Probabilidad y Estadística

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$P(T \geq t) = e^{-\alpha t}$$

$$W \sim \mathcal{U}(a, b)$$

Variables Aleatorias Continuas

y algunas

Distribuciones de Probabilidad

Raúl D. Katz

Pablo A. Sabatinelli

2013

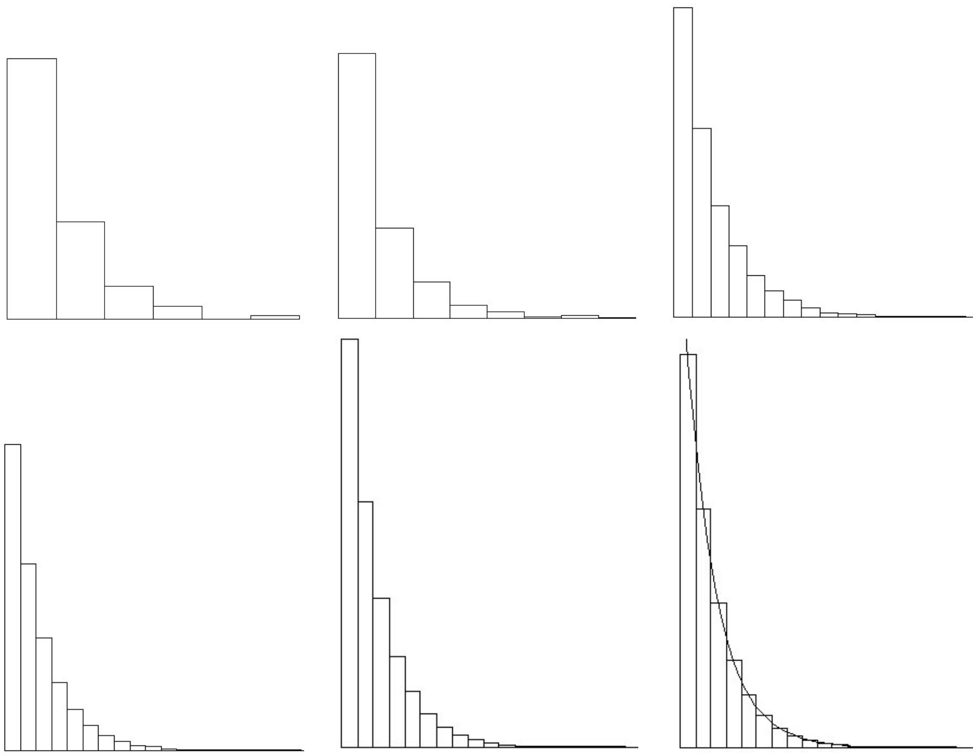
Índice

1. Variable aleatoria continua

Hemos introducido a las variables aleatorias como funciones que asignan a cada elemento de un espacio muestral un número real. Cuando la variable aleatoria es continua, es decir, cuando el recorrido de la variable aleatoria es un intervalo real, matemáticamente no es factible asignar una probabilidad positiva a cada elemento del recorrido de la variable aleatoria y al mismo tiempo verificar que la suma de las probabilidades de los distintos valores sea igual a uno. Esto nos obliga a desarrollar un método diferente para describir la distribución de probabilidad para una variable aleatoria continua.

Supongamos que nos interesa conocer el comportamiento de la variable aleatoria X: 'vida útil de una herramienta'. Desde un punto de visto teórico esta variable aleatoria puede asumir cualquier valor real no negativo. Sin embargo, esto no significa que podamos encontrar cada uno de esos valores entre los datos de una muestra.
Supongamos que se dispone de 100 datos correspondientes a la experiencia de medir la vida útil de 100 herramientas seleccionadas al azar de una población. Utilizamos un histograma y un polígono de frecuencias relativas, como representaciones gráficas para describir la distribución de esos valores medidos.

Si continuáramos indefinidamente midiendo la vida útil de nuevas herramientas del mismo tipo, obtendríamos más datos y sería factible construir histogramas y polígonos sobre la base de más intervalos de clase y de longitud menor. Estos histogramas y polígonos, basados en una mayor cantidad de mediciones, presentarían una forma cada vez más suavizada y reflejarían de manera muy aproximada la distribución de la variable en la población constituida por la totalidad de las mediciones posibles. Surgiría de este modo una curva, con propiedades análogas a los histograma y a los polígonos de frecuencias relativas.



Así como el área encerrada por un histograma sobre un intervalo representa la proporción o frecuencia relativa de datos que se encuentran en dicho intervalo, el área encerrada por la curva sobre un intervalo representa la probabilidad de que al realizar una observación de la variable aleatoria, la misma asuma un valor en dicho intervalo.

Formalizaremos estas ideas a través de la siguiente definición.

Sea X una variable aleatoria continua. Decimos que f_X es la función de densidad de probabilidad asociada a X cuando se verifica

- $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$.
- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\,dx = 1.$
- $\displaystyle P(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$

1. Sea X una variable aleatoria continua y x un número real. Entonces $P(X = x) = 0$. ¿Por qué? ¿Significa esto que es *imposible* que la variable X asuma el valor x?
2. Del apartado anterior, resulta $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.

2. Función de distribución acumulada

Sea X una variable aleatoria continua y f_X su función de densidad de probabilidad. La función

$$\begin{aligned} F_X : \quad & \mathbb{R} \to [0,1] \\ & x \mapsto F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \end{aligned}$$

se denomina función de distribución acumulada de la variable aleatoria X.

Propiedades.

1. $F_X(x) = P(X \leq x)$.
2. $0 \leq F_X(x) \leq 1$, $\forall x \in \mathbb{R}$.
3. F_X es monótona no decreciente.
4. $F'_X(x) = f_X(x)$ en los puntos de continuidad de f_X.

Ejemplo: Sea X una variable aleatoria cuya función de densidad de probabilidad es

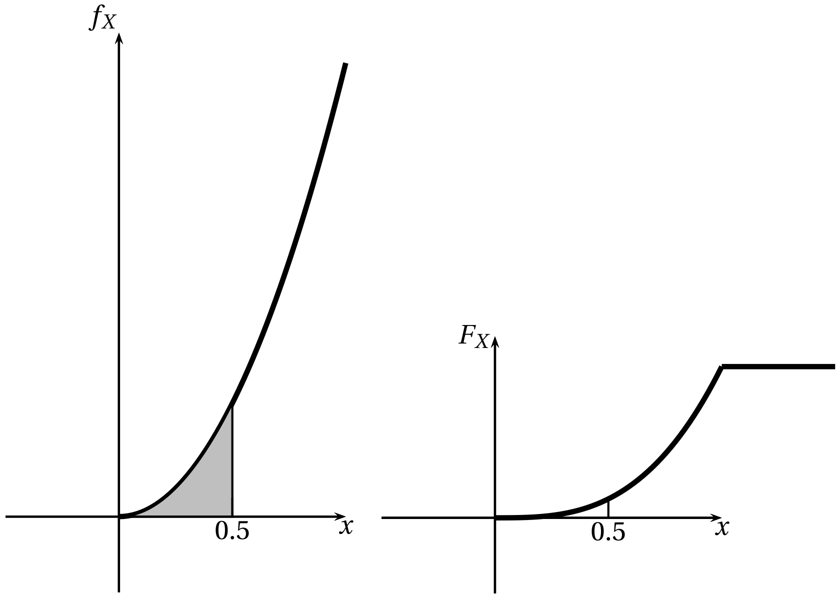
$$f_X(x) = \begin{cases} 3x^2 & 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Calcule

1. $P(X < 0.50)$.
2. $P(-1 < X < 0.20)$.
3. Determine la función de distribución acumulada para la variable aleatoria X.

1. $P(X < 0.50) = \int_{-\infty}^{0.50} f_X(x)\,dx = \int_0^{0.50} 3x^2\,dx = 0.50^3 = 0.125.$
2. $P(-1 < X < 0.20) = \int_{-1}^{0.20} f_X(x)\,dx = \int_0^{0.20} 3x^2\,dx = 0.20^3 = 0.008.$

3. Para $x < 0$, $F_X(x) = 0$. Para $x \in [0,1]$, $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x 3t^2\,dt = x^3$. Para $x > 1$, $F_X(x) = 1$. Luego
$$F_X(x) = \begin{cases} 0 & x \leq 0, \\ x^3 & 0 < x < 1, \\ 1 & x \geq 1. \end{cases}$$



3. Esperanza matemática y variancia de una variable aleatoria continua

Hemos definido la esperanza matemática de una variable aleatoria discreta X con función de probabilidad p_X como $\sum_{\mathscr{R}_X} x p_X(x)$. Cuando X es una variable aleatoria continua, con función de densidad de probabilidad f_X, se define esperanza matemática de X, a través de una integral.

> Sea X una variable aleatoria continua y f_X su función de densidad de probabilidad. La esperanza matemática o media poblacional de X, que se nota $E(X)$ o μ_X es el número
>
> $$E(X) = \mu_X = \int_{-\infty}^{+\infty} x f_X(x)\,dx,$$
>
> siempre que la integral exista y sea finita.

Hemos definido la variancia de una variable aleatoria discreta X con función de probabilidad p_X como $\sum_{\mathscr{R}_X} \left(x - \mu_X\right)^2 p_X(x)$. Cuando X es una variable aleatoria continua, con función de densidad de probabilidad f_X, se define la variancia de X, a través de una integral.

> Sea X una variable aleatoria continua y f_X su función de densidad de probabilidad. La variancia de X que se nota $V(X)$ o σ_X^2 es el número no negativo
>
> $$V(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} \left(x - \mu_X\right)^2 f_X(x)\,dx,$$
>
> siempre que la integral exista y sea finita.

A partir de la definición de variancia, se prueba que

$$V(X) = E\left(X^2\right) - (E(X))^2,$$

donde $E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx$. Este resultado facilita el cálculo de la variancia de una variable aleatoria.

Ejemplo: Consideramos la variable aleatoria Y con función de densidad

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Veamos primero que f_Y es una función de densidad de probabilidad.

$$\int_{-\infty}^{+\infty} f_Y(y)\,dy = \int_0^1 2y\,dy = \left. y^2\right|_0^1 = 1^2 - 0^2 = 1.$$

Es claro que $f_Y(y) \ge 0$ para todo $y \in \mathbb{R}$. La esperanza de Y es

$$\mu_Y = \int_{-\infty}^{+\infty} y f_Y(y)\,dy = \int_0^1 2y^2\,dy = \left.\frac{2}{3}y^3\right|_0^1 = \frac{2}{3}.$$

Si realizamos *un gran número* de observaciones de la variable aleatoria Y, entonces la media aritmética de esos valores será un número que tiene *grandes chances* de diferir *poco* de $\mu_Y = \frac{2}{3}$.

La variancia de Y es

$$\sigma_Y^2 = \int_{-\infty}^{+\infty} \left(y - \mu_Y\right)^2 f_Y(y)\,dy = \int_0^1 2y\left(y - \frac{2}{3}\right)^2\,dy = \left.\frac{1}{2}y^4 - \frac{8}{9}y^3 + \frac{4}{9}y^2\right|_0^1 = \frac{1}{18}.$$

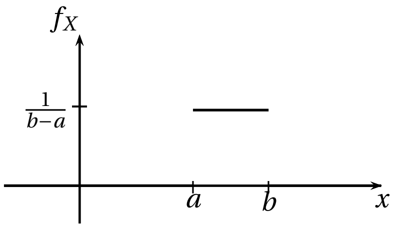
La variancia cuatifica la variación de los valores de la variable Y respecto de μ_Y.

4. Distribución uniforme

Decimos que una variable aleatoria X se distribuye uniformemente en el intervalo $[a,b]$ y notamos $X \sim \mathscr{U}(a,b)$, cuando su función de densidad de probabilidad es

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b], \\ 0 & \text{en otro caso.} \end{cases}$$

El hecho de que una variable aleatoria tenga un comportamiento uniforme significa que intervalos de igual amplitud contenidos en el intervalo $[a,b]$ tienen la misma probabilidad de ocurrir.



4.1. Ejemplo

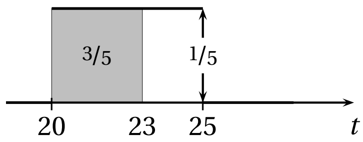
El tiempo en minutos que una persona demora en ir de su casa a una estación de tren es una variable aleatoria T con distribución uniforme en el intervalo $[20,25]$.

1. Calcule la probabilidad de que si la persona deja su casa a las 7:05, alcance el tren que parte a las 7:28.

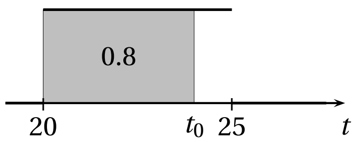
2. ¿A qué hora debe salir de su casa para tener una probabilidad de 0.8 de alcanzar el tren?

Sea T:'tiempo en minutos que una persona demora en ir de su casa a una estación de tren'. $T \sim \mathscr{U}(20,25)$.

1. $P(T < 23) = 3 \times \frac{1}{5} = \frac{3}{5}$.



2. $P(T < t_0) = 0.8 \Leftrightarrow (t_0 - 20) \times \frac{1}{5} = 0.8$. Luego $t_0 = 0.8 \times 5 + 20 = 24'$. Debe salir de su casa a las 7:04



4.2. Propuestas

1. Sea $X \sim \mathscr{U}(a,b)$. Determine su función de distribución acumulada y grafíquela.
2. El error que se comete al medir la densidad de una sustancia es una variable aleatoria continua X con distribución uniforme en el intervalo $[-0.02, 0.02]$.
   a) Calcule la probabilidad de que el error de una medición se encuentre entre 0.010 y 0.014.
   b) Calcule la probabilidad de que el error de una medición sea mayor que 0.015.
   c) Calcule la probabilidad de que el error de una medición sea a lo sumo -0.016.
   d) Calcule la probabilidad de que en una medición no se cometa error.

4.3. Esperanza matemática y variancia

Sea $X \sim \mathscr{U}(a,b)$. Entonces

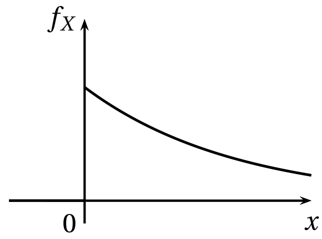
- $E(X) = \mu_X = \displaystyle\int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2}$.
- $V(X) = \sigma_X^2 = \displaystyle\int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a}\,dx = \frac{(b-a)^2}{12}$.

5. Distribución exponencial

Decimos que una variable aleatoria X tiene una distribución exponencial de parámetro $\alpha > 0$ y notamos $X \sim \mathscr{E}(\alpha)$ cuando su función de densidad de probabilidad es

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Si $x > 0$, entonces $F_X(x) = \int_0^x \alpha e^{-\alpha t}\,dt = 1 - e^{-\alpha x}$. Consecuentemente $P(X > x) = 1 - (1 - e^{-\alpha x}) = e^{-\alpha x}$.

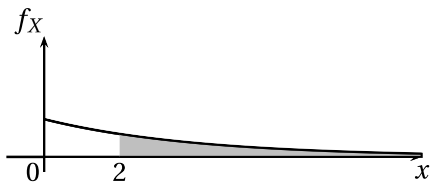


5.1. Ejemplo

La duración en años de un fusible es una variable aleatoria T con distribución exponencial con $\alpha = 0.25$. Es decir que $f_T(t) = 0.25e^{-0.25t}$, $t > 0$.

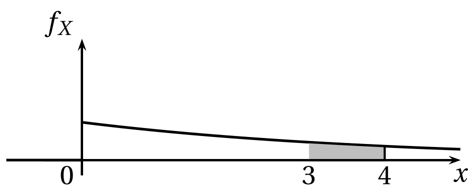
1. La probabilidad de que un fusible dure más de 2 años es

$$P(T > 2) = e^{-0.25\times 2} = e^{-0.5} \approx 0.6065.$$



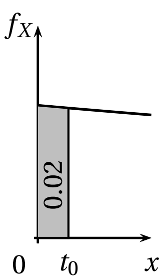
2. La probabilidad de que un fusible dure entre 3 años y 4 años es

$$P(3 < T < 4) = F_T(4) - F_T(3) = 1 - e^{-0.25\times 4} - \left(1 - e^{-0.25\times 3}\right) = e^{-0.25\times 3} - e^{-0.25\times 4} \approx 0.1045.$$



3. El fabricante quiere reponer a lo sumo el 2% de los fusibles dentro del tiempo de garantía. Entonces debe ofrecer una garantía de de a lo sumo t_0.

$$P(T < t_0) = F_T(t_0) = 1 - e^{-0.25t_0} = 0.02 \Rightarrow t_0 \approx 0.0808 \text{ años} \approx 1 \text{ mes}.$$



5.2. Esperanza matemática y variancia

Si $X \sim \mathscr{E}(\alpha)$, entonces

$$E(X) = \mu_X = \int_0^{+\infty} x\alpha e^{-\alpha x}\, dx = \frac{1}{\alpha}.$$

$$V(X) = \sigma_X^2 = \int_0^{+\infty} \left(x - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha x}\, dx = \frac{1}{\alpha^2}.$$

5.3. Ejemplos

1. El tiempo en minutos que transcurre entre las llegadas consecutivas de dos automóviles a una estación de peaje, es una variable aleatoria X con distribución exponencial con $\alpha = 4$ min^{-1}. Es decir que $f_X(x) = 4e^{-4t}$, $x > 0$.

   a) El tiempo medio transcurrido entre las llegadas consecutivas de automóviles es

   $$E(X) = \mu_X = \frac{1}{4} \text{ minutos} = 15 \text{ segundos.}$$

   b) La probabilidad de que el tiempo transcurrido entre dos llegadas consecutivas sea inferior a un minuto, si se sabe que al cabo de 30 segundos aún no ha llegado el segundo automóvil es

   $$P(X < 1/X > 0.5) = \frac{P(0.5 < X < 1)}{P(X > 0.5)} = \frac{F_X(1) - F_X(0.5)}{1 - F_X(0.5)} \approx 0.8647.$$

2. La duración en años de un fusible es una variable aleatoria T con distribución exponencial con $\lambda = 0.25$ años^{-1}.

   - Calcule la probabilidad de que un fusible dure más de tres años.
   - Calcule la probabilidad de que un fusible dure más de 4 años, si ya ha durado 1 año.

   a) $P(T > 3) = e^{-0.25\times 3} = e^{-0.75} \approx 0.472$.

   b) $P(T > 4/T > 1) = \frac{P(T>4)}{P(T>1)} = \frac{e^{-0.25\times 4}}{e^{-0.25\times 1}} = e^{-0.75} \approx 0.472$. Es decir, la probabilidad de que un fusible dure más de cuatro horas cuando se conoce que ya ha durado más de una hora, coincide con la probabilidad de que el fusible dure más de tres horas. Esto es, la probabilidad de que el fusible “sobreviva” depende tan sólo de la amplitud del intervalo de tiempo y no del tiempo en que ha estado fucionando. Con mayor generalidad enunciamos la propiedad.

5.4. Propiedad de no memoria

Sea X una variable aleatoria con distribución exponencial. Sean t y s números positivos. Entonces

$$P(X > t + s/X > s) = P(X > t).$$

$$P(X > t + s/X > s) = \frac{P((X > t + s) \cap (X > s))}{P(X > s)} = \frac{P(X > t + s)}{P(X > s)} = \frac{e^{-\alpha(t+s)}}{e^{-\alpha s}} = e^{-\alpha t} = P(X > t).$$

Esta propiedad sólo la satisfacen aquellas variables que tengan comportamiento exponencial sin importar su parámetro.

5.5. Vinculación con la distribución de Poisson

Sea X_t la cantidad de automóviles que llegan a un puesto de peaje en un periodo de tiempo t. Supongamos que $X_t \sim \mathscr{P}(\alpha t)$, es decir

$$P(X_t = k) = \frac{(\alpha t)^k e^{-\alpha t}}{k!}, \quad k = 0, 1, 2, \ldots.$$

Sea T: ‘tiempo que transcurre hasta que se produce la llegada de un automóvil al puesto de peaje’ o equivalentemente ‘tiempo que transcurre entre la llegada de un automóvil y el siguiente’. La probabilidad de que el tiempo entre la llegada de dos automóviles sea mayor que t, es igual a la probabilidad de que no lleguen automóviles al puesto de peaje en un periodo de tiempo t, puesto que estos sucesos son equivalentes. En símbolos

$$P(T > t) = P(X_t = 0) = e^{-\alpha t}.$$

Esta relación nos indica que $T \sim \mathscr{E}(\alpha)$.

5.6. Ejemplo

La distancia entre imperfecciones en un rollo de alambre se distribuye exponencialmente con una distancia media de 3 m. Sea X la distancia (medida en m.) entre dos imperfecciones consecutivas.

1. ¿Cuál es la media de la cantidad de imperfecciones por metro?
2. ¿Cuál es la probabilidad de que en 5 m. de alambre haya solo 2 imperfecciones?

Sea $X \sim \mathscr{E}\left(\frac{1}{3}\right)$.

1. Sea Y_1: 'número de imperfecciones en 1 m. de alambre'. Entonces $Y_1 \sim \mathscr{P}\left(\frac{1}{3}\right)$. $E(Y_1) = \frac{1}{3}$ imperfecciones.
2. Sea Y_5: 'número de imperfecciones en 5 m. de alambre'. Entonces $Y_5 \sim \mathscr{P}\left(\frac{5}{3}\right)$.

$$P(Y_5 = 2) = \frac{\left(\frac{5}{3}\right)^2 e^{-\frac{5}{3}}}{2!} = \frac{25}{18} e^{-\frac{5}{3}} \approx 0.2623.$$

6. Distribución normal

La distribución normal fue introducida en 1733 por Abraham De Moivre, al obtenerla como una forma límite de la distribución binomial.

Medio siglo después fue redescubierta por Laplace y Gauss para describir el comportamiento de los errores en las mediciones astronómicas, que seguían una distribución simétrica en forma de campana.

Decimos que una variable aleatoria X tiene una distribución normal de parámetros μ y σ, donde $\sigma > 0$, y notamos $X \sim \mathscr{N}(\mu, \sigma)$ cuando su función de densidad de probabilidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La vida útil de las baterías, el diámetro interior de un anillo de pistón, los errores en instrumentos para medir longitudes, la resistencia a la ruptura de una soga, el nivel de agua en un lago, el valor de glucosa en sangre y el espesor de pintura en láminas sin recubrimiento de zinc, son variables aleatorias que pueden modelizarse a través de una distribución normal. La posibilidad de poder explicar la variabilidad en situaciones tan diversas a partir de un mismo modelo, hace que la distribución normal sea una de las más importantes de la estadística.

6.1. Propiedades

1. La función $f_X(x) > 0$, para todo x real y además $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$.
2. La gráfica de f_X es simétrica respecto de la recta de ecuación $x = \mu$. En símbolos

$$f_X(\mu - x) = f_X(\mu + x), \quad \forall x \in \mathbb{R}.$$

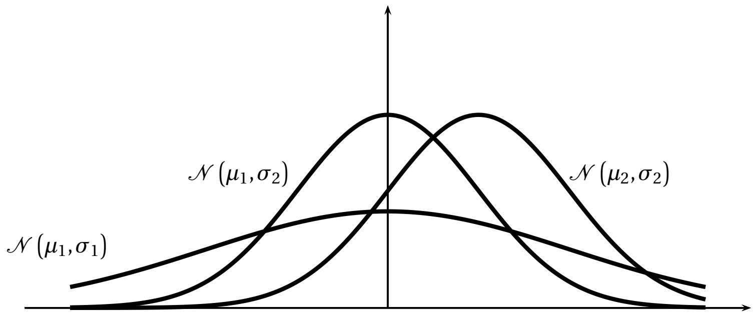
3. La función f_X tiene un máximo absoluto para $x = \mu$ y presenta puntos de inflexión en los puntos de abcisa $x = \mu \pm \sigma$.

Se demuestra que si $X \sim \mathcal{N}(\mu, \sigma)$, entonces

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

$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\, dx = \sigma^2.$$

Mostramos tres gráficos de curvas normales con diferentes medias y variancias. En este gráfico es $\mu_1 < \mu_2$ y $\sigma_1 > \sigma_2$.



6.2. Cálculo de probabilidades

Sea $X \sim \mathcal{N}(\mu, \sigma)$. Entonces la probabilidad de que una variable aleatoria con dicha distribución tome un valor en el intervalo $[a, b]$ viene dada por la integral

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt.$$

En la práctica, estas integrales se calculan con la ayuda de tablas. La función $f(t) = e^{-\frac{(t-\mu)^2}{2\sigma^2}}$ "no se puede integrar", en el sentido de que no existe ninguna función F que se exprese en términos de polinomios, raíces, senos, cosenos, exponenciales, logaritmos, etc., tal que $F' = f$.

Para el cálculo a través de tablas, resulta útil saber que si $X \sim \mathcal{N}(\mu, \sigma)$, entonces para $Z = \frac{X-\mu}{\sigma}$, resulta $Z \sim \mathcal{N}(0,1)$. En efecto, veamos que el cambio de variable $z = \frac{t-\mu}{\sigma}$, permite que la evaluación no dependa de los particulares valores de μ y σ.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = P\left(Z \leq \frac{x-\mu}{\sigma}\right)$$

donde $Z \sim \mathcal{N}(0,1)$.

Es gracias a esta propiedad de la distribución normal que en la práctica solamente existen tablas para la distribución normal de parámetros 0 y 1, conocida como distribución normal estándar, o reducida.

Con mayor generalidad:

> Si $X \sim \mathcal{N}(\mu_X, \sigma_X)$ entonces $Y = aX + b$, también tiene una distribución normal, con parámetros $\mu_Y = a\mu_X + b$ y $\sigma_Y = |a|\sigma_X$.

6.2.1. Ejemplos

La resistencia a la compresión de una muestra de cemento es una variable aleatoria que se distribuye normalmente con media 6000 kg/cm^2 y desviación estándar de 100 kg/cm^2.

1. ¿Cuál es la probabilidad de que la resistencia de una muestra de cemento tomada al azar, sea menor a 6250 kg/cm^2?

2. ¿Cuál es la probabilidad de que la resistencia de una muestra de cemento se encuentre entre 5800 y 5900 kg/cm^2?

3. ¿Cuál es la probabilidad de que la resistencia de una muestra de cemento tomada al azar, supere los 5950 kg/cm^2?

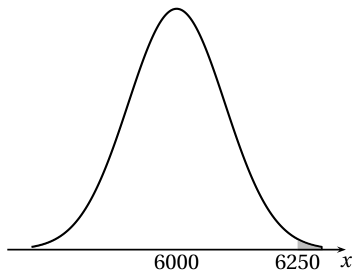
4. ¿Cuál es el valor de la resistencia que es superado por el 95 % de las muestras de cemento?

Llamamos X:'resistencia a la compresión de una muestra de cemento'. $X \sim \mathcal{N}(6000, 100)$.

1.

$$\begin{aligned} P(X < 6250) &= P\left(\frac{X-6000}{100} < \frac{6250-6000}{100}\right) \\ &= P\,(Z < 2.5) \approx 0.9938, \end{aligned}$$

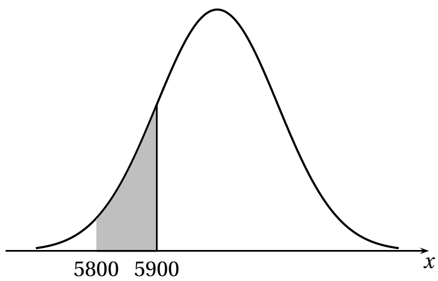
según resulta de la tabla de la distribución normal.



2.

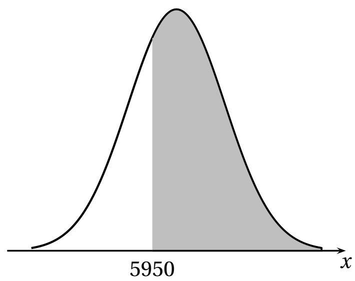
$$\begin{aligned} P(5800 < X < 5900) &= P\left(\frac{5800-6000}{100} < \frac{X-6000}{100} < \frac{5900-6000}{100}\right) \\ &= P\,(-2 < Z < -1) \\ &= P(Z < -1) - P(Z < -2) \approx 0.1587 - 0.0228 = 0.1359, \end{aligned}$$

según resulta de la tabla de la distribución normal.



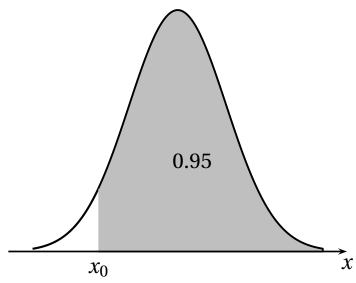
3.

$$\begin{aligned} P(X > 5950) &= P\left(\frac{X-6000}{100} > \frac{5950-6000}{100}\right) \\ &= P(Z > -0.50) \\ &= 1 - P(Z < -0.50) \approx 1 - 0.3085 = 0.6915, \end{aligned}$$



4.

$$\begin{aligned} P(X > x_0) = 0.95 \quad &\Rightarrow \quad P\left(Z > \frac{x_0 - 6000}{100}\right) = 0.95 \\ &\Rightarrow \quad P\left(Z < \frac{x_0 - 6000}{100}\right) = 0.05 \\ &\Rightarrow \quad \frac{x_0 - 6000}{100} \approx -1.64 \Rightarrow x_0 \approx 5836 \text{ kg/cm}^2. \end{aligned}$$



6.3. Problemas

1. Indique en el siguiente gráfico de una variable con distribución normal, la media y la mediana. Justifica.

2. Una empresa produce un cierto tipo de piezas metálicas, para lo que cuenta con dos máquinas. Las longitudes de las piezas producidas por la máquina 1 se distribuyen normalmente con $\mu = 87.5$ cm y $\sigma = 0.5$ cm; mientras que las longitudes de las piezas producidas por la máquina 2 se distribuyen normalmente con $\mu = 85$ cm y $\sigma = 0.8$ cm. Un cliente requiere piezas con longitudes mayores que 86 cm. ¿Con qué máquina le conviene producir piezas para este cliente? Explique por qué es posible responder sin realizar cálculos.

3. El voltaje de salida de una fuente de energía eléctrica se distribuye normalmente, con media 12 V y desviación estándar de 0.10 V. Las especificaciones para el voltaje es 12 ± 0.15 V. Calcule la probabilidad de que una fuente de voltaje de energía elegida al azar no cumpla con las especificaciones.

4. La duración de una batería de automóvil, es una variable aleatoria con distribución normal de media 900 días. Calcule para qué valor del desvío estándar se verifica que con probabilidad 0.95, una batería elegida al azar dure al menos 825 días.

5. Los ejes fabricados para el uso de dispositivos de almacenamiento óptico tienen diámetros que se distribuyen normalmente con media $\mu = 0.652$ cm y desviación estándar $\sigma = 0.003$ cm. La especificación para el diámetro de los ejes es 0.650 ± 0.005 cm.

   a) ¿Qué proporción de los ejes fabricados por este proceso no cumple con la especificación?

   b) Sea Y: 'cantidad de ejes que no cumplen con la especificación en una muestra de tamaño 10, seleccionada de una producción muy grande'. Calcule e interprete el valor de $E(Y)$.

   c) Se ha recalibrado el proceso de tal forma que la media de los diámetros es $\mu = 0.65$ cm. ¿Cuál debería ser el máximo valor de la desviación estándar para que el 99 % de los ejes cumpla con la especificación?

   d) Explique qué mide la desviación estándar.

6. La densidad del grosor de un suelo se define como la masa de sólidos secos por unidad de volumen del grosor. La densidad del grosor de un suelo es una variable aleatoria X con distribución normal, media igual a 1.5 y desviación estándar 0.2 g/cm^3.

a) Calcule $P(1.1 < X < 1.9)$. Esboce además una gráfica de la función densidad de X y muestre sobre la misma el valor de la probabilidad calculado.

b) ¿Cuál es el valor de la densidad que posee la propiedad de que el 10% de las muestras de suelo tienen densidades mayores?

c) Calcule la probabilidad de que en 10 muestras del suelo seleccionadas al azar, a lo sumo una tenga una densidad menor que 1.0 $\mathrm{g/cm^3}$.

7. Los tiempos hasta la primera avería de cierta marca de impresora se distribuyen normalmente con una media de 1500 horas y una desviación estándar de 200 horas.

   a) ¿Qué proporción de las impresoras fallarán antes de las 1000 horas?

   b) Calcule la esperanza matemática de la variable aleatoria X: 'cantidad de impresoras que fallan antes de las 1000 horas en una muestra aleatoria de 50 impresoras'.

   c) Si el fabricante desea que sólo presente averías dentro del período de garantía el 0.5% de las impresoras, ¿cuál debería ser el tiempo máximo de garantía para esas impresoras?

8. La duración de ciertos chips para computadoras varía aleatoriamente según una distribución normal, con media 4.4×10^6 horas y desviación estándar de 0.13×10^6 horas. Un fabricante de computadoras necesita que por lo menos el 90% de los chips, de un gran lote, tengan una duración de por lo menos 4×10^6 horas.

   a) ¿Cumplen los chips con los requerimientos del fabricante?

   b) ¿Cuál es la probabilidad de que un lote de 100 chips contenga a lo sumo 2 cuya duración sea menor que 3×10^6 horas?

9. Para una variable aleatoria X con distribución $\mathcal{N}(\mu, \sigma)$, calcule

   - $P(\mu - \sigma < X < \mu + \sigma) =$ ____________________
   - $P(\mu - 2\sigma < X < \mu + 2\sigma) =$ ____________________
   - $P(\mu - 3\sigma < X < \mu + 3\sigma) =$ ____________________

   Intente justificar a partir de las probabilidades calculadas, por qué la tabla de distribución normal tiene valores tabulados entre -3.99 y 3.99.

7. Distribución log-normal

Sea X una variable aleatoria. Sea la variable aleatoria $Y = e^X$ (Y es una variable aleatoria que depende de X.) Cuando X asume el valor x, entones Y asume el valor $y = e^x$, para cada $x \in \mathscr{R}_X$.)

¿Cómo se distribuye la variable aleatoria Y cuando $X \sim \mathcal{N}(\mu, \sigma)$? Observamos que para $y > 0$,

$$F_Y(y) = P(Y \leq y) = P\left(e^X \leq y\right) = P\left(X \leq \ln y\right) = \int_{-\infty}^{\ln y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

Por el teorema fundamental del cálculo

$$F'_Y(y) = \frac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}.$$

Sea $X \sim \mathcal{N}(\mu, \sigma)$ y sea $Y = e^X$. Entonces la función de densidad de la variable aleatoria Y es

$$f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} & y > 0, \\ 0 & y \leq 0. \end{cases}$$

En este caso, decimos que la variable aleatoria Y tiene una distribución log-normal con parámetros μ y σ.

GRÁFICA DE DISTRIBUCIÓN LOG-NORMAL.
La distribución log-normal se utiliza generalmente para modelar o describir procesos con asimetría a derecha. Las muestras provenientes de poblaciones log-normales presentan a menudo valores atípicos hacia la derecha. Tal es el caso de la concentración de anticuerpos en suero sanguíneo humano, el tamaño de partículas de las gotas formados por los nebulizadores utilizados en espectroscopía de llama, el tamaño de partículas que resultan de un proceso de molienda, la concentración de impurezas en el agua, el grado de acidez de una sustancia.

Se demuestra que si Y tiene una distribución log-normal de parámetros μ y σ, entonces

$$E(Y) = e^{\mu + \frac{\sigma^2}{2}}$$
$$V(Y) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2}$$

7.1. Ejemplo

Cuando un pesticida entra en contacto con la piel, se absorbe cierto porcentaje de éste. El porcentaje del pesticida que será absorbido durante cierto espacio de tiempo puede modelarse con una distribución log-normal. Suponga que para cierto pesticida, la cantidad absorbida (en porcentaje) durante un periodo de dos horas sigue una distribución log-normal con $\mu = 1.5$ y $\sigma = 0.5$

1. Determine la media y desviación estándar del porcentaje absorbido.
2. Determine la probabilidad de que el porcentaje absorbido sea mayor que 10.
3. Determine la probabilidad de que el porcentaje absorbido sea menor que 5.
4. ¿Que porcentaje de pesticida es absorbido por a lo sumo el 75 % de las personas?
5. Determina la mediana del porcentaje absorbido.

Sea Y: 'cantidad de pesticida absorbido (en porcentaje) durante un periodo de dos horas de pesticida'. $Y = e^X$, donde $X \sim \mathcal{N}(1.5, 0.5)$.

1. $E(Y) = e^{1.5 + \frac{0.5^2}{2}} = e^{0.1875} \approx 1.206\,\%$, $V(Y) = \left(e^{0.5^2} - 1\right) e^{2\times 1.5 + 0.5^2} \approx 7.325\,\%^2$.
2. $P(Y > 10) = P\left(e^X > 10\right) = P(X > \ln 10) = P\left(Z > \frac{\ln 10 - 1.5}{0.5}\right) \approx 0.054$.
3. $P(Y < 5) = P\left(e^X < 5\right) = P(X < \ln 5) = P\left(Z < \frac{\ln 5 - 1.5}{0.5}\right) \approx 0.587$.
4.

$$\begin{aligned} P\left(Y < y_0\right) = 0.75 \quad &\Leftrightarrow \quad P\left(e^X < y_0\right) = 0.75 \\ &\Leftrightarrow \quad P\left(X < \ln y_0\right) = 0.75 \\ &\Leftrightarrow \quad P\left(Z < \frac{\ln y_0 - 1.5}{0.5}\right) = 0.75 \\ &\Leftrightarrow \quad \frac{\ln y_0 - 1.5}{0.5} \approx 0.67. \end{aligned}$$

En consecuencia $y_0 \approx 6.26\,\%$.

5.

$$\begin{aligned} P\left(Y < \tilde{y}\right) = 0.50 \quad &\Leftrightarrow \quad P\left(e^X < \tilde{y}\right) = 0.50 \\ &\Leftrightarrow \quad P\left(X < \ln \tilde{y}\right) = 0.50 \\ &\Leftrightarrow \quad P\left(Z < \frac{\ln \tilde{y} - 1.5}{0.5}\right) = 0.50 \\ &\Leftrightarrow \quad \frac{\ln \tilde{y} - 1.5}{0.5} = 0. \end{aligned}$$

En consecuencia $\tilde{y} \approx 4.48\,\%$.

7.2. Problemas

1. La vida útil de un semiconductor láser tiene una distribución log-normal con $\mu = 10$ horas y $\sigma = 1.5$ horas.

   a) ¿Cuál es la probabilidad de que la vida útil de un semiconductor láser elegido al azar supere las 10000 horas?

   b) ¿Qué tiempo de vida excede el 99 % de los láser?

2. El diámetro de bolitas de telgopor para embalajes es una variable aleatoria con distribución log-normal con parámetros $\mu = 0.8$ y $\sigma = 0.1$ mm. Calcule la probabilidad de que una bolita elegida al azar tenga un diámetro mayor a 2.7 mm.

8. Miscelánea

1. El peso que soporta una varilla especial usada en la construcción es una variable aleatoria con distribución normal, con media igual a 25 toneladas. Se conoce que el 80 % de las varillas soportan un peso mayor a 23 toneladas.

   a) Las especificaciones dadas por un comprador requieren que las varillas aguanten un peso de por los menos 22 toneladas. ¿Qué proporción de las varillas no satisface las especificaciones?

   b) Calcule la probabilidad de que en una muestra aleatoria de 5 varillas a lo sumo una no cumpla con las especificaciones.

   c) ¿Cuál es el número promedio de varillas que no satisface las especificaciones, en todos los posibles lotes de 400 varillas?

2. La cantidad de averías diarias para un proceso automatizado de producción tiene una distribución de Poisson con media igual a 2 averías en jornadas de 8 horas.

   a) Calcule la probabilidad de que el proceso trabaje más de una hora sin averías.

   b) Calcule la probabilidad de que el proceso trabaje un turno de ocho horas sin averías.

   c) Explique qué significa que la media, en jornadas de 8 horas, sea igual a 2 averías.

   d) ¿Cuál es el tiempo medio entre averías?

3. El tiempo necesario para terminar determinada operación de ensamblado se distribuye uniformemente entre 30 y 40 minutos.

   a) Calcule la probabilidad de que la operación de ensamblado requiera más de 38 minutos.

   b) Analice la veracidad de la siguiente afirmación: El 70 % de las veces el armado termina después de los 33 minutos de operación.

4. El error que se comete al pesar un objeto varía aleatoriamente con distribución normal, media y desviación estándar igual a 0 y 0.15 g respectivamente.

   a) Calcule la probabilidad de que el error, en valor absoluto, de una pesada resulte inferior a 0.20 g.

   b) ¿Cuál es la esperanza matemática de la variable aleatoria Y: 'número de pesadas sobre un total de 10 en que el error, en valor absoluto, resulta inferior a 0.20 g'?

5. El tiempo en minutos que un auto permanece en una playa de estacionamiento es una variable aleatoria X normalmente distribuida con $E(X) = 50$ minutos. Se conoce además que la probabilidad de que un auto permanezca en la playa por más de 54 minutos es 0.0668.

   a) ¿Cuál es la probabilidad de que un auto permanezca en la playa por más de 52 minutos?

   b) ¿Cuál es la probabilidad de que, sobre un total de 10 autos, por lo menos dos permanezcan por más de 52 minutos?

6. Un fabricante de aviones desea obtener remaches para montar los propulsores de sus aviones. El esfuerzo a la tensión mínima necesario de cada remache es de 25000 lb. Se pide a tres fabricantes de remaches que proporcionen toda la información pertinente con respecto a los remaches que producen. Los tres fabricantes aseguran que la resistencia a la tensión de sus remaches se encuentra distribuida normalmente con medias 28000, 30000 y 29000 libras, respectivamente. ¿Tiene el fabricante suficiente información para hacer una elección? Explique.

7. El número de pedidos de asistencia que recibe un servicio de remolque de vehículos es una variable aleatoria con distribución de Poisson a razón de cuatro pedidos por hora.

   a) Calcule la probabilidad de que en un período de 2 horas se reciban 10 pedidos.

   b) Si los operadores de las grúas de remolque descansan durante 30 minutos, ¿cuál es la probabilidad de que en ese período de tiempo no se desatienda ninguna llamada de asistencia?

   c) ¿Qué distribución de probabilidad tiene la variable aleatoria T: 'tiempo transcurrido hasta que se produce la primera solicitud (o tiempo transcurrido entre dos solicitudes sucesivas)'?

8. La vida útil T de un componente es una variable aleatoria con distribución $\mathcal{N}(100, \sigma)$. Se conoce que $P(T < 120) = 0.90$.

   a) Calcule $P(T > 110)$.

   b) Un sistema consta de 5 de tales componentes. Al fallar una cualquiera de las componentes se desconecta el sistema. ¿Cuál es la probabilidad de que se desconecte el sistema antes de las 110 horas de funcionamiento? Indique qué supuestos debió realizar para el cálculo de la probabilidad anterior.

9. Un sistema consta de dos dispositivos (A y B) que funcionan simultánea e independiente. La duración en horas del dispositivo A es una variable aleatoria con distribución exponencial de parámetros $\alpha = 0.02$ horas^{-1}, en cambio la duración en horas para el dispositivo B es una variable aleatoria con distribución normal de parámetros $\mu = 10$ horas, y $\sigma = 1$ hora.

   a) Calcule la probabilidad de que falle al menos un dispositivo antes de las 12 horas de funcionamiento.

   b) Si al menos uno de los dispositivos ha fallado antes de las 12 horas de funcionamiento, calcule la probabilidad de que sea el dispositivo B.

   c) Calcule la probabilidad de que el dispositivo A funcione al cabo de las 20 horas sabiendo que funciona al cabo de las 8 horas.

10. La duración en años de cierto tipo de heladera es una variable aleatoria $D \sim \mathcal{N}(4.8, 1.3)$.

    a) El aparato está garantizado por dos años. ¿Cuál es la probabilidad de que el artefacto deba ser reemplazado durante el período cubierto por la garantía?

    b) ¿Cuánto tiempo de garantía debe otorgar el fabricante si desea reemplazar sólo el 0.5 % de las heladeras?

    c) La empresa fabricante recibe el pedido de 10 heladeras. ¿Cuál es la probabilidad de que por lo menos 8 duren más de 5 años?

11. El salario que paga una compañía a sus empleados por hora de trabajo es una variable aleatoria S con distribución normal, media igual a \$80 y desviación estándar igual a \$12. Considere las siguientes dos posibilidades:

    - se aumenta el salario por hora de cada empleado en un 10 %;
    - se aumenta el salario por hora de cada empleado en \$8.

    ¿En qué caso la proporción de empleados que ganan más de \$100 la hora es mayor?

12. El costo de reparación (medido en pesos) de un equipo es una variable aleatoria C que depende del tiempo T (medido en horas) que insume la reparación: $C = 10 + 100T$. La variable aleatoria $T \sim \mathcal{U}(0, 20)$. Calcule la probabilidad de que el costo de reparación de un equipo supere el valor \$1030.

13. *a*) Los artículos dispuestos en una cinta transportadora son analizados por un robot que detecta los artículos defectuosos. Si el número de artículos defectuosos que detecta el robot en un período de una hora sigue una ley de Poisson a razón de 5 artículos por hora, calcule la probabilidad de que en un cuarto de hora el robot detecte a lo sumo un artículo defectuoso.

    b) ¿Qué distribución de probabilidad tiene la variable aleatoria tiempo que transcurre desde que el robot detecte un artículo defectuoso hasta que detecte el siguiente? Justifique.

    c) Calcule la probabilidad de que dicho tiempo sea superior a 10 minutos.

9. Bibliografía

1. Devore, J. (2005). *Probabilidad y Estadística para Ingeniería y Ciencia.* México: Thomson Editores.

2. García, R. y Mermoz, O. (2006). *Distribuciones univariantes de probabilidad. Modelos y su identificación.* Buenos Aires: Nueva Librería.

3. Meyer, P. (1992). *Probabilidad y aplicaciones estadísticas* México: Addison-Wesley Iberoamericana.

4. Milton, J. y Arnold, J. (1986). *Probabilidad y Estadística con Aplicaciones para Ingeniería y Ciencias Computacionales* México: McGraw-Hill.

5. Montgomery, D. y Runger, G. (1996). *Probabilidad y Estadística aplicadas a la Ingeniería.* México: Mc-Graw-Hill.

6. Walpole, R., Myers, R. y Myers, S. (1999) *Probabilidad y estadística para ingenieros.* México: Pearson Educación.