Distribuciones continuas

Definición

La función $f(x)$ es una **función de densidad de probabilidad** (fdp) para la variable aleatoria continua X, definida en el conjunto de números reales, si

1. $f(x) \geq 0$, para toda $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Definición

$F(x)$, de una variable aleatoria continua X, con función de densidad $f(x)$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \text{ para } -\infty < x < \infty.$$

Como consecuencia inmediata se pueden escribir las siguientes igualdades

$$P(a < X < b) = F(b) - F(a) \text{ y } f(x) = \frac{dF(x)}{d(x)},$$

si existe la derivada.

Definición

Sea X una variable aleatoria con distribución de probabilidad $f(x)$. La **media** o **valor esperado** de X es

$$\mu = E(X) = \sum_{x} xf(x)$$

si X es discreta, y

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

si X es continua.

Teorema

Sea X una variable aleatoria con distribución de probabilidad $f(x)$. El valor esperado de la variable aleatoria $g(X)$ es

$$\mu_{g(X)} = E[g(X)] = \sum_{x} g(x)f(x)$$

si X es discreta, y

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

si X es continua.

Teorema

La varianza de una variable aleatoria X es

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

Distribución uniforme continua.

La función de densidad de la variable aleatoria uniforme continua X en el intervalo uniforme $[A, B]$ es

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B, \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución acumulada $F(x)$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{A} \underbrace{f(t)\,dt}_{0} + \int_{A}^{x} f(t)\,dt$$

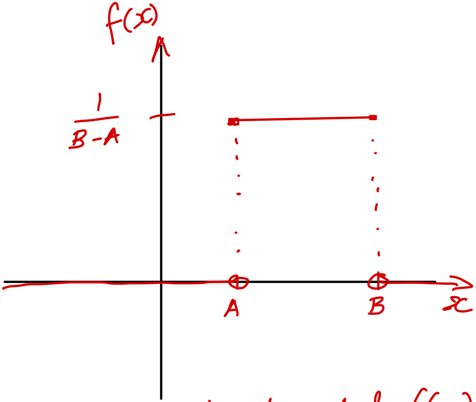
$$= \cancel{\int_{-\infty}^{A} 0\,dt} + \int_{A}^{x} \frac{1}{B-A}\,dt$$

$$= \frac{1}{B-A}\int_{A}^{x} dt = \frac{1}{B-A}\, t\,\Big]_{A}^{x}$$

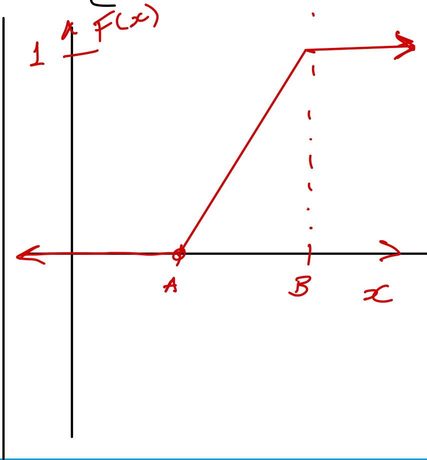
$$= \frac{1}{B-A}(x-A) = \frac{x-A}{B-A}; \quad A \leq x \leq B.$$

Gráfica de $f(x)$ y $F(x)$

$$F(x) = \begin{cases} 0 & \text{si } x < A \\ \dfrac{x-A}{B-A} & \text{si } A \leq x \leq B \\ 1 & \text{si } x > B \end{cases}$$



función de densidad. $f(x)$



$$E[x] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \cancel{\int_{-\infty}^{A} x\, f(x)\, dx}^{\,0} + \int_{A}^{B} x\, f(x)\, dx + \cancel{\int_{B}^{\infty} x\, f(x)\, dx}^{\,0}$$

$$= \int_{A}^{B} x\, f(x)\, dx = \int_{A}^{B} x \left(\frac{1}{B-A}\right) dx$$

$$= \frac{1}{B-A} \int_{A}^{B} x\, dx = \frac{1}{B-A} \left.\frac{x^2}{2}\right]_{A}^{B} = \frac{1}{2(B-A)} (B^2 - A^2)$$

$$= \frac{\cancel{(B-A)}(B+A)}{2\cancel{(B-A)}} = \frac{B+A}{2} ;$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \cancel{\int_{-\infty}^{A} x^2 f(x)\,dx}^{\,0} + \int_{A}^{B} x^2 f(x)\,dx + \cancel{\int_{B}^{\infty} x^2 f(x)\,dx}^{\,0}$$

$$= \int_A^B x^2 f(x)\,dx = \int_A^B x^2 \left(\frac{1}{B-A}\right) dx$$

$$= \frac{1}{B-A}\int_A^B x^2\,dx = \frac{1}{B-A}\,\frac{x^3}{3}\Bigg]_A^B = \frac{1}{3(B-A)}\,B^3 - A^3$$

$$= \frac{1}{3\cancel{(B-A)}}\cancel{(B-A)}(B^2+BA+A^2) = \frac{B^2+BA+A^2}{3}$$

$$V(X) = E(X^2) - \mu^2 = \frac{B^2+BA+A^2}{3} - \left(\frac{B+A}{2}\right)^2$$

$$= \frac{B^2+BA+A^2}{3} - \frac{B^2+2BA+A^2}{4} = \frac{\cancel{4B^2}+\cancel{4BA}+\cancel{4A^2}-\cancel{3B^2}-\cancel{6BA}-\cancel{3A^2}}{12}$$

$$= \frac{B^2-2BA+A^2}{12} = \frac{(B-A)^2}{12}$$

$$Var[x] = \frac{(B-A)^2}{12}$$

Distribución normal.

La densidad de la variable aleatoria normal X, con media μ y varianza σ^2, es

$$n(x;\, \mu,\, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \; -\infty < x < \infty,$$

donde $\pi = 3{,}14159\cdots$ y $e = 2{,}71828\cdots$

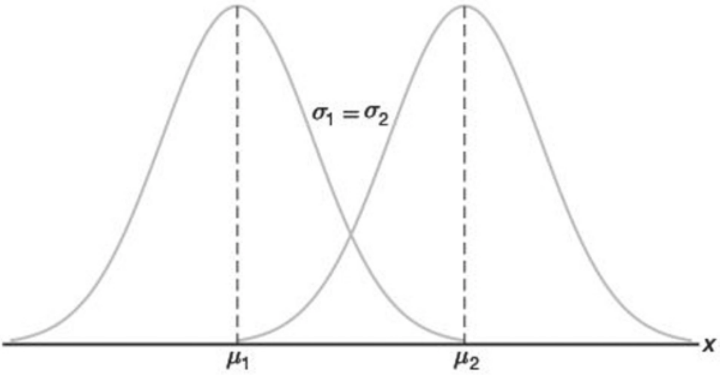


Figura 6.3: Curvas normales con $\mu_1 < \mu_2$ y $\sigma_1 = \sigma_2$.

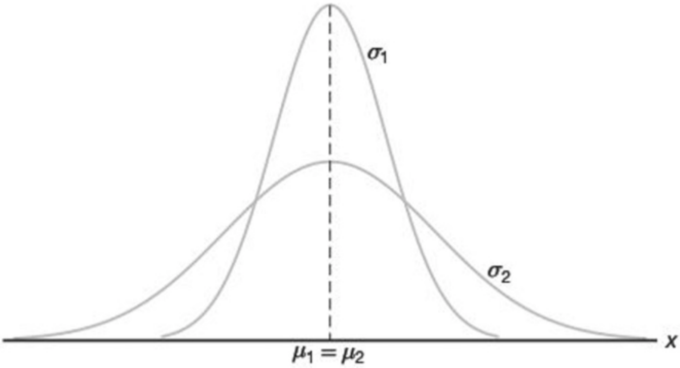


Figura 6.4: Curvas normales con $\mu_1 = \mu_2$ y $\sigma_1 < \sigma_2$.

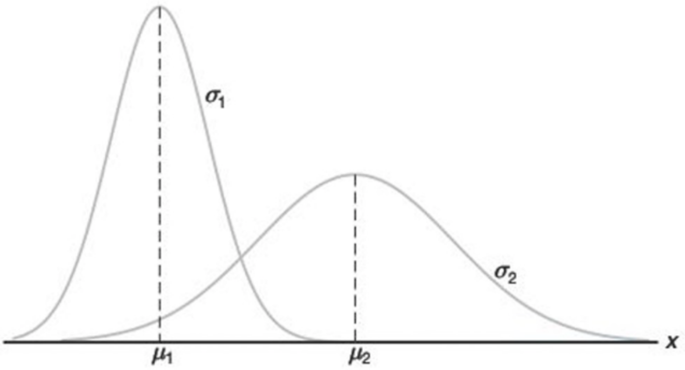
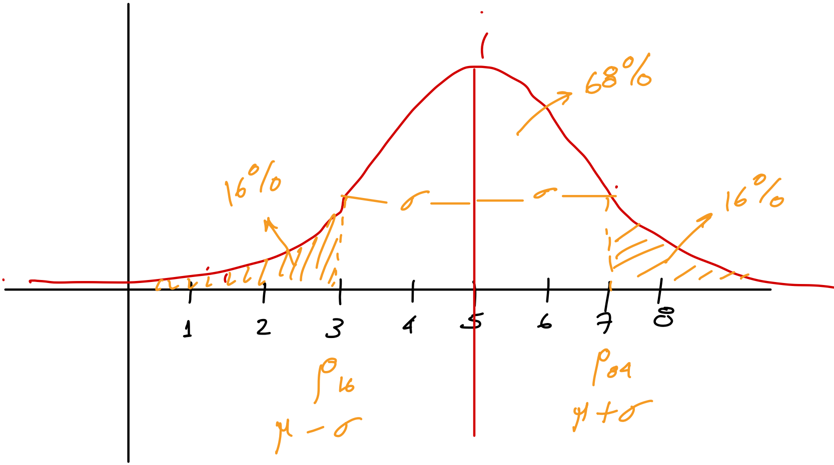


Figura 6.5: Curvas normales con $\mu_1 < \mu_2$ y $\sigma_1 < \sigma_2$.

Podemos listar las siguientes propiedades de la curva normal:

1. La moda, que es el punto sobre el eje horizontal donde la curva su punto máximo, ocurre en $x = \mu$.
2. La curva es simétrica alrededor de un eje vertical a través de la media μ.
3. La curva tiene sus puntos de inflexión en $x = \mu \pm \sigma$, es cóncava hacia abajo si $\mu - \sigma < X < \mu + \sigma$, y es cóncava hacia arriba en otro caso.
4. La curva normal se aproxima al eje horizontal de manera asintótica, conforme nos alejamos de la media en cualquier dirección.
5. El área total bajo la curva y sobre el eje horizontal es igual a uno.

Gráfica: $\mu = 5,\ \sigma = 2;\ P_{16} = 3,\ P_{84} = 7$



Teorema

La media y la varianza de $n(x; \mu, \sigma)$ son μ y σ^2, respectivamente. Por lo tanto, la desviación estándar es σ.

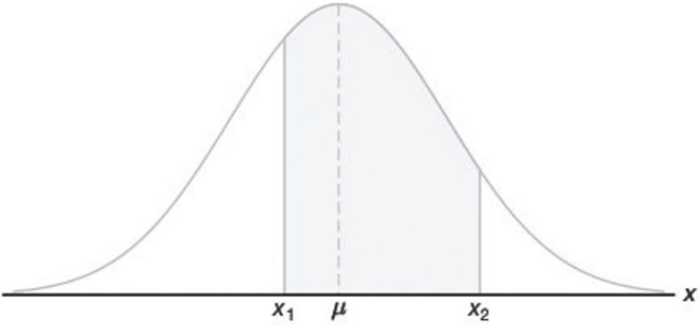


Figura 6.6: $P(x_1 < X < x_2)$ = área de la región sombreada.

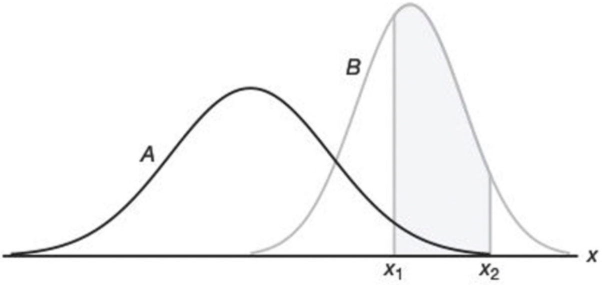


Figura 6.7: $P(x_1 < X < x_2)$ para diferentes curvas normales.

Definición

La distribución de una variable aleatoria normal con media 0 y varianza 1 se llama **distribución normal estándar.**

Teorema

Si X es una variable aleatoria binomial con media $\mu = np$ y varianza $\sigma^2 = npq$, entonces la forma limitante de la distribución de

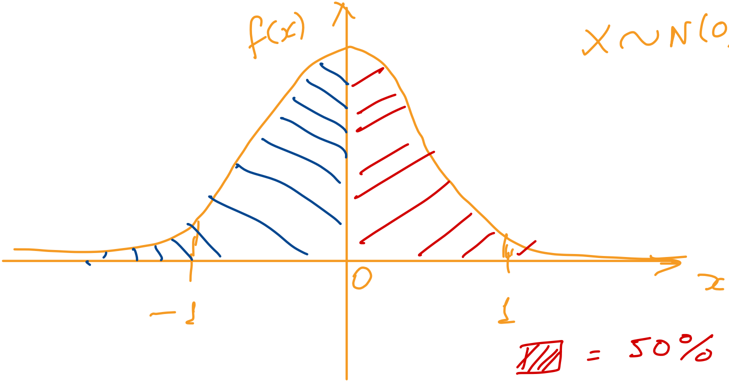
$$Z = \frac{X - np}{\sqrt{npq}},$$

conforme $n \to \infty$, es la distribución normal estándar $n(z; 0, 1)$.

Sea X una variable aleatoria binomial con parámetros n y p. Para una n grande, X tiene aproximadamente una distribución normal con $\mu = np$ y $\sigma^2 = npq = np(1-p)$ y

$$\begin{aligned}
P(X \leq x) &= \sum_{k=0}^{x} b(k;\, n,\, p) \\
&\approx \text{área bajo la curva normal a la izquierda de } x + 0{,}5 \\
&= P\left(Z \leq \frac{x + 0{,}5 - np}{\sqrt{npq}}\right),
\end{aligned}$$

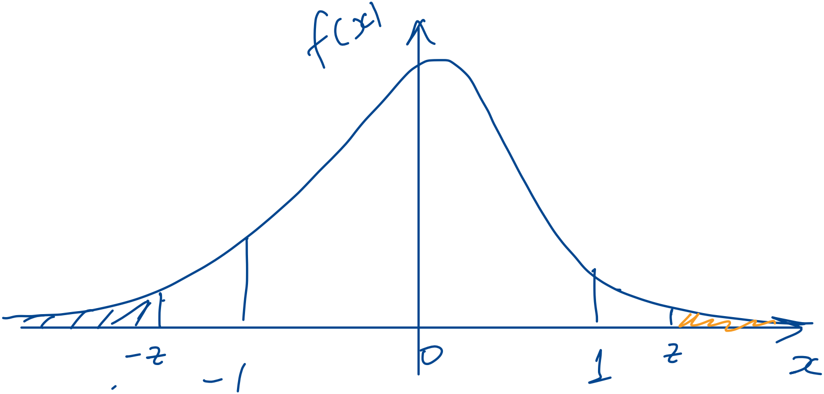
y la aproximación srá buena si np y $n(1-p)$ son mayores o iguales a 5.



$X \sim N(0,1)$

$= 50\%$

$= 50\% = 0.5 = \frac{1}{2}$



$=$ Gracias a la simetría de la distribución normal

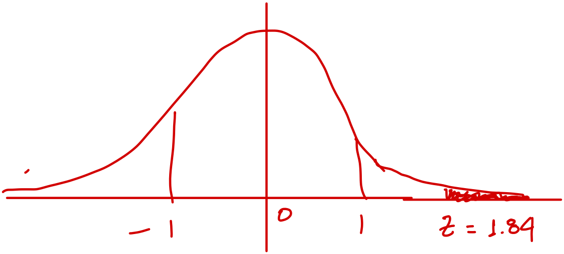
$= P(Z < -z)$; $= P(Z > z)$

Ejemplo

Dada una distribución normal estándar, calcule el área bajo la curva que se localiza.

1. a la derecha de $z = 1{,}84$.
2. entre $z = -1{,}97$ y $z = 0{,}86$.

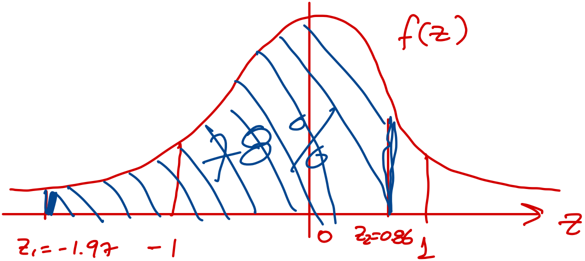
1.



$$P(Z > 1.84) = 1 - P(Z < 1.84)$$
$$= 1 - \text{pnorm}(1.84)$$
$$= 1 - 0.9671$$
$$= 0.0329$$

$$P(Z > 1.84) = P(Z < -1.84)$$
$$= \text{pnorm}(-1.84)$$
$$= 0.0329$$

2



$$P(-1.97 < Z < 0.86)$$

$$= F(0.86) - F(-1.97)$$

$$= \text{pnorm}(0.86) - \text{pnorm}(-1.97)$$

$$= 0.8051 - 0.0244$$

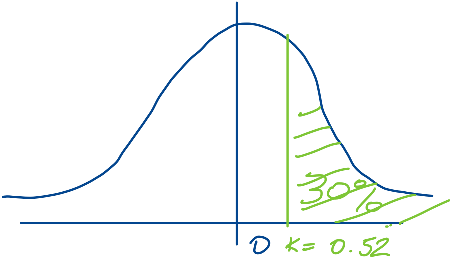
$$= 0.7807$$

Ejemplo

Dada una distribución normal estándar, calcule el valor de k tal que:

1. $P(z > k) = 0{,}3015$.
2. $P(k < z < -0{,}18) = 0{,}4197$.

1



$$P(Z > k) = 0.3015$$

$$\Rightarrow P(Z < -k) = 0.3015$$

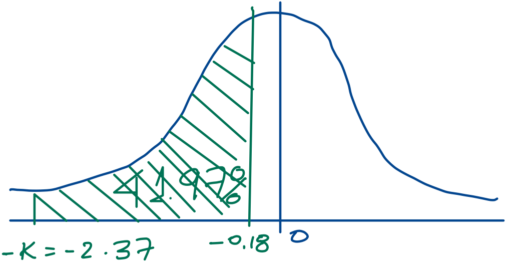
$$\Rightarrow \text{qnorm}(0.3015) = -k$$

$$-0.5200912 = -k$$

$$\Rightarrow$$

$$\Rightarrow k = 0.5200912$$

2.



$$P(Z < -0.18) - 0.4197 = P(Z < -K)$$

$$F(-0.18) - 0.4197 = F(-K)$$

pnorm(-0.18)

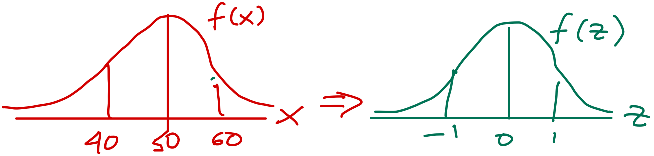
$$0.4285 - 0.4197 = F(-K)$$

$$0.0088 = F(-K)$$

$$\Rightarrow \quad -K = \text{qnorm}(0.0088)$$

$$-K = -2.37$$

Ejemplo



Dada $X \sim n(50, 10)$ es decir $E(X) = 50$ y $Var(X) = 10^2 = 100$.

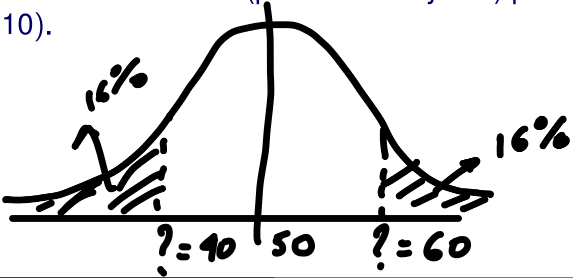
$$
\begin{aligned}
P(45 < X < 62) &= P\left(\frac{45-50}{10} < \frac{X-50}{10} < \frac{62-50}{10}\right) \\
&= P\left(\frac{-5}{10} < z < \frac{12}{10}\right) \\
&= P(-0{,}5 < z < 1{,}2) \\
&= F(1{,}2) - F(-0{,}5) \\
&= \underset{\text{pnor}(1.2)}{0{,}8849} - \underset{\text{pnor}(-0.5)}{0{,}3085} \\
&= 0{,}5764
\end{aligned}
$$

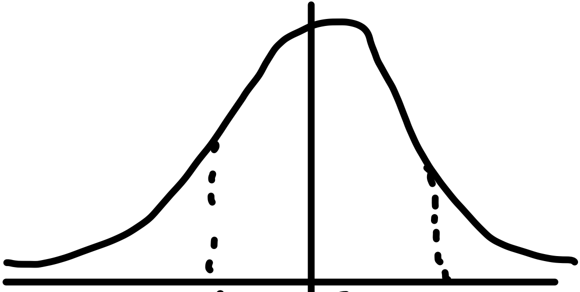
Si $X \sim N(100, 15)$ ¿$P(80 < X < 120)$?

pnorm(120,100,15) - pnorm(80,100,15) = 0.8175

Ejemplo

Hallar los puntos de inflexión (percentil 16° y 84°) para $X \sim n(50, 10)$.





qnorm(0.16, 50, 10)   50   qnorm(0.84, 50, 10)

$= 40.05$ $= 59.94$

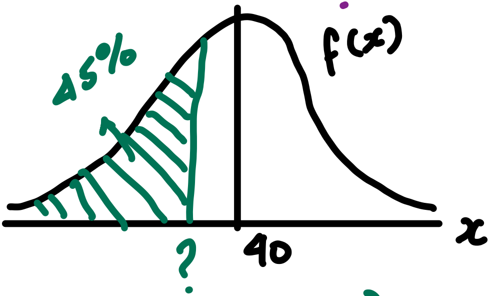
≈ 40 ≈ 60

Ejemplo

Dada una distribución normal con $\mu = 40$ y $\sigma = 6$, calcular el valor de x que:

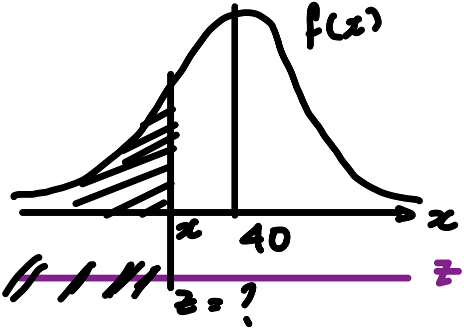
1. tiene 45 % del área a la izquierda y
2. 14 % del área a la derecha.

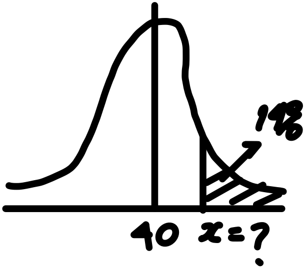
1



$$? = \text{qnorm}(0.45, 40, 6) = \underline{39.24}$$

Si: $$z = \frac{x-\mu}{\sigma} \iff x = z\sigma + \mu$$





$z = \text{qnorm}(0.45) = -0.1256613$

$\Rightarrow x = z\sigma + \mu$
$= (-0.1256613)\,6 + 40$
$= 39.24603$

$x = \text{qnorm}(0.86, 40, 6)$
$= 46.48$

Definición

Se dice que X tiene una **distribución exponencial** con parámetro $\lambda\,(\lambda > 0)$ si la función de densidad de probabilidad de X es

$$f(x;\,\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{de lo contrario} \end{cases}$$

cuya función de distribución acumulada es

$$F(x;\,\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Además $\mu = \frac{1}{\lambda}$ y $\sigma^2 = \frac{1}{\lambda^2}$, así $\mu = \sigma$.

Ejemplo

Sea X igual al tiempo entre dos llegadas sucesivas a la ventanilla de autopago de un banco local. Si X tiene una distribución esponencial con $\lambda = 1$ (la cual es idéntica a la sitribución gamma estándar con $\alpha = 1$), calcule lo siguiente:

1. El tiempo esperado entre dos llegadas sucesivas.
2. La desviación estándar del tiempo entre llegadas sucesivas.
3. $P(X \leq 4)$
4. $P(2 \leq X \leq 5)$

1) $E(X) = \frac{1}{\lambda} = \frac{1}{1} = 1$. El tiempo esperado entre dos llegadas sucesivas es 1 unidad de tiempo.

2) Como $X \sim Exp(\lambda) \Rightarrow E(X) = \sigma$, así que $\sigma = 1$.

3) $P(X \leq 4) = F(4) = 1 - \exp\{-4\}$
$= 0.9817$

4) $P(2 \leq X \leq 5) = F(5) - F(2)$
$= pexp(5,1) - pexp(2,1)$
$= 0.1286$

Ejemplo

Los datos recogidos en el Aeropuerto Internacional Toronto Pearson sugiere que una istribución exponencial con valor medio de 2.725 horas es un modelo para la duración de la lluvia (*Urban Stormwater Management Planning with Analytical Probabilistic Models*, 2000, p. 69).

¿Cuál es la probabilidad de que la duración de un evento de lluvia en este lugar particular, sea por lo menos 2 horas? ¿A lo más 3 horas? ¿Entre 2 y 3 horas?

$E(X) = 2.725$ horas

$X \sim Exp\left(\frac{1}{2.725}\right)$

$X \sim Exp(0.3669)$

$P(X \geq 2) = 1 - P(X < 2)$

$= 1 - F(2)$

$= 1 - \text{pexp}(2, 0.3669)$

$= 0.48$

$P(X \leq 3) = F(3)$

$\text{pexp}(3, 0.3669)$

$= 0.667$

$P(2 < X < 3)$

$= F(3) - F(2)$

$= \text{pexp}(3, 0.3669) - \text{pexp}(2, 0.3669)$

$= 0.147$

¿Cuál es la probabilidad de que la duración de la lluvia supere el valor medio por más de dos desviaciones estándar? ¿Cuál es la probabilidad de que sea menor que el valor medio en más de una desviación estándar?

$$P(X > \mu + 2\sigma) = P\left(X > \frac{1}{\lambda} + 2\left(\frac{1}{\lambda}\right)\right)$$

$$= P\left(X > 3\frac{1}{\lambda}\right)$$

$$= P\left(X > \frac{3}{0.3669}\right)$$

$$= P(X > 8.175)$$

$$= 1 - F(8.175)$$

$$= 1 - \text{pexp}(8.175, 0.369)$$

¿Cuál es la probabilidad de que la duración de la lluvia supere el valor medio por más de dos desviaciones estándar? ¿Cuál es la probabilidad de que sea menor que el valor medio en más de una desviación estándar?

$$P(X > \mu + 2\sigma) = 1 - 0.9501834$$
$$= 0.0498 \approx 0.05$$

$$P(X < \mu + \sigma) = P\left(X < \frac{1}{\lambda} + \frac{1}{\lambda}\right)$$
$$= P\left(X < 2\left(\frac{1}{\lambda}\right)\right)$$
$$= P\left(X < \frac{2}{\lambda}\right)$$

¿Cuál es la probabilidad de que la duración de la lluvia supere el valor medio por más de dos desviaciones estándar? ¿Cuál es la probabilidad de que sea menor que el valor medio en más de una desviación estándar?

$$P\left(X < \frac{2}{\lambda}\right) = P\left(X < \frac{2}{0.3669}\right)$$

$$= P(X < 5.451077)$$

$$= F(5.451077)$$

$$= 0.9957083$$