transmisión de bits)

En un canal binario simétrico (ejemplo 1.6.3), el 90% de los bits transmitidos se reciben correctamente. Se considera la variable aleatoria:

$$X : \text{“número de bits transmitidos antes del primero alterado”} \sim G(0.1).$$

Esta variable cuenta fracasos antes del primer éxito en pruebas independientes y con probabilidad de éxito constante, que vale 0.1. La probabilidad de que el primer bit alterado sea el 5º vale:

$$p(X = 4) = 0.9^4 \cdot 0.1 \simeq 0.0656.$$

La probabilidad de recibir menos de 5 bits antes del primero alterado vale:

$$p(X < 5) = 1 - p(X \geq 5) = 1 - \frac{0.9^5 \cdot 0.1}{1 - 0.9} = 1 - 0.9^5 \simeq 0.4095.$$

El número esperado de bits recibidos antes del primero alterado es la media de X :

$$\mu(X) = \frac{1}{p} - 1 = \frac{1}{0.1} - 1 = 9.$$

La varianza y el coeficiente de variación del número anterior valen:

$$\sigma^2(X) = \frac{1-0.1}{0.1^2} = 90, \quad \text{CV}(X) = \frac{\sqrt{90}}{9} = \frac{\sqrt{10}}{3} \simeq 105.409\%.$$

Si ya se han recibido 3 bits correctos, la probabilidad de recibir 9 bits antes del primero alterado vale:

$$p(X = 9 | X \geq 3) = p(X = 9 - 3) = 0.9^6 \cdot 0.1 \simeq 0.0531.$$

Se ha usado la falta de memoria de la distribución geométrica. ◀

2.6 Distribución de Poisson

Ciertas variables aleatorias cuentan eventos por unidad de tiempo, de longitud, de superficie o algo similar. Por ejemplo:

- X : “número de defectos en la superficie de un disco duro”,
- Y : “número de aviones que aterrizan en un aeropuerto cada día”,
- Z : “número de llamadas a una central telefónica cada minuto”,
- T : “número de peticiones a un servidor web cada hora”,
- U : “número de mutaciones en una secuencia de ADN sometida a radiación”,
- V : “número de chubascos en una zona geográfica cada año”.

Bajo ciertas condiciones, estas variables se pueden modelar mediante la distribución llamada de Poisson, como se muestra en el ejemplo siguiente.

Ejemplo 2.6.1 (Número de defectos en un disco duro)

Se considera la variable aleatoria:

X : “número de defectos en la superficie de un disco duro”.

Su imagen es $\text{Im}(X) = \{0, 1, 2, 3, \dots\} = \mathbb{N}$, luego X es discreta. Para hallar su función de masa, se suponen dos características:

- La presencia de un defecto en una zona del disco no influye en la probabilidad de que haya defectos a su alrededor. Es decir, los defectos son independientes.
- El promedio de defectos por unidad de superficie es constante en todo el disco.

Estas propiedades son análogas a las que caracterizan las pruebas de Bernoulli. Por tanto, si se divide la superficie del disco en k zonas de igual tamaño, cada una con un defecto a lo sumo, X sigue una distribución binomial de parámetros $X \sim B(k, \lambda/k)$, donde $\lambda = \mu(X)$ es el número esperado de defectos en todo el disco. En consecuencia, para todo $x \in \mathbb{N}$, se verifica que:

$$p(X = x) = \binom{k}{x} \left(\frac{\lambda}{k}\right)^x \left(1 - \frac{\lambda}{k}\right)^{k-x}.$$

Ahora bien, el valor que debe tomar k no se conoce de antemano. Para resolver este problema, se hace k “infinitamente grande”, es decir, se calcula el límite de la expresión anterior cuando k tiende a infinito:

$$\begin{aligned} p(X = x) &= \lim_{k \rightarrow \infty} \frac{k(k-1)\dots(k-x+1)}{x!} \frac{\lambda^x}{k^x} \frac{\left(1 - \frac{\lambda}{k}\right)^k}{\left(1 - \frac{\lambda}{k}\right)^x} = \\ &= \frac{\lambda^x}{x!} \lim_{k \rightarrow \infty} \frac{k(k-1)\dots(k-x+1)}{k^x} \frac{\lim_{k \rightarrow \infty} \left(1 - \frac{\lambda}{k}\right)^k}{\lim_{k \rightarrow \infty} \left(1 - \frac{\lambda}{k}\right)^x} = \\ &= \frac{\lambda^x}{x!} \frac{1}{1} \frac{e^{-\lambda}}{(1-0)^x} = e^{-\lambda} \frac{\lambda^x}{x!}. \end{aligned}$$

La expresión resultante solo depende de λ y de x . ◀

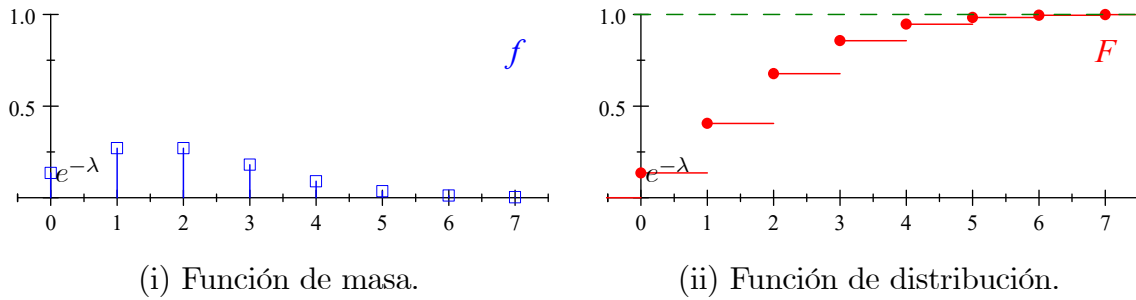


Figura 2.7: Distribución de Poisson $P(\lambda)$.

Generalizando lo observado en el ejemplo anterior, se obtiene la distribución de Poisson.

Definición 2.6.2 (Distribución de Poisson)

Sea $\lambda \in \mathbb{R}$, con $\lambda \geq 0$. La variable aleatoria X sigue una *distribución de Poisson*⁶ de parámetro λ si y solo si $\text{Im}(X) = \mathbb{N}$ y $p(X = x) = e^{-\lambda} \lambda^x / x!$ para todo $x \in \mathbb{N}$. Se nota $X \sim P(\lambda)$. ◀

Teorema 2.6.3 (Propiedades de la distribución de Poisson)

Sea $X \sim P(\lambda)$.

(i) Función de masa (figura 2.7.i): $p(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$

(ii) Función de distribución (figura 2.7.ii): $F(x) = \begin{cases} 0 & x < 0 \\ e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!} & 0 \leq x \end{cases}$

Esta expresión es poco manejable, por lo que se halla tabulada.

(iii) Media: $\mu(X) = \lambda$.

(iv) Varianza: $\sigma^2(X) = \lambda$.

(v) Reproductividad⁷: si $X \sim P(\lambda_X)$ e $Y \sim P(\lambda_Y)$ son independientes, entonces $X + Y \sim P(\lambda_X + \lambda_Y)$.

Demostración: Apartado (i)

6



Siméon Denis Poisson (Pithiviers, France, 1781 – Sceaux, France, 1840).

Matemático, geómetra y físico.

Poisson estudió en la École Polytechnique de Paris, en la que posteriormente fue profesor.

Trabajó en múltiples áreas, por lo que su nombre está vinculado a diversos conceptos matemáticos y físicos: la integral de Poisson; la ecuación de Poisson, en teoría del potencial; los corchetes de Poisson, en ecuaciones diferenciales; la razón de Poisson, en elasticidad, y la constante de Poisson, en electricidad. Introdujo también la expresión *ley de los grandes números*.

La distribución de Poisson aparece por primera vez en su trabajo *Recherches sur la probabilité des jugements en matière criminelle et matière civile*, publicado en 1837.

⁷El recíproco también es cierto (teorema de Raikov): si X e Y son variables aleatorias independientes y $X + Y$ sigue una distribución de Poisson, entonces tanto X como Y siguen distribuciones de Poisson.

La función f es de masa, porque⁸:

$$\sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

Apartado (iii)

La media de X vale:

$$\mu(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

Apartado (iv)

Análogamente, la media de X^2 vale:

$$\begin{aligned} \mu(X^2) &= \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_{n=0}^{\infty} (n+1) \frac{\lambda^n}{n!} = \\ &= e^{-\lambda} \lambda \left(\sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} + \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \right) = e^{-\lambda} \lambda (\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda, \end{aligned}$$

Por tanto, la varianza de X vale:

$$\sigma^2(X) = \mu(X^2) - \mu(X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

Apartado (v)

Para todo $x \in \text{Im}(X + Y) = \mathbb{N}$ se verifica que:

$$\begin{aligned} p(X + Y = x) &= p\left(\bigcup_{i=0}^x (\{X = i\} \cap \{Y = x - i\})\right) = \sum_{i=0}^x p(X = i) p(Y = x - i) = \\ &= \sum_{i=0}^x e^{-\lambda_X} \frac{\lambda_X^i}{i!} e^{-\lambda_Y} \frac{\lambda_Y^{x-i}}{(x-i)!} = e^{-(\lambda_X + \lambda_Y)} \frac{1}{x!} \sum_{i=0}^x \frac{x!}{i!(x-i)!} \lambda_X^i \lambda_Y^{x-i} = \\ &= e^{-(\lambda_X + \lambda_Y)} \frac{(\lambda_X + \lambda_Y)^x}{x!}. \end{aligned}$$

Por tanto, $X + Y$ sigue una distribución de Poisson de parámetro $\lambda_X + \lambda_Y$. ◀

Ejemplo 2.6.4 (Número de correos basura recibidos en 1 hora)

Una persona recibe una media de 2 *correos basura* cada hora. Los mensajes se reciben de modo independiente y su promedio por unidad de tiempo es constante. Se trata de calcular la

⁸Para todo $x \in \mathbb{R}$ la serie siguiente converge y su suma vale:

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x.$$

probabilidad de que la persona reciba más de 3 correos basura y menos de 8 en 4 horas. Para ello, se considera la variable aleatoria:

$$X : \text{“número de correos basura recibidos en 4 horas”} \sim P(4 \cdot 2) = P(8).$$

Esta variable cuenta eventos independientes por unidad de tiempo y con media constante, luego sigue una distribución de Poisson de parámetro $4 \cdot 2$. Se trata de calcular:

$$\begin{aligned} p(3 < X < 8) &= \sum_{i=4}^7 e^{-8} \frac{8^i}{i!} \simeq 0.41058 \\ &= p(X \leq 7) - p(X \leq 3) \simeq 0.4530 - 0.0424 = 0.4106 \quad (\text{tablas}). \quad \blacktriangleleft \end{aligned}$$

Variables aleatorias continuas

En el tema anterior se han estudiado las variables aleatorias discretas. Esencialmente, se trata de variables aleatorias cuya imagen es finita o numerable, como las de los ejemplos 2.1.1 y 2.1.2. Sin embargo, en ciertas aplicaciones prácticas es necesario usar variables aleatorias cuya imagen llena uno o varios intervalos de \mathbb{R} . Estas variables aleatorias se llaman continuas y su estudio es el objeto del presente tema.

3.1 Variables aleatorias continuas

Ejemplo 3.1.1 (Variable aleatoria continua)

Una fresadora fabrica tornillos de cualquier longitud entre 1 cm y 6 cm. Se considera el fenómeno aleatorio que consiste en tomar un tornillo al azar y observar su longitud (en cm). El espacio muestral está formado por todos los números reales entre 1 y 6, es decir, es el intervalo¹ $E := [1, 6]$. Se considera la variable aleatoria:

$$X : \text{“longitud de un tornillo (cm)”}.$$

Su imagen coincide con el espacio muestral: $\text{Im}(X) = [1, 6]$.

Se considera un suceso vinculado a X , por ejemplo:

$$A : \text{“el tornillo mide entre 2 y 5”} = \{2 \leq X \leq 5\}.$$

¿Cómo definir su probabilidad? Una asignación razonable es la proporción entre las longitudes de los intervalos $[2, 5]$ y $[1, 6]$, es decir:

$$p(2 \leq X \leq 5) := \frac{5 - 2}{6 - 1} = \frac{3}{5}.$$

Generalizando esta idea, si $1 \leq a \leq b \leq 6$, se puede definir:

$$p\{a \leq X \leq b\} := \frac{b - a}{6 - 1}.$$

¹Este conjunto es continuo, por lo que $\mathcal{P}(E)$ no es numerable y tiene demasiados elementos para tomarlo como espacio de sucesos.

Esta definición de probabilidad cumple los axiomas de Kolmogorov. Por ejemplo, la probabilidad del espacio muestral vale:

$$p(E) = p(1 \leq X \leq 6) = \frac{6-1}{5} = 1.$$

Para obtener la función de masa de X , hay que calcular $p(X = x)$ para todo $x \in \text{Im}(X)$:

$$p(X = x) = p(x \leq X \leq x) = \frac{x-x}{5} = 0.$$

Este fenómeno es curioso: todos los sucesos elementales tienen probabilidad 0. Es decir, la función de masa es idénticamente nula, por lo que carece de utilidad. Sin embargo, esto no impide que haya sucesos con probabilidad mayor que 0.

Para obtener la función de distribución de X , se observa lo siguiente:

$$\begin{aligned} x < 1 &\Rightarrow p(X \leq x) = p(\emptyset) = 0, \\ 1 \leq x < 6 &\Rightarrow p(X \leq x) = p(1 \leq X \leq x) = \frac{x-1}{5}, \\ x \geq 6 &\Rightarrow p(X \leq x) = p(1 \leq X \leq 6) = 1. \end{aligned}$$

El resultado se muestra en la figura 3.1.i. Se observa que F no es escalonada, sino continua.

Por otra parte, F es derivable en $\mathbb{R} - \{1, 6\}$. Su derivada se denota $f := F'$ (figura 3.1.ii) y se denomina *función de densidad* de X . Tanto la función de distribución como la función de densidad permiten calcular probabilidades relativas a X . Por ejemplo:

$$p(2 \leq X < 5) = \begin{cases} F(5) - F(2) = 4/5 - 1/5 & = 3/5. \\ \int_2^5 f = [x/5]_2^5 = 5/5 - 2/5 \end{cases}$$

La causa es que f se obtiene derivando F , luego F se obtiene integrando f . ◀

La siguiente definición formaliza los conceptos expuestos en el ejemplo anterior.

Definición 3.1.2 (Variable aleatoria continua, función de distribución y función de densidad)

Sea (E, S, p) un espacio probabilístico. Una *variable aleatoria continua* es una función $X : E \rightarrow \mathbb{R}$ tal que:

1. La *función de distribución* de X , $F : \mathbb{R} \rightarrow \mathbb{R}$, dada por $F(x) := p(X \leq x)$, está definida y es continua.
2. La *función de densidad* de X , $f : \mathbb{R} \rightarrow \mathbb{R}$, dada por $f(x) := F'(x)$, está definida y es continua salvo en un conjunto finito de puntos.

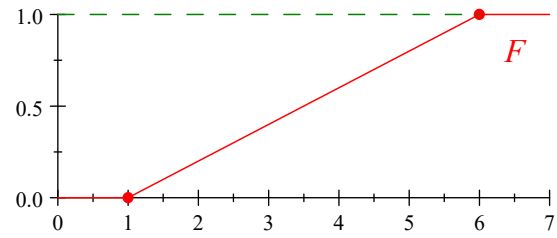
En tal caso, se dice que F es *absolutamente continua*. ◀

Teorema 3.1.3 (Propiedades de la función de distribución de una variable aleatoria continua)

Sea X una variable aleatoria continua. Su función de distribución, F , verifica:

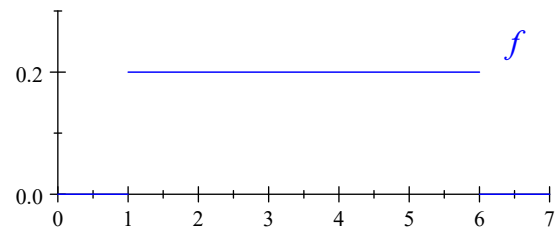
- (i) Es creciente: $a \leq b \Rightarrow F(a) \leq F(b)$.
- (ii) Empieza en 0 y termina en 1, es decir, $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{5} & 1 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$



(i) Función de distribución.

$$f(x) := F'(x) = \begin{cases} 0 & x < 1 \\ 1/5 & 1 < x < 6 \\ 0 & 6 < x \end{cases}$$



(ii) Función de densidad.

Figura 3.1: Probabilidades de una variable aleatoria continua.

(iii) Es absolutamente continua.

Demostración: Las propiedades (i) y (ii) son comunes a las variables aleatorias discretas y la propiedad (iii) forma parte de la definición de variable aleatoria continua. ◀

Teorema 3.1.4 (Propiedades de una función de densidad)

Sea X una variable aleatoria continua. Su función de densidad, f , verifica:

(i) Es positiva: $f(x) \geq 0$ para todo $x \in \mathbb{R}$ donde existe.

(ii) Abarca área 1: $\int_{-\infty}^{\infty} f = 1$.

Demostración: Apartado (i)

La función de distribución de X es creciente. Por tanto, su derivada, que es f , tiene que ser positiva donde exista.

Apartado (ii)

Por el teorema fundamental del cálculo, $\int_{-\infty}^b f = F(b)$. Haciendo que b tienda a ∞ y usando que F es función de distribución, resulta:

$$\int_{-\infty}^{\infty} f = \lim_{b \rightarrow \infty} \int_{-\infty}^b f = \lim_{b \rightarrow \infty} F(b) = 1. \quad \blacktriangleleft$$

Teorema 3.1.5 (Cálculo de probabilidades de una variable aleatoria continua)

Sea X una variable aleatoria continua. Sean F su función de distribución y f su función de densidad. Si $a, b \in \mathbb{R}$ con $a \leq b$, entonces se verifica:

(i) $p(a \leq X \leq b) = p(a \leq X < b) = p(a < X \leq b) = p(a < X < b) = F(b) - F(a) = \int_a^b f$.

(ii) $p(X = a) = 0$.

(iii) $p(X < a) = p(X \leq a) = F(a) = \int_{-\infty}^a f$.

$$(iv) p(X \geq a) = p(X > a) = 1 - F(a) = \int_a^\infty f.$$

Demostración: Apartado (i)

Como F es continua, los cuatro intervalos de extremos a y b son equiprobables:

$$\begin{aligned} p(a \leq X \leq b) &= F(b) - F(a^-) = F(b) - F(a) = p(a < X \leq b), \\ p(a < X < b) &= F(b^-) - F(a) = F(b) - F(a), \\ p(a \leq X < b) &= F(b^-) - F(a^-) = F(b) - F(a). \end{aligned}$$

Por otro lado, f es continua salvo en un conjunto finito de puntos, luego es integrable. Además, F es una primitiva de f . Aplicando la regla de Barrow:

$$\int_a^b f = F(b) - F(a).$$

Apartado (ii)

Es un caso particular del apartado (i): $p(X = a) = p(a \leq X \leq a) = F(a) - F(a) = 0$.

Apartado (iii)

De nuevo por la continuidad de F :

$$p(X < b) = F(b^-) = F(b) = p(X \leq b).$$

Partiendo de la igualdad $\int_a^b f = F(b) - F(a)$ y tomando límites cuando a tiende a $-\infty$, se deduce que:

$$\int_{-\infty}^b f = \lim_{a \rightarrow -\infty} \int_a^b f = \lim_{a \rightarrow -\infty} (F(b) - F(a)) = F(b) - \lim_{a \rightarrow -\infty} F(a) = F(b) - 0 = F(b),$$

ya que F es función de distribución.

Apartado (iv)

Usando la regla del complementario, se verifica:

$$\begin{aligned} p(X \geq a) &= 1 - p(X < a) = 1 - p(X \leq a), \\ p(X > a) &= 1 - p(X \leq a) = 1 - F(a) = \int_{-\infty}^\infty f - \int_{-\infty}^a f = \int_a^\infty f. \end{aligned} \quad \blacktriangleleft$$

3.2 Medidas de una variable aleatoria continua

Como en el caso discreto, las *medidas* de una variable aleatoria continua tienen por objetivo resumir las propiedades esenciales de su distribución de probabilidad. La diferencia más importante es la forma de definir la media, que requiere el cálculo de una integral.

Definición 3.2.1 (Medidas de una variable aleatoria continua)

Sea X una variable aleatoria continua.

- La *media* (o esperanza) de X es la integral² de los valores de X multiplicados por su densidad:

$$\mu(X) = E(X) := \int_{-\infty}^{\infty} x f(x) dx.$$

La media de X se expresa en las mismas unidades que X .

- La *varianza*, la *desviación típica* y el *coeficiente de variación* de X se definen igual que en el caso discreto (definición 2.2.3). ◀

Estas medidas tienen propiedades análogas a las de las variables aleatorias discretas, que se resumen en el siguiente teorema.

Teorema 3.2.2 (Propiedades de la media y de la varianza)

Sean X una variable aleatoria discreta y $a, b \in \mathbb{R}$ dos números reales.

- (i) Si la función de densidad de X es simétrica respecto de la recta $x = a$, entonces $\mu(X) = a$.
- (ii) Media de una combinación afín: $\mu(aX + b) = a\mu(X) + b$.
- (iii) Positividad de la varianza: $\sigma^2(X) \geq 0$.
- (iv) La varianza es igual a la media del cuadrado menos el cuadrado de la media: $\sigma^2(X) = \mu(X^2) - \mu(X)^2$.
- (v) Varianza de una combinación afín: $\sigma^2(aX + b) = a^2\sigma^2(X)$.

Demostración: Las demostraciones son análogas a las del caso discreto. ◀

Teorema 3.2.3 (Propiedades de la media y de la varianza en el caso de dos variables)

Sean X e Y dos variables aleatorias continuas definidas sobre el mismo espacio probabilístico.

- (i) Media de una suma: $\mu(X + Y) = \mu(X) + \mu(Y)$.
- (ii) Varianza de una suma: si X e Y son independientes, entonces $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$.

Demostración: Las demostraciones son análogas a las del caso discreto. ◀

En general, las variables aleatorias discretas se transforman en discretas, mientras que las variables aleatorias continuas se pueden transformar en discretas o en continuas.

Ejemplo 3.2.4 (Longitud de un tornillo: medidas)

Se considera de nuevo la variable aleatoria del ejemplo 3.1.1:

X : “longitud de un tornillo (cm)”.

- La longitud esperada de un tornillo, que es la media de X , vale:

$$\mu(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_1^6 x \frac{1}{5} dx = \left[\frac{x^2}{10} \right]_1^6 = \frac{36 - 1}{10} = \frac{7}{2} = 3.5.$$

²Para que exista la media, se exige que la integral $\int_{-\infty}^{\infty} |x| f(x) dx$ converja.

Es decir, un tornillo mide 3.5 cm en promedio. Esto era de esperar, porque la función de densidad de X es simétrica respecto del eje $x = 3.5$ (figura 3.1.ii).

- Para saber cuánto se aparta la longitud de un tornillo de su valor medio, se calcula la desviación típica de X :

$$\begin{aligned}\mu(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^6 x^2 \frac{1}{5} dx = \left[\frac{x^3}{15} \right]_1^6 = \frac{216 - 1}{15} = \frac{43}{3}, \\ \sigma^2(X) &= \mu(X^2) - \mu(X)^2 = \frac{43}{3} - \left(\frac{7}{2}\right)^2 = \frac{25}{12}, \\ \sigma(X) &= \sqrt{\frac{25}{12}} = \frac{5\sqrt{3}}{6} \simeq 1.44338.\end{aligned}$$

La desviación promedio es 1.44 cm, aproximadamente.

- Para conocer la dispersión relativa de X , se calcula su coeficiente de variación:

$$\text{CV}(X) = \frac{\sigma(X)}{|\mu(X)|} = \frac{5\sqrt{3}/6}{7/2} = \frac{5\sqrt{3}}{21} \simeq 41.2393\%.$$

Este número es bastante grande, por lo que X es bastante dispersa.

Supongamos que un tornillo es aceptable si su longitud está comprendida entre 2 cm y 5 cm, en cuyo caso se ganan 0.14 €. Si el tornillo mide menos de 2 cm, no se puede vender y se pierden 0.03 €. Si el tornillo mide más de 5 cm, se puede reajustar y vender, ganándose 0.06 €. Se trata de calcular el beneficio medio en cada tornillo. Para ello, se considera la variable aleatoria:

Y : “beneficio obtenido en un tornillo (€)”.

Su imagen es $\text{Im}(Y) = \{-0.03, 0.06, 0.14\}$. Por tanto, Y es una variable aleatoria discreta, cuya función de masa es:

$\text{Im}(Y)$	$p(Y = y)$
-0.03	$p(X < 2) = F(2) = 1/5$
0.06	$p(X > 5) = 1 - F(5) = 1 - 4/5 = 1/5$
0.14	$p(2 \leq X \leq 5) = F(5) - F(2) = 4/5 - 1/5 = 3/5$

El beneficio medio en cada tornillo es la media de Y , que vale:

$$\mu(Y) = -0.03 \cdot \frac{1}{5} + 0.06 \cdot \frac{1}{5} + 0.14 \cdot \frac{3}{5} = 0.09.$$

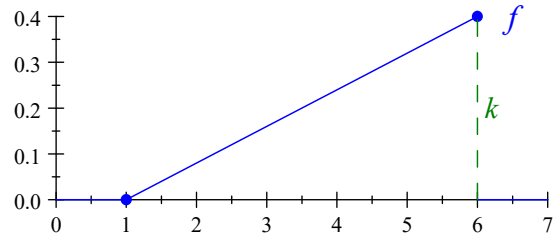
Es decir, en promedio se ganan 0.09 € en cada tornillo.

Consideremos también la variable aleatoria:

Z : “cantidad de metal desperdiciado al fabricar un tornillo (g)”.

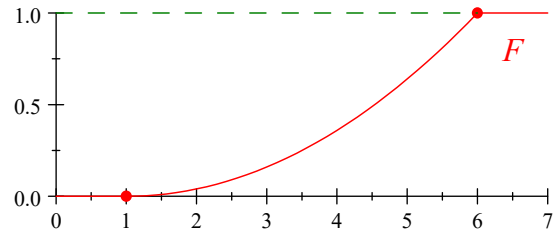
Supongamos que Z está relacionada con la longitud del tornillo por la fórmula $Z = (7 - X)/2$. Se trata de calcular las medidas de Z .

$$f(x) := \begin{cases} 0 & x < 1 \\ 2(x-1)/25 & 1 \leq x \leq 6 \\ 0 & 6 \leq x \end{cases}$$



(i) Función de densidad.

$$F(x) = \begin{cases} 0 & x < 1 \\ (x-1)^2/25 & 1 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$



(ii) Función de distribución.

Figura 3.2: Longitud de un tornillo con distribución no uniforme.

- La cantidad media de metal desperdiciado al fabricar un tornillo es la media de Z :

$$\mu(Z) = \mu\left(\frac{7-X}{2}\right) = \frac{7-\mu(X)}{2} = \frac{7-7/2}{2} = \frac{7}{4} = 1.75.$$

En promedio, se pierden 1.75 g de metal con cada tornillo.

- La desviación típica de Z vale:

$$\sigma^2(Z) = \sigma^2\left(\frac{7-X}{2}\right) = \frac{\sigma^2(X)}{2^2} \Rightarrow \sigma(Z) = \frac{\sigma(X)}{2} = \frac{5\sqrt{3}/6}{2} = \frac{5\sqrt{3}}{12} \simeq 0.721688.$$

- La dispersión relativa de Z vale:

$$CV(Z) = \frac{\sigma(Z)}{|\mu(Z)|} = \frac{5\sqrt{3}/12}{7/4} = \frac{5\sqrt{3}}{21} \simeq 41.2393\%.$$

Ejemplo 3.2.5 (Longitud de un tornillo con distribución no uniforme)

Volvamos otra vez al ejemplo 3.1.1, donde se estudia la variable aleatoria:

X : “longitud de un tornillo (cm)”,

cuya imagen es el intervalo $[1, 6]$. En ese ejemplo se ha supuesto que la probabilidad se distribuye uniformemente a lo largo del intervalo $[1, 6]$, de modo que los subintervalos de la misma longitud tienen la misma probabilidad. En realidad, puede ocurrir que los tornillos cortos, de longitud próxima a 1 cm, se fabrican con menos frecuencia que los tornillos largos, de longitud próxima a 6 cm. Esto significa que $p(1 \leq X \leq 2)$ debe ser menor que $p(5 \leq X \leq 6)$, aunque los dos intervalos tienen la misma longitud.

Para modelar esta situación hay que modificar la función de densidad f . Una posibilidad es que f tenga forma triangular, como en la figura 3.2.i. De este modo, las probabilidades crecen a

lo largo del intervalo $[1, 6]$. En dicha figura hay que determinar k , que es la altura del triángulo, para que f sea una función de densidad. Esto exige que $k \geq 0$, para que f sea positiva, y que el área bajo la curva sea 1. Usando la fórmula del área de un triángulo, se obtiene:

$$\int_{-\infty}^{\infty} f = \frac{(6-1)k}{2} = \frac{5k}{2} = 1 \Rightarrow k = \frac{2}{5}.$$

Ajustando los coeficientes de la recta que forma la hipotenusa del triángulo, se obtiene para f la expresión de la figura 3.2.i.

La función de distribución de X es $F(x) = p(X \leq x) = \int_{-\infty}^x f$. Para calcular este valor, se consideran 3 casos:

- Si $x < 1$, entonces $F(x) = \int_{-\infty}^x 0 = 0$.
- Si $1 \leq x < 6$, entonces:

$$F(x) = \int_{-\infty}^1 f + \int_1^x f = 0 + \int_1^x \frac{2(t-1)}{25} dt = \left[\frac{(t-1)^2}{25} \right]_1^x = \frac{(x-1)^2}{25}.$$

- Si $x \geq 6$, entonces:

$$F(x) = \int_{-\infty}^6 f + \int_6^x f = F(6) + 0 = 1.$$

En resumen, la expresión de F y su gráfica son las de la figura 3.2.ii.

La función de distribución de X permite calcular probabilidades. Por ejemplo, la probabilidad de que un tornillo mida entre 2 cm y 5 cm, sabiendo que mide más de 3 cm, vale:

$$\begin{aligned} p(2 \leq X \leq 5 \mid X > 3) &= \frac{p(\{2 \leq X \leq 5\} \cap \{X > 3\})}{p(X > 3)} = \frac{p(3 < X \leq 5)}{1 - p(X \leq 3)} = \\ &= \frac{F(5) - F(3)}{1 - F(3)} = \frac{16/25 - 4/25}{1 - 4/25} = \frac{12/25}{21/25} = \frac{4}{7}. \end{aligned}$$

Las medidas de la variable X son las siguientes:

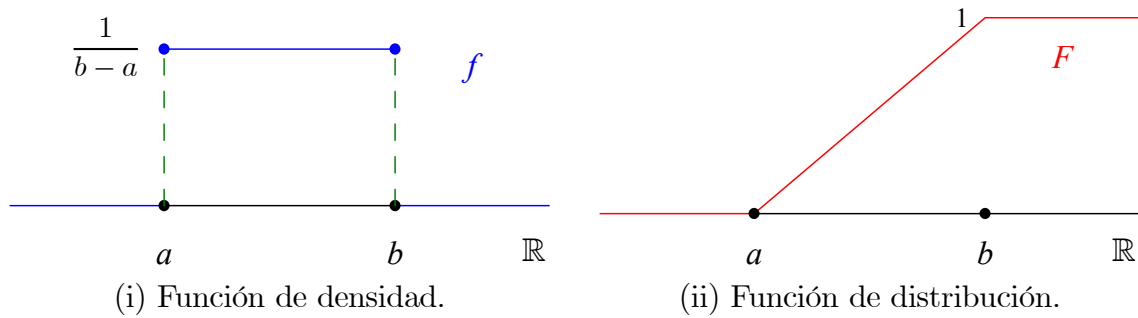
- La longitud media de un tornillo vale:

$$\begin{aligned} \mu(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_1^6 x \frac{2(x-1)}{25} dx = \frac{2}{25} \int_1^6 (x^2 - x) dx = \frac{2}{25} \left[\frac{x^3}{3} - \frac{x^2}{2} \right]_1^6 = \\ &= \frac{2}{25} \left(54 - \frac{-1}{6} \right) = \frac{2}{25} \cdot \frac{325}{6} = \frac{13}{3} \simeq 4.3333. \end{aligned}$$

En promedio, un tornillo mide 4.33 cm, aproximadamente. Este punto ya no es el centro del intervalo $[1, 6]$.

- La desviación típica de la longitud de un tornillo vale:

$$\begin{aligned} \mu(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^6 x^2 \frac{2(x-1)}{25} dx = \frac{2}{25} \left[\frac{x^4}{4} - \frac{x^3}{3} \right]_1^6 = \frac{121}{6}, \\ \sigma^2(X) &= \mu(X^2) - \mu(X)^2 = \frac{121}{6} - \left(\frac{13}{3} \right)^2 = \frac{25}{18}, \\ \sigma(X) &= \sqrt{\frac{25}{18}} = \frac{5\sqrt{2}}{6} \simeq 1.17851. \end{aligned}$$

Figura 3.3: Distribución uniforme $U(a, b)$.

- La dispersión relativa de X vale:

$$\text{CV}(X) = \frac{5\sqrt{2}/6}{13/3} = \frac{5\sqrt{2}}{26} \simeq 27.1964\%.$$

3.3 Distribución uniforme

Hay ciertas distribuciones de probabilidad continuas que aparecen con frecuencia en la práctica. A continuación se estudian algunas de las más usadas. Empezamos generalizando la distribución uniforme observada en el ejemplo 3.1.1.

Se consideran las variables aleatorias:

$$\begin{aligned} X &: \text{“longitud de un tornillo (cm)”}, & \text{Im}(X) &= [1, 6], \\ Y &: \text{“número real obtenido al azar en el intervalo } [2, 5]\text{”}, & \text{Im}(Y) &= [2, 5], \\ Z &: \text{“error de redondeo en una medición”}. & \text{Im}(Z) &= [-0.5, 0.5]. \end{aligned}$$

En todos estos casos:

- La imagen de la variable es un intervalo acotado, $[a, b] \subset \mathbb{R}$.
- Los subintervalos de $[a, b]$ de igual longitud tienen igual probabilidad.

Esto caracteriza un tipo de distribución de probabilidad.

Definición 3.3.1 (Distribución uniforme)

La variable aleatoria X sigue una *distribución uniforme en el intervalo* $[a, b]$ ($a < b$) si y solo si su función de densidad vale:

$$f(x) := \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

Se nota $X \sim U(a, b)$.

Teorema 3.3.2 (Propiedades de la distribución uniforme)

Sea $X \sim U(a, b)$.

(i) La función f es de densidad (figura 3.3.i).

$$(ii) \text{ Función de distribución (figura 3.3.ii): } F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & b \leq x \end{cases}$$

$$(iii) \text{ Media: } \mu(X) = \frac{a+b}{2}.$$

$$(iv) \text{ Varianza: } \sigma^2(X) = \frac{(b-a)^2}{12}.$$

Demostración: Apartado (i)

La función f es positiva, continua en $\mathbb{R} - \{a, b\}$ y abarca área: $\int_{-\infty}^{\infty} f = (b-a) \frac{1}{b-a} = 1$.

Apartado (ii)

Si $a \leq x < b$, la función de distribución en x vale:

$$F(x) = \int_{-\infty}^x f = \int_a^x \frac{1}{b-a} dt = \left[\frac{t}{b-a} \right]_a^x = \frac{x-a}{b-a}.$$

Apartado (iii)

La función de densidad de X es simétrica respecto de la recta $x = (a+b)/2$.

Apartado (iv)

La media de X^2 vale:

$$\mu(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \left[\frac{x^3}{3(b-a)} \right]_a^b = \frac{b^3 - a^3}{3(b-a)}.$$

Por tanto, la varianza de X vale:

$$\begin{aligned} \sigma^2(X) &= \mu(X^2) - \mu(X)^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{4(b^3 - a^3) - 3(b-a)(a^2 + 2ab + b^2)}{12(b-a)} = \\ &= \frac{b^3 - 3ab^2 + 3a^2b - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}. \end{aligned}$$

3.4 Distribución exponencial

Se consideran las variables aleatorias:

X : “duración de un transistor (años)” ,

Y : “tiempo de reparación de un ordenador (h)” ,

Z : “tiempo de espera de un proceso en la cola de un sistema operativo (ms)” ,

T : “tiempo entre dos peticiones a un servidor web (s)” .

En todos estos casos:

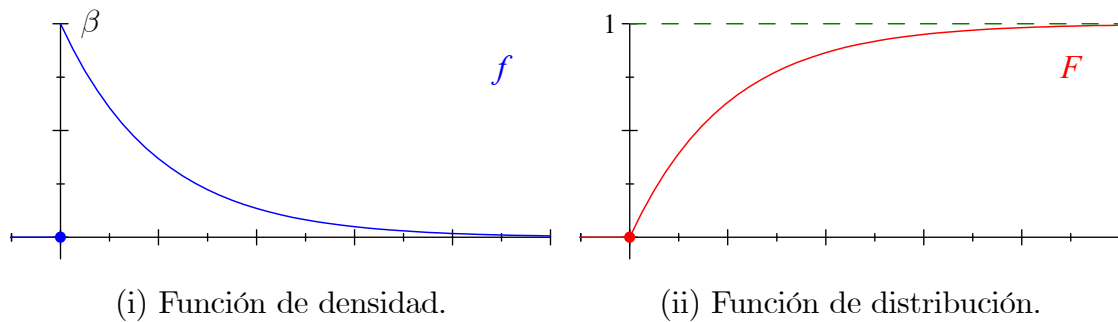


Figura 3.4: Distribución exponencial $\text{Exp}(\beta)$.

- La imagen de la variable es el intervalo $[0, \infty)$.
- Los subintervalos de $[0, \infty)$ de igual longitud tienen menor probabilidad cuanto más se alejan de 0.

Una forma de modelar esta situación es usar la distribución exponencial.

Definición 3.4.1 (Distribución exponencial)

La variable aleatoria X sigue una *distribución exponencial de parámetro* $\beta > 0$ si y solo si su función de densidad vale:

$$f(x) := \begin{cases} 0 & x \leq 0 \\ \beta e^{-\beta x} & 0 < x \end{cases}$$

Se nota $X \sim \text{Exp}(\beta)$. ◀

Teorema 3.4.2 (Propiedades de la distribución exponencial)

Sea $X \sim \text{Exp}(\beta)$.

(i) La función f es de densidad (figura 3.4.i).

(ii) Función de distribución (figura 3.4.ii): si $x \geq 0$, entonces $p(X > x) = e^{-\beta x}$; por tanto:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\beta x} & 0 \leq x \end{cases}$$

(iii) Media: $\mu(X) = 1/\beta$.

(iv) Varianza: $\sigma^2(X) = 1/\beta^2$.

(v) Carece de memoria: si $t, x \geq 0$, entonces:

$$p(X > t + x \mid X > t) = p(X > x).$$

Es decir, la probabilidad de que X tome un mayor que $t + x$, sabiendo que toma un valor mayor que t , coincide con la probabilidad de que tome un valor mayor que x .

Demostración: Apartado (i)

La función f es positiva, continua en $\mathbb{R} - \{0\}$ y abarca área:

$$\int_{-\infty}^{\infty} f = \int_0^{\infty} \beta e^{-\beta x} dx = [-e^{-\beta x}]_0^{\infty} = 0 - (-1) = 1.$$

Apartado (ii)

Si $x \geq 0$, entonces:

$$F(x) = \int_{-\infty}^x f = \int_0^x \beta e^{-\beta t} dt = [-e^{-\beta t}]_0^x = -e^{-\beta x} - (-1) = 1 - e^{-\beta x}.$$

Apartados (iii) y (iv) Ver el teorema 3.6.3.

Apartado (v)

Se trata de calcular:

$$\begin{aligned} p(X > t+x | X > t) &= \frac{p(\{X > t+x\} \cap \{X > t\})}{p(X > t)} = \frac{p(X > t+x)}{p(X > t)} = \\ &= \frac{e^{-\beta(t+x)}}{e^{-\beta t}} = e^{-\beta x} = p(X > x). \end{aligned} \quad \blacktriangleleft$$

En general, la distribución exponencial:

- Aparece en el estudio de los tiempos de vida de ciertos procesos (tiempos de espera).
- Desempeña el papel de la distribución geométrica cuando el tiempo se mide de forma continua.
- Aparece en problemas de tráfico en redes.

La afirmación de que una variable aleatoria sigue una distribución exponencial no se puede probar matemáticamente. En todos los casos, requiere comprobación experimental.

Ejemplo 3.4.3 (Distribución exponencial)

Se sabe, por experiencia, que la duración de los transistores de cierta marca sigue una distribución exponencial, pero se desconoce el valor del parámetro β . También se sabe que la mitad de dichos transistores dura más de un año. Para calcular β , se considera la variable aleatoria:

$$X : \text{“duración de un transistor (años)”} \sim \text{Exp}(\beta).$$

La información disponible indica que $p(X > 1) = 0.5$, luego:

$$p(X > 1) = \int_1^{\infty} \beta e^{-\beta x} dx = [-e^{-\beta x}]_1^{\infty} = e^{-\beta} = \frac{1}{2} \Rightarrow -\beta = -\ln(2) \Rightarrow \beta = \ln(2).$$

Para despejar β se toman logaritmos neperianos.

Si un transistor lleva 4 años funcionando, ¿cuál es la probabilidad de que dure más de 7 años? Se trata de calcular una probabilidad condicionada:

$$\begin{aligned} p(X > 7 | X \geq 4) &= p(X > 7 - 4) = p(X > 3) = \int_3^{\infty} \beta e^{-\beta x} dx = \\ &= [-e^{-\beta x}]_3^{\infty} = e^{-3 \ln(2)} = \frac{1}{8} = 0.125. \end{aligned}$$

Se ha usado que la distribución exponencial carece de memoria y que X es continua.

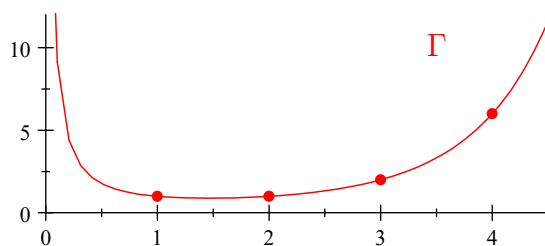


Figura 3.5: Función Gamma $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt$.

¿Cuál es la duración esperada de un transistor de esta marca? Eso es la media de X , que vale:

$$\mu(X) = \frac{1}{\beta} = \frac{1}{\ln(2)} \simeq 1.4427.$$

En promedio, cada transistor dura 1.4427 años. La varianza de la duración de un transistor es:

$$\sigma^2(X) = \frac{1}{\beta^2} = \frac{1}{\ln(2)^2} \simeq 0.721348.$$

Esta cantidad está expresada en años². El coeficiente de variación vale:

$$\text{CV}(X) = \frac{\sigma(X)}{|\mu(X)|} = \frac{\sqrt{1/\beta^2}}{|1/\beta|} = 100\%.$$

Este coeficiente vale 1 para cualquier distribución exponencial. ◀

3.5 Distribución normal

La función Gamma aparece en el estudio de la distribución normal. A continuación se recuerdan su definición y sus principales propiedades.

Teorema 3.5.1 (Propiedades de la función Γ)

La función Gamma (figura 3.5) se define como:

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt \quad \text{para todo } x > 0.$$

Para todo $x > 0$ se verifica:

- (i) $\Gamma(x) > 0$.
- (ii) $\Gamma(x+1) = x\Gamma(x)$.
- (iii) $\Gamma(n+1) = n!$ para todo $n \in \mathbb{N}$.
- (iv) $\Gamma(1/2) = \sqrt{\pi}$. ◀

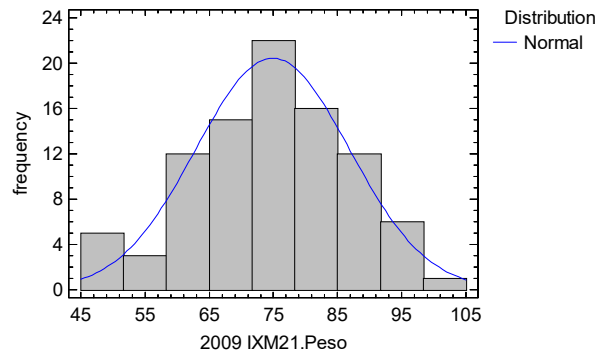


Figura 3.6: Histograma del peso de una persona.

Cuando se toman muestras de ciertas variables aleatorias, se obtienen histogramas acampanados. Por ejemplo:

- X : “nota de un alumno en un examen”,
- Y : “peso de una persona (kg)”,
- Z : “altura de un árbol de cierta especie (m)”,
- T : “diámetro de una pieza (mm)”.

Por tanto, las funciones de densidad correspondientes deben ser acampanadas (figura 3.6). Para modelar este tipo de variables, el matemático Abraham de Moivre³ propuso la distribución normal, que tiene la siguiente función de densidad.

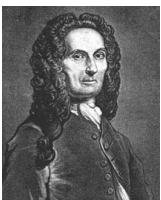
Definición 3.5.2 (Distribución normal)

La variable aleatoria X sigue una *distribución normal de parámetros* $\mu \in \mathbb{R}$ y $\sigma > 0$ si y solo si su función de densidad vale:

$$f(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Esta función se llama *campana de Gauss*⁴. Se nota $X \sim N(\mu, \sigma)$. ◀

3



Abraham de Moivre (Vitry-le-François, Francia, 1667 – London, Inglaterra, 1754).

Matemático. De religión calvinista, de Moivre emigró a Inglaterra tras la revocación del Edicto de Nantes (1685). Allí, se ganó la vida como profesor particular de matemáticas.

De Moivre contribuyó a la geometría analítica. Por ejemplo, la llamada fórmula de De Moivre, $(\cos(x) + i \operatorname{sen}(x))^n = \cos(nx) + i \operatorname{sen}(nx)$, liga los números complejos con la trigonometría.

De Moivre es conocido, sobre todo, por su libro *The Doctrine of Chances: A method of calculating the probabilities of events in play*, publicado por primera vez en 1711. En esta obra sobre probabilidad, introduce el concepto de independencia estadística y la distribución normal. Además, aproxima la distribución binomial por la normal cuando el número de pruebas es grande.

4

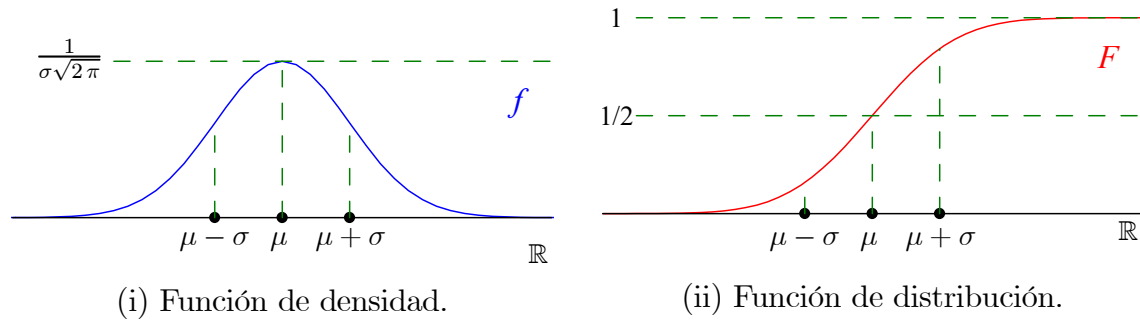


Johann Carl Friedrich Gauss (Brunswick, Alemania, 1777 – Göttingen, Alemania, 1855).

Matemático y científico. Gauss fue profesor en la universidad de Göttingen y director de su observatorio astronómico.

Gauss realizó importantes contribuciones en teoría de números, análisis, geometría diferencial, geodesia, magnetismo, astronomía y óptica. Escribió su obra más importante, *Disquisitiones Arithmeticae*, a los 21 años. La aritmética modular es una de sus invenciones.

Gauss se hizo famoso por predecir la órbita del asteroide Ceres, descubierto en su época. En sus trabajos sobre errores desarrolló la distribución normal.

Figura 3.7: Distribución normal $N(\mu, \sigma)$.**Teorema 3.5.3** (Propiedades de la campana de Gauss)

La campana de Gauss (figura 3.7 i) tiene las siguientes propiedades:

- (i) Es estrictamente positiva e indefinidamente derivable.
- (ii) Es simétrica respecto de la recta vertical $x = \mu$.
- (iii) Tiene como asíntota horizontal el eje de abscisas, $y = 0$.
- (iv) Tiene máximo en el punto $\left[\mu, \frac{1}{\sigma\sqrt{2\pi}}\right]$ y puntos de inflexión en $x = \mu - \sigma$ y $x = \mu + \sigma$. ◀

Teorema 3.5.4 (Propiedades de la distribución normal)

Sea $X \sim N(\mu, \sigma)$.

- (i) La función f es de densidad (figura 3.7.i).
- (ii) Función de distribución (figura 3.7.ii): $F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dt$.

F no es expresable en términos elementales. Sus valores se aproximan por métodos numéricos. F es estrictamente creciente, porque su derivada, f , es estrictamente positiva.
- (iii) Media: $\mu(X) = \mu$.
- (iv) Varianza: $\sigma^2(X) = \sigma^2$.
- (v) Reproductividad respecto de ambos parámetros: si $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$ son independientes, entonces $X + Y \sim N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$.
- (vi) Transformación lineal: si $a \neq 0$ y $b \in \mathbb{R}$, entonces $aX + b \sim N(a\mu + b, |a|\sigma)$.
- (vii) Tipificación:

$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$

La distribución $N(0, 1)$ se llama normal estándar y está tabulada.

Demostración: Apartado (i)

La función f es positiva y continua en \mathbb{R} . Por la simetría de f respecto de $x = \mu$, el área bajo la curva vale $\int_{-\infty}^{\infty} f = 2 \int_{\mu}^{\infty} f$. Para calcular esta integral, se usa el cambio de variable:

$$t = \left(\frac{x - \mu}{\sqrt{2}\sigma}\right)^2 \Rightarrow dt = 2 \frac{x - \mu}{\sqrt{2}\sigma} \frac{1}{\sqrt{2}\sigma} dx = 2\sqrt{t} \frac{1}{\sqrt{2}\sigma} dx \Rightarrow dx = \frac{\sqrt{2}\sigma}{2\sqrt{t}} dt.$$

El resultado es:

$$\begin{aligned} \int_{-\infty}^{\infty} f &= 2 \int_{\mu}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} dx = 2 \int_0^{\infty} \frac{1}{\sigma\sqrt{2}\sqrt{\pi}} e^{-t} \frac{\sqrt{2}\sigma}{2\sqrt{t}} dt = \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} t^{1/2-1} e^{-t} dt = \frac{1}{\sqrt{\pi}} \Gamma(1/2) = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1. \end{aligned}$$

Apartado (iii)

La campana de Gauss es simétrica respecto de la recta $x = \mu$. Por tanto, $\mu(X) = \mu$.

Apartado (v)

Ver Fernández-Abascal [8, página 443, Proposición 10.5].

Apartado (vii)

Aplicando el apartado anterior, se obtiene:

$$X \sim N(\mu, \sigma) \Rightarrow X - \mu \sim N(\mu - \mu, \sigma) = N(0, \sigma) \Rightarrow \frac{X - \mu}{\sigma} \sim N\left(\frac{0}{\sigma}, \frac{\sigma}{\sigma}\right) = N(0, 1). \quad \blacktriangleleft$$

La distribución normal aproxima gran cantidad de variables aleatorias, entre ellas:

- Propiedades físicas de la materia, como la resistencia a la fractura de una pieza.
- Errores cometidos al medir ciertas magnitudes.
- Caracteres morfológicos, como el peso, la estatura, la longitud o la envergadura de un ejemplar de cierta especie.
- Caracteres fisiológicos, como el efecto de un fármaco.
- Caracteres psicológicos, como el cociente intelectual de una persona.
- Caracteres sociológicos, como el consumo de cierto producto por un mismo grupo de personas.

En general, la distribución normal modela variables aleatorias que son suma de muchas causas independientes, cada una de ellas con poca influencia en el resultado final. Además, la distribución normal es básica en Estadística Inferencial. Por todo ello, constituye el modelo continuo más importante.

El adjetivo “normal” tiene carácter histórico, porque se llegó a creer que todas las distribuciones se pueden aproximar por una normal, salvo unas pocas que serían “anormales”. Esto no es cierto y dicho adjetivo debe entenderse como un rasgo de tradición.

Ejemplo 3.5.5 (Distribución normal)

En un ordenador, el tiempo de espera de un proceso, hasta que se le asigna el procesador, sigue una distribución normal de media 10s y desviación típica 2s. En un momento dado hay 4 procesos independientes en la cola del procesador. ¿Cuál es tiempo medio total de espera de los 4 procesos? Para calcularlo, se consideran las variables aleatorias:

$$\begin{aligned} X_i &: \text{“tiempo de espera de un proceso (s)”} \sim N(10, 2), \quad i = 1, 2, 3, 4, \\ X &: \text{“tiempo total de espera de 4 procesos (s)”}. \end{aligned}$$

Es obvio que $X = X_1 + \dots + X_4$, por lo que el tiempo medio total de espera vale:

$$\mu(X) = \mu(X_1 + \dots + X_4) = \mu(X_1) + \dots + \mu(X_4) = 4 \cdot 10 = 40.$$

La espera media de los 4 procesos será 40 s.

¿Cuál es la probabilidad de que los 4 procesos esperen más de 45 s en total? Esto es $p(X > 45)$. Como los procesos son independientes, las variables X_i también son independientes. Usando la reproductividad de la distribución normal, se verifica:

$$\begin{aligned} X &\sim N\left(\mu(X_1) + \dots + \mu(X_4), \sqrt{\sigma^2(X_1) + \dots + \sigma^2(X_4)}\right) = \\ &= N\left(4 \cdot 10, \sqrt{4 \cdot 2^2}\right) = N(40, 4). \end{aligned}$$

Se puede calcular una probabilidad relativa a cualquier distribución normal usando la tipificación y la tabla de la distribución $N(0, 1)$. El proceso es el siguiente:

$$\begin{aligned} p(X > 45) &= p\left(\frac{X - \mu}{\sigma} > \frac{45 - 40}{4}\right) = p(N(0, 1) > 1.25) = \\ &= 1 - p(N(0, 1) \leq 1.25) \simeq 1 - 0.8944 = 0.1056 \quad (\text{tabla de } N(0, 1)). \end{aligned}$$

Primero se ha tipificado la variable X , usando que si $X \sim N(\mu, \sigma)$, entonces $\frac{X - \mu}{\sigma} \sim N(0, 1)$. En este caso, $\mu = 40$ y $\sigma = 4$. Después, se ha usado la regla del complementario. Finalmente, se ha consultado la tabla de la distribución $N(0, 1)$ en la intersección de la fila 1.2 con la columna 0.05. Allí figura el valor 0.8944. ◀

3.6 Distribución gamma

La distribución gamma es útil porque su función de densidad es muy versátil. En concreto:

- Sustituye a la distribución normal para modelar variables aleatorias continuas que toman valores positivos.
- Modela tiempos de espera medidos de modo continuo.
- Modela tiempos de vida de componentes.

Definición 3.6.1 (Distribución gamma)

La variable aleatoria X sigue una *distribución gamma de parámetros* $\alpha > 0$ y $\beta > 0$ si y solo si su función de densidad vale:

$$f(x) := \begin{cases} 0 & x \leq 0 \\ \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & 0 < x \end{cases}$$

Se nota $X \sim \gamma(\alpha, \beta)$. ◀

La gráfica de f exhibe perfiles muy diferentes, que dependen del valor de α , llamado parámetro de forma. El valor de β solo afecta a la escala. La figura 3.8 muestra los casos principales, todos ellos con $\beta = 1$. A medida que α aumenta, la gráfica de f se asemeja a una campana de Gauss. El teorema siguiente detalla estas propiedades.

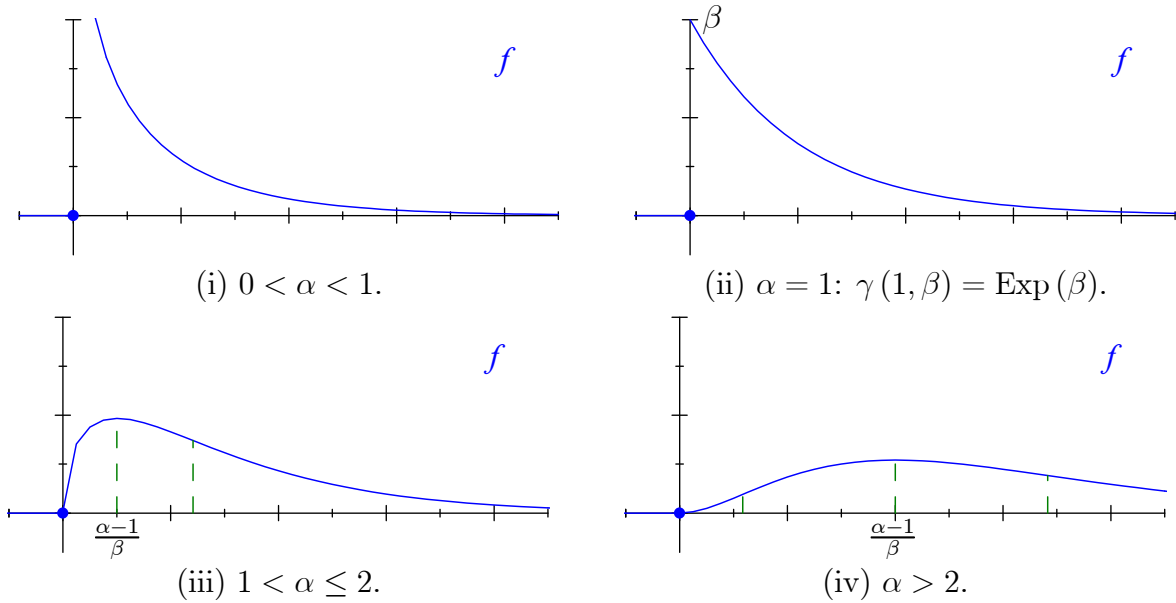


Figura 3.8: Función de densidad gamma $\gamma(\alpha, \beta)$.

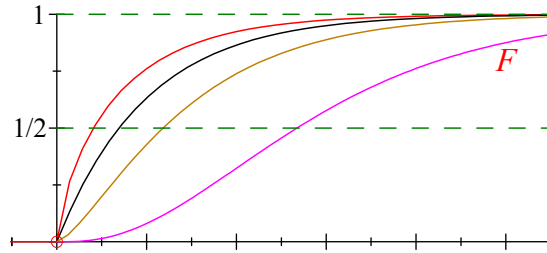


Figura 3.9: Función de distribución gamma $\gamma(\alpha, \beta)$.

Teorema 3.6.2 (Propiedades de la densidad γ)

La función de densidad gamma (figura 3.8) tiene las siguientes propiedades:

- (i) Es nula en $(-\infty, 0]$, estrictamente positiva en $(0, \infty)$ e indefinidamente derivable en $\mathbb{R} - \{0\}$.
- (ii) Tiene como asíntota horizontal el eje de abscisas, $y = 0$.
- (iii) Si $0 < \alpha < 1$ (figura 3.8.i), entonces f tiene como asíntota vertical el eje de ordenadas, $x = 0$; además, f es estrictamente decreciente y convexa en $(0, \infty)$.
- (iv) Si $\alpha = 1$ (figura 3.8.ii), entonces se obtiene $\gamma(1, \beta) = \text{Exp}(\beta)$.
- (v) Si $1 < \alpha \leq 2$ (figura 3.8.iii), entonces f tiene tangente vertical en $x = 0^+$, máximo en $x = \frac{\alpha-1}{\beta}$ y un punto de inflexión en $x = \frac{\alpha-1+\sqrt{\alpha-1}}{\beta}$.
- (vi) Si $\alpha > 2$ (figura 3.8.iv), entonces f tiene tangente horizontal en $x = 0$, máximo en $x = \frac{\alpha-1}{\beta}$ y dos puntos de inflexión en $x = \frac{\alpha-1 \pm \sqrt{\alpha-1}}{\beta}$. ◀

Teorema 3.6.3 (Propiedades de la distribución γ)

Sea $X \sim \gamma(\alpha, \beta)$.

- (i) La función f es de densidad (figura 3.8).

$$(ii) \text{ Función de distribución (figura 3.9): } F(x) = \begin{cases} 0 & x < 0 \\ \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-\beta t} dt & 0 \leq x \end{cases}$$

Si α es entero, se puede calcular $F(x)$ de modo exacto. En caso contrario, se aproxima la integral por métodos numéricos.

$$(iii) \text{ Media y varianza: } \mu(X) = \alpha/\beta \text{ y } \sigma^2(X) = \alpha/\beta^2.$$

(iv) *Reproductividad respecto del primer parámetro:* si $X \sim \gamma(\alpha_X, \beta)$ e $Y \sim \gamma(\alpha_Y, \beta)$ son independientes, entonces $X + Y \sim \gamma(\alpha_X + \alpha_Y, \beta)$.

Demostración: Apartado (i)

La función f es positiva y continua en $\mathbb{R} - \{0\}$. Para calcular el área bajo la curva, se usa el cambio de variable:

$$t = \beta x \Rightarrow x = \frac{t}{\beta} \Rightarrow dx = \frac{1}{\beta} dt.$$

Teniendo en cuenta la definición de la función Gamma, el resultado es:

$$\begin{aligned} \int_{-\infty}^{\infty} f &= \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-t} \frac{1}{\beta} dt = \\ &= \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{t^{\alpha-1}}{\beta^{\alpha-1}} e^{-t} \frac{1}{\beta} dt = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha-1} e^{-t} dt = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1. \end{aligned}$$

Apartado (iii)

Haciendo el cambio de variable $t = \beta x$, la media de X vale:

$$\begin{aligned} \mu(X) &= \int_0^{\infty} x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \int_0^{\infty} \frac{t^\alpha}{\Gamma(\alpha)} e^{-t} \frac{1}{\beta} dt = \frac{1}{\beta \Gamma(\alpha)} \int_0^{\infty} t^\alpha e^{-t} dt = \\ &= \frac{1}{\beta \Gamma(\alpha)} \Gamma(\alpha + 1) = \frac{\alpha \Gamma(\alpha)}{\beta \Gamma(\alpha)} = \frac{\alpha}{\beta}. \end{aligned}$$

Análogamente, la media de X^2 vale:

$$\begin{aligned} \mu(X^2) &= \int_0^{\infty} x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \int_0^{\infty} \frac{t^{\alpha+1}}{\beta \Gamma(\alpha)} e^{-t} \frac{1}{\beta} dt = \frac{1}{\beta^2 \Gamma(\alpha)} \int_0^{\infty} t^{\alpha+1} e^{-t} dt = \\ &= \frac{1}{\beta^2 \Gamma(\alpha)} \Gamma(\alpha + 2) = \frac{(\alpha + 1) \alpha \Gamma(\alpha)}{\beta^2 \Gamma(\alpha)} = \frac{\alpha^2 + \alpha}{\beta^2}. \end{aligned}$$

Por tanto, la varianza de X vale:

$$\sigma^2(X) = \mu(X^2) - \mu(X)^2 = \frac{\alpha^2 + \alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

Apartado (iv) Ver Fernández-Abascal [8, página 451, Proposición 10.7]. ◀

Ejemplo 3.6.4 (Tiempo ejecución de 100 procesos)

En un ordenador, el tiempo de ejecución de un proceso sigue una distribución exponencial de

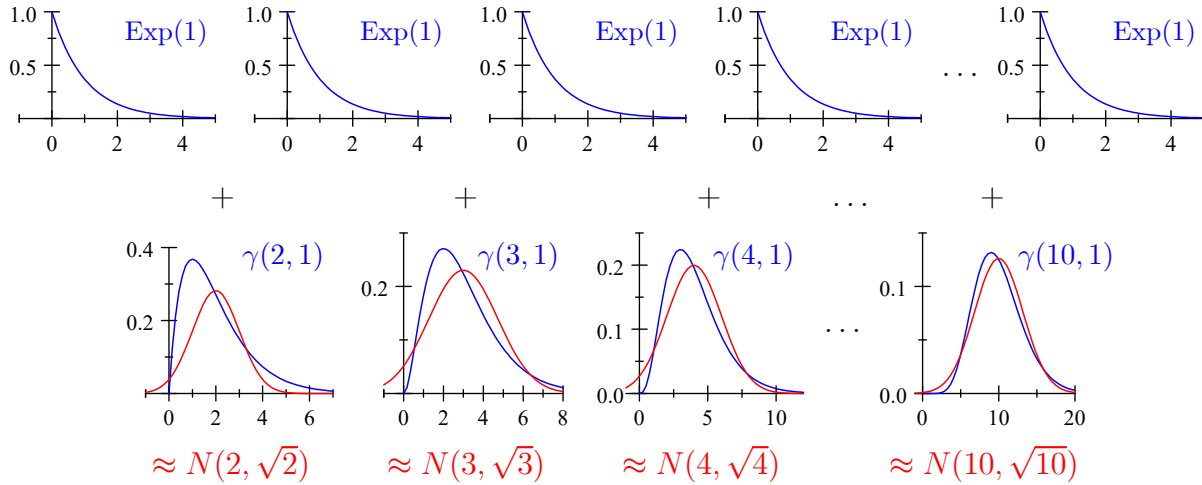


Figura 3.10: Suma de exponenciales comparada con normal.

media 2 s. ¿Qué distribución sigue el tiempo total de ejecución de 100 procesos independientes? Para saberlo, se consideran las variables aleatorias:

$$X_i : \text{“tiempo de ejecución del proceso } i \text{ (s)”} \sim \text{Exp}(\beta), \quad i = 1..100,$$

$$X : \text{“tiempo total de ejecución de 100 procesos (s)”}.$$

El parámetro de una distribución exponencial es el inverso de su media, que vale 2. Por tanto, $X_i \sim \text{Exp}(1/2) = \gamma(1, 1/2)$. Como estas variables son independientes y la distribución gamma es reproductiva respecto de su primer parámetro, se deduce que:

$$X = X_1 + \dots + X_{100} \sim \gamma\left(\overbrace{1 + \dots + 1}^{100}, 1/2\right) = \gamma(100, 1/2).$$

Esta distribución ya no es exponencial. ◀

3.7 Teorema del límite central

En esta sección se analiza lo que sucede cuando se suman variables aleatorias *equidistribuidas* (con la misma distribución) e independientes.

Tomemos como ejemplo la distribución exponencial. Sean $X_1, \dots, X_n \sim \text{Exp}(\beta)$ independientes. Sus funciones de densidad (figura 3.10) son muy diferentes de una campana de Gauss. Sin embargo, como $\text{Exp}(\beta) = \gamma(1, \beta)$ y la distribución gamma es reproductiva respecto de su primer parámetro, se verifica lo siguiente:

$$X_1 + X_2 \sim \gamma(2, \beta), \quad X_1 + X_2 + X_3 \sim \gamma(3, \beta), \quad X_1 + X_2 + X_3 + X_4 \sim \gamma(4, \beta),$$

$$X_1 + \dots + X_5 \sim \gamma(5, \beta), \quad \dots, \quad X_1 + \dots + X_n \sim \gamma(n, \beta).$$

Las funciones de densidad de las sumas se parecen a una campana de Gauss (figura 3.10). ¿Con qué parámetros? Si dos distribuciones son similares, su media y su varianza también deben

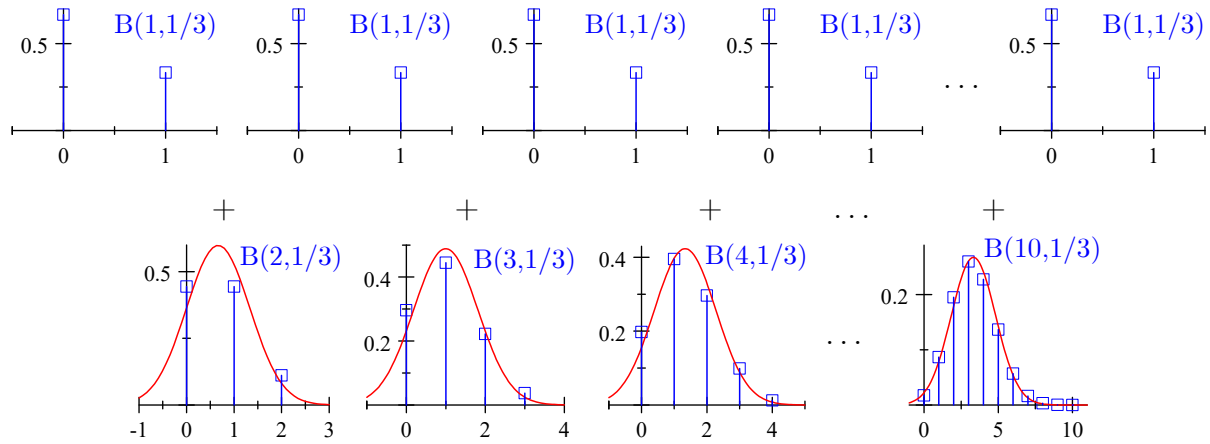


Figura 3.11: Suma de binomiales comparada con normal.

serlo. Por tanto, es razonable comparar $\gamma(n, \beta)$ con la distribución normal que tiene la misma media y la misma desviación típica, es decir, $N(n/\beta, \sqrt{n}/\beta)$. La figura 3.10 muestra que ambas distribuciones son muy similares cuando n , el número de sumandos, es grande. Esto se expresa escribiendo:

$$X_1 + \dots + X_n \underset{n \rightarrow \infty}{\approx} N(n/\beta, \sqrt{n}/\beta) = N(n\mu(X_i), \sqrt{n}\sigma(X_i)).$$

Lo que ocurre con la distribución exponencial no es casualidad. Ni siquiera hace falta que las variables que se suman sean continuas. Por ejemplo, sean $X_1, \dots, X_n \sim B(1, p)$ independientes. Sus funciones de masa (figura 3.11) son muy diferentes de una campana de Gauss. Sin embargo, como la distribución binomial es reproductiva respecto de su primer parámetro, se verifica lo siguiente:

$$X_1 + X_2 \sim B(2, p), \quad X_1 + X_2 + X_3 \sim B(3, p), \quad X_1 + X_2 + X_3 + X_4 \sim B(4, p), \\ X_1 + \dots + X_5 \sim B(5, p), \quad \dots, \quad X_1 + \dots + X_n \sim B(n, p).$$

La función de masa resultante se parece a una campana de Gauss (figura 3.11). Concretamente, a la distribución normal que tiene la misma media y la misma desviación típica que $B(n, p)$, que es $N(np, \sqrt{np(1-p)})$. La figura 3.11 muestra que ambas distribuciones son muy similares cuando n , el número de sumandos, es grande. Esto se expresa escribiendo:

$$X_1 + \dots + X_n \underset{n \rightarrow \infty}{\approx} N(np, \sqrt{np(1-p)}) = N(n\mu(X_i), \sqrt{n}\sigma(X_i)).$$

En general, sean X_1, \dots, X_n variables aleatorias equidistribuidas e independientes. Sean $\mu(X_i) = \mu$ y $\sigma(X_i) = \sigma$, $i = 1..n$. Si n es suficientemente grande, entonces la distribución de $X_1 + \dots + X_n$ coincide, aproximadamente, con la distribución normal de la misma media y la misma desviación típica, es decir, $N(n\mu, \sqrt{n}\sigma)$. Esto se expresa escribiendo:

$$X_1 + \dots + X_n \underset{n \rightarrow \infty}{\approx} N(n\mu, \sqrt{n}\sigma).$$

O bien, dividiendo por n y aplicando la propiedad de transformación lineal de la distribución normal:

$$\bar{X} := \frac{X_1 + \dots + X_n}{n} \underset{n \rightarrow \infty}{\approx} N\left(\frac{n\mu}{n}, \frac{\sqrt{n}\sigma}{n}\right) = N\left(\mu, \sigma/\sqrt{n}\right).$$

Las variables aleatorias \bar{X} se llaman *medias muestrales*. Tipificando, la ecuación anterior se puede escribir:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

Este resultado es el *teorema del límite central*, que se formaliza a continuación.

Teorema 3.7.1 (del límite central - Lévy-Lindeberg)

Sea $[X_1, X_2, \dots, X_n, \dots]$ una sucesión de variables aleatorias equidistribuidas e independientes. Sean μ y σ su media y su desviación típica, respectivamente. Para todo $x \in \mathbb{R}$ se verifica que:

$$\lim_{n \rightarrow \infty} p\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) = p(N(0, 1) \leq x).$$

Dicho de otro modo, las funciones de distribución de las variables aleatorias $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ convergen a la función de distribución de $N(0, 1)$.

Demostración: Ver Quesada [13, páginas 575..576, Teorema 15.4]. ◀

Nota 3.7.2 (Velocidad de la convergencia)

El teorema del límite central se aplica a variables aleatorias equidistribuidas e independientes, sin importar su modelo de distribución, ya sea discreto o continuo. Sin embargo, el valor de n a partir del cual la aproximación es buena sí que depende del modelo de distribución:

- Si dicha distribución es continua, acampanada y aproximadamente simétrica, es decir, similar a una normal, puede ser suficiente tomar $n \geq 6$.
- Si es acampanada y unimodal, pero poco simétrica, puede bastar tomar $n \geq 20$.
- En otras condiciones, por ejemplo cuando la distribución es discreta, puede ser necesario tomar $n \geq 100$. ◀

El teorema del límite central explica por qué muchos fenómenos aleatorios tienen distribución aproximadamente normal: son suma de muchas otras variables aleatorias más o menos equidistribuidas y más o menos independientes. Así, la estatura de las personas depende de factores genéticos, nutricionales y sociales, entre otros, lo cual justifica que se comporte como una distribución normal.

Ejemplo 3.7.3 (Tiempo ejecución de 100 procesos)

Volvamos al ejemplo 3.6.4, sobre el tiempo total de ejecución de 100 procesos independientes. ¿Cuál es la probabilidad de que dicho tiempo sea menor que 180 s? Se trata de calcular $p(X < 180)$, siendo $X \sim \gamma(100, 1/2)$. Integrando la función de densidad correspondiente, se obtiene:

$$p(X < 180) = \int_0^{180} \frac{0.5^{100}}{\Gamma(100)} x^{99} e^{-0.5x} dx \simeq 0.158221.$$

Esta integral se ha obtenido por métodos numéricos usando un ordenador. Si no se dispone de esta herramienta, se puede estimar el resultado usando el teorema del límite central, ya que X es suma de un número grande ($n = 100$) de variables aleatorias independientes y con

distribución $X_i \sim \gamma(1, 1/2)$. Por tanto:

$$X \approx N(n\mu(X_i), \sqrt{n}\sigma(X_i)) = N\left(100 \cdot \frac{1}{1/2}, \sqrt{100} \cdot \sqrt{\frac{1}{(1/2)^2}}\right) = N(200, 20).$$

Tipificando, se obtiene:

$$\begin{aligned} p(X < 180) &= p\left(\frac{X - \mu}{\sigma} < \frac{180 - 200}{20}\right) \simeq p(N(0, 1) < -1) = p(N(0, 1) > 1) = \\ &= 1 - p(N(0, 1) \leq 1) \simeq 1 - 0.8413 = 0.1587 \quad (\text{tabla}). \end{aligned}$$

En este cálculo se ha usado la simetría de la distribución $N(0, 1)$, la regla del complementario y la tabla de $N(0, 1)$. Comparando el resultado estimado (0.1587) con el valor real (0.158 221) se aprecia una buena aproximación. ◀

Para calcular este valor, es necesario extraer todas las bolas de la urna y a veces esto no es posible o no es deseable. Sin embargo, supongamos que se conoce el resultado de una *muestra*, por ejemplo $[1, 0, 1, 0, 0]$, donde un 1 indica que la bola extraída es blanca y un 0 lo contrario. Si dicha muestra es representativa de lo que pasa en la urna, p se debe parecer a la proporción de bolas blancas en la muestra, que vale $2/5 = 0.4$. Como este valor puede coincidir con p o no, se escribe $\hat{p} = 0.4$ y se dice que se ha obtenido una *estimación puntual* de p . Si se toma otra muestra, se obtiene otra estimación puntual, que puede coincidir con la primera o no. Por ejemplo:

- La muestra $[1, 0, 0, 1]$, que tiene tamaño $n = 4$, produce la estimación:

$$\hat{p} = \frac{1 + 0 + 0 + 1}{4} = \frac{2}{4} = 0.5.$$

- La muestra $[0, 1, 0, 1, 1]$, de tamaño $n = 5$, produce la estimación:

$$\hat{p} = \frac{0 + 1 + 0 + 1 + 1}{5} = \frac{3}{5} = 0.6.$$

- La muestra genérica $[X_1, X_2, \dots, X_n]$, de tamaño n , produce la estimación:

$$\hat{p} = \frac{X_1 + \dots + X_n}{n}.$$

En una muestra genérica, cada extracción de la urna es una variable aleatoria:

$$X_i : \text{“número de bolas blancas en la extracción } i\text{”} \sim B(1, p_i), \quad i = 1..n,$$

donde p_i es la proporción de bolas blancas en la urna en la extracción i . Pero si el muestreo se hace *con reemplazamiento*, devolviendo la bola a la urna después de cada extracción, entonces $p_1 = \dots = p_i = \dots = p_n = p$. En ese caso, todas las variables X_i tienen la misma distribución que:

$$X : \text{“número de bolas blancas en una extracción”} \sim B(1, p).$$

y son independientes. Además, la expresión

$$\hat{p} = \frac{X_1 + \dots + X_n}{n} = \bar{X}.$$

es función de las variables aleatorias $[X_1, X_2, \dots, X_n]$ que componen la muestra, luego es otra variable aleatoria, denominada *media muestral*. ◀

Las ideas contenidas en el ejemplo anterior se formalizan en la siguiente definición.

Definición 4.1.2 (Muestra, estadístico, estimador y estimación)

Sea X una variable aleatoria.

1. Una *muestra aleatoria simple* de X (en adelante, muestra) es una lista de variables aleatorias con la misma distribución que X e independientes: $[X_1, X_2, \dots, X_n]$. El número n se llama *tamaño muestral*.
2. Un *estadístico* es una variable aleatoria construida como función de las que componen la muestra: $T(X_1, \dots, X_n)$.

3. Sea θ un parámetro desconocido de la distribución de X . Un *estimador* de θ es un estadístico de la forma $\hat{\theta} = T(X_1, \dots, X_n)$ usado para estimar θ .
4. Una *estimación* de θ es el valor que toma un estimador $\hat{\theta}$ cuando se observa una muestra concreta. ◀

Una muestra verifica que:

1. Todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra.
2. Cada elemento de la muestra es independiente de los demás.

Ambas cualidades son imprescindibles para que la muestra represente bien a la población de la que procede.

Cada elemento de la muestra, antes de ser observado, es una variable aleatoria X_i con la misma distribución que la población X . Después de observado, dicho elemento es un número x_i . Por tanto, hay que distinguir entre la lista de n variables aleatorias que constituye la muestra aleatoria simple, $[X_1, \dots, X_n]$, y la lista de n números $[x_1, \dots, x_n]$, que constituye una *observación* de la muestra. Tampoco se debe confundir un estimador, $T(X_1, \dots, X_n)$, que es una fórmula, con una *estimación*, $T(x_1, \dots, x_n) \in \mathbb{R}$, que es el valor numérico que toma un estimador ante una observación de la muestra. Concretamente, en el ejemplo 4.1.1, la primera muestra es $[X_1, X_2, X_3, X_4, X_5]$, donde $X_i \sim B(1, p)$, mientras que la primera observación de la muestra es $[1, 0, 1, 0, 0]$. La fórmula $\hat{p} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$ es un estimador de p , mientras que el valor $\hat{p} = \frac{1 + 0 + 1 + 0 + 0}{5} = 0.4$ es una estimación de p .

Ejemplo 4.1.3 (Estadísticos)

El estadístico más usado es la *media muestral*:

$$\bar{X} := \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Se suele emplear como estimador de la media de la variable. Parece lógico estimar la varianza de la variable mediante la *varianza muestral*, que es:

$$V(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Sin embargo, por razones que se verán más adelante, dicha estimación se hace con la *cuasivarianza muestral*, definida cuando $n > 1$:

$$S^2(X_1, \dots, X_n) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} V(X_1, \dots, X_n).$$

Otros estadísticos usados con frecuencia son el mínimo y el máximo de las variables muestrales:

$$\min(X_1, \dots, X_n) \quad \text{y} \quad \max(X_1, \dots, X_n). \quad \blacktriangleleft$$

4.2 Estimación por momentos

La estimación puntual de un parámetro desconocido requiere un estimador del mismo. Una forma sencilla de obtenerlo es el método empleado en el ejemplo 4.1.1, que se describe a continuación.

Ejemplo 4.2.1 (Distribución binomial)

Consideremos una prueba aleatoria que se repite un número fijo de veces de modo independiente y con probabilidad de éxito constante. En estas condiciones, se tiene la variable aleatoria:

$$X : \text{“número de éxitos en } k \text{ pruebas”} \sim B(k, p).$$

En general, el valor de k es conocido, pero el valor de p es desconocido y hay que estimarlo. Para ello, sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Suponiendo que la media de la variable coincide con la media muestral, se resuelve la siguiente ecuación en p :

$$\mu(X) = kp = \bar{X} \Rightarrow p = \frac{\bar{X}}{k}.$$

Dependiendo de la muestra observada, esta igualdad puede ser cierta o no, por lo que se escribe $\hat{p} = \bar{X}/k$, que es un estimador de p .

Supongamos que el número de pruebas es $k = 4$ y que se observa la muestra $[2, 1, 3, 2, 0]$. En este caso, la media muestral y la estimación de p son:

$$\bar{x} = \frac{2 + 1 + 3 + 2 + 0}{5} = 1.6 \Rightarrow \hat{p} = \frac{\bar{x}}{k} = \frac{1.6}{4} = 0.4. \quad \blacktriangleleft$$

Otro caso similar es el de la distribución geométrica.

Ejemplo 4.2.2 (Distribución geométrica)

Sea $X \sim G(p)$ con p desconocido. Para estimar p , sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Igualando la media de la variable a la media muestral y resolviendo la ecuación en p , resulta:

$$\mu(X) = \frac{1-p}{p} = \frac{1}{p} - 1 = \bar{X} \Rightarrow \frac{1}{p} = \bar{X} + 1 \Rightarrow p = \frac{1}{\bar{X} + 1}.$$

Esta igualdad no será cierta para todas las muestras, por lo que se escribe $\hat{p} = \frac{1}{\bar{X} + 1}$, que es un estimador de p . ◀

El siguiente ejemplo es un poco más complejo, porque involucra dos parámetros.

Ejemplo 4.2.3 (Distribución gamma)

Sea $X \sim \gamma(\alpha, \beta)$ con α y β desconocidos. Para estimar α y β , sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Igualando la media de la variable a la media muestral, se obtiene una ecuación en α y β :

$$\mu(X) = \frac{\alpha}{\beta} = \bar{X}.$$

No se puede resolver esta ecuación, que tiene dos incógnitas, por lo que se forma otra ecuación igualando la varianza de la variable a la varianza muestral. Resolviendo el sistema de ecuaciones en α y β , resulta:

$$\begin{cases} \mu(X) = \alpha/\beta = \bar{X} \\ \sigma^2(X) = \alpha/\beta^2 = V \end{cases} \Rightarrow \begin{cases} \frac{\alpha^2/\beta^2}{\alpha/\beta^2} = \alpha = \frac{\bar{X}^2}{V} \\ \frac{\alpha/\beta}{\alpha/\beta^2} = \beta = \frac{\bar{X}}{V} \end{cases}$$

Para despejar α se divide el cuadrado de la primera ecuación entre la segunda y para despejar β se divide la primera ecuación entre la segunda. Estas igualdades no serán ciertas para todas las muestras, por lo que se escribe $\hat{\alpha} = \bar{X}^2/V$ y $\hat{\beta} = \bar{X}/V$, que son los estimadores de α y de β , respectivamente. ◀

El algoritmo empleado en los ejemplos anteriores se denomina *estimación por momentos*.

Algoritmo 4.2.4 (Estimación por momentos)

Sea X una variable aleatoria cuya distribución depende de uno o varios parámetros desconocidos. Para estimar dichos parámetros, se hace lo siguiente:

1. Tomar una muestra aleatoria simple de X , llamada $[X_1, \dots, X_n]$.
2. Formar una ecuación o un sistema con tantas ecuaciones como parámetros desconocidos. Para ello, igualar la media de X a la media muestral y, si hay dos parámetros desconocidos, igualar la varianza de X a la varianza muestral:

$$\mu(X) = \bar{X} \quad \text{o} \quad \begin{cases} \mu(X) = \bar{X} \\ \sigma^2(X) = V \end{cases}$$

3. Resolver la ecuación o el sistema de ecuaciones.
4. Devolver la solución como conjunto de estimadores. ◀

Este método fue formalizado por Karl Pearson¹ a fines del siglo XIX. Es un método muy sencillo, pero tiene el inconveniente de que puede generar estimaciones incompatibles con los valores de la muestra, como sucede en el ejemplo siguiente.

Ejemplo 4.2.5 (Distribución uniforme)

Sea $X \sim U(a, b)$ con a y b desconocidos. Para estimar a y b por el método de los momentos,



¹ Karl Pearson (London, England, 1857 – Coldharbour, Surrey, England, 1936).

Es uno de los fundadores de la estadística. Pearson se graduó en el King's College de Cambridge, y completó estudios en Heidelberg y Berlín. Enseñó matemáticas en el University College, London. Allí fundó y dirigió el Departamento de Estadística Aplicada, donde se formaron multitud de científicos de la época, tanto británicos como extranjeros.

Sus primeros trabajos estadísticos son 18 artículos titulados *Mathematical Contributions to the Theory of Evolution*, que contienen contribuciones a la regresión, el coeficiente de correlación y el test chi-cuadrado. Pearson acuñó el término *desviación típica* y cofundó, con Weldon y Galton, la revista *Biometrika*.

sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Se resuelve el sistema de ecuaciones:

$$\begin{cases} \mu(X) = \frac{a+b}{2} = \bar{X} \\ \sigma^2(X) = \frac{(b-a)^2}{12} = V \end{cases} \Rightarrow \begin{cases} a+b = 2\bar{X} \\ b-a = \sqrt{12V} = 2\sqrt{3V} \end{cases}$$

Restando y sumando las dos ecuaciones, resulta que los estimadores de a y de b son, respectivamente:

$$\begin{cases} 2a = 2\bar{X} - 2\sqrt{3V} \\ 2b = 2\bar{X} + 2\sqrt{3V} \end{cases} \Rightarrow \begin{cases} \hat{a} = \bar{X} - \sqrt{3V} \\ \hat{b} = \bar{X} + \sqrt{3V} \end{cases}$$

Supongamos que se observa la muestra $[12, 9, 22, 7, 0, 10]$. La media y la varianzas muestrales valen:

$$\begin{aligned} \bar{x} &= \frac{12 + 9 + 22 + 7 + 0 + 10}{6} = 10, \\ v &= \frac{12^2 + 9^2 + 22^2 + 7^2 + 0^2 + 10^2}{6} - 10^2 = 43. \end{aligned}$$

Por tanto, las estimaciones son:

$$\hat{a} = 10 - \sqrt{3 \cdot 43} \simeq -1.35782 \quad \text{y} \quad \hat{b} = 10 + \sqrt{3 \cdot 43} \simeq 21.3578.$$

El valor de \hat{b} es menor que un elemento de la muestra (22), lo cual es absurdo, ya que $\text{Im}(X) = [a, b]$. Por tanto, esta estimación no es aceptable. De hecho, los parámetros de una distribución uniforme se estiman como el mínimo y el máximo de los valores muestrales, respectivamente:

$$\hat{a} = \min(12, 9, 22, 7, 0, 10) = 0 \quad \text{y} \quad \hat{b} = \max(12, 9, 22, 7, 0, 10) = 22.$$

Estos valores son compatibles con los elementos de la muestra. ◀

Existe otro método de estimación que no presenta el problema observado en el ejemplo anterior, pero no lo estudiaremos en este curso.

4.3 Método del pivote

Sea X una variable aleatoria cuya distribución depende del parámetro desconocido θ . La estimación puntual de θ proporciona un número $\hat{\theta}$ que puede coincidir o no con el verdadero valor de θ . En consecuencia, es deseable tener una medida del error que se comete al reemplazar θ por $\hat{\theta}$. Algo similar a la cota de error en un método numérico, la cual indica que el valor exacto pertenece a cierto intervalo. El siguiente ejemplo muestra cómo hacerlo.

Ejemplo 4.3.1 (Intervalo para la media de una población normal con varianzas conocida)

Se sabe que la variable aleatoria X sigue una distribución $N(\mu, \sigma)$, con $\sigma = 1$ y μ desconocida, y se trata de estimar μ . Para ello, se obtiene una muestra de X y resulta $[1.5, 3.5, 3, 1.6]$. La media muestral vale:

$$\bar{x} = \frac{1.5 + 3.5 + 3 + 1.6}{4} = \frac{9.6}{4} = 2.4.$$

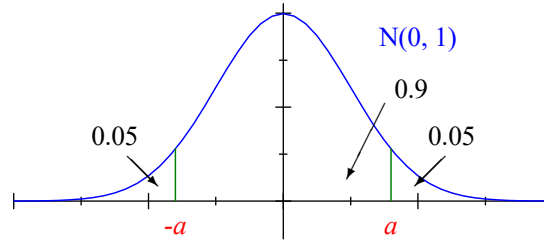


Figura 4.2: $p(-a \leq N(0, 1) \leq a) = 0.9$.

Por tanto, es de esperar que μ , que es la media de X , esté próxima a 2.4. Dicho de otro modo, μ debe estar en algún intervalo alrededor de 2.4 y la probabilidad de que esto ocurra aumentará con el tamaño del intervalo. Se trata de fijar una probabilidad, por ejemplo 90 %, y hallar dicho intervalo.

Para ello, se utiliza la reproductividad de la distribución normal:

$$\begin{aligned} X_1, \dots, X_n \sim N(\mu, \sigma) \text{ e independientes} &\Rightarrow X_1 + \dots + X_n \sim N(n\mu, \sqrt{n}\sigma) \Rightarrow \\ &\Rightarrow \bar{X} = \frac{X_1 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1). \end{aligned}$$

Considerando la distribución $N(0, 1)$, cuya función de densidad es simétrica (figura 4.2), se puede hallar un número $a \in \mathbb{R}$ tal que:

$$p(-a \leq N(0, 1) \leq a) = 0.9 \Leftrightarrow p(N(0, 1) \leq a) = 0.95 \Leftrightarrow a \simeq 1.645.$$

Teniendo en cuenta que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \Rightarrow p\left(-a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq a\right) = 0.9.$$

Por tanto, el intervalo buscado se obtiene despejando μ de las inecuaciones:

$$\begin{aligned} -a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq a &\Leftrightarrow a \frac{\sigma}{\sqrt{n}} \geq -\bar{X} + \mu \geq -a \frac{\sigma}{\sqrt{n}} \Leftrightarrow \\ &\Leftrightarrow \bar{X} + a \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X} - a \frac{\sigma}{\sqrt{n}} \Leftrightarrow \mu \in \left[\bar{X} - a \frac{\sigma}{\sqrt{n}}, \bar{X} + a \frac{\sigma}{\sqrt{n}}\right]. \end{aligned}$$

Ambos extremos del intervalo dependen de σ y de a , que son conocidos, pero también de la muestra, luego son estadísticos. Sustituyendo los valores muestrales en la expresión anterior, se obtiene un *intervalo, al 90 % de confianza, para μ* :

$$\mu \in \left[2.4 - 1.645 \frac{1}{\sqrt{4}}, 2.4 + 1.645 \frac{1}{\sqrt{4}}\right] \simeq [1.5775, 3.2225] \text{ al } 90 \% \text{ de confianza.}$$

Dicho intervalo ya no es aleatorio, por lo que no es correcto escribir $p(1.85 \leq \mu \leq 2.95) = 0.9$. Pero si se generan muchos intervalos de confianza con la fórmula anterior, se puede esperar que el 90 % de ellos contendrá al verdadero valor de μ .

Si se aumenta la probabilidad de éxito, por ejemplo a 95 %, entonces debe ocurrir que:

$$p(N(0, 1) \leq a) = 0.975 \Rightarrow a \simeq 1.96.$$

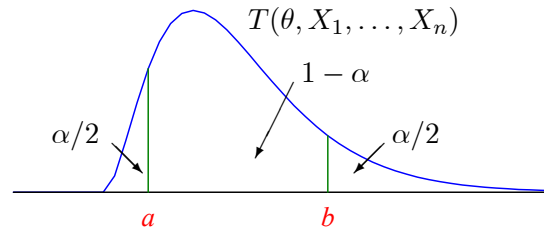


Figura 4.3: Método del pivote: $p(a \leq T \leq b) = 1 - \alpha$.

En tal caso, se obtiene un intervalo de confianza para μ al 95 %:

$$\mu \in \left[2.4 - 1.96 \frac{1}{\sqrt{4}}, 2.4 + 1.96 \frac{1}{\sqrt{4}} \right] = [1.42, 3.38] \text{ al } 95\% \text{ de confianza.}$$

Ambos intervalos de confianza están centrados en el punto 2.4, que es la media muestral, pero el radio del segundo es mayor que el del primero, como corresponde a un nivel de confianza mayor. ◀

El ejemplo anterior conduce a la siguiente definición.

Definición 4.3.2 (Intervalo de confianza y nivel de significación)

Sea X una variable aleatoria cuya distribución depende del parámetro desconocido θ . Sea $0 < \alpha < 1$ un número real. Sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Un *intervalo de confianza para θ , al nivel $1 - \alpha$* tiene por extremos un par de estadísticos, T_1 y T_2 , tales que:

$$p(T_1(X_1, \dots, X_n) \leq \theta \leq T_2(X_1, \dots, X_n)) = 1 - \alpha.$$

El número $1 - \alpha$ se llama *nivel de confianza*, mientras que α se llama *nivel de significación*. ◀

Lo ideal sería tomar $\alpha = 0$ y encontrar un intervalo I tal que $p(\theta \in I) = 1$. Pero esta condición generalmente hace que I se extienda a todo \mathbb{R} , por lo que carece de interés. En la práctica, hay que buscar un equilibrio entre el nivel de confianza y la longitud del intervalo.

Sistematizando el proceso del ejemplo 4.3.1, se obtiene el siguiente algoritmo.

Algoritmo 4.3.3 (Método del pivote)

Sea X una variable aleatoria cuya distribución depende del parámetro desconocido θ . Dados:

- Un nivel de confianza $1 - \alpha$, con $0 < \alpha < 1$.
- Una muestra aleatoria simple de X , llamada $[X_1, \dots, X_n]$.

Para construir un intervalo de confianza para θ , al nivel $1 - \alpha$, se hace lo siguiente (figura 4.3):

1. Hallar un *pivote* T , que es un estadístico tal que:
 - Solo depende del parámetro θ y de la muestra: $T(\theta, X_1, \dots, X_n)$.
 - Tiene distribución conocida.

2. Hallar dos constantes, $a, b \in \mathbb{R}$, tales que:

$$p(T < a) = \alpha/2 \quad \text{y} \quad p(T \leq b) = 1 - \alpha/2.$$

Esto es posible, pues T tiene distribución conocida, y hace que $p(a \leq T \leq b) = 1 - \alpha$.

3. *Pivotar*, que consiste en despejar θ de las dos inecuaciones $a \leq T(\theta, X_1, \dots, X_n) \leq b$ para obtener el intervalo de confianza:

$$\theta \in [T_1(X_1, \dots, X_n), T_2(X_1, \dots, X_n)] \text{ al nivel } 1 - \alpha \text{ de confianza.} \quad \blacktriangleleft$$

En cada una de las siguientes secciones se resuelve un problema típico aplicando el método del pivote.

4.4 Intervalo de confianza para la varianza

Sea $X \sim N(\mu, \sigma)$, con μ y σ desconocidas. Sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Se trata de construir un intervalo de confianza para σ^2 , para lo cual se necesita un pivote.

Ya hemos visto que $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, pero esta expresión no sirve como pivote, porque tanto μ como σ son desconocidas. Por otro lado, la cuasivarianza muestral es un estimador centrado de σ^2 (la varianza muestral no lo es). Por tanto, es natural partir de ella:

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

Se trata de tipificar las variables X_i y \bar{X} que aparecen en S^2 . Desarrollando el cuadrado de la diferencia, se obtiene:

$$\begin{aligned} (n-1)S^2 &= \sum_{i=1}^n \left((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2 \right) = \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2 = \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

Esto es así porque:

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

A continuación, se divide por σ^2 :

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

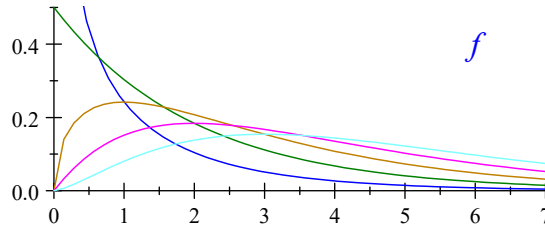


Figura 4.4: Densidad de χ_n^2 para $n = 1.5$.

Puesto que X sigue una distribución normal, se verifica que:

$$X_i \sim N(\mu, \sigma) \Rightarrow \frac{X_i - \mu}{\sigma} \sim N(0, 1) \quad \text{y también} \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Por tanto, la expresión $(n-1)S^2/\sigma^2$ es suma y resta de variables aleatorias $N(0, 1)$ elevadas al cuadrado. Si se conociese su distribución, serviría como pivote. El objetivo es hallarla.

Teorema 4.4.1 (Cuadrado de una distribución normal estándar)

Si $X_1, \dots, X_n \sim N(0, 1)$ son independientes, entonces:

- (i) $X_i^2 \sim \gamma(1/2, 1/2)$, para todo $i = 1..n$.
- (ii) $X_1^2 + \dots + X_n^2 \sim \gamma(n/2, 1/2)$.

Demostración: Apartado (ii)

Si $X_1, \dots, X_n \sim N(0, 1)$ son independientes, entonces $X_1^2, \dots, X_n^2 \sim \gamma(1/2, 1/2)$ y también son independientes. Como la distribución gamma es reproductiva respecto de su primer parámetro, se verifica que:

$$X_1^2 + \dots + X_n^2 \sim \gamma(1/2 + \dots + 1/2, 1/2) = \gamma(n/2, 1/2). \quad \blacktriangleleft$$

La distribución $\gamma(n/2, 1/2)$ recibe un nombre particular.

Definición 4.4.2 (Distribución chi cuadrado)

Si $n \in \mathbb{N} - \{0\}$, la distribución $\gamma(n/2, 1/2)$ se denomina *chi cuadrado (de Pearson) de parámetro n* . Se nota χ_n^2 o $\chi^2(n)$. \blacktriangleleft

La función de densidad de la distribución χ_n^2 tiene las siguientes propiedades (figura 4.4):

- Si $n = 1, 2$, es estrictamente decreciente y convexa.
- Si $n = 3, 4$, tiene máximo en $x = n - 2$ y punto de inflexión en $x = n - 2 + \sqrt{2n - 4}$.
- Si $n \geq 5$, tiene máximo en $x = n - 2$ y puntos de inflexión en $x = n - 2 \mp \sqrt{2n - 4}$.

Teorema 4.4.3 (Fisher)

Sea $X \sim N(\mu, \sigma)$. Si $[X_1, \dots, X_n]$ es una muestra aleatoria simple de X , entonces:

- (i) Los estadísticos \bar{X} y S^2 son independientes.

$$(ii) \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Demostración: Ver Freund-Walpole [10, página 287, teorema 8.10] o bien Rohatgi [15, página 340, Theorem 1]. ◀

El teorema de Fisher proporciona un pivote para construir un intervalo de confianza para σ^2 , al nivel $1 - \alpha$:

1. Normalidad: $X \sim N(\mu, \sigma)$.

2. Pivote: $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.

3. Constantes: $p(\chi_{n-1}^2 \leq a) = \alpha/2$ y $p(\chi_{n-1}^2 \leq b) = 1 - \alpha/2$.

4. Pivotar:

$$a \leq \frac{(n-1)S^2}{\sigma^2} \leq b \Leftrightarrow \frac{1}{a} \geq \frac{\sigma^2}{(n-1)S^2} \geq \frac{1}{b} \Leftrightarrow \sigma^2 \in \left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a} \right].$$

Ejemplo 4.4.4 (Montaje de ordenadores: IC para σ)

En una fábrica de ordenadores, se quiere saber si hay mucha variación en el tiempo de montaje de un ordenador. Para ello, se toma una muestra de la variable aleatoria:

X : “tiempo de montaje de un ordenador (min)”

y se obtiene:

38 50 37 33 44 39 36 42 41

Se trata de construir un intervalo de confianza para σ , la desviación típica de X , al 98%. Las medidas de esta muestra son:

$$\bar{x} = \frac{38 + 50 + 37 + 33 + 44 + 39 + 36 + 42 + 41}{9} = 40,$$

$$s^2 = \frac{9}{8} \left(\frac{38^2 + 50^2 + 37^2 + 33^2 + 44^2 + 39^2 + 36^2 + 42^2 + 41^2}{9} - 40^2 \right) = 25.$$

La cuasidesviación típica muestral es $s = 5$, pero la desviación típica verdadera, σ , puede ser algo menor o algo mayor. El intervalo de confianza se construye del siguiente modo:

1. Normalidad: STATGRAPHICS ajusta $X \approx N(40, 5)$ con p -valor = 0.996713, luego el ajuste es bueno.²

2. Pivote: $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 = \chi_8^2$.

3. Constantes: $p(\chi_8^2 \leq a) = 0.01 \Rightarrow a \simeq 1.6465$ y $p(\chi_8^2 \leq b) = 0.99 \Rightarrow b \simeq 20.0902$.

4. Pivotar: $\sigma \in \left[\sqrt{\frac{8 \cdot 5^2}{20.0902}}, \sqrt{\frac{8 \cdot 5^2}{1.6465}} \right] \simeq [3.1552, 11.021]$ al 98% de confianza.

(STATGRAPHICS obtiene los mismos resultados.)

Este intervalo no contiene al punto 3, pero sí al punto 3.2, por ejemplo. Por tanto, al 98% de confianza, se rechaza que la desviación típica del tiempo de montaje de un ordenador es $\sigma = 3$ min, pero se acepta que $\sigma = 3.2$ min. ◀

²En la sección 5.5.2 se estudiarán los detalles de esta comprobación.

4.5 Intervalo de confianza para la media

Sea X una variable aleatoria con media μ y varianza σ^2 , ambas desconocidas. Se trata de construir un intervalo de confianza para μ . Hay varios casos posibles, que se estudian a continuación.

4.5.1 Variable normal

Sea $X \sim N(\mu, \sigma)$, con μ y σ desconocidas. Sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Ya se ha visto que:

$$Y := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Esta expresión no sirve como pivote para μ , porque también contiene el valor desconocido σ . Para eliminarlo, se considera el estadístico:

$$Z := \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Por el teorema de Fisher, $Z \sim \chi_{n-1}^2$ y es independiente de Y . Dividiendo Y por $\sqrt{Z/(n-1)}$, se obtiene:

$$\frac{Y}{\sqrt{\frac{Z}{n-1}}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2/\sigma^2}{n-1}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \frac{\sigma}{S} = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Esta expresión es un pivote para μ , ya que William Gosset³ halló su distribución.

Definición 4.5.1 (Distribución t de Student)

Sean $Y \sim N(0, 1)$ y $Z \sim \chi_n^2$ independientes. La variable aleatoria:

$$X := \frac{Y}{\sqrt{Z/n}}$$

sigue una *distribución t de parámetro n* . Se nota $X \sim t_n$ o $X \sim t(n)$. ◀

La distribución t_n tiene función de densidad similar⁴ a la de $N(0, 1)$ (figura 4.5):

3



William Sealey Gosset (Canterbury, England, 1876 – Beaconsfield, England, 1937).

Matemático y químico. Gosset trabajó como químico para la compañía cervecera *Guinness*, que no permitía a sus empleados publicar artículos científicos. Por este motivo, Gosset publicó sus estudios estadísticos con el seudónimo *Student*.

Durante un año, investigó en el laboratorio de Pearson sobre la distribución límite de la binomial (distribución de Poisson) y sobre la distribución muestral de la media, la desviación típica y el coeficiente de correlación.

Gosset inventó el test t para muestras pequeñas, que aplicó al control de calidad en la industria cervecera.

⁴Ver Fernández-Abascal [8, página 479, proposición 11.4].

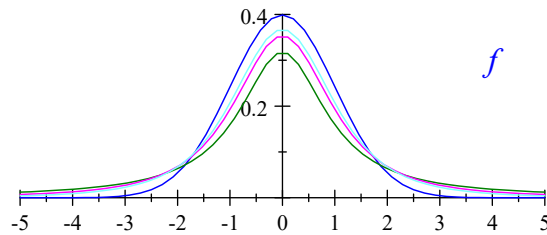


Figura 4.5: Densidades de t_1, t_2 y t_3 comparadas con $N(0, 1)$.

1. Es simétrica respecto de la recta $x = 0$.
2. Tiene por asíntota horizontal la recta $y = 0$.
3. Tiene máximo en $x = 0$.
4. Converge a la función de densidad de la distribución $N(0, 1)$ cuando n tiende a infinito: $t_n \underset{n \rightarrow \infty}{\approx} N(0, 1)$. La aproximación es buena cuando $n \geq 30$.

La figura 4.5 compara las funciones de densidad de t_1, t_2, t_3 y $N(0, 1)$. Las gráficas son similares entre sí, aunque las de t_n son más achatadas que la de $N(0, 1)$ para cualquier valor de n .

La distribución t proporciona un pivote para construir un intervalo de confianza para μ , al nivel $1 - \alpha$:

1. Normalidad: $X \sim N(\mu, \sigma)$.
2. Pivote: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$.
3. Constantes: la distribución t es simétrica respecto de $x = 0$. Por tanto, basta hallar $a \in \mathbb{R}$ tal que $p(t_{n-1} \leq a) = 1 - \alpha/2$.
4. Pivotar:

$$-a \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq a \Leftrightarrow a \frac{S}{\sqrt{n}} \geq -\bar{X} + \mu \geq -a \frac{S}{\sqrt{n}} \Leftrightarrow \mu \in \left[\bar{X} - a \frac{S}{\sqrt{n}}, \bar{X} + a \frac{S}{\sqrt{n}} \right].$$

Ejemplo 4.5.2 (Montaje de ordenadores: IC para μ)

En la fábrica de ordenadores del ejemplo 4.4.4, interesa saber, al 98 % de confianza, cuál es el tiempo medio de montaje de un ordenador. Ya se ha visto que las medidas de la muestra de X son: $n = 9, \bar{x} = 40$ y $s = 5$. La media muestral es 40, pero la media de X puede ser algo menor o algo mayor. Para saber lo que sucede, se construye un intervalo de confianza para μ al 98 %:

1. Normalidad: STATGRAPHICS ajusta $X \approx N(40, 5)$ con p -valor = 0.996713, luego el ajuste es bueno.
2. Pivote: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} = t_8$.
3. Constantes: $p(t_8 \leq a) = 0.99 \Rightarrow a = 2.8965$.
4. Pivotar: $\mu \in 40 + 2.8965 \frac{5}{\sqrt{9}} [-1, 1] \simeq [35.173, 44.828]$ al 98 % de confianza.

(STATGRAPHICS obtiene los mismos resultados.)

El punto 36 pertenece al intervalo de confianza. Por tanto, al 98 % de confianza, se acepta que el tiempo medio de montaje de un ordenador es $\mu = 36$ min. Por el contrario, con el mismo nivel de confianza, se rechaza que $\mu = 35$ min, por ejemplo.

Fijado el nivel de confianza $1 - \alpha$, el intervalo de confianza para μ está centrado en la media muestral, \bar{x} . Si se toma dicho valor como estimación de μ , entonces el error absoluto cometido está acotado, al nivel de confianza $1 - \alpha$, por el radio del intervalo, que es $a \cdot s/\sqrt{n}$. En este caso, dicho radio vale $2.8965 \cdot 5/\sqrt{9} \simeq 4.828$.

Si se quiere reducir dicho radio a 2, por ejemplo, hay que resolver la inecuación $a \cdot s/\sqrt{n} \leq 2$. Además del tamaño muestral n , la inecuación contiene los valores desconocidos a , que depende del nivel de confianza $1 - \alpha$ y de n , y la cuasivarianza muestral s , que depende de la muestra observada. Para realizar un cálculo aproximado, se elige a como el valor límite para $1 - \alpha$ y se supone que s no cambia con la muestra. En estas condiciones, se verifica:

$$a \frac{s}{\sqrt{n}} \leq 2 \Rightarrow \frac{a s}{2} \leq \sqrt{n} \Rightarrow n \geq \left(\frac{2.3264 \cdot 5}{2} \right)^2 \simeq 33.8.$$

Como n debe ser entero, se requiere una muestra de tamaño $n = 34$. ◀

4.5.2 Variable no normal con muestra grande

Sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de la variable aleatoria X . Supongamos que X no es normal, pero tiene varianza finita y n es suficientemente grande ($n \geq 100$). En ese caso, la desviación típica de X se puede aproximar por la cuasidesviación típica muestral⁵: $S \underset{n \rightarrow \infty}{\simeq} \sigma$.

Usando el teorema del límite central (TLC), se obtiene la aproximación:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{n \rightarrow \infty}{\simeq} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

Este pivote aproximado permite construir un *intervalo de confianza aproximado* para μ , al nivel de confianza $1 - \alpha$:

1. Normalidad: X no es normal, pero su muestra es grande, luego \bar{X} es aproximadamente normal (TLC) y se puede usar un pivote aproximado.
2. Pivote aproximado: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx N(0, 1)$.
3. Constantes: la distribución $N(0, 1)$ es simétrica respecto de $x = 0$. Por tanto, basta hallar $a \in \mathbb{R}$ tal que $p(N(0, 1) \leq a) = 1 - \alpha/2$.
4. Pivotar: $\mu \in \left[\bar{X} - a \frac{S}{\sqrt{n}}, \bar{X} + a \frac{S}{\sqrt{n}} \right]$.

⁵Ver Rohatgi [15, página 357, Example 4].

Ejemplo 4.5.3 (Aerolínea: IC para μ)

La aerolínea A afirma que el precio medio de sus billetes es 120 €. Se trata de verificar dicha afirmación al 96 % de confianza. Para ello, se toma una muestra de la variable aleatoria:

$$X : \text{“precio de un billete de la aerolínea } A \text{ (€)”},$$

que figura en la columna *Billete A* del fichero 4-Intervalos - Datos.sgd, y cuyas medidas son: $n = 400$, $\bar{x} = 124$ y $s = 40$. Se trata de construir un intervalo de confianza para μ , la media de X , al 96 %:

1. Normalidad: STATGRAPHICS ajusta $X \approx N(124, 40)$ con p -valor = 0, luego X no es normal. No obstante, la muestra de X es grande ($n = 400$), luego \bar{X} es aproximadamente normal (TLC) y se puede usar un pivote aproximado.
2. Pivote aproximado: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx N(0, 1)$.
3. Constantes: $p(N(0, 1) \leq a) = 0.98 \Rightarrow a = 2.0538$.
4. Pivotar: $\mu \in 124 + 2.0538 \frac{40}{\sqrt{400}} [-1, 1] \simeq [119.89, 128.11]$ al 96 % de confianza. STATGRAPHICS obtiene $[119.879, 128.121]$ usando como pivote t_{399} en lugar de $N(0, 1)$.

El punto 120 pertenece a este intervalo de confianza. Por tanto, al 96 % de confianza, se acepta que el precio medio de un billete de A es 120 €.

Si el tamaño muestral fuese $n = 490$, en lugar de $n = 400$, entonces el intervalo de confianza resultaría:

$$\mu \in 124 + 2.0538 \frac{40}{\sqrt{490}} [-1, 1] \simeq [120.29, 127.71] \text{ al 96 \% de confianza (STATGRAPHICS =)}.$$

En tal caso, habría que rechazar la afirmación de A . ◀

4.5.3 Proporción con muestra grande

Sea $X \sim B(1, p)$, con p desconocido⁶. Sea $[X_1, \dots, X_n]$ una muestra aleatoria simple de X . Se trata de construir un intervalo de confianza para p . En este caso, $\mu = p$ y un estimador de p es $\hat{P} = \hat{\mu} = \bar{X}$. Además, $X_i^2 = X_i$, luego la varianza muestral vale:

$$V = \frac{X_1^2 + \dots + X_n^2}{n} - \bar{X}^2 = \frac{X_1 + \dots + X_n}{n} - \bar{X}^2 = \bar{X} - \bar{X}^2 = \bar{X}(1 - \bar{X}).$$

Por tanto, si n es suficientemente grande, el pivote de la sección anterior toma la forma:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{n \rightarrow \infty}{\approx} N(0, 1) \Rightarrow \frac{\bar{X} - \mu}{S/\sqrt{n}} \simeq \frac{\bar{X} - \mu}{\sqrt{V}/\sqrt{n}} = \frac{\hat{P} - p}{\sqrt{\hat{P}(1 - \hat{P})}/n} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

Este pivote aproximado permite construir un *intervalo de confianza aproximado* para p , al nivel de confianza $1 - \alpha$:

⁶El parámetro p es la proporción de elementos de la población para los cuales X vale 1, es decir, que verifican la propiedad en estudio.

1. Normalidad: $X \sim B(1, p)$ no es normal, pero su muestra es grande, luego \bar{X} es aproximadamente normal (TLC) y se puede usar un pivote aproximado.
2. Pivote aproximado:
$$\frac{\hat{P} - p}{\sqrt{\hat{P}(1 - \hat{P})/n}} \approx N(0, 1).$$
3. Constantes: la distribución $N(0, 1)$ es simétrica respecto de $x = 0$. Por tanto, basta hallar $a \in \mathbb{R}$ tal que $p(N(0, 1) \leq a) = 1 - \alpha/2$.
4. Pivotar: $p \in \left[\hat{P} - a\sqrt{\hat{P}(1 - \hat{P})/n}, \hat{P} + a\sqrt{\hat{P}(1 - \hat{P})/n} \right]$.

Ejemplo 4.5.4 (Lectora de marcas ópticas: IC para p)

Se quiere estimar, al 97 % de confianza, la proporción de errores cometidos por una lectora de marcas ópticas. Para ello, se observa una muestra de 200 marcas y resulta que la lectora lee mal 3 de ellas. Por tanto, se considera la variable aleatoria:

$$X : \text{“número de errores cometidos por la lectora al leer una marca”} \sim B(1, p).$$

Se trata de construir un intervalo de confianza para p al 97 %:

1. Normalidad: $X \sim B(1, p)$ no es normal, pero su muestra es grande ($n = 200$), luego \bar{X} es aproximadamente normal (TLC) y se puede usar un pivote aproximado.
2. Pivote aproximado:
$$\frac{\hat{P} - p}{\sqrt{\hat{P}(1 - \hat{P})/n}} \approx N(0, 1).$$
3. Constantes: $p(N(0, 1) \leq a) = 0.985 \Rightarrow a = 2.1701$.
4. Pivotar: $p \in \frac{3}{200} + 2.1701\sqrt{\frac{3}{200}\left(1 - \frac{3}{200}\right)}/200 [-1, 1] \simeq [-0.00365213, 0.0336521]$ al 97 % de confianza. STATGRAPHICS usa otro pivote y obtiene $[0.00254944, 0.0466711]$.

Por tanto, se puede estimar que la lectora lee mal entre el 0 % y el 3.37 % de las marcas ópticas.

El error de la estimación está acotado por el radio del intervalo, que es 0.0186521. Si se quiere reducir dicho radio a 0.01, el tamaño muestral necesario, suponiendo que la proporción muestral no varía, es:

$$\begin{aligned} a\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq 0.01 &\Rightarrow a^2 \frac{\hat{p}(1 - \hat{p})}{n} \leq 0.01^2 \Rightarrow \frac{a^2 \hat{p}(1 - \hat{p})}{0.01^2} \leq n \Rightarrow \\ &\Rightarrow n \geq \frac{a^2 \hat{p}(1 - \hat{p})}{0.01^2} \simeq \frac{2.1701^2}{0.01^2} \frac{3}{200} \left(1 - \frac{3}{200}\right) \simeq 695.804. \end{aligned}$$

Como n debe ser entero, hay que tomar una muestra de $n = 696$ marcas ópticas. ◀

Contrastes de hipótesis

5.1 Conceptos generales

Muchos problemas prácticos, como el del ejemplo siguiente, requieren tomar decisiones sobre una población en base a la información contenida en una muestra.

Ejemplo 5.1.1 (Tiempo de acceso en un disco duro)

Se quiere determinar si un disco duro es de marca A o de marca B . Se sabe que el tiempo medio de acceso a un dato es 60 microsegundos, si la marca es A , y $64 \mu\text{s}$, si la marca es B . Para decidir, se considera la variable aleatoria:

X : “tiempo de acceso a un dato en el disco duro (μs)”.

Si $\mu(X) = 60$, entonces el disco es de marca A , mientras que si $\mu(X) = 64$, entonces el disco es de marca B . Esta cuestión se formula como un *contraste de hipótesis* estadísticas:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 64 \end{cases}$$

La hipótesis H_0 se denomina *nula*, mientras que la hipótesis H_1 se denomina *alternativa*. Ambas hipótesis son excluyentes.

¿Cómo aceptar o rechazar H_0 ? Si se toma una muestra de X que represente bien a la variable, entonces la media muestral se parecerá a μ . Si \bar{X} está más cerca de 60 que de 64, entonces es poco verosímil que $\mu = 64$ y se acepta H_0 . Por el contrario, si \bar{X} está más cerca de 64 que de 60, entonces es poco verosímil que $\mu = 60$ y se rechaza H_0 . Es decir, se adopta como *criterio de decisión* aceptar el valor de μ más próximo a \bar{X} (figura 5.1):

- Si $\bar{X} < 62$, se acepta H_0 . Se dice que \bar{X} está en la *región de aceptación*.
- Si $\bar{X} \geq 62$, se rechaza H_0 y en consecuencia se acepta H_1 . Se dice que \bar{X} está en la *región de rechazo* o *región crítica*. En este caso, el punto frontera $\bar{X} = 62$ se ha incluido arbitrariamente en la región de rechazo, pero también se puede incluir en la región de aceptación.

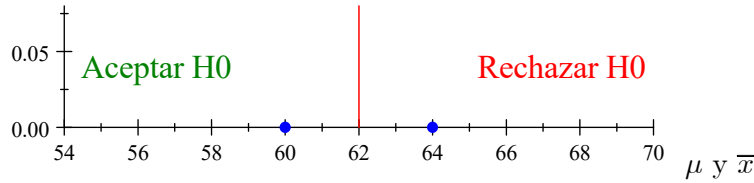


Figura 5.1: Criterio de decisión para $H_0 : \mu = 60$.

Por ejemplo, si se toma una muestra de X con $n = 100$, $\bar{x} = 61$ y $s = 20$, entonces se acepta H_0 y se considera que el disco es de marca A .

Al tomar la decisión sobre H_0 , se puede acertar o fallar, por lo que hay cuatro situaciones posibles:

	Aceptar H_0	Rechazar H_0
H_0 cierta	Acierto	Error de tipo I (con probabilidad α)
H_0 falsa	Error de tipo II (con probabilidad β)	Acierto

Es imposible saber si la decisión tomada es correcta o no, pero se puede calcular la probabilidad de cometer cada uno de los dos tipos de error. Para ello, se aplica el teorema del límite central, cosa que es posible, porque el tamaño muestral $n = 100$ es bastante grande, y se verifica (figura 5.2):

$$\bar{X} \underset{n \rightarrow \infty}{\approx} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{n \rightarrow \infty}{\simeq} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

También se ha usado que $S \underset{n \rightarrow \infty}{\simeq} \sigma$. En estas condiciones:

- La probabilidad de cometer un error de tipo I se denota α y, con el criterio de rechazo $\bar{X} \geq 62$, vale:

$$\begin{aligned} \alpha &= p(\text{error de tipo I}) = p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\bar{X} \geq 62 \mid \mu = 60) = \\ &= p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{62 - 60}{20/\sqrt{100}}\right) \simeq p(N(0, 1) \geq 1) = 1 - p(N(0, 1) < 1) \simeq \\ &\simeq 1 - 0.8413 = 0.1587. \end{aligned}$$

Este valor es el área bajo la función de densidad de \bar{X} , centrada en $\mu = 60$, a la derecha del punto crítico $\bar{x} = 62$ (figura 5.2.i).

- Por su parte, la probabilidad de cometer un error de tipo II se denota β y vale:

$$\begin{aligned} \beta &= p(\text{error de tipo II}) = p(\text{aceptar } H_0 \mid H_0 \text{ falsa}) = p(\bar{X} < 62 \mid \mu = 64) = \\ &= p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{62 - 64}{20/\sqrt{100}}\right) \simeq p(N(0, 1) < -1) = p(N(0, 1) > 1) \simeq 0.1587. \end{aligned}$$

Este valor es el área bajo la función de densidad de \bar{X} , centrada en $\mu = 64$, a la izquierda del punto crítico $\bar{x} = 62$ (figura 5.2.ii).

Se puede reducir la probabilidad de cometer uno de los dos errores a costa de aumentar la del otro. Por ejemplo, para conseguir que $\alpha = 0.05$, se desplaza el punto crítico de $\bar{x} = 62$ a

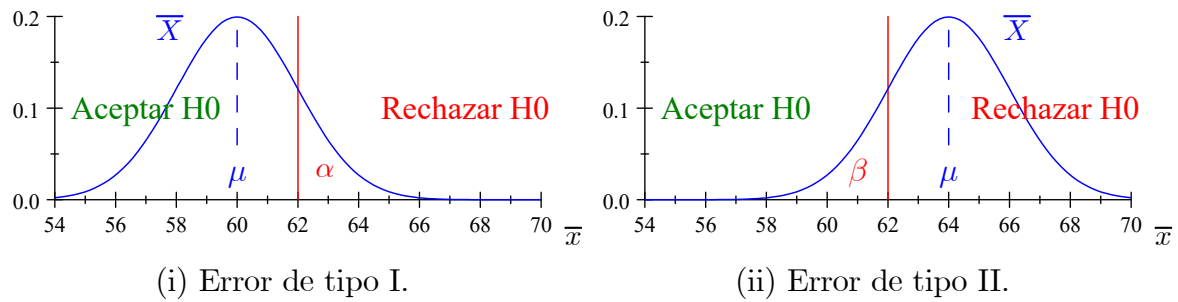


Figura 5.2: Errores en un contraste de hipótesis.

$\bar{x} = k$:

$$\begin{aligned}
 0.05 &= p(\bar{X} \geq k \mid \mu = 60) = p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{k - 60}{20/\sqrt{100}}\right) \simeq p\left(N(0, 1) \geq \frac{k - 60}{2}\right) \Rightarrow \\
 &\Rightarrow p\left(N(0, 1) < \frac{k - 60}{2}\right) \simeq 1 - 0.05 = 0.95 \Rightarrow \frac{k - 60}{2} \simeq 1.6449 \Rightarrow \\
 &\Rightarrow k \simeq 60 + 1.6449 \cdot 2 = 63.29.
 \end{aligned}$$

Este desplazamiento del punto crítico a $\bar{x} = 63.29$ hace que el valor de β aumente:

$$\begin{aligned}
 \beta &= p(\bar{X} \leq 63.29 \mid \mu = 64) = p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \frac{63.29 - 64}{20/\sqrt{100}}\right) \simeq p(N(0, 1) \leq -0.355) = \\
 &= 1 - p(N(0, 1) \leq 0.355) \simeq 1 - 0.6368 = 0.3632. \quad \blacktriangleleft
 \end{aligned}$$

La generalización del ejemplo anterior conduce a las siguientes definiciones.

Definición 5.1.2 (Contraste de hipótesis)

Sea X una variable aleatoria.

- Una *hipótesis estadística* es una conjetura sobre la distribución de X . Puede ser:
 - *No paramétrica* cuando se refiere al modelo de distribución de X (binomial, de Poisson, normal, ...).
 - *Paramétrica* cuando se refiere al valor de un parámetro del modelo.
- Un *contraste de hipótesis* es un algoritmo para aceptar o rechazar una hipótesis estadística en base a la información contenida en una muestra de X .
- Todo contraste se formula mediante dos hipótesis excluyentes:

$$\begin{cases} H_0 : \text{hipótesis nula: es la hipótesis a contrastar.} \\ H_1 : \text{hipótesis alternativa.} \end{cases}$$

- Para contrastar H_0 se adopta un *criterio de decisión*: partir el conjunto de muestras de X en dos regiones:
 - *Región de aceptación*: está formada por las muestras para las cuales se acepta H_0 .
 - *Región de rechazo* (o *crítica*): es el complementario de la región de aceptación.

- El *estadístico de contraste* es la función de la muestra que determina para qué valores de la misma se acepta H_0 y para cuáles se rechaza. ◀

Definición 5.1.3 (Errores en un contraste de hipótesis)

El contraste de la hipótesis H_0 está sujeto a dos tipos de errores, denominados:

- *Error de tipo I*: consiste en rechazar H_0 siendo cierta. Su probabilidad se denomina *nivel de significación del contraste*:

$$\alpha := p(\text{error de tipo I}) = p(\text{rechazar } H_0 \mid H_0 \text{ cierta}).$$

Es decir, α es la probabilidad de observar una muestra dentro de la región de rechazo cuando H_0 es cierta.

- *Error de tipo II*: consiste en aceptar H_0 siendo falsa. Su probabilidad vale:

$$\beta := p(\text{error de tipo II}) = p(\text{aceptar } H_0 \mid H_0 \text{ falsa}).$$

Es decir, β es la probabilidad de observar una muestra dentro de la región de aceptación cuando H_0 es falsa. ◀

Los errores de tipos I y II conducen a decisiones erróneas. Por tanto, el contraste ideal debería minimizar ambos a la vez. Pero, en general, fijada la muestra, cuando uno de los errores disminuye, el otro aumenta. Es decir, no existe un contraste óptimo. El único modo de reducir ambos errores a la vez es aumentar el tamaño muestral.

La clave de un contraste de hipótesis es definir las regiones de aceptación y de rechazo. En el caso del disco duro (ejemplo 5.1.1), el número que separa ambas regiones (*valor crítico*) se fijó arbitrariamente en $\bar{x} = 62$. Se trata de elegirlo de un modo razonado. Es decir, está pendiente la cuestión de seleccionar un contraste adecuado, aunque no sea óptimo, para resolver un problema dado. Esto se verá en la siguiente sección.

5.2 Construcción de contrastes paramétricos

Sea X una variable aleatoria cuya distribución depende de un parámetro desconocido θ . Las hipótesis paramétricas más comunes toman la forma:

- *Simple*: $\theta = \theta_0$.
- *Unilateral*: $\theta < \theta_0$, $\theta \leq \theta_0$, $\theta \geq \theta_0$ o $\theta > \theta_0$.
- *Bilateral*: $\theta \neq \theta_0$.

Ejemplo 5.2.1 (Tiempo de acceso a un disco duro)

En el ejemplo 5.1.1, relativo al disco duro, se pueden formular diversos contrastes:

- Para decidir si el tiempo medio de acceso es $60 \mu\text{s}$ o $64 \mu\text{s}$, se formula el contraste:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 64 \end{cases}$$

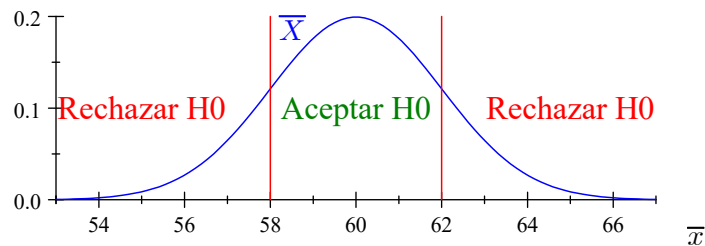


Figura 5.3: Contraste bilateral: $\bar{X} \approx N(\mu, \sigma/\sqrt{n})$.

Las hipótesis H_0 y H_1 se refieren al valor del parámetro μ , luego son hipótesis paramétricas. Ambas son simples.

- Para decidir si el tiempo medio de acceso no supera $60 \mu s$, se formula el contraste:

$$\begin{cases} H_0 : \mu \leq 60 \\ H_1 : \mu > 60 \end{cases}$$

Las hipótesis H_0 y H_1 son paramétricas y unilaterales. En este caso, se debe rechazar H_0 si \bar{X} está muy por encima de 60. Por eso, el criterio de decisión tiene una forma similar a la de la figura 5.2.i.

- Para decidir si el tiempo medio de acceso es exactamente $60 \mu s$, se formula el contraste:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu \neq 60 \end{cases}$$

También H_0 y H_1 son hipótesis paramétricas, pero H_0 es simple y H_1 es bilateral. En este caso, se debe rechazar H_0 si \bar{X} está muy lejos de 60, por exceso o por defecto. Por ejemplo, el criterio de decisión puede ser rechazar H_0 si y solo si $\bar{X} \notin [58, 62]$ (figura 5.3). ◀

Para construir un *contraste paramétrico*, se adopta la siguiente filosofía:

- Formular las hipótesis de modo que el error más serio sea el de tipo I.
- Fijar un nivel de significación, α , y mantener acotado el error de tipo I por α mientras se trata de minimizar el error de tipo II.

Por este motivo, H_0 juega con ventaja frente a H_1 . En consecuencia, H_0 nunca se considera probada, sino que se acepta a menos que los datos muestrales proporcionen evidencia suficiente para descartarla. De ahí el adjetivo nula. En tal caso, se acepta implícitamente H_1 . La decisión de rechazar H_0 no significa que sea falsa, sino que los datos muestrales proporcionan evidencia para el rechazo con el grado de confianza especificado.

Definición 5.2.2 (*p*-valor de un contraste de hipótesis)

El *p*-valor de un contraste de hipótesis es el mínimo nivel de significación que conduce a rechazar la hipótesis nula con la muestra observada.

Dicho de otro modo, el *p*-valor es el tamaño del error de tipo I que se comete al rechazar la hipótesis nula con la muestra observada. ◀

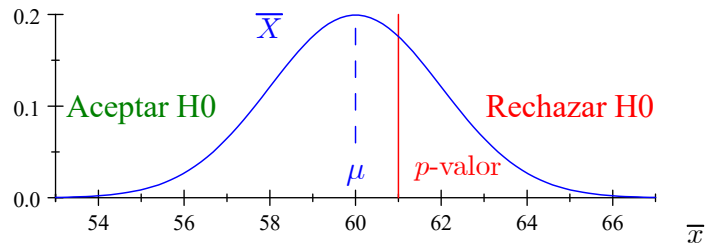


Figura 5.4: El p -valor de un contraste.

Ejemplo 5.2.3 (p -valor en el tiempo de acceso en un disco duro)

En el ejemplo 5.1.1, relativo al disco duro, se contrasta:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 64 \end{cases}$$

Se toma una muestra de X con $n = 100$, $\bar{x} = 61$ y $s = 20$. Si 61 es el mínimo valor de \bar{X} que conduce a rechazar la hipótesis nula, entonces el p -valor del contraste es (figura 5.4):

$$\begin{aligned} p\text{-valor} &= p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\bar{X} \geq 61 \mid \mu = 60) = \\ &= p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{61 - 60}{20/\sqrt{100}}\right) \simeq p(N(0, 1) \geq 0.5) \simeq 0.3085. \end{aligned}$$

Si se rechaza H_0 con esta muestra, entonces la probabilidad de cometer un error de tipo I es 30.85 %. Esta penalización es muy alta, por lo que se debe aceptar H_0 . ◀

Cuanto más próximo a 0 está el p -valor de un contraste, más evidencia aporta la muestra para rechazar H_0 . Por el contrario, cuanto más próximo a 1 se encuentra el p -valor, más evidencia aporta la muestra para aceptar H_0 . En todo caso, la decisión se toma comparando el p -valor con el nivel de significación α fijado de antemano.

Algoritmo 5.2.4 (Construcción de un contraste de hipótesis paramétrico)

Sea X una variable aleatoria cuya distribución depende del parámetro desconocido θ . Dados:

- Un contraste de hipótesis relativo al valor de θ , que debe ser de uno de los 4 tipos siguientes:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

- Un nivel de significación α , con $0 < \alpha < 1$.
- Una muestra de X : $[X_1, \dots, X_n]$.

Para realizar el contraste, se hace lo siguiente:

1. Comprobar la normalidad de X .
2. Hallar un *pivote*, que es un estadístico tal que:

- Solo depende del parámetro θ y de la muestra: $T(\theta, X_1, \dots, X_n)$.
- Tiene distribución conocida cuando H_0 es cierta.

3. Calcular el p -valor del contraste para la muestra observada, que es:

$$p\text{-valor} = \begin{cases} p(T \leq t) & \text{si } H_1 : \theta < \theta_0 \\ p(T \geq t) & \text{si } H_1 : \theta > \theta_0 \\ 2p(T \leq t) & \text{si } H_1 : \theta \neq \theta_0 \text{ y } t < \theta_0 \\ 2p(T \geq t) & \text{si } H_1 : \theta \neq \theta_0 \text{ y } t > \theta_0 \end{cases}$$

donde $t = T(\theta_0, x_1, \dots, x_n)$ es el valor del pivote, para la muestra observada, cuando H_0 es cierta.

4. Decidir: aceptar H_0 si y solo si $p\text{-valor} > \alpha$. ◀

Este método es análogo al empleado para construir intervalos de confianza. Para ilustrarlo, se aplicará a continuación para realizar ciertos contrastes de hipótesis.

5.3 Contraste para la media

Sea X una variable aleatoria con media μ y desviación típica σ , ambas desconocidas. Se trata de realizar un contraste de hipótesis relativo a μ . Hay varias situaciones posibles, que se estudian a continuación.

5.3.1 Variable normal

Sea $X \sim N(\mu, \sigma)$ con μ y σ desconocidas. El ejemplo siguiente muestra el proceso a seguir.

Ejemplo 5.3.1 (Montaje de ordenadores: contrastes unilateral y bilateral para μ)

En la fábrica de ordenadores del ejemplo 4.4.4, el tiempo medio de montaje de un ordenador era 36 min durante la última revisión realizada. Para saber si dicho valor ha aumentado, al 2% de significación, se toma una muestra de la variable aleatoria:

X : “tiempo de montaje de un ordenador (min)”

y se obtienen los resultados: $n = 9$, $\bar{x} = 40$ y $s = 5$. La media muestral es mayor que 36, pero puede que la media verdadera no lo sea. Hay que contrastar, al 2% de significación, que la media de X verifica $\mu > 36$:

$$\begin{cases} H_0 : \mu \leq 36 \\ H_1 : \mu > 36 \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu = 36 \\ H_1 : \mu > 36 \end{cases}$$

1. Normalidad: STATGRAPHICS ajusta $X \approx N(40, 5)$ con $p\text{-valor} = 0.996713$, luego el ajuste es bueno.

2. Pivote: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} = t_8.$

3. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\bar{X} \geq 40 \mid \mu = 36) = p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{40 - 36}{5/\sqrt{9}}\right) =$
 $= p(t_8 \geq 2.4) = 1 - p(t_8 \leq 2.4) \begin{cases} < 1 - 0.975 = 0.025 \\ > 1 - 0.98 = 0.02 \end{cases} \quad (\text{STATGRAPHICS } 0.0215884).$

4. Decisión: p -valor $> 0.02 = \alpha \Rightarrow$ aceptar H_0 . Al 2% de significación, se acepta que $\mu \leq 36$ min, es decir, que el tiempo medio de montaje de un ordenador no ha aumentado.

Visto que $\mu \leq 36$ min, se quiere contrastar, al 2% de significación, que $\mu = 35$ min:

$$\begin{cases} H_0 : \mu = 35 \\ H_1 : \mu \neq 35 \end{cases}$$

1 y 2. Normalidad y pivote: como en el contraste anterior.

3. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\{\bar{X} \leq 30\} \cup \{\bar{X} \geq 40\} \mid \mu = 35) =$
 $= p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \frac{30 - 35}{5/\sqrt{9}}\right) + p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq \frac{40 - 35}{5/\sqrt{9}}\right) = p(t_8 \leq -3) + p(t_8 \geq 3) =$
 $= 2p(t_8 \geq 3) = 2(1 - p(t_8 \leq 3)) \begin{cases} < 2(1 - 0.99) = 0.02 \\ > 2(1 - 0.995) = 0.01 \end{cases} \quad (\text{STATGRAPHICS } 0.0170717).$

4. Decisión: p -valor $< 0.02 = \alpha \Rightarrow$ rechazar H_0 . Al 2% de significación, se rechaza que $\mu = 35$ min.

Hay que señalar que este contraste es bilateral, por lo que se llega a la misma conclusión construyendo un intervalo de confianza para μ al 98% y viendo si el punto 35 pertenece o no a dicho intervalo de confianza (ver el ejemplo 4.5.2). ◀

5.3.2 Variable no normal con muestra grande

Sea $[X_1, \dots, X_n]$ una muestra de la variable aleatoria X . Supongamos que X no es normal, pero tiene varianza finita y n es suficientemente grande ($n \geq 100$). En este caso,

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

Este pivote aproximado permite realizar un *contraste de hipótesis aproximado* relativo a μ . El ejemplo siguiente muestra el proceso a seguir.

Ejemplo 5.3.2 (Aerolínea: contraste unilateral para μ)

La aerolínea A del ejemplo 4.5.3 afirma que el precio medio de sus billetes es menor que 127 €. Para comprobarlo, al 4% de significación, se toma una muestra de la variable aleatoria:

X : “precio de un billete de la aerolínea A (€)”

y se obtienen los resultados: $n = 400$, $\bar{x} = 124$ y $s = 40$. La media muestral es menor que 127, pero puede que la media verdadera no lo sea. Hay que contrastar, al 4% de significación, que la media de X verifica $\mu < 127$:

$$\begin{cases} H_0 : \mu \geq 127 \\ H_1 : \mu < 127 \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu = 127 \\ H_1 : \mu < 127 \end{cases}$$

1. Normalidad: STATGRAPHICS ajusta $X \approx N(124, 40)$ con p -valor = 0, luego X no es normal. No obstante, su muestra es grande ($n = 400$), luego \bar{X} es aproximadamente normal y se puede usar un pivote aproximado.

2. Pivote aproximado: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx N(0, 1)$.

3. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\bar{X} \leq 124 \mid \mu = 127) =$

$$\begin{aligned} &= p\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \frac{124 - 127}{40/\sqrt{400}}\right) \simeq p(N(0, 1) \leq -1.5) = 1 - p(N(0, 1) \leq 1.5) \simeq \\ &\simeq 1 - 0.9332 = 0.0668. \end{aligned}$$

STATGRAPHICS usa como pivote t_{399} en lugar de $N(0, 1)$ y obtiene p -valor = 0.0672027.

4. Decisión: p -valor $\simeq 0.0668 > 0.04 = \alpha \Rightarrow$ aceptar H_0 . Al 4% de significación, se rechaza que el precio medio de un billete de la aerolínea A es menor que 127 €. ◀

5.3.3 Proporción con muestra grande

Sea $X \sim B(1, p)$, con p desconocido¹. Sea $[X_1, \dots, X_n]$ una muestra de X . En este caso, $\mu = p$, $\sigma = \sqrt{p(1-p)}$ y un estimador de p es $\hat{P} = \bar{X}$. Si n es suficientemente grande, el teorema del límite central proporciona la aproximación:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\hat{P} - p}{\sqrt{p(1-p)}/\sqrt{n}} = \frac{\hat{P} - p}{\sqrt{p(1-p)}/n} \underset{n \rightarrow \infty}{\approx} N(0, 1).$$

Este pivote aproximado permite realizar un *contraste de hipótesis aproximado* relativo a p . El ejemplo siguiente muestra el proceso a seguir.

Ejemplo 5.3.3 (Lectora de marcas ópticas: contraste unilateral para p)

Se quiere comprobar que la proporción de errores que comete una lectora de marcas ópticas no supera el 1%. Para ello, se toma una muestra de 200 marcas ópticas y resulta que la lectora lee mal 3 de ellas. Se trata de decidir al 3% de significación, para lo cual se considera la variable aleatoria:

$$X : \text{“número de errores cometidos por la lectora al leer una marca óptica”} \sim B(1, p).$$

Hay de contrastar, al 3% de significación, que $p \leq 0.01$:

$$\begin{cases} H_0 : p \leq 0.01 \\ H_1 : p > 0.01 \end{cases} \Leftrightarrow \begin{cases} H_0 : p = 0.01 \\ H_1 : p > 0.01 \end{cases}$$

¹El parámetro p es la proporción de elementos de la población para los cuales X vale 1, lo cual significa que verifican cierta propiedad.

1. Normalidad: $X \sim B(1, p)$ no es normal, pero su muestra es grande ($n = 200$), luego \bar{X} es aproximadamente normal y se puede usar un pivote aproximado.

2. Pivote aproximado: $\frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1)$.

3. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(\hat{P} \geq 3/200 \mid p = 0.01) =$
 $= p\left(\frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \geq \frac{3/200 - 0.01}{\sqrt{0.01(1-0.01)/200}}\right) \simeq p(N(0, 1) \geq 0.71067) =$
 $= 1 - p(N(0, 1) \leq 0.71067) \simeq 1 - 0.7611 = 0.2389$.

STATGRAPHICS usa otro pivote² y obtiene p -valor = 0.323321.

4. Decisión: p -valor $\simeq 0.2389 > 0.03 = \alpha \Rightarrow$ aceptar H_0 . Al 3% de significación, se acepta que la lectora comete un 1% de errores, como mucho. ◀

5.4 Contraste para la varianza

Sea $X \sim N(\mu, \sigma)$, con σ desconocida. Se trata de realizar un contraste de hipótesis relativo a σ . El siguiente ejemplo muestra el proceso a seguir.

Ejemplo 5.4.1 (Montaje de ordenadores: contraste unilateral para σ)

En la fábrica de ordenadores del ejemplo 5.3.1, se quiere contrastar, al 2% de significación, que la desviación típica del tiempo de montaje de un ordenador supera los 3 min. Para ello, se observa una muestra de la variable aleatoria:

X : “tiempo de montaje de un ordenador (min)”

y se obtiene: $n = 9$, $\bar{x} = 40$ y $s = 5$. La cuasidesviación típica muestral es mayor que 3, pero σ puede ser algo menor o algo mayor. Hay que contrastar, al 2% de significación, que la desviación típica de X vale $\sigma > 3$. Este contraste se plantea y resuelve así:

$$\begin{cases} H_0 : \sigma \leq 3 \\ H_1 : \sigma > 3 \end{cases} \Leftrightarrow \begin{cases} H_0 : \sigma^2 \leq 3^2 \\ H_1 : \sigma^2 > 3^2 \end{cases} \Leftrightarrow \begin{cases} H_0 : \sigma^2 = 3^2 \\ H_1 : \sigma^2 > 3^2 \end{cases}$$

1. Normalidad: STATGRAPHICS ajusta $X \approx N(40, 5)$ con p -valor = 0.996713, luego el ajuste es bueno.

²STATGRAPHICS calcula:

$$\begin{aligned} p\text{-valor} &= p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(B(200, p) \geq 3 \mid p = 0.01) = \\ &= p(B(200, 0.01) \geq 3) = \sum_{i=3}^{200} \binom{200}{i} 0.01^i \cdot 0.99^{200-i} \simeq 0.32332. \end{aligned}$$

2. Pivote: $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 = \chi_8^2$.
3. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(S^2 \geq 5^2 \mid \sigma^2 = 3^2) =$
 $= p\left(\frac{(n-1)S^2}{\sigma^2} \geq \frac{8 \cdot 5^2}{3^2}\right) = p(\chi_8^2 \geq 22.222) = 1 - p(\chi_8^2 \leq 22.222) <$
 $< 1 - 0.995 = 0.005$ (STATGRAPHICS 0.00452042).
4. Decisión: $p\text{-valor} < 0.005 < 0.02 = \alpha \Rightarrow$ rechazar H_0 . Al 2% de significación, se acepta que $\sigma > 3$ min. ◀

5.5 Contrastes no paramétricos

En los problemas de estimación y contraste de hipótesis considerados hasta ahora, se ha supuesto que los datos provienen de una distribución de tipo conocido, aunque se ignora el valor de algunos de sus parámetros. Concretamente, en las secciones anteriores se han estudiado los contrastes de hipótesis paramétricos. En ellos, H_0 y H_1 se refieren a los valores que toman uno o más parámetros desconocidos en la distribución en estudio. En esta sección se analizan problemas en los que se infiere la forma de la distribución, determinando la bondad del ajuste del modelo propuesto a los datos observados.

Definición 5.5.1 (Hipótesis no paramétrica)

Sea X una variable aleatoria. Una hipótesis estadística sobre X es *no paramétrica* cuando define el modelo de distribución de X (binomial, de Poisson, normal, etc.). ◀

Ejemplo 5.5.2 (Tiempo de acceso a un disco duro)

En el ejemplo 5.1.1, relativo a un disco duro, para decidir si el tiempo de acceso a un dato sigue una distribución $N(60, 10)$, se formula:

$$\begin{cases} H_0 : X \sim N(60, 10) \\ H_1 : X \approx N(60, 10) \end{cases}$$

Ambas hipótesis se refieren al modelo de distribución, luego son no paramétricas. ◀

Un procedimiento de esta naturaleza se denomina *contraste no paramétrico*, o *contraste de bondad de ajuste*. Si X es una variable aleatoria y F es un modelo determinado de distribución (o su función de distribución), el contraste se formula como:

$$\begin{cases} H_0 : X \sim F \\ H_1 : X \approx F \end{cases}$$

La metodología que se utiliza para construir estos contrastes es similar a la del caso paramétrico. Ahora, se trata de medir la discrepancia entre los valores observados en la muestra y los que se obtendrían si la hipótesis nula fuese cierta.

Los contrastes más utilizados son el χ^2 y el de Kolmogorov-Smirnov, que se describen a continuación.

5.5.1 Contraste χ^2

El contraste χ^2 compara las frecuencias observadas en la muestra con las frecuencias esperadas de la distribución en la hipótesis nula, para un número finito de clases. Este test, por su naturaleza discreta, solo es apropiado para variables discretas, aunque también se puede aplicar a las variables continuas.

Algoritmo 5.5.3 (Contraste χ^2)

Sea X una variable aleatoria con distribución desconocida. Dados:

- Un contraste de hipótesis relativo a la distribución de X :

$$\begin{cases} H_0 : X \sim F \\ H_1 : X \not\sim F \end{cases}$$

- Un nivel de significación α , con $0 < \alpha < 1$.
- Una muestra de X , $[X_1, \dots, X_n]$, con tamaño $n \geq 30$.

Para realizar el contraste, se hace lo siguiente:

1. Estimar los parámetros desconocidos de F , a partir de la muestra, *por máxima verosimilitud*. Llamar r al número de parámetros estimados.
2. Construir la *tabla de discrepancias*:

Im (X)	n_i	p_i	$n \cdot p_i$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
C_1	n_1	p_1	$n \cdot p_1$	$\frac{(n_1 - n \cdot p_1)^2}{n \cdot p_1}$
C_2	n_2	p_2	$n \cdot p_2$	$\frac{(n_2 - n \cdot p_2)^2}{n \cdot p_2}$
\vdots	\vdots	\vdots	\vdots	\vdots
C_m	n_m	p_m	$n \cdot p_m$	$\frac{(n_m - n \cdot p_m)^2}{n \cdot p_m}$
$\sum_{i=1}^m n_i = n \quad \sum_{i=1}^m p_i = 1 \quad \sum_{i=1}^m n \cdot p_i = n$				$D := \sum_{i=1}^m \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$

En esta tabla:

- C_1, C_2, \dots, C_m son m *clases*, que particionan la imagen de X .
 - n_i es la *frecuencia observada* en la clase C_i , con $i = 1..m$.
 - $p_i := p(X \in C_i \mid H_0 \text{ cierta})$, con $i = 1..m$.
 - $n \cdot p_i$ es la *frecuencia esperada* en C_i , con $i = 1..m$. Si en alguna clase se verifica que $n \cdot p_i < 5$, entonces se debe reagrupar las clases.
3. Tomar como pivote aproximado:

$$D := \sum_{i=1}^m \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} \underset{n \rightarrow \infty}{\approx} \chi_{m-r-1}^2.$$

El pivote D mide las discrepancias entre las frecuencias observadas y las esperadas.

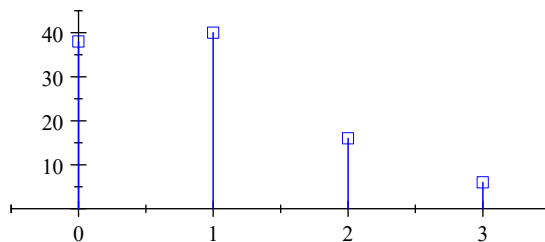


Figura 5.5: Número de coches vendidos diariamente.

4. Hallar el p -valor del contraste para la muestra observada, que es $p(\chi_{m-r-1}^2 \geq D)$.
5. Decidir: aceptar H_0 si y solo si p -valor $> \alpha$.

Demostración: Ver Rohatgi [15, página 505, Theorem 4]. ◀

El pivote D mide $n_i - n \cdot p_i$, que es la discrepancia entre las frecuencias observadas en la muestra de X y las esperadas de la distribución F . Si la distribución de X es realmente F , entonces tales diferencias deben ser pequeñas y el valor de D debe estar próximo a 0. Por este motivo, se rechaza H_0 cuando dicho valor, d , es demasiado grande. El nombre de este contraste proviene de la distribución aproximada de su pivote.

Ejemplo 5.5.4 (Variable $P(\lambda)$): contraste χ^2)

En un concesionario de automóviles, se quiere ajustar un modelo de probabilidad a la variable aleatoria:

X : “número de coches vendidos diariamente”.

La imagen de X es $\text{Im}(X) = \mathbb{N}$, luego X es discreta. Por su naturaleza, X puede seguir una distribución de Poisson. Para contrastarlo, se observa un muestra de X y se obtiene:

X	0	1	2	3
Frecuencia	38	40	16	6

(Columna *Concesionario* del fichero **6-Contrastes - Datos.sgd**.) El diagrama de barras de la muestra (figura 5.5) sugiere una distribución binomial o de Poisson. La media y la varianza muestrales valen, respectivamente:

$$\bar{x} = \frac{0 \cdot 38 + 1 \cdot 40 + 2 \cdot 16 + 3 \cdot 6}{38 + 40 + 16 + 6} = \frac{9}{10} = 0.9,$$

$$v = \frac{0^2 \cdot 38 + 1^2 \cdot 40 + 2^2 \cdot 16 + 3^2 \cdot 6}{100} - \left(\frac{9}{10}\right)^2 = \frac{77}{100} = 0.77.$$

Estos valores son similares, luego se opta por la distribución de Poisson. Es decir, se contrasta, al 5% de significación:

$$\begin{cases} H_0 : X \sim P(\lambda) \\ H_1 : X \not\sim P(\lambda) \end{cases}$$

Se trata de un contraste de hipótesis no paramétrico sobre un modelo de distribución discreta y el tamaño muestral es $n = 100 \geq 30$, por lo que se aplica el test χ^2 .

1. Estimar los parámetros por máxima verosimilitud: $\hat{\lambda} = \bar{x} = 0.9$ y $r = 1$ (STATGRAPHICS =).
2. Tabla de discrepancias:

Im (X)	n_i	p_i	$n \cdot p_i$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
0	38	0.4066	40.66	0.1740186915
1	40	0.7725 - 0.4066	36.59	0.3177944793
2	16	0.9371 - 0.7725	16.46	0.01285540704
≥ 3	6	1 - 0.9371	6.29	0.01337042925
$m = 4$	$n = 100$	1	100	$D = 0.5180390072$

Todas las clases cumplen que $n \cdot p_i \geq 5$, luego la agrupación es válida. (STATGRAPHICS obtiene $D = 0.517358$.)

3. Pivote aproximado: $D \approx \chi_{m-r-1}^2 = \chi_2^2$ (STATGRAPHICS =).
4. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(D \geq 0.518 \mid X \sim P(0.9)) \simeq p(\chi_2^2 \geq 0.518) =$
 $= 1 - p(\chi_2^2 < 0.518) \begin{cases} < 1 - 0.1 = 0.9 \\ > 1 - 0.25 = 0.75 \end{cases}$ (STATGRAPHICS 0.772071).

5. Decisión: $p\text{-valor} > 0.75 > 0.05 = \alpha \Rightarrow$ aceptar H_0 . Al 5% de significación, se acepta que el número de coches vendidos diariamente en el concesionario sigue una distribución de Poisson $P(0.9)$. ◀

5.5.2 Contraste de Kolmogorov-Smirnov

El contraste χ^2 se encuentra limitado cuando la distribución bajo la hipótesis nula es continua, por lo que en este caso es más adecuado otro contraste, llamado de *Kolmogorov-Smirnov*. Este test no requiere ninguna elección arbitraria de clases y se puede aplicar con muestras de tamaño pequeño. A cambio, solo es adecuado para variables aleatorias continuas. La idea del contraste radica en comparar la función de distribución bajo la hipótesis nula con la función de distribución llamada empírica, que se obtiene a partir de la muestra.

Definición 5.5.5 (Función de distribución empírica)

Sea X una variable aleatoria con función de distribución F . Sea $[x_1 \leq x_2 \leq \dots \leq x_n]$ una muestra observada de X . Se llama *función de distribución empírica* de la muestra de X a la función de distribución de $U(x_1, \dots, x_n)$. Se nota F_n y vale:

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ i/n & x_i \leq x < x_{i+1}, \quad i = 1..n-1 \\ 1 & x_n \leq x \end{cases} \quad \blacktriangleleft$$

Para todo $x \in \mathbb{R}$, el valor de $F_n(x)$ es la proporción de elementos de la muestra menores o iguales que x . La figura 5.6 superpone F , que es la función de distribución de X , con la

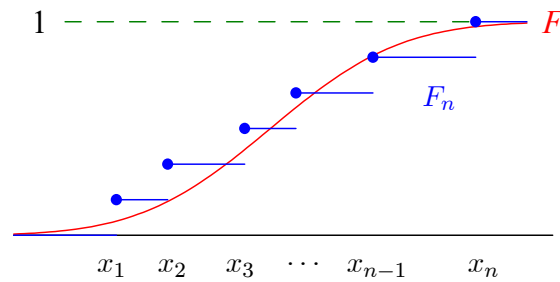


Figura 5.6: Distribuciones real, F , y empírica, F_n .

función de distribución empírica F_n . La función F_n es escalonada, con n peldaños de altura $1/n$. Cuando $x_i = x_{i+1}$, el peldaño correspondiente no existe.

Rohatgi [15, página 311, Theorem 2] demuestra que F_n converge a F cuando n tiende a infinito. Por tanto, para ver si X sigue la distribución F , se puede medir la discrepancia entre F_n y F por medio de su diferencia:

$$D_n := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Algoritmo 5.5.6 (Contraste de Kolmogorov-Smirnov)

Sea X una variable aleatoria con distribución desconocida. Dados:

- Un contraste de hipótesis relativo a la distribución de X :

$$\begin{cases} H_0 : X \sim F \\ H_1 : X \not\sim F \end{cases}$$

donde F es una distribución continua.

- Un nivel de significación α , con $0 < \alpha < 1$.
- Una muestra de X : $[X_1, \dots, X_n]$.

Para realizar el contraste, se hace lo siguiente:

1. Estimar los parámetros desconocidos de F a partir de la muestra.
2. Obtener la función de distribución de X , que es F_n .
3. Comparar F_n , la función de distribución empírica de X , con la función de distribución F . Para ello, se construye la *tabla de discrepancias*:

x_i	i/n	$F(x_i)$	$i/n - F(x_i)$	$F(x_i) - (i-1)/n$
x_1	$1/n$	$F(x_1)$	$1/n - F(x_1)$	$F(x_1) - 0$
x_2	$2/n$	$F(x_2)$	$2/n - F(x_2)$	$F(x_2) - 1/n$
\vdots	\vdots	\vdots	\vdots	\vdots
x_n	$n/n = 1$	$F(x_n)$	$1 - F(x_n)$	$F(x_n) - (n-1)/n$
n			$D_n^+ := \max_{i=1..n} \left(\frac{i}{n} - F(x_i) \right)$	$D_n^- := \max_{i=1..n} \left(F(x_i) - \frac{i-1}{n} \right)$

4. Tomar como pivote:

$$D_n := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \max(D_n^+, D_n^-) = \max_{i=1..n} \left(\frac{i}{n} - F(x_i), F(x_i) - \frac{i-1}{n} \right).$$

La distribución de D_n es independiente de la distribución F y está tabulada.

Calcular el valor del pivote para la muestra observada: d .

5. Hallar el p -valor del contraste para la muestra observada, que es $p(D_n > d)$.

6. Decidir: aceptar H_0 si y solo si p -valor $> \alpha$.

Demostración: Ver Rohatgi [15, página 609, Theorem 1]. ◀

Ejemplo 5.5.7 (Variable $\text{Exp}(\beta)$): contraste de Kolmogorov-Smirnov)

En un servidor de ficheros, se quiere ajustar un modelo de probabilidad a la variable aleatoria:

X : “tiempo de respuesta a una petición (s)”.

Para ello, se toma una muestra de X y se obtiene: [3.4, 1.9, 4.1, 0.6]. Por la naturaleza de la variable, se considera que X puede seguir una distribución exponencial y se desea contrastar esta hipótesis al 3% de significación:

$$\begin{cases} H_0 : X \sim \text{Exp}(\beta) \\ H_1 : X \not\sim \text{Exp}(\beta) \end{cases}$$

Se trata de un contraste no paramétrico sobre un modelo de distribución continua, luego se aplica el test de Kolmogorov-Smirnov.

1. Estimar parámetros: el estimador de β por momentos o por máxima verosimilitud es $\hat{\beta} = 1/\bar{X}$. Por tanto, la estimación de β con la muestra observada es:

$$\hat{\beta} = \frac{1}{\bar{X}} = \frac{4}{3.4 + 1.9 + 4.1 + 0.6} = 0.4 \quad (\text{STATGRAPHICS estima } \widehat{1/\beta} = 2.5.)$$

2. Función de distribución de X :

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-0.4x} & 0 \leq x \end{cases}$$

El cálculo correspondiente a $x \geq 0$ es:

$$F(x) = \int_0^x \beta e^{-\beta t} dt = [-e^{-\beta t}]_0^x = -e^{-\beta x} - (-1) = 1 - e^{-\beta x}.$$

3. Tabla de discrepancias:

x_i	i/n	$F(x_i)$	$i/n - F(x_i)$	$F(x_i) - (i-1)/n$
0.6	0.25	0.213 372	0.036 628	0.213 372
1.9	0.5	0.532 334	< 0	0.282 334
3.4	0.75	0.743 339	0.006 661	0.243 339
4.1	1	0.806 020	0.193 98	0.056 020
$n = 4$			$D_n^+ = 0.193 98$	$D_n^- = 0.282 334$

4. Pivote: D_4 , que toma el valor 0.282 334 (STATGRAPHICS =).

5. p -valor = $p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = p(D_4 \geq 0.282\,334 \mid X \sim \text{Exp}(0.4)) =$

$$= 1 - p(D_4 \leq 0.282\,334) \begin{cases} < 1 - 0 = 1 \\ > 1 - 0.8 = 0.2 \end{cases} \quad (\text{STATGRAPHICS } 0.907336).$$

6. Decisión: $p\text{-valor} > 0.2 > 0.03 = \alpha \Rightarrow$ aceptar H_0 . Al 3% de significación, se acepta que el tiempo de respuesta a una petición sigue una distribución $\text{Exp}(0.4)$. ◀

Apéndice A

Estadística descriptiva

En este apéndice se estudian ciertos conceptos de estadística descriptiva. Se aplican a variables definidas sobre un espacio muestral finito. Por tanto, tales variables son discretas cuando son numéricas.

A.1 Población, variable y muestra

Definición A.1.1 (Población, variable y muestra)

En todo problema estadístico hay tres ingredientes:

- *Población o espacio muestral*: es un conjunto de cuyos elementos interesa estudiar una característica. Se representa por la letra E . Por ejemplo:

$$E = \{\text{personas de un grupo}\} \quad \text{o} \quad E = \{\text{listas de 1000 números}\}.$$

- *Variable*: es una característica de la población que se quiere estudiar. Por ejemplo:

$$X : \text{“edad de una persona del grupo”} \quad \text{o} \quad X : \text{“tiempo de ordenación de una lista”}.$$

- *Muestra*: es un subconjunto finito y no vacío de la población. Por ejemplo:

$$M = \{\text{personas de un subconjunto del grupo}\} \quad \text{o} \quad M = \{20 \text{ listas de 1000 números}\}.$$

Matemáticamente, una variable es una función de la población en un conjunto de valores, $X : E \rightarrow R$ (figura A.1). Según que sus valores sean numéricos o no, hay dos tipos de variables:

- Variable *cuantitativa* o *numérica*, cuando $\text{Im}(X) \subseteq \mathbb{R}$. En este caso, X es una variable aleatoria.
- Variable *cualitativa* o *categorica*: $\text{Im}(X) \not\subseteq \mathbb{R}$. ◀

Ejemplo A.1.2 (Personas de un grupo)

Sobre la población $E = \{\text{personas de un grupo}\}$ se pueden considerar diversas variables:

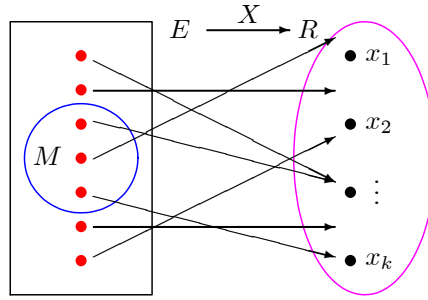


Figura A.1: Población, variable y muestra.

<i>Persona</i>	1	2	3	4	5	6	7	8	9	10
$X = edad$	20	19	20	20	21	20	20	25	21	21
$Y = estatura$	181	176	175	165	180	183	178	164	175	180
$Z = sexo$	<i>h</i>	<i>h</i>	<i>h</i>	<i>m</i>	<i>h</i>	<i>h</i>	<i>h</i>	<i>m</i>	<i>h</i>	<i>h</i>
<i>Persona</i>	11	12	13	14	15	16	17	18	19	20
$X = edad$	20	23	20	21	19	20	20	19	23	21
$Y = estatura$	184	167	178	170	168	187	185	186	168	170
$Z = sexo$	<i>h</i>	<i>m</i>	<i>h</i>	<i>h</i>	<i>m</i>	<i>h</i>	<i>h</i>	<i>h</i>	<i>m</i>	<i>m</i>

Tabla A.1: Datos de las personas de un grupo.

- X : “edad de una persona del grupo”. Su imagen $\text{Im}(X) \subseteq \mathbb{R}$ es un conjunto numérico, luego X es cuantitativa.
- Y : “estatura de una persona del grupo”. $\text{Im}(Y) \subseteq \mathbb{R}$, luego Y es cuantitativa.
- Z : “sexo de una persona del grupo”. $\text{Im}(Z) = \{hombre, mujer\}$ no es un conjunto numérico, luego Z es cualitativa. No obstante, si se codifica $hombre = 0$ y $mujer = 1$, entonces $\text{Im}(Z) = \{0, 1\}$ y Z se puede considerar como cualitativa.

Para cierto grupo de $n = 20$ personas, los valores de X , Y y Z figuran en la tabla A.1. ◀

A.2 Frecuencias y gráficos de una variable

Cuando la estadística estudia una variable, los elementos de la población no son relevantes. Solo interesan los valores que toma la variable y la frecuencia con que los toma.

Definición A.2.1 (Tabla de frecuencias y diagrama de barras)

Sea $X : E \rightarrow R$ una variable definida sobre un espacio muestral finito de tamaño $n = \text{card}(E)$.

Su *tabla de frecuencias* tiene la forma:

$\text{Im}(X)$	n_i	f_i	N_i	F_i
x_1	n_1	n_1/n	n_1	f_1
\vdots	\vdots	\vdots	\vdots	\vdots
x_i	n_i	n_i/n	$\sum_{j=1}^i n_j$	N_i/n
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	n_k/n	n	1
	n	1		

La primera columna contiene los valores de la variable. Para cada uno de estos valores, x_i , el resto de su fila contiene sus frecuencias, que se definen del siguiente modo:

- n_i , *frecuencia* (absoluta): es el número de elementos de la población cuyo valor es x_i . Se verifica que $\sum_{i=1}^k n_i = n$ (tamaño de la población).
- f_i , *frecuencia relativa*: es el cociente entre la frecuencia absoluta y el tamaño muestral. Representa la probabilidad de que la variable X tome el valor x_i y se puede expresar como un porcentaje. Se verifica que:

$$0 \leq f_i = \frac{n_i}{n} = p(X = x_i) \leq 1 \quad \text{y} \quad \sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{\sum_{i=1}^k n_i}{n} = 1.$$

- N_i , *frecuencia acumulada*: es la suma de las frecuencias de los valores x_i y precedentes, es decir, $N_i := \sum_{j=1}^i n_j$. Obviamente, $N_1 = n_1$ y $N_k = n$.
- F_i , *frecuencia acumulada relativa*: es el cociente entre la frecuencia acumulada de x_i y el tamaño muestral. Coincide con la suma de las frecuencias relativas de los valores x_i y precedentes, es decir:

$$F_i := \frac{N_i}{n} = \frac{\sum_{j=1}^i n_j}{n} = \sum_{j=1}^i \frac{n_j}{n} = \sum_{j=1}^i f_j.$$

Obviamente, $F_1 = f_1$ y $F_k = 1$.

Además:

- La *distribución de frecuencias* de X está formada por los pares [valor, frecuencia], es decir: $[x_1, n_1], [x_2, n_2], \dots, [x_k, n_k]$.
- El *diagrama de barras* de X se construye levantando sobre cada valor $x_i \in \text{Im}(X)$ una barra de altura su frecuencia, es decir, uno de los valores n_i, f_i, N_i o F_i . Este diagrama es análogo al gráfico de la función de masa de X considerada como variable aleatoria.

En el caso de que X sea una variable cuantitativa, la tabla de frecuencias relativas es la *función de masa* de X , considerada como variable aleatoria, y el diagrama de barras es su gráfica. Las frecuencias acumuladas relativas son los valores de la *función de distribución* de X . ◀

Ejemplo A.2.2 (Frecuencias y barras de *Edad*)

Consideremos como espacio muestral el grupo de 20 personas del ejemplo A.1.2 y como variable

<i>Edad</i>	n_i	f_i	N_i	F_i
19	2	0.1	2	0.1
20	10	0.5	12	0.6
21	5	0.25	17	0.85
23	2	0.1	19	0.95
25	1	0.05	20	1
<hr/>				
$k = 5$	$n = 20$	1		



Figura A.2: Tabla de frecuencias y diagrama de barras de *Edad*.

<i>Estatura</i>	n_i	f_i	N_i	F_i
164	1	0.05	1	0.05
165	1	0.05	2	0.1
167	1	0.05	3	0.15
168	2	0.1	5	0.25
170	2	0.1	7	0.35
175	2	0.1	9	0.45
176	1	0.05	10	0.5
178	2	0.1	12	0.6
180	2	0.1	14	0.7
181	1	0.05	15	0.75
183	1	0.05	16	0.8
184	1	0.05	17	0.85
185	1	0.05	18	0.9
186	1	0.05	19	0.95
187	1	0.05	20	1
<hr/>				
$k = 15$	$n = 20$	1		

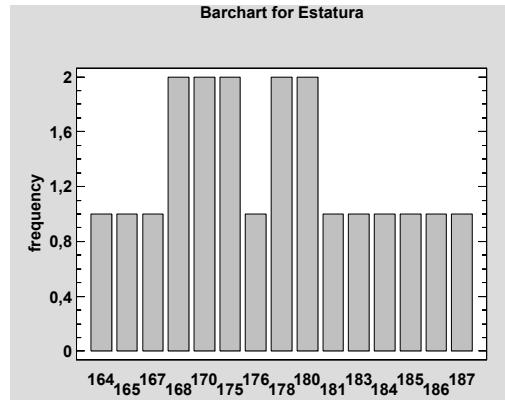


Figura A.3: Tabla de frecuencias y diagrama de barras de *Estatura*.

aleatoria X , que es la edad de una persona del grupo. La tabla de frecuencias y el diagrama de barras de X aparecen en la figura A.2. ◀

Cuando una variable toma muchos valores distintos, su tabla de frecuencias tiene demasiadas filas para ser operativa. Además, su diagrama de barras tiene muchas barras, cuyas alturas tienden a ser similares.

Ejemplo A.2.3 (Frecuencias y barras de *Estatura*)

Volviendo al ejemplo A.1.2, la tabla de frecuencias y el diagrama de barras de la variable aleatoria Y , que es la estatura de una persona del grupo, aparecen en la figura A.3. Esta variable toma 15 valores distintos, por lo que la su tabla de frecuencias es poco operativa. ◀

<i>Estatura</i>	n_i	f_i	N_i	F_i
(160, 165]	2	0.1	2	0.1
(165, 170]	5	0.25	7	0.35
(170, 175]	2	0.1	9	0.45
(175, 180]	5	0.25	14	0.7
(180, 185]	4	0.2	18	0.9
(185, 190]	2	0.1	20	1
$m = 6$	$n = 20$	1		

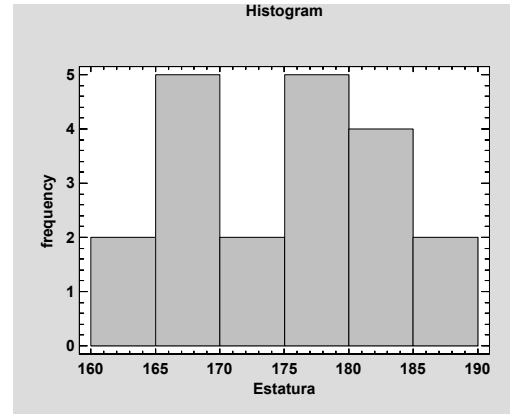


Figura A.4: Tabla de frecuencias e histograma de *Estatura*.

Los problemas detectados en el ejemplo anterior se resuelven agrupando los valores de la variable en intervalos de clase. Esto requiere que dicha variable sea cuantitativa.

Definición A.2.4 (Intervalos de clase e histograma)

Dada una variable cuantitativa X , definida sobre un espacio muestral de tamaño n , una agrupación de X en *intervalos de clase* es un partición de $\text{Im}(X)$ de la forma:

$$\text{Im}(X) \subseteq (a_0, a_1] \cup (a_1, a_2] \cup \cdots \cup (a_{i-1}, a_i] \cup \cdots \cup (a_{m-2}, a_{m-1}] \cup (a_{m-1}, a_m],$$

donde $-\infty < a_0 < a_1 < a_2 < \cdots < a_i < \cdots < a_m < \infty$. La frecuencia de cada intervalo es la suma de las frecuencias de los valores de X pertenecientes al mismo.

Si todos los intervalos tienen la misma longitud, entonces el *histograma* de X se construye levantando sobre cada intervalo, $(a_{i-1}, a_i]$, un rectángulo con base dicho intervalo y altura su frecuencia. ◀

Ejemplo A.2.5 (Histograma de *Estatura*)

En el grupo de personas del ejemplo A.1.2, la variable *Estatura* toma 15 valores distintos, comprendidos entre 164 y 187. Como son muchos, conviene agruparlos en intervalos. Si se divide el intervalo $(160, 190]$ en subintervalos de amplitud 5, se obtienen la tabla de frecuencias y el histograma de la figura A.4. ◀

A.3 Medidas de una variable

A.3.1 Medidas de posición

Se recuerda aquí la definición de media y se definen otras medidas de posición.

Definición A.3.1 (Medidas de posición)

Sean $X : E \rightarrow \mathbb{R}$ una variable aleatoria definida sobre un espacio muestral de tamaño n . Si $[x_1, n_1], \dots, [x_k, n_k]$ es la distribución de frecuencias de X , se definen:

- *Media* de X (en STATGRAPHICS, *average*): es la suma de los valores de la variable dividida por el número de ellos, es decir:

$$\mu(X) := \sum_{i=1}^k x_i p(X = x_i) = \frac{x_1 n_1 + \cdots + x_k n_k}{n}.$$

- *Mediana* de X (en STATGRAPHICS, *median*), $me(X)$: es un número mayor o igual que la mitad de los valores de X y menor o igual que el resto.
- *Primer cuartil* de X (en STATGRAPHICS, *lower quartile*), $Q_1(X)$: es un número mayor o igual que $1/4$ de los valores de X y menor o igual que el resto.
- *Tercer cuartil* de X (en STATGRAPHICS, *upper quartile*), $Q_3(X)$: es un número mayor o igual que $3/4$ de los valores de X y menor o igual que el resto.

Todas las medidas de posición de X se expresan en las mismas unidades que X . ◀

Ejemplo A.3.2 (Medidas de posición de *Edad*)

En el grupo de personas del ejemplo A.1.2, las medidas de posición de $X = \textit{Edad}$ son las siguientes:

- La media vale:

$$\mu(X) = \frac{19 \cdot 2 + 20 \cdot 10 + 21 \cdot 5 + 23 \cdot 2 + 25 \cdot 1}{20} = \frac{207}{10} = 20.7.$$

- Para calcular la mediana, se ordenan los $n = 20$ valores de X de menor a mayor y se busca el que ocupa la posición central:

$$\overbrace{19, 19, 20, 20, 20, 20, 20, 20, 20, 20}^{n/2=10}, \overbrace{20, 20, 21, 21, 21, 21, 21, 21, 23, 23, 25}^{n/2=10}.$$

- Cuando n es impar, solo hay un valor central y la mediana es dicho valor.
- Cuando n es par, como en este caso, hay 2 valores centrales y la mediana es su semisuma: $me(X) = \frac{20+20}{2} = 20$.
- El primer cuartil Q_1 debe ser mayor o igual que $1/4$ de los valores de X y menor o igual que el resto. En este caso, se tiene:

$$\overbrace{19, 19, 20, 20, 20}^{n/4=5}, \overbrace{20, 20, 20, 20, 20}^{n/4=5}, \overbrace{20, 20, 21, 21, 21}^{n/4=5}, \overbrace{21, 21, 23, 23, 25}^{n/4=5}.$$

Por tanto, $Q_1(X) = \frac{20+20}{2} = 20$. Análogamente, $Q_3(X) = \frac{21+21}{2} = 21$. ◀

A.3.2 Medidas de dispersión

Para complementar las medidas de posición, se definen las *medidas de dispersión*, que indican cuánto se apartan los valores de la variable de sus promedios y permiten valorar la representatividad de éstos. Recordamos aquí sus definiciones.

Definición A.3.3 (Medidas de dispersión)

Sea $X : E \rightarrow \mathbb{R}$ una variable aleatoria definida sobre un espacio muestral de tamaño n . Si $[x_1, n_1], \dots, [x_k, n_k]$ es la distribución de frecuencias de X , se definen:

- La *varianza* de X es la media de los cuadrados de las desviaciones respecto de $\mu(X)$:

$$\sigma^2(X) := \mu((X - \mu(X))^2) = \frac{1}{n} \sum_{i=1}^k (x_i - \mu(X))^2 n_i.$$

Es más cómodo calcular la varianza como la media del cuadrado menos el cuadrado de la media:

$$\sigma^2(X) = \mu(X^2) - \mu(X)^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - \mu(X)^2.$$

La *cuasivarianza* de X (en STATGRAPHICS, *variance*) se define para $n \geq 2$ por la fórmula:

$$S^2(X) := \frac{1}{n-1} \sum_{i=1}^k (x_i - \mu(X))^2 n_i = \frac{n}{n-1} \sigma^2(X).$$

La varianza y la cuasivarianza se expresan en el cuadrado de las unidades de X .

- La *desviación típica* de X es la raíz cuadrada de la varianza y la *cuasidesviación típica* (en STATGRAPHICS, *standard deviation*) es la raíz cuadrada de la cuasivarianza:

$$\sigma(X) := \sqrt{\sigma^2(X)} \quad \text{y} \quad S(X) := \sqrt{S^2(X)} \quad \text{para } n \geq 2.$$

Ambas tienen las mismas unidades que X y miden la dispersión absoluta de la variable.

- El *coeficiente de variación* de X es el cociente entre la desviación típica y el valor absoluto de la media (supuesto que $\mu(X) \neq 0$):

$$\text{CV}(X) := \frac{\sigma(X)}{|\mu(X)|} \quad \text{y} \quad \text{coeff. of variation}(X) := \frac{S(X)}{|\mu(X)|} \quad \text{para } n \geq 2.$$

El *cuasicoeficiente de variación* de X (en STATGRAPHICS, *coeff. of variation*) divide la cuasidesviación típica entre $|\mu(X)|$. Ambos coeficientes son adimensionales y miden la dispersión relativa de la variable, por lo que permiten comparar la dispersión de dos variables¹. ◀

Ejemplo A.3.4 (Medidas de dispersión de *Edad*)

En el grupo de personas del ejemplo A.1.2, la variable $X = \text{Edad}$ tiene media 207/10 y sus medidas de dispersión valen:

$$\sigma^2(X) = \frac{19^2 \cdot 2 + 20^2 \cdot 10 + 21^2 \cdot 5 + 23^2 \cdot 2 + 25^2 \cdot 1}{20} - \left(\frac{207}{10}\right)^2 = \frac{201}{100}.$$

¹En general, $\sigma^2(X) < S^2(X)$, $\sigma(X) < S(X)$ y $\text{CV}(X) < \text{coeff. of variation}(X)$, pero prácticamente coinciden cuando n es grande.

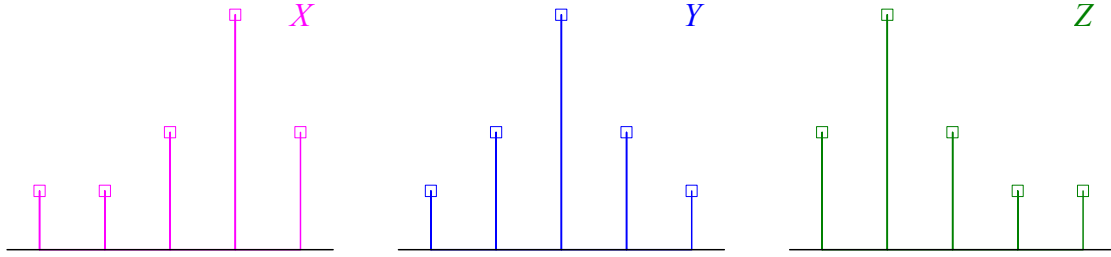


Figura A.5: X es asimétrica a la izquierda, Y es simétrica y Z es asimétrica a la derecha.

$$\sigma^2 = \frac{201}{100} \simeq 2.01, \quad \sigma = \sqrt{\frac{201}{100}} \simeq 1.41774, \quad CV = \frac{\sqrt{\frac{201}{100}}}{\frac{207}{10}} \simeq 6.84901\%,$$

$$S^2 = \frac{20}{19} \cdot \frac{201}{100} = \frac{201}{95} \simeq 2.11579, \quad S = \sqrt{\frac{201}{95}} \simeq 1.45458, \quad \frac{S}{|\mu|} = \frac{\sqrt{\frac{201}{95}}}{\frac{207}{10}} \simeq 7.02693\%.$$

Como el coeficiente de variación de *Edad* es pequeño, esta variable es poco dispersa respecto de su media. Por tanto, la media de *Edad*, que vale 20.7, representa bien a esta variable. ◀

A.3.3 Medidas de asimetría

Las *medidas de asimetría* se introducen para caracterizar la forma más o menos simétrica de la distribución de una variable, como ilustra la figura A.5.

Definición A.3.5 (Medida de asimetría)

Sea $X : E \rightarrow \mathbb{R}$ una variable aleatoria definida sobre un espacio muestral de tamaño n . Sea $[x_1, n_1], \dots, [x_k, n_k]$ la distribución de frecuencias de X . El *coeficiente de asimetría de Fisher*² de X es el cociente entre la media de los cubos de las desviaciones respecto de $\mu(X)$ y el cubo de la desviación típica:

$$\text{CAF}(X) := \frac{\mu(X - \mu(X))^3}{\sigma(X)^3} = \frac{1}{n \sigma(X)^3} \sum_{i=1}^k (x_i - \mu(X))^3 n_i.$$

Esta medida es adimensional y expresa lo siguiente:

- Si $\text{CAF}(X) < 0$, entonces la distribución es asimétrica a la izquierda.
- Si $\text{CAF}(X) = 0$, entonces la distribución es aproximadamente simétrica.

2



Ronald Aylmer Fisher (London, England 1890 – Adelaide, Australia, 1962).

Estadístico, biólogo y genetista. Fisher estudió matemáticas y astronomía en Cambridge, aunque se interesó también por la biología.

Después de trabajar en una granja, para una compañía de inversiones y como profesor de matemáticas, Fisher obtuvo el puesto de estadístico en la *Rothamsted Agricultural Experiment Station*, el más prestigioso centro de investigación agrícola del Reino Unido. Allí realizó muchas contribuciones a la estadística, en particular al diseño y análisis de experimentos, y a la genética.

Fisher es autor de *Statistical Methods for Research Workers* (1925), *The design of experiments* (1935) y *Statistical tables* (1947) y creador del método de máxima verosimilitud.

- Si $\text{CAF}(X) > 0$, entonces la distribución es asimétrica a la derecha.

En lugar del coeficiente de asimetría de Fisher, STATGRAPHICS calcula otra medida de asimetría, llamada *skewness*, y que se define para $n \geq 3$ por:

$$\text{skewness}(X) := \frac{n}{(n-1)(n-2)S^3} \sum_{i=1}^k (x_i - \mu(X))^3 n_i = \frac{\sqrt{n(n-1)}}{n-2} \text{CAF}(X).$$

Cuando n es grande, $\text{skewness}(X) \simeq \text{CAF}(X)$. ◀

Ejemplo A.3.6 (Medida de asimetría de *Edad*)

En el grupo de personas del ejemplo A.1.2, las medidas de asimetría de $X = \text{Edad}$ valen:

$$\begin{aligned} \text{CAF}(X) &= \frac{(19 - \frac{207}{10})^3 2 + (20 - \frac{207}{10})^3 10 + (21 - \frac{207}{10})^3 5 + (23 - \frac{207}{10})^3 2 + (25 - \frac{207}{10})^3}{20\sqrt{201/100}^3} = \\ &= \frac{\frac{2268}{25}}{20\sqrt{\frac{201}{100}}^3} \simeq 1.59177 > 0, \\ \text{skewness}(X) &= \frac{20}{19 \cdot 18\sqrt{\frac{201}{95}}^3} \cdot \frac{2268}{25} \simeq 1.72385 > 0. \end{aligned}$$

En ambos casos, el signo positivo indica que la distribución es asimétrica a la derecha, tal como se observa en el diagrama de barras (figura A.2). ◀

Ejemplo A.3.7 (Medidas de la variable *Estatura*)

En el grupo de personas del ejemplo A.1.2, las medidas de la variable $Y = \text{Estatura}$ son las siguientes:

- La media vale:

$$\mu(Y) = \frac{164 + 165 + 167 + 168 \cdot 2 + 170 \cdot 2 + 175 \cdot 2 + \dots + 186 + 187}{20} = 176.$$

- Para calcular la mediana y los cuartiles, se ordenan los $n = 20$ valores de menor a mayor:

$$\overbrace{164, 165, 167, 168, 168}^{n/4=5}, \overbrace{170, 170, 175, 175, 176}^{n/4=5}, \overbrace{178, 178, 180, 180, 181}^{n/4=5}, \overbrace{183, 184, \dots, 187}^{n/4=5}.$$

- $\text{me}(Y) = \frac{176+178}{2} = 177$ indica que el 50 % de las personas mide 177 cm o menos y el otro 50 % mide 177 cm o más.
- $Q_1(Y) = \frac{168+170}{2} = 169$ indica que el 25 % de las personas mide 169 cm o menos y el otro 75 % mide 169 cm o más.
- $Q_3(Y) = \frac{181+183}{2} = 182$ indica que el 75 % de las personas mide 182 cm o menos y el otro 25 % mide 182 cm o más.

- Medidas de dispersión:

$$\sigma^2(Y) = \frac{164^2 + 165^2 + \dots + 186^2 + 187^2}{20} - 176^2 = \frac{155\,137}{5} - 176^2 = \frac{257}{5}.$$

$$\sigma^2 = \frac{257}{5} = 51.4, \quad \sigma = \sqrt{\frac{257}{5}} \simeq 7.17966, \quad CV = \frac{\sqrt{\frac{257}{5}}}{176} \simeq 4.08051\%,$$

$$S^2 = \frac{20}{19} \cdot \frac{257}{5} \simeq 54.1053, \quad S = \sqrt{\frac{1028}{19}} \simeq 7.35563, \quad \frac{S}{|\mu|} = \frac{\sqrt{\frac{1028}{19}}}{176} \simeq 4.17933\%.$$

- Como el coeficiente de variación es pequeño, *Estatuta* es poco dispersa respecto de su media. Por tanto, la media de *Estatuta*, que vale 176 cm, representa bien a esta variable.
- Al comparar los coeficientes de variación de *Estatuta* y *Edad*, resulta:

$$CV(Estatuta) \simeq 4.17933\% < 7.02693\% \simeq CV(Edad),$$

luego la variable *Estatuta* es menos dispersa que la variable *Edad*.

- Medidas de asimetría:

$$CAF(Y) = \frac{(164 - 176)^3 + \dots + (187 - 176)^3}{20\sqrt{\frac{257}{5}}^3} = \frac{-1062}{20\sqrt{\frac{257}{5}}^3} \simeq -0.12353 < 0,$$

$$\text{skewness}(Y) = \frac{20}{19 \cdot 18\sqrt{\frac{1028}{19}}^3} \cdot (-1062) \simeq -0.156052 < 0.$$

El valor negativo, pero próximo a 0, indica que la distribución de Y es ligeramente asimétrica a la izquierda, como se observa en su histograma (figura A.4). ◀

A.4 Diagramas de caja y bigotes

El diagrama de caja (box-and-whisker plot) de una variable numérica es una representación gráfica que incluye varias de sus medidas y revela los valores atípicos.

Definición A.4.1 (Diagrama de caja)

Sea $X : E \rightarrow \mathbb{R}$ una variable aleatoria definida sobre un espacio muestral de tamaño n . El *diagrama de caja* de X se construye del siguiente modo:

1. Sobre un eje horizontal, se dibuja un rectángulo cuyos extremos son los cuartiles Q_1 y Q_3 .
2. Sobre el eje, se marca la mediana, me , con un trazo vertical y la media, $\mu(X)$, con un punto, que puede quedar fuera del rectángulo.
3. Se calcula el *intervalo de admisión*, que es:

$$\left[Q_1 - \frac{3}{2}(Q_3 - Q_1), Q_3 + \frac{3}{2}(Q_3 - Q_1) \right].$$

Cualquier valor de X situado fuera de dicho intervalo se considera *atípico*.

4. Sobre el eje, se traza un segmento que conecta el menor valor típico con Q_1 y otro segmento que conecta Q_3 con el mayor valor típico.

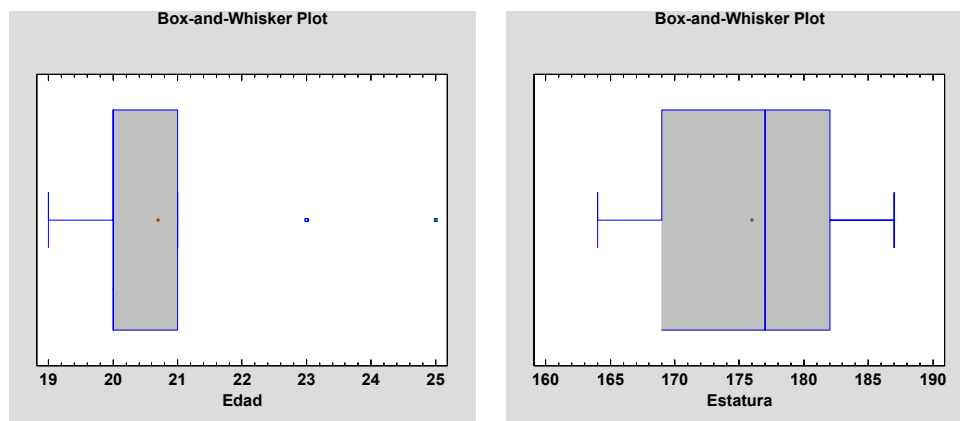


Figura A.6: Diagramas de caja de *Edad* y *Estatura*.

5. Se marcan los valores atípicos. ◀

Ejemplo A.4.2 (Diagramas de caja de *Edad* y *Estatura*)

En el grupo de personas del ejemplo A.1.2, el diagrama de caja de $X = \textit{Edad}$ (figura A.6) se construye teniendo en cuenta que:

$$Q_1(X) = 20, \quad Q_3(X) = 21, \quad \text{me}(X) = 20 \quad \text{y} \quad \mu(X) = 20.7.$$

El intervalo de admisión es:

$$\left[20 - \frac{3}{2}(21 - 20), 21 + \frac{3}{2}(21 - 20) \right] = [18.5, 22.5],$$

luego el menor valor típico es 19 y el mayor es 21. Hay dos valores atípicos, que son 23 y 25.

Para construir el diagrama de caja de $Y = \textit{Estatura}$ (figura A.6), se tiene en cuenta que:

$$Q_1(Y) = 169, \quad Q_3(Y) = 182, \quad \text{me}(Y) = 177 \quad \text{y} \quad \mu(Y) = 176.$$

El intervalo de admisión es:

$$\left[169 - \frac{3}{2}(182 - 169), 182 + \frac{3}{2}(182 - 169) \right] = [149.5, 201.5],$$

luego el menor valor típico es 164 y el mayor 187. No hay valores atípicos. ◀