Introducción a la Probabilidad y Estadística - Probabilidad y Estadística

Grisel Britos

2023

VARIABLES ALEATORIAS CONTINUAS

En la vida diaria nos encontramos con variables que no son discretas, como por ejemplo: la cantidad diaria de lluvia caída en cierta región, el tiempo de vida de cierto tipo de componentes electrónicos, la profundidad en un lago para puntos aleatoriamente elegidos, etc. En todos estos ejemplos los valores posibles de las variables no es un conjunto finito ni infinito numerables, son intervalos o semirectas.

Definición: Sea X una variable aleatoria sobre un modelo probabilístico $(\Omega, \mathcal{A}, P)$. Se dice que X es una **variable aleatoria continua** si

$$P(X = x) = 0 \quad \forall x \in \mathbb{R}.$$

Dada X variable aleatoria sobre un modelo probabilístico $(\Omega, \mathcal{A}, P)$, recordemos la definición de función de distribución acumulada (f.d.a.) de X:

$$F(x) = P(X \leq x) \quad \forall x \in \mathbb{R}.$$

Entonces, de la definición de una variable aleatoria continua, resulta que:

i) $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$ $\forall a < b$ en $\mathbb{R}$.

ii) $P(X \geq a) = P(X > a) = 1 - F(a) \quad \forall a \in \mathbb{R}$.

Observación: Algunas de estas igualdades podrían no ser verdaderas si la variable aleatoria es discreta.

Daremos una proposición que será muy útil para probar las propiedades de la f.d.a. (F).

Proposición: Sea $(\Omega, \mathcal{A}, P)$ un modelo probabilístico. Entonces:

a) Si $\{A_n\}_{n=1}^{\infty}$ es una sucesión creciente de eventos en $\mathcal{A}$, esto es $A_1 \subseteq A_2 \subseteq A_3 \subseteq ...$, entonces

$$P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} P(A_n)$$

b) Si $\{B_n\}_{n=1}^{\infty}$ es una sucesión decreciente de eventos en $\mathcal{A}$, esto es $B_1 \supseteq B_2 \supseteq B_3 \supseteq ...$, entonces

$$P(\bigcap_{i=1}^{\infty} B_i) = \lim_{n\to\infty} P(B_n)$$

Propiedades de la f.d.a. (F) de una variable aleatoria continua

Sea $(\Omega, \mathcal{A}, P)$ un modelo probabilístico y X una variable aleatoria con f.d.a. $F : \mathbb{R} \to [0, 1]$, entonces

1) F es monótona creciente, o sea, si $x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$.
2) $\lim_{x\to\infty} F(x) = 1$ y $\lim_{x\to-\infty} F(x) = 0$.
3) $\forall a \in \mathbb{R}$ entonces $\lim_{x\to a^+} F(x) = F(a)$.
4)
$$\lim_{x\to a^-} F(x) = \lim_{x\to a^+} F(x) = F(a) \quad \forall a \in \mathbb{R}$$
Luego, F es continua en $\mathbb{R}$.

La siguiente definición nos permitirá realizar el cálculo de probabilidades de algunas variables aleatorias:

Definición: Se llama **función densidad de probabilidad** (f.d.p.) a toda función $f : \mathbb{R} \to \mathbb{R}$ no negativa tal que $\int_{-\infty}^{\infty} f(t)dt = 1$.

A continuación estudiaremos un subconjunto importante de variables aleatorias que reciben el nombre de **variables aleatorias absolutamente continuas.**

Definición: Sea X una variable aleatoria continua con f.d.a. F y sea f una función de densidad de probabilidad (f.d.p.) tal que

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad \forall x \in \mathbb{R}$$

Entonces se dice que X es una **variable aleatoria absolutamente continua** o, equivalentemente, que F es función de distribución acumulada absolutamente continua.

Luego, si X es una variable aleatoria absolutamente continua entonces $\forall a, b \in \mathbb{R}$ tal que $a < b$ se tiene que

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

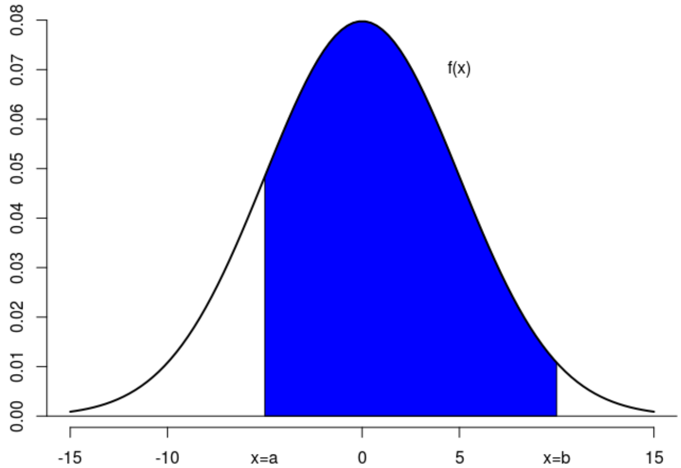
Proposición: Si X es una v.a. con f.d.a. F tal que tenga derivada continua, salvo en un conjunto finito de valores, entonces X es absolutamente continua y su f.d.p. está dada por

$$f(x) = \begin{cases} F'(x) & \forall x \text{ donde exista la derivada} \\ 0 & \text{c.c.} \end{cases}$$

En este curso nos concentraremos en trabajar con variables aleatorias absolutamente continuas.

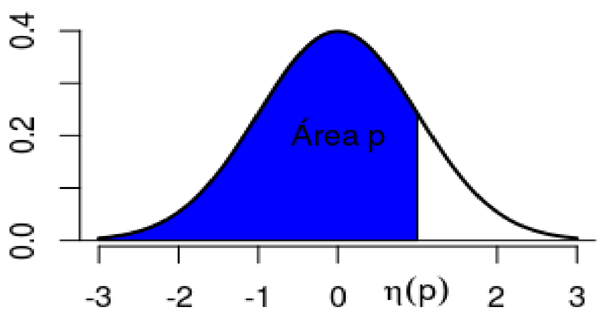
Definición: Sea $p \in (0, 1)$ y X una variable aleatoria continua sobre $(\Omega, \mathcal{A}, P)$ con F f.d.a., entonces se llama **percentil** ($p100$) o **cuantil** p de X al valor $\eta(p)$ tal que

$$p = F(\eta(p))$$



Observación: Si f es la f.d.p de X, entonces $\boldsymbol{p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(t)dt}$

Es decir, si la gráfica de la f.d.p. de X es la siguiente



el área sombreada resulta ser igual a p.

Proposición: Sea X una variable aleatoria continua con f.d.a. F y cuantil p de X, $\eta_X(p)$, con $p \in [0,1]$. Si $Y = aX + b$ con $a \neq 0$ y $b \in \mathbb{R}$, entonces

$$\eta_Y(p) = \begin{cases} a\eta_X(p) + b & \text{si} \quad a > 0 \\ a\eta_X(1-p) + b & \text{si} \quad a < 0 \end{cases}$$

ESPERANZA Y VARIANZA DE UNA VARIBLE ALEATORIA CONTINUA

Definición: Sea X una variable aleatoria continua sobre $(\Omega, \mathcal{A}, P)$ con función densidad de probabilidad f. Se define el **valor esperado o valor medio o esperanza de X** a

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

siempre que $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$

Notación: $\mu = E(X)$.

¿Cómo calcular la esperanza de una función de una variable aleatoria continua X sin necesidad de hallar su f.d.p.?

Proposición: Sea X una variable aleatoria continua sobre $(\Omega, \mathcal{A}, P)$ con f.d.p. f. Si $h(X)$ es una variable aleatoria continua entonces

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

siempre que $\int_{-\infty}^{\infty} |h(x)| f(x) dx < \infty$

Definición: Sea X una variable aleatoria continua con f.d.p. f entonces se define la **varianza de X** como

$$V(X) = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

siempre que $E(X^2) < \infty$ y donde $\mu = E(X)$.

Se llama **desviación estándar de X** a la raíz cuadrada de la varianza.
Notación: denotamos con σ^2 a la varianza y con σ al desvío estándar.

Proposición: Sea X una variable aleatoria continua con f.d.p. f y $E(X^2) < \infty$ entonces

a) $0 \leq V(X) = E(X^2) - \mu^2$ donde $\mu = E(X)$.

b) $E(aX + b) = aE(X) + b$ y $V(aX + b) = a^2 V(X)$; para todo $a, b \in \mathbb{R}$

Estudiaremos ahora algunas de las distribuciones más conocidas para las variables aleatorias continuas.

Distribución Uniforme

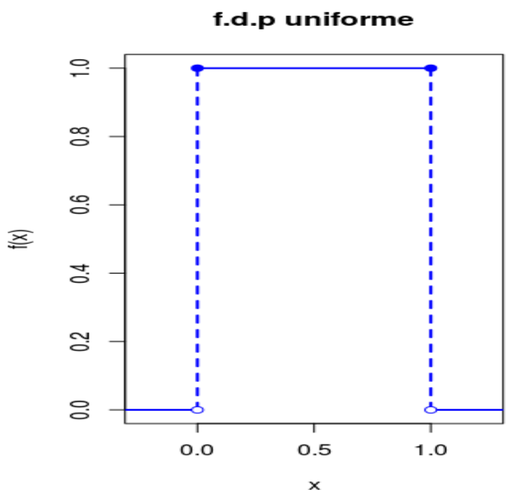
Definición: Sea X una variable aleatoria continua, diremos que tiene **distribución Uniforme en el intervalo $[a, b]$** si su función densidad de probabilidad es

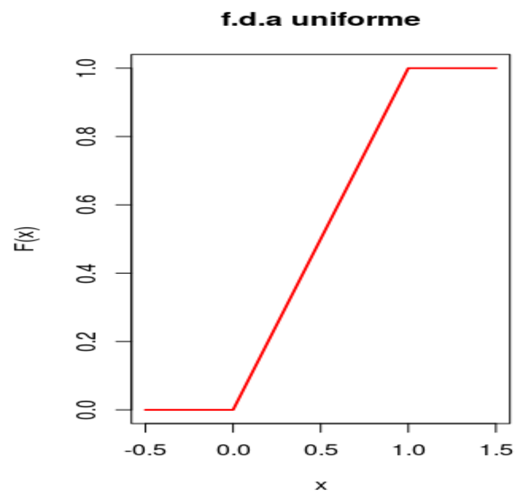
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a, b] \\ 0 & \text{c.c.} \end{cases}$$

Notación: $X \sim U[a, b]$

Si $X \sim U[a,b]$, su f.d.a. está dada por

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





Un caso particular de gran interés es cuando $a = 0$ y $b = 1$ ($X \sim U[0,1]$)

Proposición: Si $X \sim U[0,1]$ entonces:

a) $\eta_X(p) = p$ con $p \in (0,1)$

b) $E(X) = \frac{1}{2}$ y $V(X) = \frac{1}{12}$

c) Para cualquier $a < b$ números reales, sea la v.a. $Y = (b-a)X + a$. Entonces

   i) $Y \sim U[a,b]$

   ii) $\eta_Y(p) = (b-a)p + a$ para cualquier $p \in (0,1)$

   iii) $E(X) = \frac{a+b}{2}$ y $V(X) = \frac{(b-a)^2}{12}$

Ejemplo 1: El tiempo (en minutos) que tarda un camión en realizar un viaje (ida y vuelta), para transportar concreto hacia una obra en construcción, tiene distribución uniforme en el intervalo $[50,70]$.

a) ¿Cuál es la probabilidad de que la duración del viaje de un camión sea menor de 65 minutos?

b) ¿Cuál es la probabilidad de que la duración del viaje de un camión sea mayor de 65 minutos dado que se sabe que fue mayor a 55 minutos?

c) Dar el valor medio y desviación estándar del tiempo que tarda un camión en realizar un viaje (ida y vuelta) para transportar concreto hacia la obra en construcción.

d) Suponga que los tiempos que tardan cada uno de tres camiones en realizar un viaje (ida y vuelta) para transportar concreto hacia la obra en construcción son independientes entre sí. ¿Cuál es la probabilidad de que por lo menos uno de ellos tarde más de 65 minutos?

Distribución Normal o Gaussiana

Esta distribución es una de las más importantes dentro de la Teoría de la Probabilidad y Estadística. Muchas variables pueden ser aproximadas por esta distribución, como ser: errores de medición en experimentos científicos, calificaciones en diversas pruebas, mediciones antropométricas, tiempo de reacción en experimentos, etc.

Definición: Dada una variable aleatoria continua X, diremos que tiene **distribución Normal de parámetros μ y σ^2** si su función densidad de probabilidad es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$

Notación: $X \sim N(\mu, \sigma^2)$.

Propiedades de la f.d.p. de una $N(\mu, \sigma^2)$

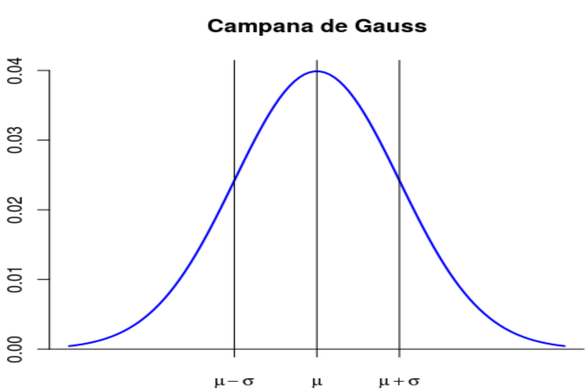
a) $f(x) \geq 0\ \forall x \in \mathbb{R}$ y $\int_{-\infty}^{\infty} f(x)dx = 1$

b) f es simétrica respecto de μ, o sea $f(\mu - x) = f(\mu + x)\ \forall x \in \mathbb{R}$.

c) f tiene un punto de máximo en μ y puntos de inflexión en $(\mu \pm \sigma)$.

d) $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$

e) La gráfica tiene forma de campana.

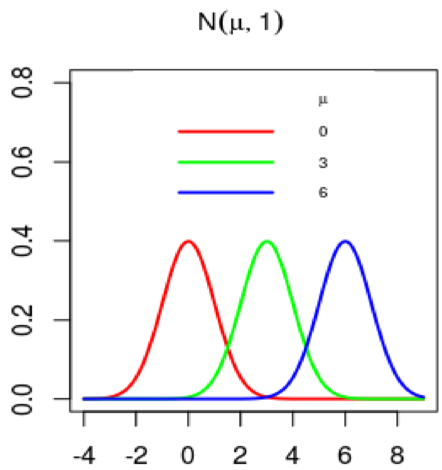


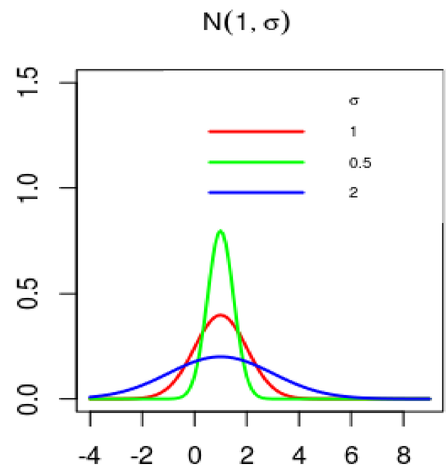
Observación: Dentro de esta familia de distribuciones, un caso particular es cuando $\mu = 0$ y $\sigma = 1$, distribución que recibe el nombre de **normal estándar** y se simboliza como $Z \sim N(0, 1)$.
La f.d.a. de Z se la denota con Φ y es

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \forall z \in \mathbb{Z}$$

Proposición: Si $Z \sim N(0, 1)$ entonces $E(Z) = 0$ y $V(Z) = 1$.

Para el cálculo de probabilidades de una normal estándar hay tablas generadas por métodos de integración numérica.





En la tabla que usaremos se encuentra tabulado el área sombreada, $\Phi(z) = P(Z \leq z)$, para valores de $z \in \mathbb{R}_{\geq 0}$.
Entonces $P(a < Z < b) = \Phi(b) - \Phi(a)$, $\forall a < b$.

Ejercicios para el manejo de la tabla $N(0,1)$

1) Calcular las siguientes probabilidades:

   a) $P(Z \leq 0)$
   b) $P(0 \leq Z \leq 2,57)$
   c) $P(-2,57 \leq Z \leq 2,57) = P(|Z| \leq 2,57)$
   d) $P(Z \geq -1,75)$
   e) $P(Z \geq 2,57)$

2) Determinar el valor de la constante c tal que:

   a) $\Phi(c) = 0,9838$
   b) $P(0 \leq Z \leq c) = 0,291$
   c) $P(|Z| \leq c) = 0,668$

3) Determinar los percentiles 25; 50 y 75.

Proposición: Si $X \sim N(\mu, \sigma^2)$ entonces

a) $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ (**Estandarización de X**)

b) $E(X) = \mu = \tilde{\mu}$ y $V(X) = \sigma^2$

Observaciones:

1. Análogamente al inciso (a) de la proposición anterior se prueba que si $Z \sim N(0,1)$ entonces

$$X = \sigma Z + \mu \sim N(\mu, \sigma^2)$$

2. Si $X \sim N(\mu, \sigma^2)$ entonces para calcular $P(a \leq X \leq b)$ primero se debe estandarizar la v.a. X y después usar la tabla de la $N(0,1)$:

$$P(a \leq X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

para todo $a < b$ números reales, donde Φ es la f.d.a. de una v.a. $N(0,1)$.

Ejemplo 2: En una empresa hay dos máquinas disponibles A y B que cortan corchos, para ser usados en botellas de vino, tal que el 40 % son cortados con la máquina A y el resto con la B. Se sabe que la máquina A corta corchos con diámetro que están normalmente distribuido con una media de 3 cm y una desviación estándar de 0,1 cm y la máquina B corta corchos con diámetro que están normalmente distribuido con una media de 3,04 cm y una desviación estándar de 0,02 cm. Para que un corcho sea aceptable, para su uso, debe cumplir que su diámetro esté comprendido entre 2,9 y 3,1 cm.

a) ¿Cuál de las dos máquinas producen mayor porcentaje de corchos aceptables?

b) Se selecciona al azar un corcho de la producción, ¿cuál es la probabilidad de que el corcho sea aceptable?

c) Para la máquina que produce mayor porcentaje de corchos aceptables:

- *i) ¿cuál es la probabilidad de que el diámetro esté comprendido entre 2,99 y 3,05 cm?*
- *ii) Si se seleccionan al azar 10 corchos de la producción ¿cuál es la probabilidad de que el diámetro esté comprendido entre 2,99 y 3,05 cm en por lo menos dos de los 10?*
- *iii) Hallar los percentiles 30 y 70 para la variable diámetro del corcho cortado con la máquina elegida.*

Distribución Gamma

No todas las variables aleatorias continuas tienen una función de densidad de probabilidad (f.d.p.) simétrica, un ejemplo de ello es la familia de distribuciones Gamma, las cuales son sesgadas a derecha. Antes de dar la definición de la f.d.p. Gamma, definiremos la función gamma, muy importante en muchas ramas de la matemática.

Definición: Para $\alpha > 0$, la función gamma evaluada en α está definida como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Propiedades de la función gamma

a) $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ para cualquier $\alpha > 1$.

b) $\Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}$

c) $\Gamma(1/2) = \sqrt{\pi}$

Definición: Se dice que una variable aleatoria X tiene **distribución Gamma con parámetros α y β** si la f.d.p. es

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{si} \quad x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$

para α y β números positivos.
Notación: $X \sim \Gamma(\alpha, \beta)$.

Proposición: Sea X una v.a. con distribución $\Gamma(\alpha, \beta)$ entonces:

$$E(X) = \mu = \alpha\beta \quad \text{y} \quad V(X) = \sigma^2 = \alpha\beta^2$$

Dentro de esta familia cabe mencionar dos distribuciones, una de ellas tiene múltiples aplicaciones y la otra es usada en inferencia estadística, ellas son la distribución exponencial y la Chi-cuadrado.

Distribución Exponencial

Definición: Se dice que una variable aleatoria X tiene **distribución exponencial de parámetro λ** si la f.d.p. es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$

para $\lambda > 0$.
Notación: $X \sim \mathcal{E}(\lambda) = \Gamma(1, 1/\lambda)$.

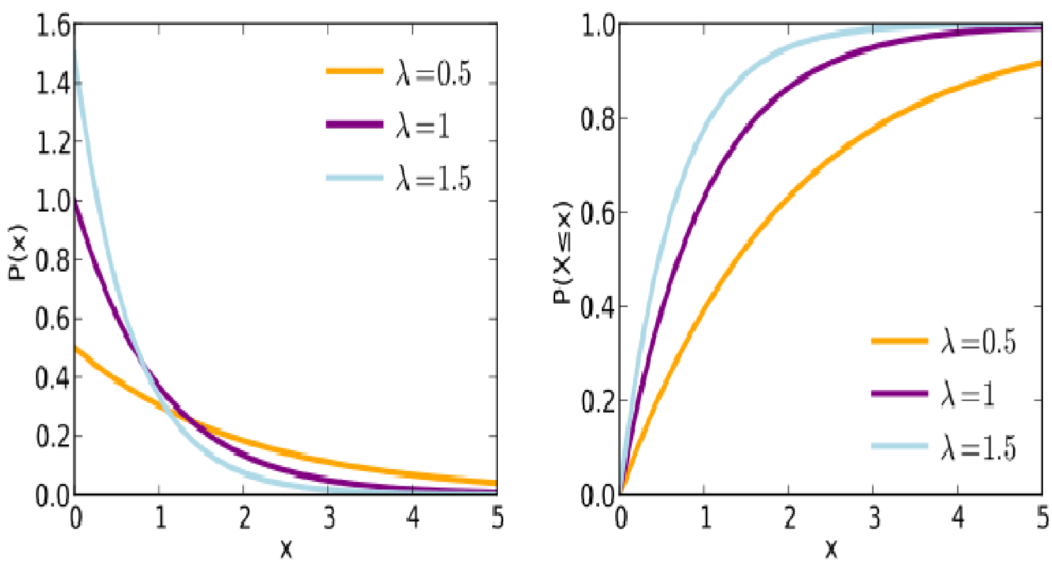


Figura 1: Izq: funciones de densidad para v.a. exponenciales. Der: funciones de distribucuón acumulada para v.a. exponenciales

Consecuencias:
Si $X \sim \mathcal{E}(\lambda)$ entonces

a) $E(X) = \mu = \frac{1}{\lambda}$ y $V(X) = \sigma^2 = \frac{1}{\lambda^2}$

b)
$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{si} \quad x \geq 0 \end{cases}$$

c) Si $t_0 \geq 0$ entonces $P(X \geq t + t_0 | X \geq t_0) = P(X \geq t) \quad \forall t > 0$.

Esta distribución es muy utilizada para modelar, por ejemplo, el tiempo de vida o duración de determinados componentes, con lo cual la propiedad enunciada en el ítem (c) (conocida como **“carencia**

de memoria") se refiere que no se tiene en cuenta el efecto desgaste producido por el paso del tiempo y, por lo tanto, puede no ser adecuada en algunas situaciones. Sin embargo, existen otras distribuciones alternativas como la distribución Weilbull y la lognormal.

Distribución Chi-cuadrado

Definición: Se dice que una variable aleatoria X tiene **distribución Chi-cuadrado de parámetro** k si la f.d.p. es

$$f(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & \text{si} \quad x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$

para $k > 0$.
Notación: $X \sim \chi^2(k) = \Gamma(k/2, 2)$.

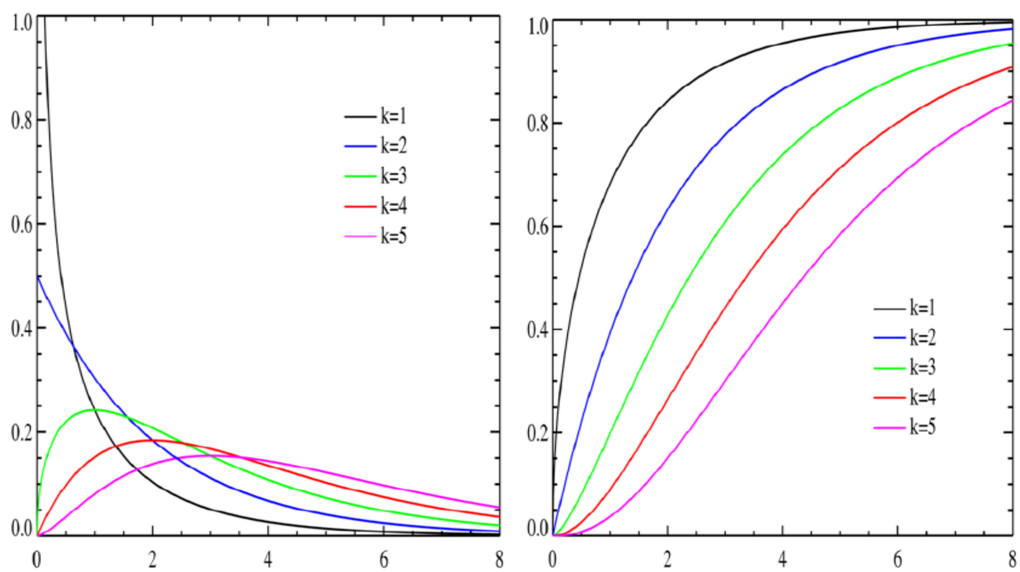


Figura 2: Izq: funciones de densidad para v.a. chi cuadrado. Der: funciones de distribucuón acumulada para v.a. chi cuadrado

Consecuencias: Si $X \sim \chi^2(k)$ entonces

$$E(X) = \mu = k \quad \text{y} \quad V(X) = \sigma^2 = 2k$$

Esta distribución es muy importante en procedimientos estadísticos (Intervalos de confianza y Pruebas de hipótesis para la varianza poblacional).

Proposición: Si $Z \sim N(0, 1)$ entonces $Z^2 \sim \chi^2(1)$.

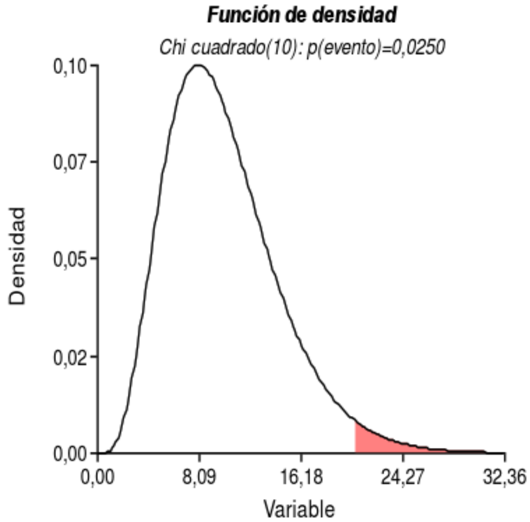
Ejercicios:

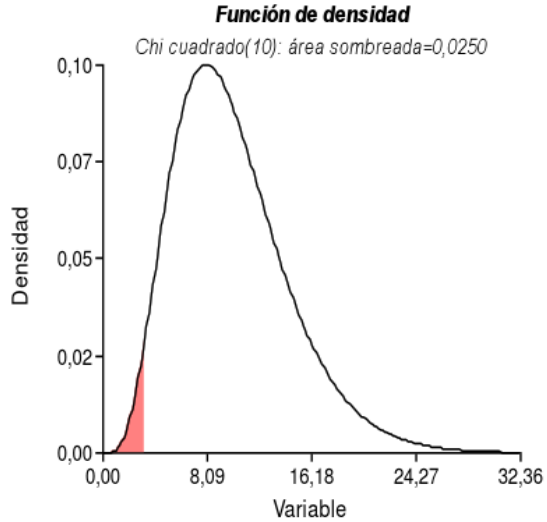
1) Hallar los puntos críticos para una $X \sim \chi^2(10)$ tales que

   a) el área a cola superior sea igual a 0,025.

   b) el área a cola inferior sea igual a 0,025

2) Calcular $P(3,25 \leq X \leq 23,2)$ con $X \sim \chi^2(10)$





Otras distribuciones continuas

Hay muchas situaciones prácticas donde ninguna de las distribuciones mencionadas ajustan adecuadamente al conjunto de datos observados. Vamos a mencionar dos que podrían ser apropiadas en algunas situaciones: distribución Weibull y la Log-normal.

Distribución Weibull

Definición: Se dice que una variable aleatoria X tiene **distribución Weibull con parámetros α y β** si su f.d.p. es

$$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(\frac{x}{\beta})^\alpha} & \text{si} \quad x > 0 \\ 0 & \text{c.c.} \end{cases}$$

para α y β números positivos.
Notación: $X \sim W(\alpha, \beta)$.

Proposición: Sea X una v.a. con distribución $W(\alpha, \beta)$ entonces:

a) Si $\alpha = 1$ entonces $W(1, \beta) = \mathcal{E}(1/\beta)$.

b) La función de distribución acumulada es:

$$F(x) = \begin{cases} 1 - e^{-(\frac{x}{\beta})^\alpha} & \text{si} \quad x > 0 \\ 0 & \text{c.c.} \end{cases}$$

c) $E(X) = \beta\Gamma(1 + \frac{1}{\alpha})$ y $V(X) = \beta^2[\Gamma(1 + \frac{2}{\alpha}) - \Gamma(1 + \frac{1}{\alpha})^2]$

Distribución Log-normal

Definición: Sea X una variable aleatoria continua entonces diremos que tiene **distribución Log-normal de parámetros μ y σ^2** si la variable aleatoria $Y = ln(X) \sim N(\mu, \sigma^2)$. Luego la f.d.p. de X es

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}} & \text{si} \quad x > 0 \\ 0 & \text{c.c.} \end{cases}$$

donde $\mu \in \mathbb{R}$ y $\sigma > 0$.
Notación: $X \sim lognormal(\mu, \sigma^2)$.

Proposición: Sea X una v.a. con distribución Log-normal de parámetros μ y σ^2 entonces:

a) La función de distribución acumulada de X es:

$$F(x) = \begin{cases} \Phi(\frac{ln(x)-\mu}{\sigma}) & \text{si} \quad x > 0 \\ 0 & \text{c.c.} \end{cases}$$

donde Φ es la función de distribución acumulada de la normal estándar.

b) $E(X) = e^{\mu + \frac{\sigma^2}{2}}$ y $V(X) = e^{2\mu+\sigma^2}[e^{\sigma^2} - 1]$