Probabilidad

Daniel Fraiman

Maestría en Ciencia de Datos, Universidad de San Andrés

Variables aleatorias

Existen

- variables aleatorias discretas.
- variables aleatorias continuas.
- variables aleatorias mixtas.

VARIABLES ALEATORIAS CONTINUAS

Variables aleatorias continuas

Ejemplos

- Tiempo de espera en el banco.
- Tiempo de duración de una bombita de luz.
- Máxima temperatura que se va a registrar el 1 de julio.

Variables aleatorias continuas

V.A. DISCRETAS	$\Longleftrightarrow$	V.A.CONTINUAS
$X = \{0, 1, 2\}$	$\Longleftrightarrow$	$X \in [0, 2]$
$X = \{14, 20, 23\}$	$\Longleftrightarrow$	$X \in [9, 13]$
$X \in \mathbb{N}$	$\Longleftrightarrow$	$X \in \mathbb{R}$
$P(X = k)$[proba puntual]	$\Longleftrightarrow$	$f_X(x)$[función de densidad]
$\sum_k P(X = k) = 1$	$\Longleftrightarrow$	$\int_{-\infty}^{\infty} f_X(t)dt = 1$

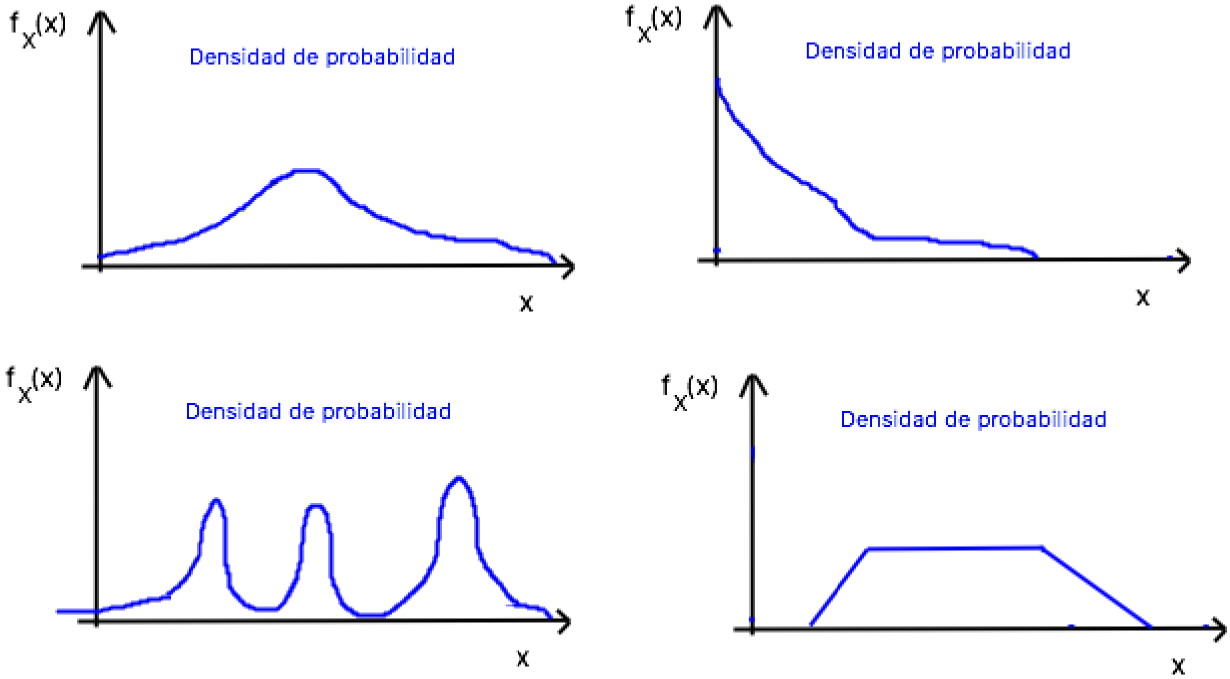
En variables aleatorias continuas, NO EXISTE mas la funcion de probabilidad puntual, sino una de DENSIDAD

Variables aleatorias continuas

¿Qué necesitamos saber de X para realmente comprenderla?

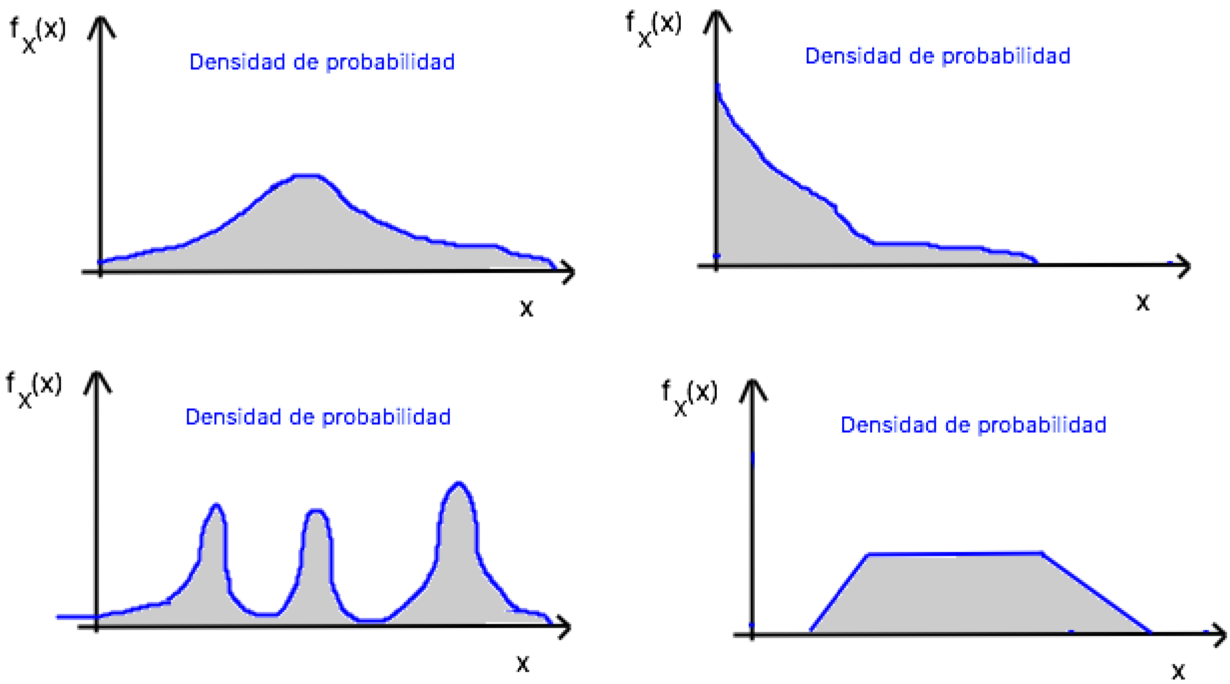
En este caso alcanza conocer la función de densidad $f_X(x)$, esta función contiene los valores que puede tomar la variable y con qué chances la v.a. puede caer en cualquier intervalo prefijado.

Variables aleatorias continuas



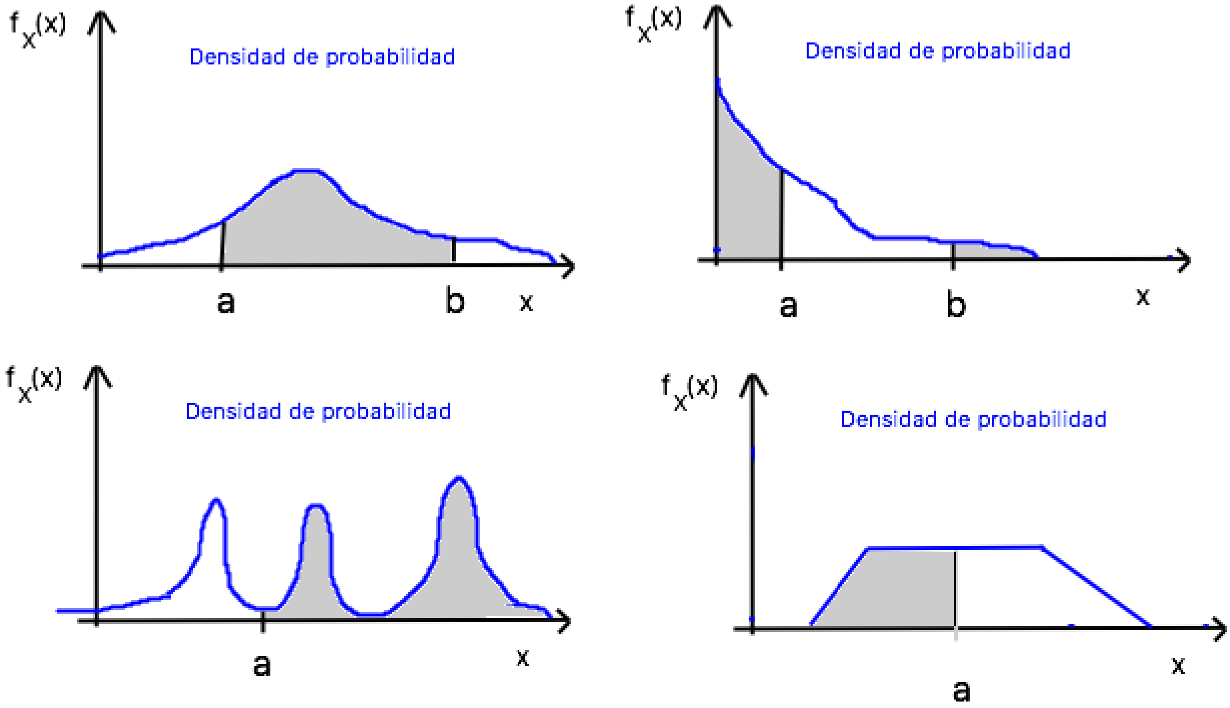
Lo que tienen en comun estas funciones de densidad, es que el AUC es 1 y siempre son positivas

Variables aleatorias continuas



AREA TOTAL =1

Variables aleatorias continuas



Variables aleatorias continuas

Si soy capaz de hacer el grafico de densidad de esa v.a. y no pego saltos, entonces esa v.a. es CONTINUA, si pega saltos, es DISCRETA

Definición:

Una v.a. X es continua si $F_X(x) = P(X \leq x)$ es continua $\forall x \in \mathbb{R}$.

- Dada una $F_X(x)$ existe una función $f_X : \mathbb{R} \to \mathbb{R}^{\geq 0}$ que cumple:

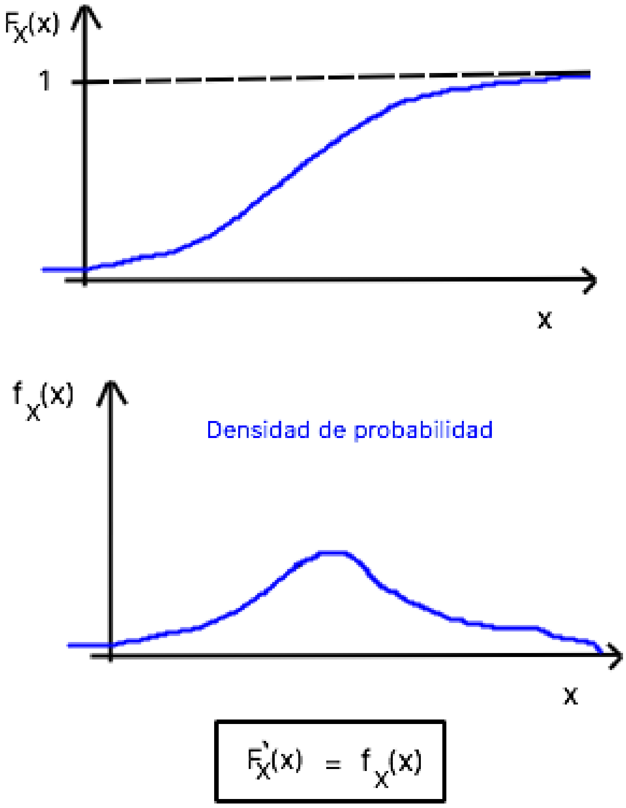
$$F_X(x) = \mathbb{P}\left(X \leq x\right) = \int_{-\infty}^{x} f_X(t)dt$$

- donde f_X es la función de densidad de la v.a. X.
- Notar: $f_X(x) \geq 0$ y puede ser mayor a 1.
- Por el Teorema Fundamental del Cálculo

$$F_X'(x) = f_X(x)$$

para todo punto x donde F sea derivable.

Variables aleatorias continuas



Variables aleatorias continuas

Propiedades:

- $\int_{-\infty}^{\infty} f_X(t)dt = 1 \quad$ (porque $\lim\limits_{x\to\infty} F_X(x) = 1$)
- $\mathbb{P}\left(a < X < b\right) = \mathbb{P}\left(a \leq X < b\right) = \mathbb{P}\left(a < X \leq b\right) = \mathbb{P}\left(a \leq X \leq b\right) = \int_a^b f_X(t)dt = F_X(b) - F_X(a)$
- $\mathbb{P}\left(X = c\right) = 0$
- Notar: En las variables continuas vamos a calcular probabilidades sobre intervalos.

DISCRETAS —> P(2<x<8)=P(3<=x<=7)=Fx(7)-Fx(3) [Fx—>Acumulada]
CONTINUAS—> P(2<x<8)=Fx(8) - Fx(2)

Ya que la prob puntual es 0, no convierto como en la discreta

Variables aleatorias continuas

Variables muy conocidas

- Uniforme
- Exponencial
- Gamma
- Normal (campana de Gauss)

Uniforme

Variables Uniforme

$X \sim$ Uniforme[a,b] si

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

Ejercicio

- Graficar $f_X(x)$
- Graficar $F_X(x)$
- Calcular $F_X(x)$

Uniforme

Variables Uniforme

$X \sim$ Uniforme[a,b] si

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } b \leq x \leq a \\ 0 & \text{si } x \notin [a,b] \end{cases}$$

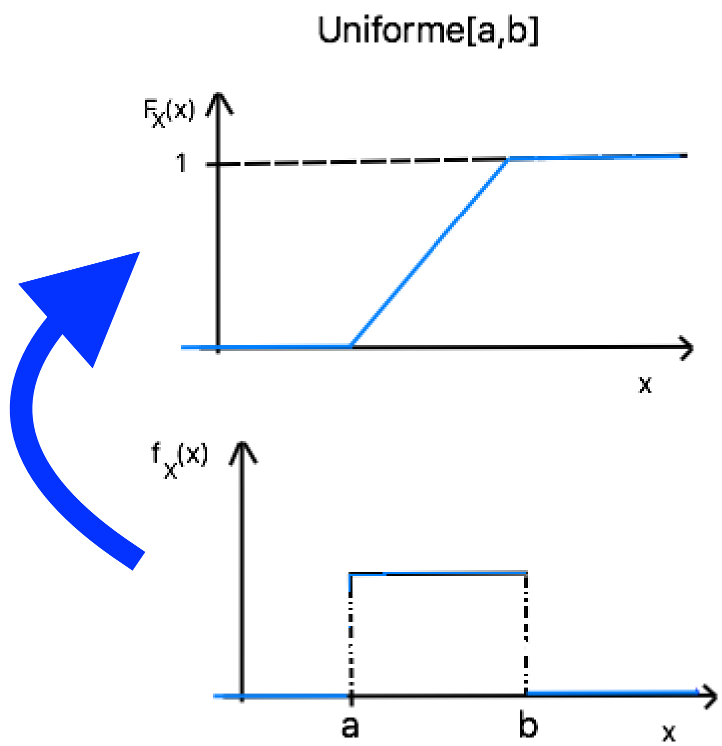
Por lo tanto:

integrando

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

Uniforme

integrando



Funcion de distribucion

Funcion de densidad de probabilidad

Uniforme

Ejemplo

El contenido de la botella de aceite de oliva (en mililitros) que tengo en la alacena sigue una ley Uniforme[0,500]. ¿Cuál es la probabilidad de que cuando vaya a buscar aceite de oliva quede entre 100ml y 200ml?

Llamemos X =contenido en mililitros de la botella de aceite de oliva.

¿$\mathbb{P}(100 < X < 200)$?

$$f_X(x) = \begin{cases} \frac{1}{500} & \text{si } 0 \leq x \leq 500 \\ 0 & \text{si } x \notin [0, 500] \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{500} & \text{si } 0 \leq x \leq 500 \\ 1 & \text{si } x > 500 \end{cases}$$

Uniforme

Ejemplo

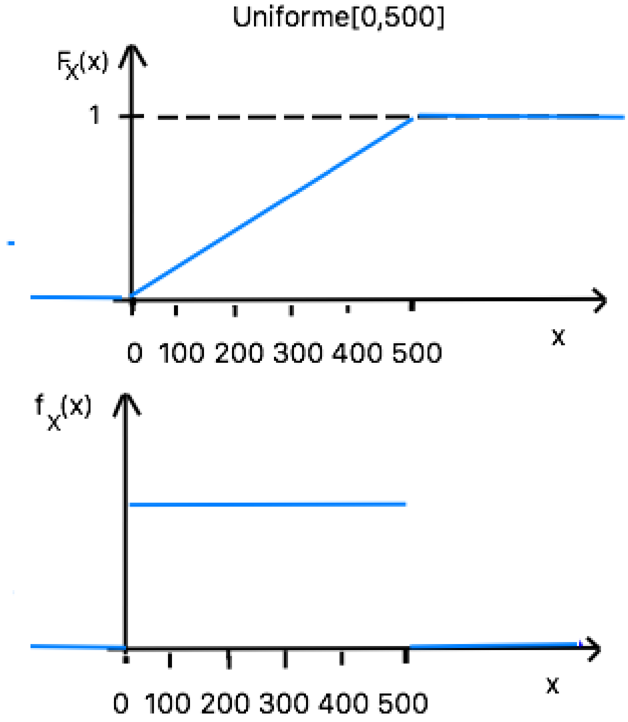
$X \sim U[0, 500]$, ¿$\mathbb{P}(100 < X < 200)$?

$$f_X(x) = \begin{cases} \frac{1}{500} & \text{si } 0 \leq x \leq 500 \\ 0 & \text{si } x \notin [0, 500] \end{cases}$$

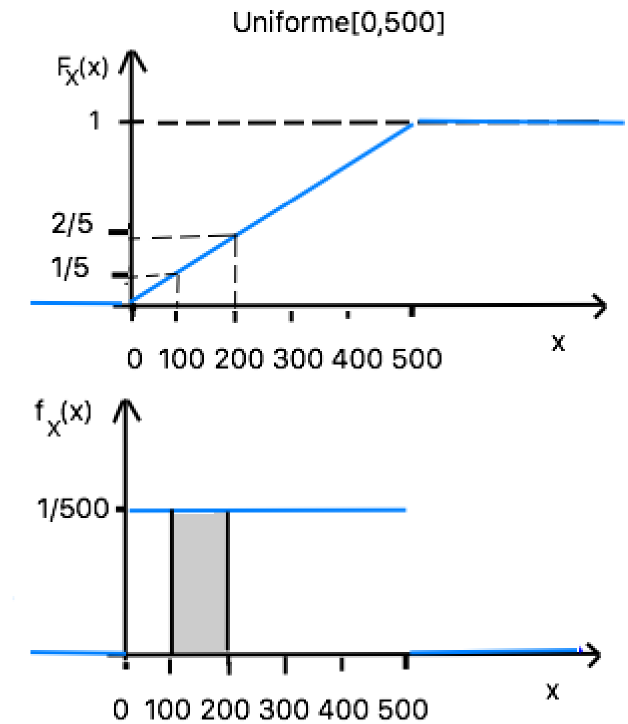
$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{500} & \text{si } 0 \leq x \leq 500 \\ 1 & \text{si } x > 500 \end{cases}$$

- $\mathbb{P}(100 < X < 200) = \int_{100}^{200} f_X(t)dt = \int_{100}^{200} \frac{1}{500} dt = \frac{1}{500} t|_{100}^{200} = \frac{1}{500}(200 - 100) = \frac{1}{5}$
- $\mathbb{P}(100 < X < 200) = F_X(200) - F_X(100) = \frac{200}{500} - \frac{100}{500} = \frac{1}{5}$

Uniforme



Uniforme



Exponencial

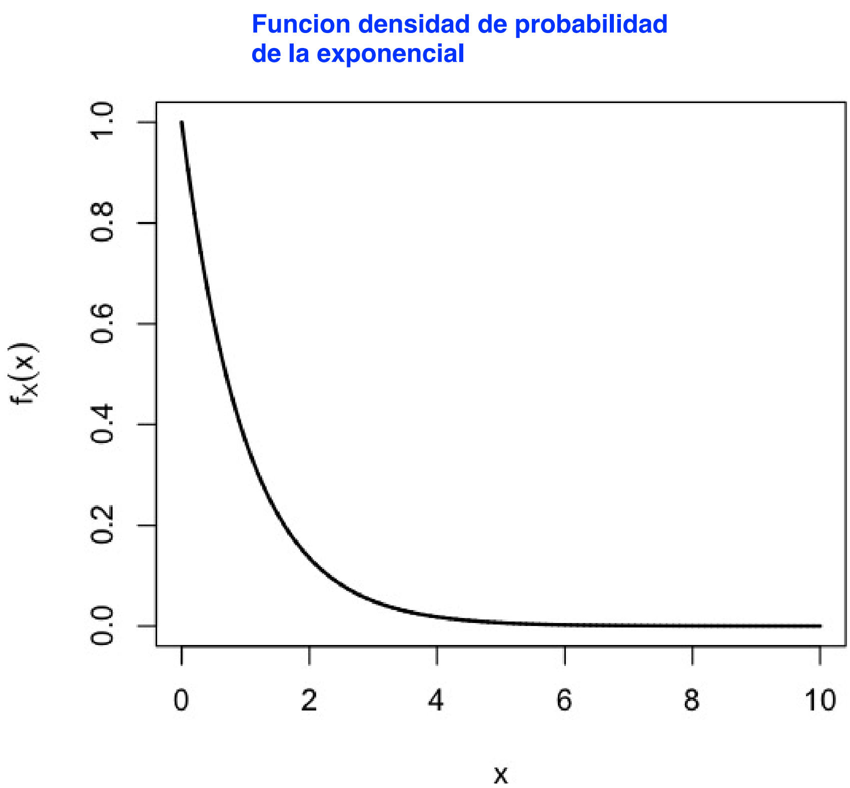
Variable Exponencial

$X \sim \text{Exponencial}(\lambda)$ si

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

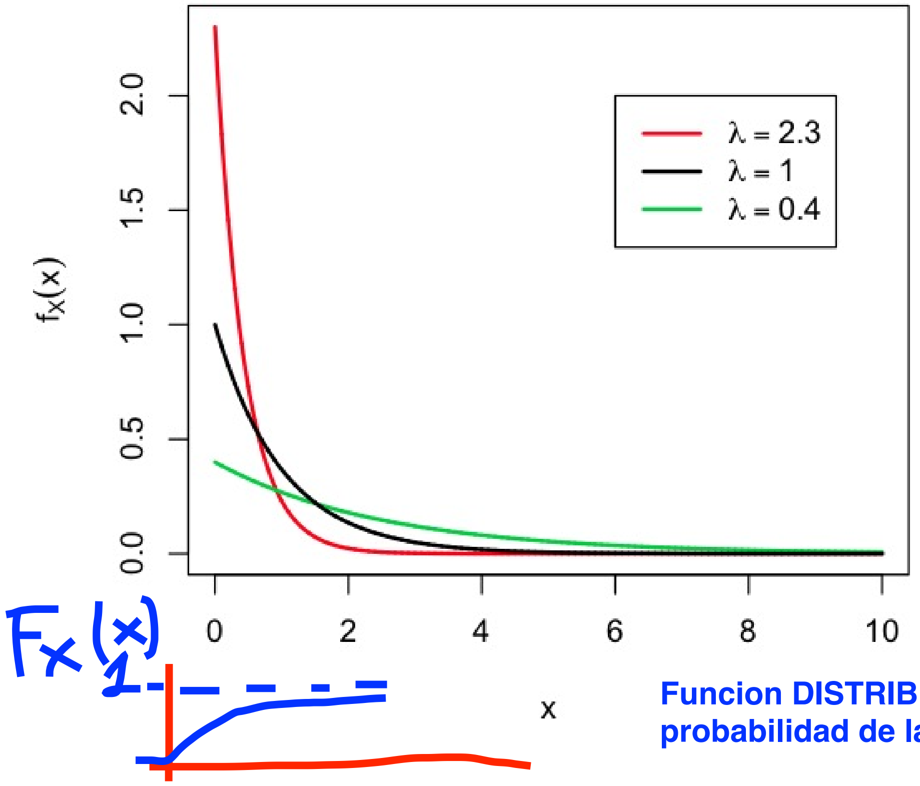
donde $\lambda > 0$.

Exponencial



Exponencial

Funcion DENSIDAD de
probabilidad de la exponencial



Funcion DISTRIBUCION de
probabilidad de la exponencial

Exponencial

Variable Exponencial

$X \sim$ Exponencial(λ) si

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

donde $\lambda > 0$.

Ejercicio

- Hacer un gráfico aproximado de $F_X(x)$
- Calcular $F_X(x)$

Exponencial

Propiedades

$X \sim$ Exponencial(λ) [Recordar $\lambda > 0$]

- **Funcion DISTRIBUCION de probabilidad EXPONENCIAL**

$$\mathbb{P}(X \leq x) = F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

- $$\mathbb{P}(X > x) = \begin{cases} 1 & \text{si } x \leq 0 \\ e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

- Cumple la propiedad de falta de memoria:

$$\mathbb{P}(X > t + a | X > t) = \mathbb{P}(X > a) \text{ con } t > 0, a > 0$$

Exponencial

Variable Exponencial

El tiempo de duración (en años) de una bombita de luz, X, tiene una distribución Exponencial con $\lambda = 0{,}5$.

$$X \sim Exp(0{,}5)$$

Hallar la probabilidad de que la bombita dure más de 3 años.

- ¿$\mathbb{P}(X > 3)$?
- $\mathbb{P}(X > 3) = e^{-0{,}5 \cdot 3} \approx 0{,}2231$

Distribución Gamma: una generalización de la Exponencial

alfa—> Parametro de forma
lambda—> Parametro de escala

Una variable aleatoria continua X tiene distribución Gamma de parámetros α y λ ($\alpha, \lambda > 0$) si su función de densidad está dada por:

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

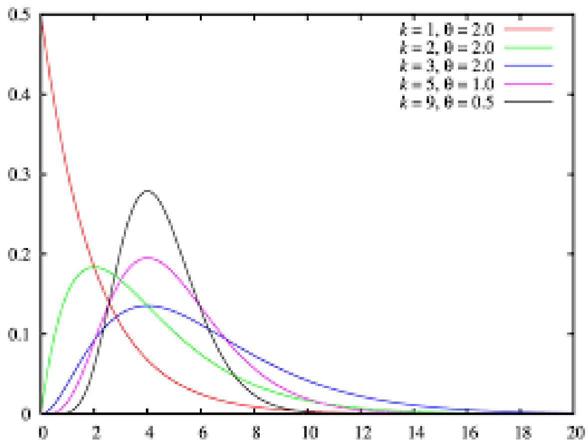
donde $\Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} \lambda^\alpha e^{-\lambda y}\, dy = \int_0^{+\infty} z^{\alpha-1} e^{-z}\, dz$. es la cte. de normalización.
Notación: $X \sim \Gamma(\alpha, \lambda)$.

Observación: Si $\alpha = 1$ resulta $\frac{\lambda^1}{\Gamma(1)} x^{1-1} e^{-\lambda x} = \lambda e^{-\lambda x}$. Entonces $\Gamma(1, \lambda) = Exp(\lambda)$.

la gamma con Alfa ==1 es la Exponencial

Gamma

Gráfico de densidades Gamma:



- λ es el parámetro de escala.
- α es el parámetro de forma. Si $\alpha > 1$ la densidad es no decreciente (tiene un máximo). Si $\alpha \leq 1$ la densidad es decreciente en el dominio.

Gamma

Propiedades de Función Gamma

Recordemos que para todo $\alpha > 0$,

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

1. $\Gamma(1) = 1$.
2. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, para todo $\alpha > 0$.
3. $\Gamma(n) = (n-1)!$ para todo $n \in \mathbb{N}$.
4. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Normal

Variable Normal

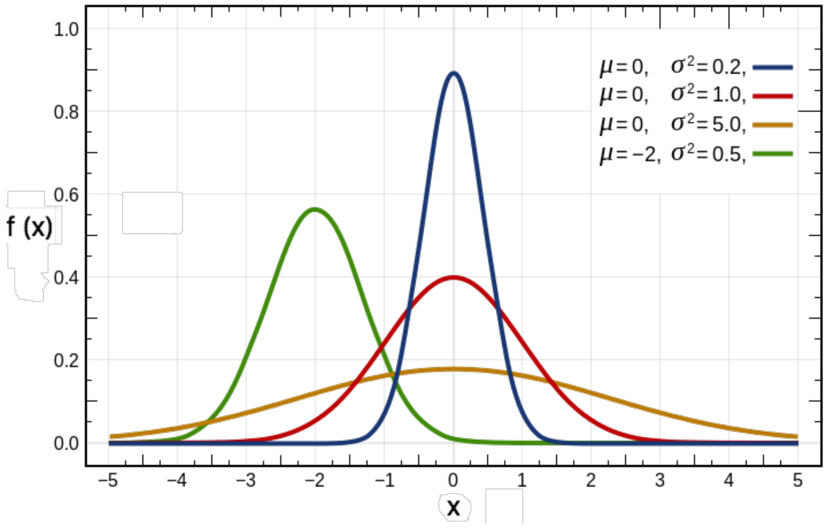
La distribución Normal juega un rol fundamental en la teoría de Probabilidades y en Estadísitica.
$X \sim \text{Normal}(\mu,\sigma^2)$ si

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

con $\mu \in \mathbb{R}$ y $\sigma > 0$.

Normal

El gráfico de la función de densidad Normal tiene forma de campana centrada en μ: *campana de Gauss*.



f_X es simétrica respecto de μ: $f_X(\mu + x) = f_X(\mu - x)$ $\forall x \in \mathbb{R}$.

Normal

Observación

- La densidad de la Normal no tiene primitiva $\rightarrow$ No existe una $F_X(x)$ explícita tal que $F'_X(x) = f_X(x)$.

Teorema: Corrimientos y cambios de escala

(1) $X \sim N(\mu, \sigma^2) \leftrightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

(2) $X \sim N(\mu, \sigma^2) \leftrightarrow Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

Normal

Notación

Si $\mu = 0$ y $\sigma = 1 \to Z \sim N(0,1)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \forall z \in \mathbb{R}$$

- $\varphi(z) := f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
- $\Phi(z) := F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$

Propiedad

- $\Phi(-a) = 1 - \Phi(a)$ con $a \in \mathbb{R}$ (por simetría)

Normal

Ejemplos

1. Sea $X \sim N(-3, 25)$. Hallar $\mathbb{P}(X > -2{,}4)$.

   $\mathbb{P}(X > -2{,}4) = \mathbb{P}\left(\frac{X+3}{5} > \frac{-2{,}4+3}{5}\right) = \mathbb{P}(Z > 0{,}12) =$
   $= 1 - \Phi(0{,}12) = 1 - 0{,}5478 = 0{,}4522$.

2. Sea $Y \sim N(2, 9)$. Hallar $\mathbb{P}\left(|Y - 1| < \frac{1}{2}\right)$.

   $\mathbb{P}\left(|Y - 1| < \frac{1}{2}\right) = \mathbb{P}\left(-\frac{1}{2} < Y - 1 < \frac{1}{2}\right) =$
   $= \mathbb{P}\left(-\frac{3}{2} < Y - 2 < -\frac{1}{2}\right) = \mathbb{P}\left(-\frac{1}{2} < \frac{Y-2}{3} < -\frac{1}{6}\right) =$
   $\mathbb{P}\left(-\frac{1}{2} < Z < -\frac{1}{6}\right)$
   $= \Phi(-\frac{1}{6}) - \Phi(-\frac{1}{2}) = \Phi(\frac{1}{2}) - \Phi(\frac{1}{6}) = 0{,}6915 - 0{,}5675 = 0{,}124$.

3. Sea $W \sim N(1, 4)$, hallar el valor c tal que $\mathbb{P}(W \leq c) = 0{,}1$.

   $\mathbb{P}(W \leq c) = \mathbb{P}\left(\frac{W-1}{2} \leq \frac{c-1}{2}\right) = \mathbb{P}\left(Z \leq \frac{c-1}{2}\right) = \Phi(\frac{c-1}{2}) = 0{,}1$.
   Entonces $\Phi(-\frac{c-1}{2}) = 1 - 0{,}1 = 0{,}9$. Luego, $-\frac{c-1}{2} = 1{,}28$ lo cual implica $c = -1{,}56$.

   En R —> qnorm(0.1, 1, 2) = -1.56
   Area, media, desvio

¿Por qué la Normal es tan conocida?

Un poco de historia

Fue presentada por primera vez por Abraham de Moivre en un artículo del año 1733. En el siglo XIX se utilizó esta distribución (entre otros, por Gauss) para modelar los errores que se cometen en mediciones físicas.

Teorema Central del Límite

La Distribución Normal es un atractor para la distribuciones de promedios (o sumas). Informalmente, lo que dice este teorema es que si sumamos *muchas* variables aleatorias independientes idénticamente distribuidas la variable que obtenemos *se parece* a una Normal.

Cuantiles

Definición

Para una variable aleatoria continua X definimos el cuantil-p de X como el valor x_p que verifica

$$F_X(x_p) = P(X \leq x_p) = p.$$

Es decir, para $0 < p < 1$, el cuantil-p de X es $F_X^{-1}(p)$.

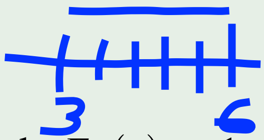
Algunos cuantiles tienen nombres particulares:

- cuantil-$\frac{1}{2}$ = "mediana".
- cuantil-$\frac{1}{4}$ = "cuartil inferior".
- cuantil-$\frac{3}{4}$ = "cuartil superior".

Cuantiles

Ejemplos

1. Hallar la mediana de $X \sim Exp(4)$.
2. Hallar el cuantil-$\frac{1}{5}$ de $Y \sim U[3,6]$.



X(0.2) = 3 + 3/5

1. Buscamos m tal que $F_X(m) = \frac{1}{2}$ siendo $F_X(x) = 1 - e^{-4x}$ si $x \geq 0$.
   Entonces $1 - e^{-4m} = \frac{1}{2}$ de donde se despeja $m = \frac{\ln 2}{4}$.
2. Recordemos que $f_Y(x) = \frac{1}{3}$ si $x \in [3,6]$. Buscamos c tal que

$$F_Y(c) = \int_{-\infty}^{c} f_Y(x)\,dx = \int_3^c \frac{1}{3}\,dx = \frac{1}{3}(c-3) = \frac{1}{5}.$$

   Despejando queda $c = \frac{18}{5}$.