

# Capítulo 1

## Introducción

### 1.1. Estadísticos

En muchas ocasiones se plantea como objetivo de la inferencia determinar cierto valor (**parámetro**),  $\theta$ , asociado al comportamiento en la población de una característica de interés,  $X$ , como la edad, el salario, la situación laboral, el empleo, etc.

Entre los parámetros de interés suelen destacar, por su gran uso: la media, la probabilidad de “éxito”, la varianza...

Usualmente, se dirá que la variable  $X$  sigue una distribución de probabilidad  $F_\theta$ , dependiente del parámetro  $\theta$ . Este último varía en el **espacio paramétrico**  $\Theta$ , es decir,  $\theta \in \Theta \subseteq \mathbb{R}^k$ .

El procedimiento habitual para aproximar el valor del parámetro es recoger información sobre el comportamiento de la variable  $X$  mediante una *muestra aleatoria* de esa variable,  $(X_1, \dots, X_n)$  siendo  $n$  el tamaño muestral, y resumir la información recogida sobre el comportamiento de la variable  $X$  utilizando diferentes transformaciones medibles, o **estadísticos**.

**Definición 1.1.1.** *Un **estadístico** es cualquier aplicación medible*

$$T : (X_1, \dots, X_n) \longrightarrow \mathbb{R}^p$$

con  $p \leq n$ .

Con el manejo de estadísticos se pretende simplificar la estructura del problema, pasando a trabajar en un espacio de menor dimensión. Los estadísticos permiten trasladar la distribución de probabilidades muestral, sobre  $\mathbb{R}^n$ , a un espacio de dimensión menor,  $\mathbb{R}^p$ . Generalmente en este curso se considera  $p = 1$  o  $p = 2$ .

El manejo de los estadísticos, que son variables aleatorias (v.a.), y sus propiedades necesitará del cálculo de la distribución de los mismos. Para ello, se recurrirá al uso de propiedades como la **reproductividad**, propiedad asociada a la suma de variables independientes con un determinado tipo de distribución, como Binomial, Poisson, Normal, Gamma, o al uso de nuevas variables aleatorias obtenidas por **transformaciones funcionales**. En estas últimas, su comportamiento se caracterizará por la función de distribución o, en el caso continuo, se utilizará el Teorema del Cambio de Variable para calcular su función de densidad.

**Teorema 1.1.1. (de cambio de variable)** Sea  $X$  una v.a. (unidimensional) continua con soporte  $(a, b)$  y función de densidad  $f_X(x)$ . Sea  $g$  una función diferenciable en  $(a, b) \subset \mathbb{R}$  con derivada distinta de cero (es decir, inyectiva) entonces la función de densidad de la variable  $Y = g(X)$  viene dada por:

$$f_Y(y) = f_X(g^{-1}(y)) |(g^{-1})'(y)| \quad y \in (\min\{g(a), g(b)\}, \max\{g(a), g(b)\}).$$

Así, por ejemplo, si  $X \equiv \mathcal{U}(0, 1)$  y considerando la transformación de  $X$  por la función  $g(x) = -\log(x)$ , es decir,  $Y = -\log(X)$ , definida en el intervalo  $(0, 1)$ , entonces  $Y$  tiene función de densidad

$$f_Y(y) = f_X(e^{-y}) |-e^{-y}| = 1 e^{-y} = e^{-y}, \quad y \in (0, \infty).$$

Obteniéndose, en este caso, una función de densidad conocida;  $Y \equiv \exp(1)$ .

Las condiciones exigidas a la función de transformación  $g$  considerada en el teorema (derivabilidad y monotonía estricta) equivalen a la exigencia de tales condiciones a la función  $g^{-1}$ . Este teorema es válido en condiciones más generales: puede extenderse al caso en que la función  $g$  no sea inyectiva pero su soporte,  $S$ , se pueda dividir en un conjunto numerable de subconjuntos disjuntos  $A_i$  (es decir,  $S = \cup_i A_i$ ,  $i \in I$  con  $I$  numerable, y  $A_i \cap A_j = \emptyset \quad \forall i \neq j$ ), donde  $g$  tenga inversa,  $g_i^{-1}$  para cada  $i \in I$ , y cada una de ellas satisfaga condiciones del enunciado del teorema anterior. En este caso, la función de densidad de  $Y = g(X)$  es:

$$f_Y(y) = \sum_{i \in I} f_X(g_i^{-1}(y)) |(g_i^{-1})'(y)|, \quad y \in g(S)$$

Así, si  $X \equiv \mathcal{N}(0, 1)$  y se quiere calcular la distribución de su cuadrado  $Y = X^2$ , se pueden considerar las funciones  $g_1(x) = x^2$  si  $x > 0$  y  $g_2(x) = x^2$  si  $x \leq 0$ , con inversas respectivas  $g_1^{-1}(y) = \sqrt{y}$  para  $y > 0$  y  $g_2^{-1}(y) = -\sqrt{y}$  para  $y \geq 0$ , de donde

$$f_Y(y) = \frac{f_{\mathcal{N}(0,1)}(\sqrt{y}) + f_{\mathcal{N}(0,1)}(-\sqrt{y})}{2\sqrt{y}}, \quad y \in (0, \infty).$$

Es decir,

$$f_Y(y) = \frac{2e^{-y/2}}{\sqrt{2\pi}} \frac{I(y > 0)}{2\sqrt{y}} = \frac{(1/2)^{1/2}}{\sqrt{\pi}} e^{-y/2} y^{-\frac{1}{2}} I(y > 0)$$

donde  $I_A(y) = I(y \in A)$  es la **función indicatriz** que toma el valor 1 si  $y \in A$  y 0 en caso contrario. Esto es,  $Y$  sigue una distribución gamma  $Y \equiv \gamma(\frac{1}{2}, \frac{1}{2}) \equiv \chi_1^2$ .

Entre todas las transformaciones posibles de una variable, una ampliamente usada es la tipificación:

**Definición 1.1.2.** Sea  $X$  una v.a. con media  $\mu$  y varianza  $\sigma^2$  finitas, se define su **variable tipificada** como  $Z_X = \frac{X - \mu}{\sigma}$ . Esta transformación da lugar a una variable adimensional, con media 0 y desviación típica 1.

Otra transformación importante de una v.a. continua es la transformación por su función de distribución, que da lugar a una variable  $\mathcal{U}(0, 1)$ :

**Teorema 1.1.2.** Sea  $X$  una v.a. absolutamente continua, con función de distribución  $F(x)$ . La v.a  $Y = F(X)$  tiene distribución  $\mathcal{U}(0, 1)$ .

## 1.2. Momentos muestrales

Los momentos muestrales son estadísticos ampliamente utilizados en la inferencia estadística, en particular los momentos muestrales de orden 1 y 2, aunque también se pueden utilizar momentos de orden superior.

**Definición 1.2.1.** Dada una muestra aleatoria simple  $(X_1, \dots, X_n)$  de una variable aleatoria  $X$ , se definen los estadísticos:

▪ **momento muestral de orden  $r$**  como  $\overline{X^r} = \sum_{i=1}^n X_i^r / n$ ;

▪ **momento muestral centrado de orden  $r$**  como

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

**Definición 1.2.2.** Dada una muestra aleatoria simple  $(X_1, \dots, X_n)$  de una variable aleatoria  $X$  con media  $\mu$  y varianza  $\sigma^2$ , se consideran los estadísticos **media muestral**, **varianza muestral** y **cuasivarianza muestral**, definidos respectivamente como

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}; \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}; \quad \hat{S}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}.$$

Entre las propiedades de estos estadísticos se deben resaltar las siguientes:

- [1]  $E(\bar{X}) = \mu; \quad \text{Var}(\bar{X}) = \sigma^2/n.$
- [2] Si  $X \equiv \mathcal{N}(\mu, \sigma)$  entonces  $\bar{X} \equiv \mathcal{N}(\mu, \sigma/\sqrt{n}).$
- [3] **Teorema Central del Límite:** Sea  $X_1, X_2, \dots, X_n, \dots$  una sucesión de variables aleatorias, independientes e idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$  distinta de cero, entonces el estadístico **suma muestral**,  $S_n = X_1 + \dots + X_n$ , verifica:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

donde  $\xrightarrow{\mathcal{L}}$  denota la *convergencia en ley* o distribución. El resultado anterior equivale a:

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

- [4] Convergencias asintóticas: Dada  $X_1, X_2, \dots, X_n, \dots$  una sucesión de variables aleatorias, independientes e idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$  distinta de cero, se verifica que:

$$[4.1] \quad \bar{X} \xrightarrow{P} \mu; \quad [4.2] \quad \bar{X} \xrightarrow{c.s.} \mu; \quad [4.3] \quad \sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma);$$

donde  $\xrightarrow{P}$  denota la *convergencia en probabilidad*,  $\xrightarrow{c.s.}$  la *convergencia casi segura* y  $\xrightarrow{\mathcal{L}}$  la *convergencia en ley* o distribución.

- [5]  $E(S^2) = (n - 1) \sigma^2/n.$

Es decir, la esperanza de  $S^2$  no es la varianza poblacional,  $\sigma^2$ , por ello en los cálculos con datos, sobre todo en los programas informáticos de estadística, suele ser muy usual trabajar con una modificación de la varianza: la **cuasivarianza muestral**. Esta también es llamada por muchos autores varianza muestral y se define como

$$\hat{S}^2 = \frac{n}{n-1} S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1},$$

cuya esperanza es  $E(\hat{S}^2) = \sigma^2$ .

- [6] Cuando  $X$  sigue una distribución  $\mathcal{N}(\mu, \sigma^2)$  se tiene que  $\frac{n-1}{\sigma^2} \hat{S}^2 \equiv \chi_{n-1}^2$ .
- [7] Cuando  $X$  sigue una distribución  $\mathcal{N}(\mu, \sigma^2)$  entonces  $\bar{X}$  y  $S^2$  son independientes. Además, por [2] y [6], el estadístico

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\hat{S}}$$

sigue una distribución *t*-Student con  $n - 1$  grados de libertad, es decir,  $T \equiv t_{n-1}$ .

Los estadísticos media muestral y cuasivarianza muestral serán muy utilizados en la inferencia, por verificar estas propiedades, para aproximar valores de los parámetros  $\mu$  y  $\sigma^2$ , respectivamente, asociados al comportamiento de una variable aleatoria  $X$ .

Los momentos muestrales son base para la estimación por el método de los momentos, que se estudiará posteriormente.

### 1.3. Estadísticos ordenados

Los estadísticos ordenados juegan un papel importante en el cálculo de la función de distribución empírica, una primera aproximación a la función de distribución a partir de la información muestral, y por tanto en el estudio de sus propiedades.

Por otra parte, cada vez es más frecuente trabajar con problemas en los que no es admisible la hipótesis de normalidad, o simplemente no se dispone de información precisa sobre la distribución de la población. En tales circunstancias es muy útil recurrir a la Inferencia no paramétrica, que trabaja con estadísticos como el mínimo, el máximo, la mediana, etc., que forman parte de la familia de los estadísticos ordenados.

Cuando se trabaja con una variable aleatoria  $X \equiv \mathcal{U}(0, \theta)$ , parece claro que el parámetro de interés es el mayor valor de la variable, por lo que es bastante intuitivo considerar el máximo de la muestra como una primera aproximación, es decir, se considera el estadístico máximo de la muestra:  $X_{(n)}$ .

**Definición 1.3.1.** *Dada una m.a.s.  $(X_1, \dots, X_n)$  se pueden ordenar sus componentes de forma no decreciente, obteniéndose  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ . A la aplicación*

$$\begin{aligned} X_{(\cdot)} : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (x_1, \dots, x_n) &\longrightarrow (x_{(1)}, \dots, x_{(n)}) \end{aligned}$$

*se le llama **muestra ordenada**.*

*Esta aplicación no es unívoca ya que cada imagen  $(x_{(1)}, \dots, x_{(n)})$  tiene, salvo cuando hay observaciones empatadas,  $n!$  contra-imágenes.*

**Definición 1.3.2.** *Se llama **estadístico ordenado de orden  $k$**  a la aplicación*

$$\begin{aligned} X_{(k)} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\longrightarrow x_{(k)} \end{aligned}$$

*que nos da la  $k$ -ésima componente de la muestra ordenada.*

Por ejemplo,  $X_{(1)} = \min\{X_{(1)}, \dots, X_{(n)}\}$  y  $X_{(n)} = \max\{X_{(1)}, \dots, X_{(n)}\}$ .

Una observación importante es que aunque las  $n$  observaciones muestrales  $(X_1, \dots, X_n)$  sean independientes, el estadístico ordenado  $(X_{(1)}, \dots, X_{(n)})$  no tiene sus componentes independientes ni tienen la misma distribución.

El cálculo de las distribuciones del máximo y el mínimo constituye un buen acercamiento a las técnicas para obtener las distribuciones de estos estadísticos, que se continuará con el cálculo de la distribución conjunta de los estadísticos ordenados y de las distribuciones marginales de una o varias componentes.

**Teorema 1.3.1.** *Sea  $X$  una v.a. con función de distribución  $F$  de la que se obtiene una m.a.s. de tamaño  $n$ . Entonces se verifica:*

- 1)  $F_{X_{(n)}}(x) = [F(x)]^n;$
- 2)  $F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n;$
- 3)  $F_{X_{(k)}}(x) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}.$

**Demostración.**

- 1)  $F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_i \leq x \ \forall i) = P((X_1 \leq x) \cap \dots \cap (X_n \leq x))$   
 $= \prod_{i=1}^n P(X_i \leq x) = [F(x)]^n.$
- 2)  $F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_i > x \ \forall i)$   
 $= 1 - P((X_1 > x) \cap \dots \cap (X_n > x)) = 1 - \prod_{i=1}^n P(X_i > x) = 1 - [1 - F(x)]^n.$
- 3)  $F_{X_{(k)}}(x) = P(X_{(k)} \leq x) = P(\text{al menos } k \text{ valores muestrales sean } \leq x)$   
 $= \sum_{j=k}^n P((\text{n}^\circ \text{ de observaciones muestrales } \leq x) = j)$   
 $= \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}.$

Así, por ejemplo en el caso  $X \equiv \mathcal{U}(0, 1)$ , se tiene que

$$F_{X_{(n)}}(y) = y^n I(0 < y < 1) + 1 I(y \geq 1) \quad \Rightarrow \quad X_{(n)} \equiv \beta(n, 1).$$

En consecuencia, si  $X \equiv \mathcal{U}(0, \theta)$ , se tiene que

$$F_{X_{(n)}}(y) = (y/\theta)^n I(0 < y < \theta) + 1 I(y \geq \theta) \quad \Rightarrow \quad X_{(n)}/\theta \equiv \beta(n, 1).$$

Otro estadístico importante que es función de los estadísticos ordenados es el **rango o recorrido** muestral, definido por  $R = X_{(n)} - X_{(1)}$ .

**Corolario 1.1.** *La función de densidad del estadístico ordenado  $X_{(k)}$  de una muestra aleatoria simple (m.a.s.) de tamaño  $n$  obtenida a partir de una v.a continua viene dado por la expresión*

$$f_{X_{(k)}}(x) = n! \frac{[F(x)]^{k-1}}{(k-1)!} f(x) \frac{[1 - F(x)]^{n-k}}{(n-k)!}.$$

**Demostración.** La demostración se obtiene por derivación de la función de distribución.

$$1) f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x).$$

$$2) f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x).$$

$$\begin{aligned} 3) f_{X_{(k)}}(x) &= \left( \sum_{j=k}^{n-1} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} + [F(x)]^n \right)' \\ &= n f(x) \left( \binom{n-1}{k-1} [F(x)]^{k-1} [1 - F(x)]^{n-k} - \binom{n-1}{k} [F(x)]^k [1 - F(x)]^{n-k-1} \right. \\ &\quad \left. + \binom{n-1}{k} [F(x)]^k [1 - F(x)]^{n-k-1} - \binom{n-1}{k+1} [F(x)]^{k+1} [1 - F(x)]^{n-k-2} \right. \\ &\quad \left. + \dots + [F(x)]^{n-1} \right) = \frac{n!}{(k-1)!(n-k)!} f(x) [F(x)]^{k-1} [1 - F(x)]^{n-k} \end{aligned}$$

**Teorema 1.3.2.** *Si  $X$  es una v.a. continua de la que se extrae una m.a.s. de tamaño  $n$ , la función de densidad conjunta de la muestra ordenada es*

$$g(x_{(1)}, \dots, x_{(n)}) = \begin{cases} n! f(x_1, \dots, x_n) & \text{si } x_{(1)} < \dots < x_{(n)} \\ 0 & \text{en otro caso} \end{cases}$$

## 1.4. Función de distribución empírica

La función de distribución empírica es un ejemplo de estimador funcional, ya que va a servir para aproximar no un parámetro sino la función de distribución de una variable.

**Definición 1.4.1.** Sea  $(X_1, \dots, X_n)$  una m.a.s. de una v.a.  $X$  con función de distribución  $F$ . Se define la **función de distribución empírica de  $X$  asociada a la m.a.s.  $(X_1, \dots, X_n)$**  como la función  $F_n : \mathbb{R} \rightarrow [0, 1]$  que asocia a cada  $x \in \mathbb{R}$  el valor

$$F_n(x) = \frac{\#(X_i \leq x)}{n} = \begin{cases} 0 & \text{si } x < X_{(1)} \\ i/n & \text{si } X_{(i)} \leq x < X_{(i+1)} \text{ para algún } i \in \{1, \dots, n-1\} \\ 1 & \text{si } x \geq X_{(n)} \end{cases}$$

donde  $\#A$  = cardinal de  $A$ .

Es decir, a cada realización muestral le asocia una función escalonada, que es función de distribución, y al cambiar la muestra, cambia la función (salvo que sólo cambie el orden de los datos).

#### 1.4.1. Propiedades

1. Fijado un valor  $x \in \mathbb{R}$ , la F. D. empírica puede expresarse alternativamente como

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) = \frac{1}{n} \sum_{i=1}^n I_{[X_i, \infty)}(x).$$

Esta última expresión es muy útil para demostrar propiedades.

Si para cada  $x \in \mathbb{R}$  fijo se definen las v.a.

$$Y_i = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases}$$

se tiene que  $P(Y_i = 1) = P(X_i \leq x) = F_{X_i}(x) = F(x)$ , es decir,  $Y_i$  tiene distribución Bernoulli de parámetro  $F(x)$ . Por lo tanto,  $\sum_{i=1}^n Y_i \sim \mathcal{B}(n, F(x))$ .

Para  $x \in \mathbb{R}$  fijo,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i,$$

con lo que  $F_n(x)$  es un estimador de  $F(x)$  con

$$E(F_n(x)) = E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \frac{1}{n} \sum_{i=1}^n F(x) = F(x);$$



$$\text{Var}(F_n(x)) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) = \frac{F(x)(1-F(x))}{n}.$$

Es interesante señalar que para cualquier  $x$  fijo, la función de distribución empírica es ese punto  $F_n(x)$  es un estimador insesgado de  $F(x)$ , pero que su varianza es tanto mayor cuanto más se aproxime  $F(x)$  a 0.5.

2. La Ley Fuerte de los Grandes Números nos asegura

$$F_n(x) \xrightarrow{c.s.} F(x) \quad \text{para cada } x \in \mathbb{R}.$$

3. Por el Teorema Central del Límite se tiene que

$$\frac{F_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{para cada } x \in \mathbb{R}.$$

#### 1.4.2. Teorema de Glivenko-Cantelli

El Teorema de Glivenko-Cantelli establece de manera rigurosa la conexión entre las distribuciones teórica y la muestral (o empírica) al garantizar la convergencia uniforme, casi segura, de la función de distribución empírica a la teórica, pero no establece cuál es su velocidad de convergencia. En su enunciado, utiliza la distancia  $d_\infty$  entre funciones. En temas posteriores, se estudia su comportamiento para distribuciones en el caso continuo.

**Teorema 1.4.1. (de Glivenko-Cantelli)** Sea  $(X_n)_{n \in \mathbb{N}}$  una sucesión de v.a. independientes e igualmente distribuidas, con función de distribución común  $F$ . Sean  $(X_1, \dots, X_n)$  las  $n$  primeras componentes de la sucesión anterior (es decir, una m.a.s. de  $X$ ) y  $F_n$  el funcional función de distribución empírica. Consideremos la aplicación

$$\begin{aligned} \Delta_n : (X_1, \dots, X_n) &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\longrightarrow D_\infty(F_n, F) = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \end{aligned}$$

entonces se verifica que

$$\Delta_n \xrightarrow{c.s.} 0.$$

#### Demostración.

Para cada  $x \in \mathbb{R}$  fijo se definen las variables aleatorias  $Y_i = I_{(-\infty, x]}(X_i)$  con  $i = 1, \dots, n$  donde  $I_A(\cdot)$  es la función indicatriz. Por definición, se sabe que las  $Y_i$  siguen distribución Bernoulli con parámetro  $P(X_i \leq x)$ , es decir,  $F(x)$ . Así, se verifican las condiciones de la Ley Fuerte de los Grandes Números:

- $E(Y_i) = F(x) \in \mathbb{R}$ ;
- $\text{Var}(Y_i) = F(x) [1 - F(x)] \in \mathbb{R}$ ; y
- $\sum_{i=1}^{\infty} \frac{\text{Var}(Y_i)}{i^2} = \sum_{i=1}^{\infty} \frac{F(x) [1 - F(x)]}{i^2} = F(x) [1 - F(x)] \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$ .

Por tanto, se cumple que  $\bar{Y}_n \xrightarrow{c.s.} F(x)$ , es decir,  $F_n(x) \xrightarrow{c.s.} F(x)$ . En particular,

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \xrightarrow{c.s.} F(x) \\ F_n(x^-) &= \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x)}(X_i) \xrightarrow{c.s.} F(x^-) \end{aligned} \quad (1.1)$$

donde  $F(x^-) = \lim_{c \rightarrow x^-} F(c)$ .

Para cada  $k \in \mathbb{N}$  y  $j = 1, \dots, k$ , definimos  $x_{j,k}$  el menor cuantil poblacional de orden  $j/k$ , es decir,  $x_{j,k}$  es el primer valor en el cual la función de distribución,  $F$ , sobrepasa o iguala  $j/k$ :

$$x_{j,k} = \min \left\{ x \in \mathbb{R} : F(x) \geq \frac{j}{k} \right\}.$$

Por lo tanto, en particular por (1.1), se verifica que

$$F_n(x_{j,k}) \xrightarrow{c.s.} F(x_{j,k}) \quad \text{y} \quad F_n(x_{j,k}^-) \xrightarrow{c.s.} F(x_{j,k}^-). \quad (1.2)$$

Se consideran las sucesiones del espacio muestral  $(X_1, \dots, X_n, \dots) = (X_n)_{n \in \mathbb{N}}$ , y se denota por  $w$  una sucesión particular,  $w = (x_n)_{n \in \mathbb{N}}$ . Se definen:

$$\begin{aligned} A_{j,k} &= \{w : \lim_{n \rightarrow \infty} F_n(x_{j,k}, w) = F(x_{j,k})\} \\ B_{j,k} &= \{w : \lim_{n \rightarrow \infty} F_n(x_{j,k}^-, w) = F(x_{j,k}^-)\} \end{aligned} \quad \text{y} \quad D_k = \bigcap_{j=1}^k (A_{j,k} \cap B_{j,k}).$$

Por (1.2), se tiene  $P(A_{j,k}) = P(B_{j,k}) = 1$  y, por las propiedades de la probabilidad,

$$\begin{aligned} P(D_k) &= 1 - P\left(\bigcup_{j=1}^k (A_{j,k}^c \cup B_{j,k}^c)\right) \geq 1 - \sum_{j=1}^k P(A_{j,k}^c \cup B_{j,k}^c) \\ &= 1 - \sum_{j=1}^k [P(A_{j,k}^c) + P(B_{j,k}^c) - P(A_{j,k}^c \cap B_{j,k}^c)] = 1. \end{aligned}$$

Sea

$$D = \bigcap_{k \in \mathbb{N}} D_k = \{w : \lim_{n \rightarrow \infty} F_n(x_{j,k}, w) = F(x_{j,k}) \text{ y } \lim_{n \rightarrow \infty} F_n(x_{j,k}^-, w) = F(x_{j,k}^-) \\ \forall x_{j,k} \text{ simultáneamente con } k \in \mathbb{N} \text{ y } j = 1, \dots, k\},$$

se verifica también  $P(D) = 1$  por ser  $D$  una intersección numerable de conjunto de probabilidad 1.

Para cada sucesión del espacio muestral,  $w = (x_n)_{n \in \mathbb{N}}$ , se construye la función de distribución empírica asociada a sus  $n$  primeras componentes,  $F_n$ , y se considera

$$\delta_n^k = \sup_{x_{j,k}} \left\{ |F_n(x_{j,k}) - F(x_{j,k})|, |F_n(x_{j,k}^-) - F(x_{j,k}^-)| \right\}.$$

- Obsérvese ahora que si  $x \in [x_{j,k}, x_{j+1,k})$ , se tiene que

$$F_n(x_{j,k}) \leq F_n(x) \leq F_n(x_{j+1,k}^-); \quad (1.3)$$

$$\frac{j}{k} \leq F(x_{j,k}) \leq F(x) \leq F(x_{j+1,k}^-) \leq \frac{j+1}{k}. \quad (1.4)$$

Las desigualdades en (1.3) son ciertas por ser  $F_n$  no decreciente por definición de función de distribución empírica, mientras que las tres primeras desigualdades en (1.4) se deben, en este orden, a la definición de  $x_{j,k}$  y a que la función de distribución  $F$  es no decreciente por definición. Con respecto a la última desigualdad en (1.4), se razona por reducción al absurdo: si  $F(x_{j+1,k}^-) > \frac{j+1}{k}$  entonces, por definición de límite por la izquierda, existe un  $x < x_{j+1,k}$  tal que  $F(x) > \frac{j+1}{k}$ , lo cual contradice la definición de  $x_{j+1,k}$ .

En consecuencia, se verifican la siguientes acotaciones:

$$F_n(x_{j,k}) - F(x_{j+1,k}^-) \leq F_n(x) - F(x) \leq F_n(x_{j+1,k}^-) - F(x_{j,k}). \quad (1.5)$$

Además, dado que

$$0 \leq F(x_{j+1,k}^-) - F(x_{j,k}) \leq \frac{1}{k}$$

por la no-negatividad de la función de distribución, por ser  $F(x_{j+1,k}^-) \leq \frac{j+1}{k}$  por (1.4) y por cumplirse  $F(x_{j,k}) \geq \frac{j}{k}$  por definición de  $x_{j,k}$ , se obtiene

$$F_n(x_{j+1,k}^-) - F(x_{j,k}) \leq F_n(x_{j+1,k}^-) + \frac{1}{k} - F(x_{j+1,k}^-) \leq \delta_n^k + \frac{1}{k} \quad \text{y} \\ F_n(x_{j,k}) - F(x_{j+1,k}^-) \geq F_n(x_{j,k}) - F(x_{j,k}) - \frac{1}{k} \geq -\delta_n^k - \frac{1}{k},$$

donde la última desigualdad se debe a la definición de  $\delta_n^k$ .

De lo anterior y por (1.5) se deduce que

$$-\delta_n^k - \frac{1}{k} \leq F_n(x) - F(x) \leq \delta_n^k + \frac{1}{k}.$$

Es decir, queda demostrado que, para cualquier  $k \in \mathbb{N}$  y  $j \in \{1, \dots, k-1\}$ ,

$$|F_n(x) - F(x)| \leq \delta_n^k + \frac{1}{k} \quad \text{para todo } x \in [x_{j,k}, x_{j+1,k}).$$

- Si  $x < x_{1,k}$ , razonando como en (1.4) se cumple que  $F(x_{1,k}^-) \leq \frac{1}{k}$ , y por la no-negatividad de la función de distribución,  $F$ , y la función de distribución empírica,  $F_n$ , así como por tratarse ambas de funciones no decrecientes, se tiene que

$$\left. \begin{array}{l} 0 \leq F_n(x) \leq F_n(x_{1,k}^-) \\ 0 \leq F(x) \leq F(x_{1,k}^-) \end{array} \right\} \Rightarrow -F(x_{1,k}^-) \leq F_n(x) - F(x) \leq F_n(x_{1,k}^-).$$

Como por definición de  $\delta_n^k$

$$F_n(x_{1,k}^-) \leq F(x_{1,k}^-) + \delta_n^k \quad \text{y} \quad -F(x_{1,k}^-) \geq -\frac{1}{k} \geq -\frac{1}{k} - \delta_n^k,$$

se concluye que

$$|F_n(x) - F(x)| \leq \frac{1}{k} + \delta_n^k \quad \text{para todo } x < x_{1,k}.$$

- Si  $x \geq x_{k,k}$ , se tiene que  $|F_n(x) - F(x)| = |1 - 1| = 0$ .

Por lo tanto, para todo  $x \in \mathbb{R}$ , se cumple que

$$|F_n(x) - F(x)| \leq \frac{1}{k} + \delta_n^k,$$

con lo que

$$\Delta_n(w) = \sup_{x \in \mathbb{R}} |F_n(x, w) - F(x)| \leq \frac{1}{k} + \delta_n^k$$

para cualquier  $w = (x_n)_{n \in \mathbb{N}}$ .

Así, en los puntos  $w \in D$ , por definición del conjunto  $D$  y de  $\delta_n^k$ , se tiene que  $\lim_{n \rightarrow \infty} \delta_n^k = 0$ . En consecuencia

$$0 \leq \lim_{n \rightarrow \infty} \Delta_n(w) \leq \lim_{n \rightarrow \infty} \left( \frac{1}{k} + \delta_n^k \right) = \frac{1}{k} + \lim_{n \rightarrow \infty} \delta_n^k = \frac{1}{k} \quad \text{para todo } w \in D.$$

Como esto se verifica para todo  $k \in \mathbb{N}$ , entonces

$$\lim_{n \rightarrow \infty} \Delta_n(w) = 0 \text{ para todo } w \in D.$$

Además,  $P(D) = 1$ , concluyendo que  $\lim_{n \rightarrow \infty} \Delta_n \stackrel{c.s.}{=} 0$ .

Esto significa que  $B = \{w : \lim_{n \rightarrow \infty} \Delta_n(w) = 0\} \supset D$  con  $P(D) = 1$ , por tanto:

- $P(B) = 1$ , y
- $\forall \varepsilon > 0, \exists N_\varepsilon$  tal que  $\forall n \geq N_\varepsilon, \sup_{x \in \mathbb{R}} |F_n(x, w) - F(x)| < \varepsilon$  para todo  $w \in B$ .

En conclusión,  $P(\{w = (x_n)_{n \in \mathbb{N}} : \lim_{n \rightarrow \infty} \left( \sup_{x \in \mathbb{R}} |F_n(x, w) - F(x)| \right) = 0\}) = 1$ , es decir,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{c.s.} 0.$$

La interpretación de este resultado asegura que en un conjunto de muestras de probabilidad 1, la función de distribución teórica de la variable,  $F$ , está dentro de la banda determinada por la función de distribución empírica asociada a la muestra  $F_n(x) - \varepsilon < F(x) < F_n(x) + \varepsilon$ , es decir, se puede estimar la función de distribución teórica con la precisión que queramos, aproximándola por la empírica, siempre que tengamos suficiente tamaño muestral.

## 1.5. Generación de muestras aleatorias: simulación

Cuando la búsqueda de la distribución de un estadístico es complicada, puede recurrirse a la aproximación de su comportamiento mediante la simulación. Esta técnica permitirá obtener tantos valores aleatorios del estadístico como se quiera y, a partir de ellos, se puede aproximar la función de distribución del estadístico por la función de distribución empírica asociada a los valores generados (según nos garantiza el Teorema de Glivenko-Cantelli). Finalmente, se puede utilizar ésta para calcular de forma aproximada probabilidades relacionadas con el estadístico.

La generación de números aleatorios (o pseudo-aleatorios) con distribución  $\mathcal{U}(0, 1)$  es un paso básico en los procedimientos de simulación. La mayoría de lenguajes de programación, así como las calculadoras básicas y programas para el uso de la estadística, tienen funciones que suministran números pseudo-aleatorios con distribución  $\mathcal{U}(0, 1)$

(en el lenguaje R, `runif`) basados usualmente en algoritmos congruenciales del tipo  $n_{i+1} = (a \cdot n_i + c)(\text{Modulo}_m)$  que son valores enteros en  $[0, m - 1]$  y se pasan al intervalo  $[0, 1)$  dividiendo por  $m$  (se llaman pseudo-aleatorios porque conociendo la semilla  $(n_0)$ ,  $a$ ,  $c$  y  $m$  se puede reproducir la secuencia de números).

El **método de Montecarlo** es una técnica numérica utilizada para generar valores aleatorios de distribuciones a partir de los generadores de números aleatorios  $\mathcal{U}(0, 1)$  (se utiliza también para el cálculo de probabilidades).

Sea  $X$  una v.a. que toma los valores  $\{x_i, i = 1, \dots, k\}$  con probabilidades  $\{p_i, i = 1, \dots, k\}$  tales que  $p_i > 0$  y  $\sum_{i=1}^k p_i = 1$ , para generar un dato de la variable  $X$ :

- 1º) Se realiza una partición del intervalo  $[0, 1]$  en  $k$  clases (subintervalos)  $(C_1, \dots, C_k)$  de amplitudes  $(p_1, \dots, p_k)$ , respectivamente.
- 2º) Se genera un número aleatorio  $u$  con distribución  $\mathcal{U}(0, 1)$  y se mira en qué clase está: si  $u \in C_j$ , entonces el valor generado de  $X$  es  $x_j$ .

**Ejemplo:** Dada  $X$  una v.a. discreta que toma los valores  $\{1, 1.5, 3\}$  con probabilidades respectivas  $\{1/2, 1/3, 1/6\}$ , se quiere generar una muestra aleatoria de tamaño 5 de esta variable. Los pasos a seguir son:

- 1º) Se particiona el intervalo  $[0, 1]$  en 3 clases:  $C_1 = [0, 1/2]$ ,  $C_2 = (1/2, 5/6]$ ,  $C_3 = (5/6, 1]$ , cuyas amplitudes se corresponden con las probabilidades.
- 2º) Se generan 5 valores de la distribución uniforme  $\mathcal{U}(0, 1)$ . Por ejemplo, un resultado es:  $(0.08417384, 0.62186518, 0.75045832, 0.76283675, 0.29542565)$ .
- 3º) Se clasifican los valores anteriores en las clases dadas:  $(C_1, C_2, C_2, C_2, C_1)$ .
- 4º) Se transforman las clases en los valores de  $X$ :  $(1, 1.5, 1.5, 1.5, 1)$ .

Este procedimiento está implementado en R para las distribuciones discretas más usuales a través de las funciones `rbinom`, `rpois`, `rgeom`, etc.

El **método de transformación** se utiliza para generar valores de una v.a. continua. Está basado en los generadores de números aleatorios  $\mathcal{U}(0, 1)$  y requiere el cálculo de la inversa de la función de distribución de la variable. Está basado en el Teorema 1.1.2, según el cual si  $X$  es continua  $F(X) \equiv \mathcal{U}(0, 1)$ . Por tanto, se genera un valor  $u$  procedente de una distribución  $\mathcal{U}(0, 1)$  y se transforma por la inversa de la función de distribución para obtener el valor correspondiente de  $X$ , es decir,  $x = F^{-1}(u)$ .

**Ejemplo:** Si  $X$  es una v.a. exponencial  $X \equiv \exp(\lambda)$ , su función de distribución es  $F(x) = [1 - e^{-\lambda x}]I(x > 0)$ . La inversa de esta función en  $(0, 1)$  es  $F^{-1}(y) = -\ln(1 - y)/\lambda$ ,

por tanto, para generar un valor de  $X$  se genera un valor  $u$  procedente de  $\mathcal{U}(0, 1)$  y se transforma a  $x = -\ln(1 - u)/\lambda$ .

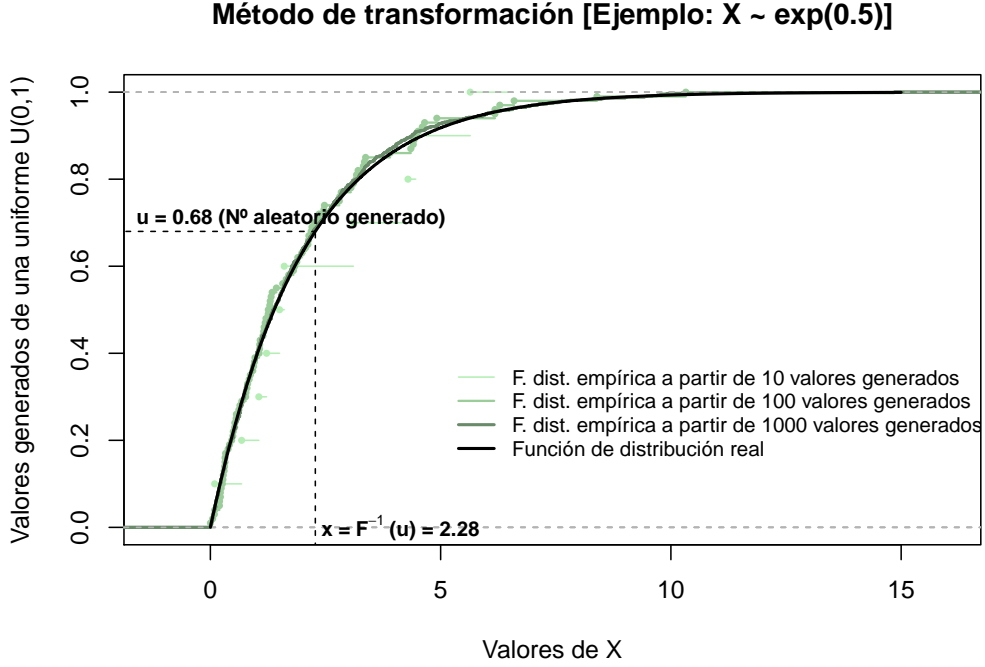


Figura 1.1: Método de transformación aplicado al ejemplo para  $X \equiv \exp(\lambda = 0.5)$ . Se han generado 10, 100 y 1000 valores  $u$  procedentes de una distribución  $\mathcal{U}(0, 1)$  y se han transformado a  $x = F^{-1}(u) = -\ln(1 - u)/\lambda$ . Posteriormente, se ha representado la función de distribución empírica para los distintos conjuntos de valores  $u$  generados, así como la función de distribución real (en negro).

El inconveniente de este procedimiento es que necesita el cálculo de la función inversa de la función de distribución, que es sabido que para algunas distribuciones como la normal no tiene expresión explícita, por lo que hay que recurrir a procedimientos alternativos.

En el caso normal, la generación de valores puede realizarse a través de la **transformación de Box-Muller**: si  $U_1$  y  $U_2$  son v.a. independientes con distribución  $\mathcal{U}(0, 1)$ , entonces

$$X = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \quad \text{e} \quad Y = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

son v.a. independientes con distribución  $\mathcal{N}(0, 1)$ .

En R existen funciones para la generación de valores de las distribuciones continuas más usuales: `runif`, `rnorm`, `rbeta`, `rgamma`, `rexp`, `rchisq`, `rt` y `rf`.

El **método de aceptación-rechazo** es otro de los métodos de generación de valores de variables continuas que no requiere el cálculo de la inversa  $F^{-1}$ , pero sí el conocimiento de la función de densidad,  $f$ , y una acotación superior de esta última a través de una función  $g$  que cumple:

- $f(x) \leq g(x) \quad \forall x \in \mathbb{R}$ ;
- la integral de  $g$  es finita, es decir,  $\int_{-\infty}^{\infty} g(x)dx = C$ ;
- se puede calcular la inversa de la función acumulada de  $g$ ,  $G^{-1}$ , definida esta última como  $G(x) = \int_{-\infty}^x g(t)dt$ .

El procedimiento de simulación de un valor  $a$  de la variable  $X$  es:

- 1º) generar un valor  $u$  de  $\mathcal{U}(0, C)$ ;
- 2º) transformarlo mediante  $a = G^{-1}(u)$ ;
- 3º) calcular  $f(a)$  y  $g(a)$ ;
- 4º) generar otro número aleatorio  $b$  en  $(0, g(a))$  a través de una  $\mathcal{U}(0, g(a))$ ;
- 5º) si  $b < f(a)$  se acepta el valor  $a$  como valor de la variable  $X$ , en otro caso se vuelve a empezar.

**Ejemplo:** Si  $X$  es una v.a. normal estándar  $X \equiv \mathcal{N}(0, 1)$ , su función de densidad es  $f(x) = e^{-x^2/2}/\sqrt{2\pi}$ . La función de distribución no tiene una expresión explícita sin recurrir a su integral, por lo que es muy conveniente utilizar un método que no requiera el cálculo de la inversa de  $F$ . Sin embargo, sí se conoce que la función de densidad tiene una forma campaniforme y alcanza su máximo en  $x^* = \mu = 0$ , punto en el cual  $f(x^*) = 1/\sqrt{2\pi}$ . Además, si  $x \notin [-1, 1]$  entonces  $x^2 > |x|$  y, por tanto,  $f(x) < e^{-|x|/2}/\sqrt{2\pi}$ . Así, utilizaremos la siguiente función  $g$  como acotación superior de  $f$ :

$$g(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} & \text{si } x \in [-1, 1] \\ \frac{e^{-|x|/2}}{\sqrt{2\pi}} & \text{en caso contrario} \end{cases}$$

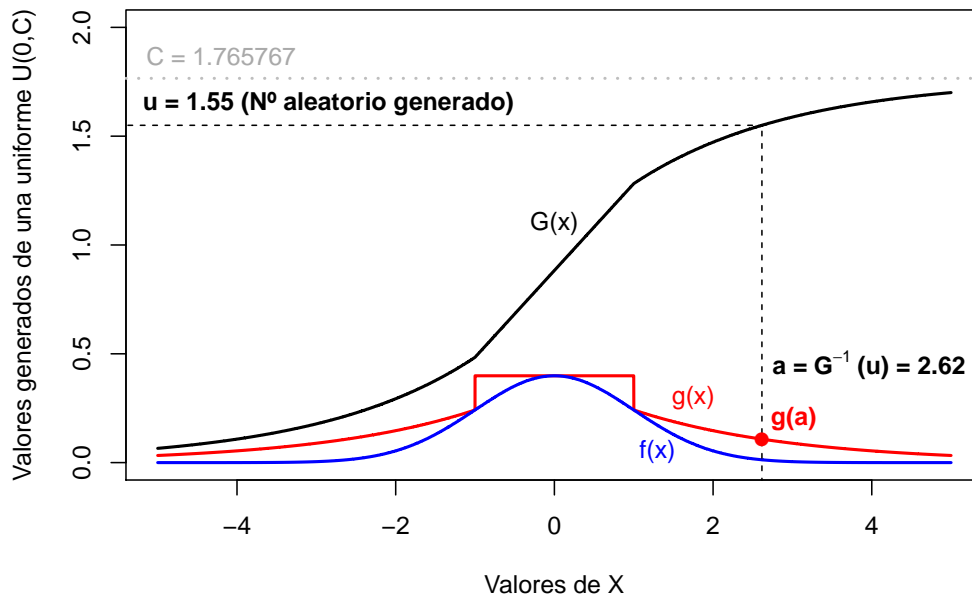


Así, se cumple que:

- $f(x) \leq g(x) \quad \forall x \in \mathbb{R}$ ;
- la integral de  $g$  es finita:  $\int_{-\infty}^{\infty} g(x)dx = \sqrt{2/\pi} [2e^{-1/2} + 1] \simeq 1.765767 = C$ ;
- se puede calcular la inversa de la función acumulada de  $g$ ,  $G^{-1}$ :

$$G^{-1}(y) = \begin{cases} \ln\left(\frac{\pi}{2}y^2\right) & \text{si } y \in \left(0, \sqrt{\frac{2}{\pi e}}\right) \\ \sqrt{2\pi}y - \frac{2}{\sqrt{e}} - 1 & \text{si } y \in \left[\sqrt{\frac{2}{\pi e}}, \sqrt{\frac{2}{\pi e}} + \sqrt{\frac{2}{\pi}}\right] \\ -2 \ln\left(1 + \frac{2}{\sqrt{e}} - \sqrt{\frac{\pi}{2}}y\right) & \text{si } y \in \left(\sqrt{\frac{2}{\pi e}} + \sqrt{\frac{2}{\pi}}, \sqrt{\frac{2}{\pi}}\left[\frac{2}{\sqrt{e}} + 1\right]\right) \end{cases}$$

### Método de aceptación–rechazo [Ejemplo: $X \sim N(0,1)$ ]



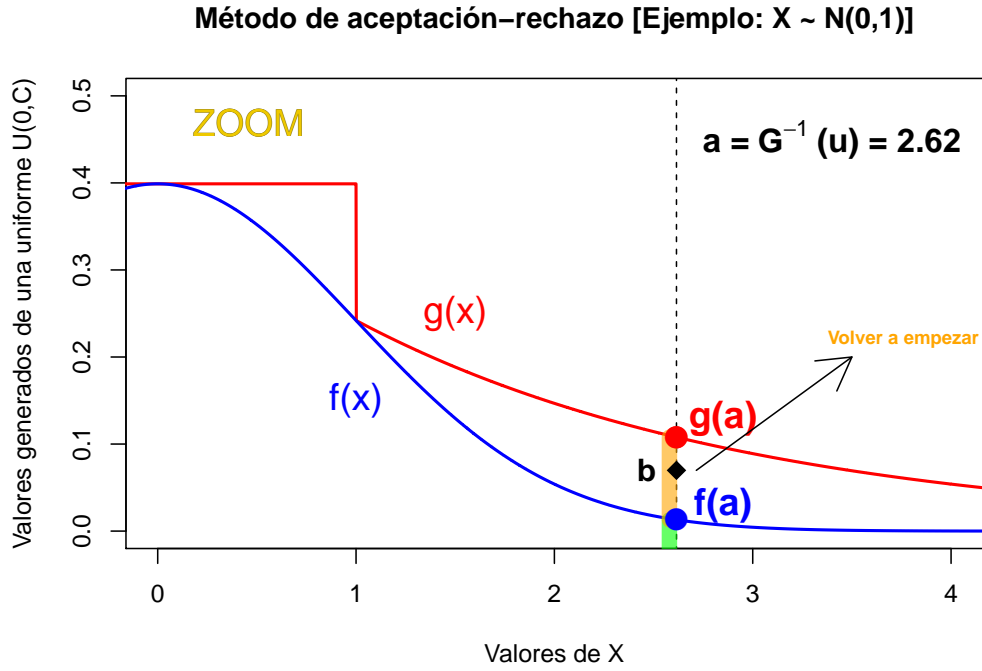


Figura 1.2: Método de aceptación-rechazo aplicado al ejemplo cuando  $X \equiv \mathcal{N}(0,1)$ . Se ha generado un valor  $u$  procedente de una distribución  $\mathcal{U}(0,C)$  con  $C = 1.765767$  y se ha calculado su inversa por la función de distribución que acota superiormente a la real,  $a = G^{-1}(u) = 2.62$  [Figura superior]. Posteriormente, se ha hecho zoom del cuadrante  $[0,4] \times [0,0.5]$  para mostrar los pasos 3) – 5) [Figura inferior]. En este caso, el número aleatorio  $b$  generado a partir de una  $\mathcal{U}(0,g(a))$  es mayor que  $f(a)$ , por lo que cae en la “región de rechazo” de  $u$  y se debe volver a empezar el proceso. En verde se señala la zona en la cual se acepta  $a$  como valor de la variable  $X$ , mientras que en naranja es la zona de “rechazo” (implica reiniciar la generación de  $u$ ).