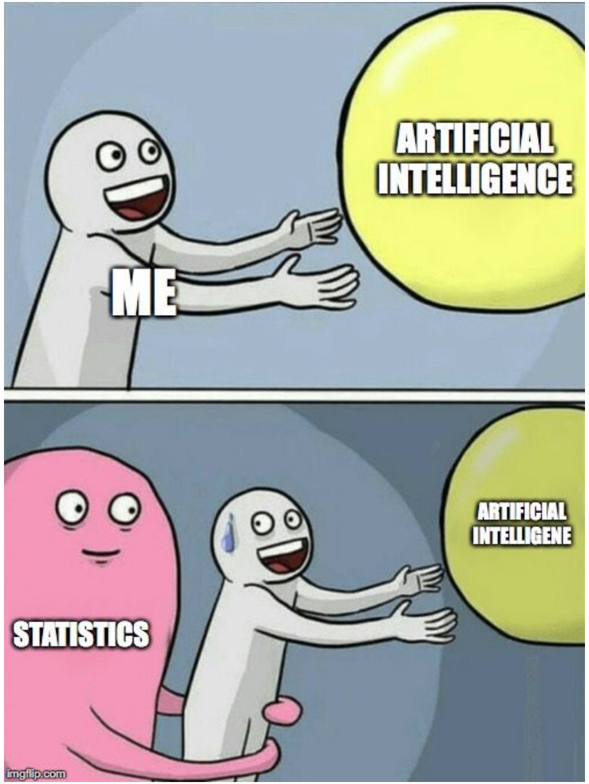
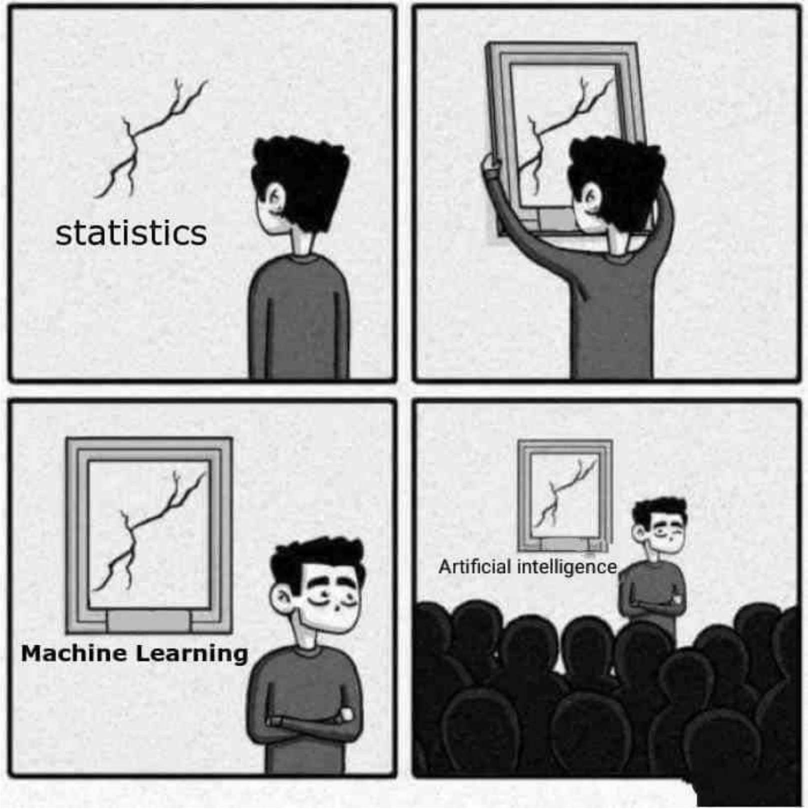
Probabilidad y Estadística

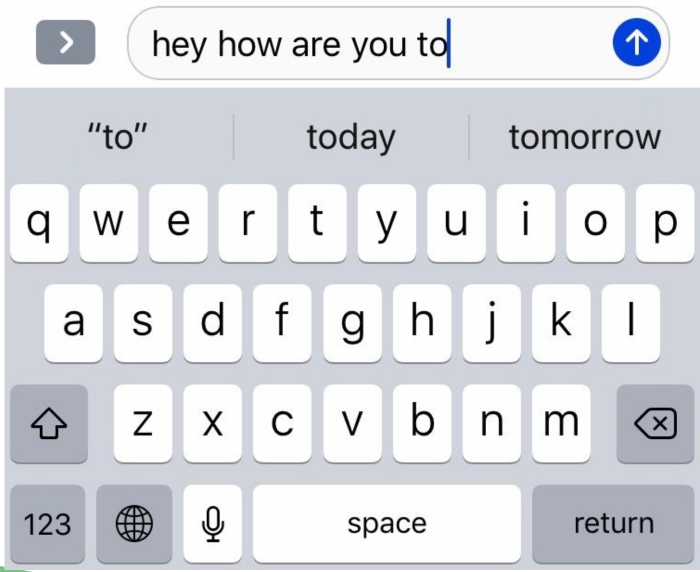
Clase 1

Ud. se encuentra aquí:



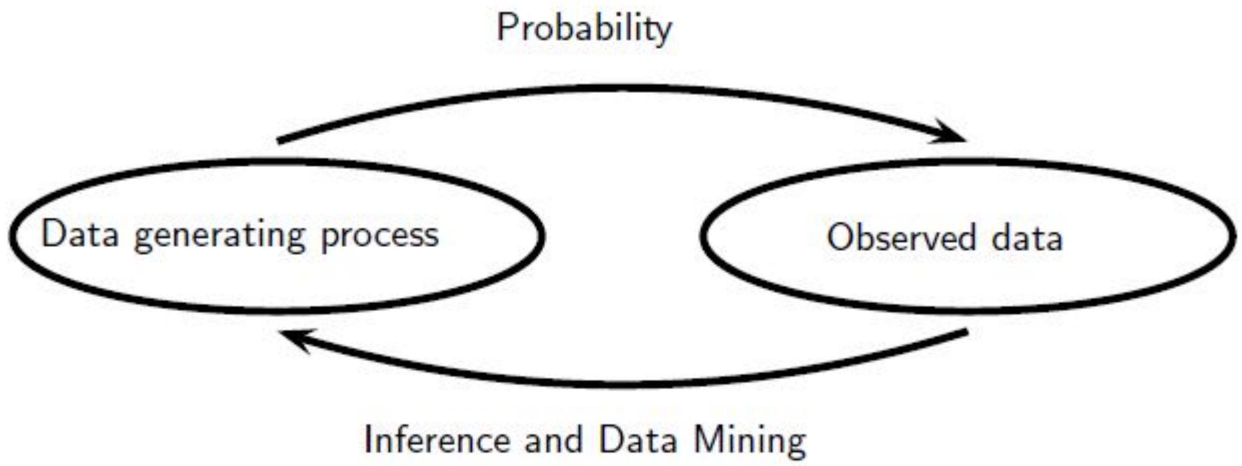
para poder estar aquí:





Ejemplos de probabilidad en IA?

Probabilidad vs. Estadística



“All of Statistics: A Concise Course in Statistical Inference”, Larry Wasserman

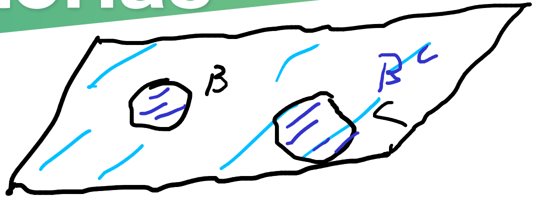
Cronograma

Clase 1	Repaso Distribuciones útiles.
✓ Clase 2	V.a. condicionadas Esperanza condicional
Clase 3	ECM y estimadores de cuadrados mínimos Estimador de máxima verosimilitud
✓ Clase 4	Estimación Bayesiana
Clase 5	Estimación no paramétrica
Clase 6	Intervalos de confianza
✓ Clase 7	Test de hipótesis
Clase 8	Examen

Repaso

Espacio de Probabilidad (Ω,A,P) y variables aleatorias

Ω ; $\Omega^c = \emptyset$



$\Omega \neq \emptyset$ $\quad \mathcal{A}$

$(\Omega, \mathcal{A})$ Esp. Medible

Alg.

1. $\Omega \in \mathcal{A}$
2. $B \in \mathcal{A} \Rightarrow B^c \in \mathcal{A}$

3* $B, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A}$

3* $\{B_i\}_{i=1}^{n} \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{n} B_i \in \mathcal{A}$ $\quad$; $B_1 \cap B_2 \sim B \cap C$

3 $\bigcup_{i=1}^{\infty} B_i \Leftarrow \{B_i\}_{i=1}^{\infty}$ $\quad | \quad \{B_i\}_{i=1}^{\infty} \Rightarrow \bigcup_{i=1}^{\infty} B_i$

$\mathcal{A}$ ($\not\ni B$)

(σ - algebra on Ω)

P

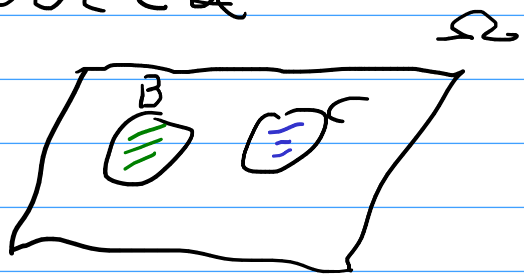
probability measure

[0, 1]

$$0 \leq \mathbb{P}(A) \leq 1, \ \forall A \in \mathcal{A}$$
$$\mathbb{P}(\Omega) = 1$$
$$A \cap B = \emptyset, \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

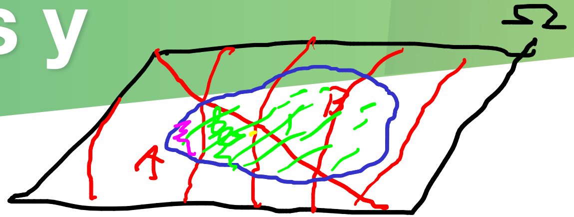
"*Probability: A Survery of the Mathematical Theory*" J. Lamperti

3) $B, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A}$



Probabilidades condicionales y proba. total

$\mathbb{P}((A\cap B_1)\cup\ldots\cup(A\cap B_n))$



Def: Se llama **probabilidad condicional** de A dado B ($\mathbb{P}(A|B)$) a la probabilidad de que ocurra el evento A sabiendo que B ha ocurrido, y está definida por

$$\mathbb{P}(A|B) = \boxed{\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}}$$

$i \neq j$

Def: Diremos que los eventos $B_1, \ldots B_n$ forman una **partición** si $B_i \cap B_j = \emptyset \; \forall\, i, j$ y $\bigcup_{i=1}^{n} B_j = \Omega$.

Luego podemos describir al evento A como $A = (A\cap B_1) \cup \ldots \cup (A \cap B_n)$ de forma que

$$\mathbb{P}(A) = \underbrace{\mathbb{P}(A\cap B_1) + \ldots + \mathbb{P}(A\cap B_n)} = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \ldots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

Fórmula de probabilidad total

Teorema de Bayes e independencia

Teorema de Bayes: Sean $B_1, \ldots B_n$ una partición de Ω , y A un evento con probabilidad positiva:

"Test 99%"

Test ~ 0,99

$$\mathbb{P}(B_i|A) = \boxed{\frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}} = P(A) > 0$$

Def: Diremos que dos eventos A y B son **independientes** si y sólo sí vale que

A, B son Indep. $\Longleftarrow$ $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

A, B son Ind. $\Rightarrow P(A \cap B) = P(A)P(B)$

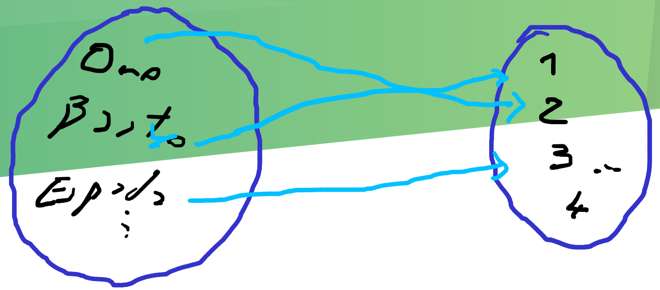
$(\Omega, \mathcal{A})$

σ-álg.

$$X : \Omega \to \mathbb{R}$$
$$\omega \mapsto X(\omega)$$

Variables aleatorias

Variables aleatorias



Una va. X es una función que mapea eventos a números.

Ejemplo: basto →1, Oro →2, espada →3, copa→4

X tiene asociada una función de distribución, definida como

$X \sim F_X$

$$F_X(x) = \mathbb{P}(X \leq x)$$

$$\left\{ \begin{array}{l} F_X(x) \in [0,1] \; \forall x \in \mathbb{R} \\ F_X(x) \text{ es monótona no decreciente} \\ F_X(x) \text{ es continua por derecha} \\ \lim_{x \to -\infty} F_X(x) = 0 \text{ y } \lim_{x \to \infty} F_X(x) = 1 \end{array} \right.$$

además se define la función de distribución inversa o función cuantil como $F_X^{-1}(q) = \inf\{x : F(x) > q\}$

Si F_X es estrictamente creciente y contínua, F_X^{-1} es su inversa.

Tipos de v.a.

1 • Discretas (v.a.d): toman valores en un conjunto discreto o numerable de puntos. Si X es **v.a.d**, tendrá además función de probabilidad dada por

$$p_X(x) = \mathbb{P}(X = x)$$

2 • Continuas (v.a.c): toman valores en un intervalo continuo. Si X es una **v.a.c**. tendrá asociada una función de densidad

$$(\mathbb{P}[X \leq x]$$

$$f_X(x) = \frac{dF_X(x)}{dx} \qquad F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

¿Qué propiedades debían cumplir cada una de estas funciones?

$f_X(x) \geq 0$; $\int_{-\infty}^{\infty} f_X(u)\, du = 1$

Espacio de Probabilidad (Ω,A,P) y variables aleatorias

$$\Omega \in \mathcal{A}$$
$$B \in \mathcal{A} \Rightarrow B^c \in \mathcal{A}$$
$$B, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A}$$

3) $\{A_i\}_{i=1}^{\infty} \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

$$0 \leq \mathbb{P}(A) \leq 1, \ \forall A \in \mathcal{A}$$
$$\mathbb{P}(\Omega) = 1$$
$$A \cap B = \emptyset, \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

$$\begin{tikzcd}
\overset{\mathcal{A}}{(\sigma\text{ - algebra on } \Omega)} \arrow[rr, "\mathbf{X}"] \arrow[dr, "\mathbf{P}"'] & & \overset{\mathbb{R}}{(\sigma\text{ - algebra on } \mathbb{R})} \arrow[dl, "\mathbf{P}_X"] \\
& {[0,\ 1]} &
\end{tikzcd}$$

probability measure

$X \sim F_X$

$\mathbf{P}_X \sim F_X$

induced measure on **R**
by $P_X(A) = P(X^{-1}(A))$
called "distribution of X"

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

"*Probability: A Survery of the Mathematical Theory*" J. Lamperti

$$X \sim F_X$$

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$X : \Omega \longrightarrow \mathbb{R}^2$$

$$\omega \longmapsto X(\omega)$$

Vectores aleatorios

Distribución conjunta y marginales

	$Y=0$	$Y=1$	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	1

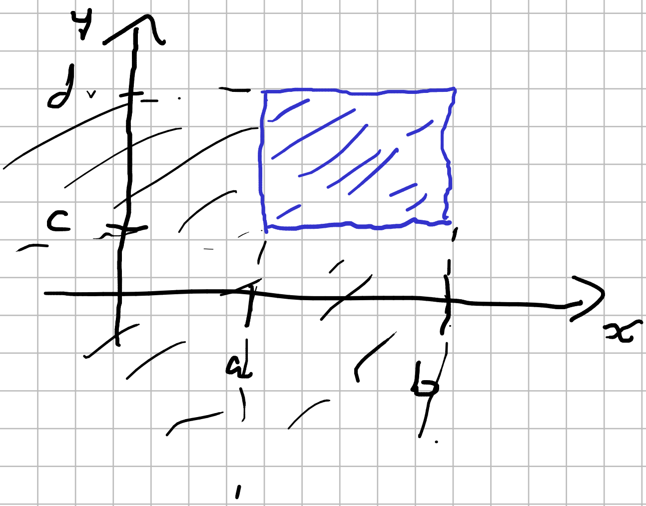
P_{XY} P_X P_Y

Si tenemos dos variables X e Y se define su **función de distribución conjunta** como $F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$ y vale la regla del rectángulo:

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

Caso continuo: $f_{X,Y}(x,y)$ es la función de densidad conjunta y se definen las funciones de densidad marginales como $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$ y
$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$

Caso discreto: $p_{X,Y}(x,y)$ es función de probabilidad conjunta y se definen las funciones de probabilidad marginal como $p_X(x) = \sum_y p_{X,Y}(x,y)$ y
$p_Y(x) = \sum_x p_{X,Y}(x,y)$


y
d
c
a
b
x

$$\begin{array}{ccc} & F_{XY} & \\ \swarrow & & \searrow \\ F_X & & F_Y \end{array} \qquad \begin{array}{ccc} & f_{XY} & \\ \swarrow & & \searrow \\ f_X & & f_Y \end{array}$$

Independencia de v.a.

Diremos que dos v.a. X e Y son independientes si vale que

$$F_{X.Y}(x,y) = F_X(x)F_Y(y) \quad \forall\, x,y \in \mathbb{R}$$

Caso discreto:

$$p_{X.Y}(x,y) = p_X(x)p_Y(y) \quad \forall\, x,y \in \mathbb{R}$$

Caso continuo:

$$f_{X.Y}(x,y) = f_X(x)f_Y(y) \quad \forall\, x,y \in \mathbb{R}$$

En el caso de que la intersección sea 0,
La independencia implicaba inter-sección cero pero no al revés.

Momentos

Momentos

$E(x) = \int x f_X(x)dx$

Esperanza (o media):

$$\mathbb{E}[g(X)] = \int g(x) f_X(x) dx$$

$$\text{Caso Disc.} = \sum g(x) p_X(x)$$

Varianza:

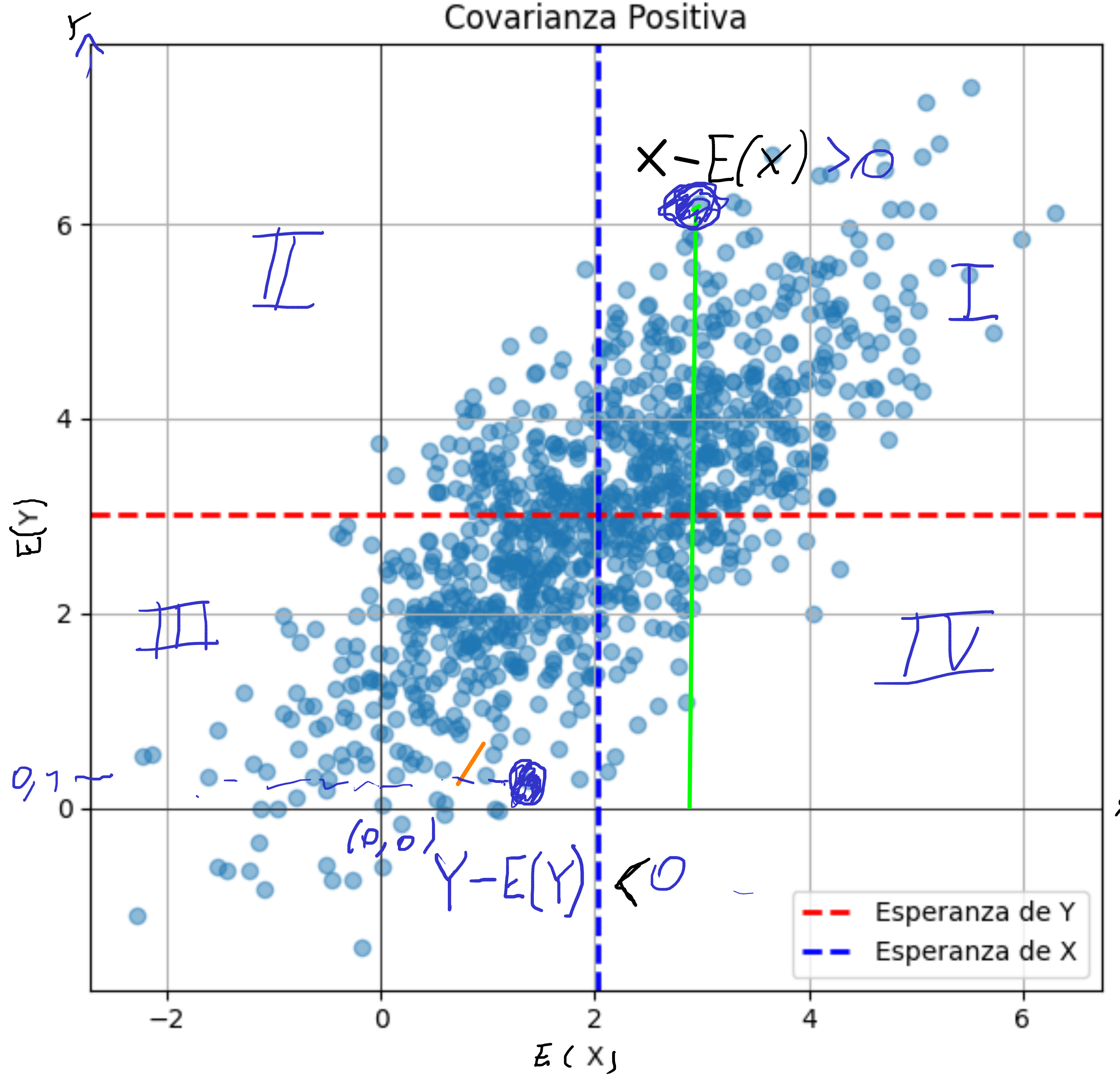
$$var(X) = \mathbb{E}[\underbrace{(X - \mathbb{E}[X])^2}] = \mathbb{E}[X^2] - (\mathbb{E}(X))^2$$

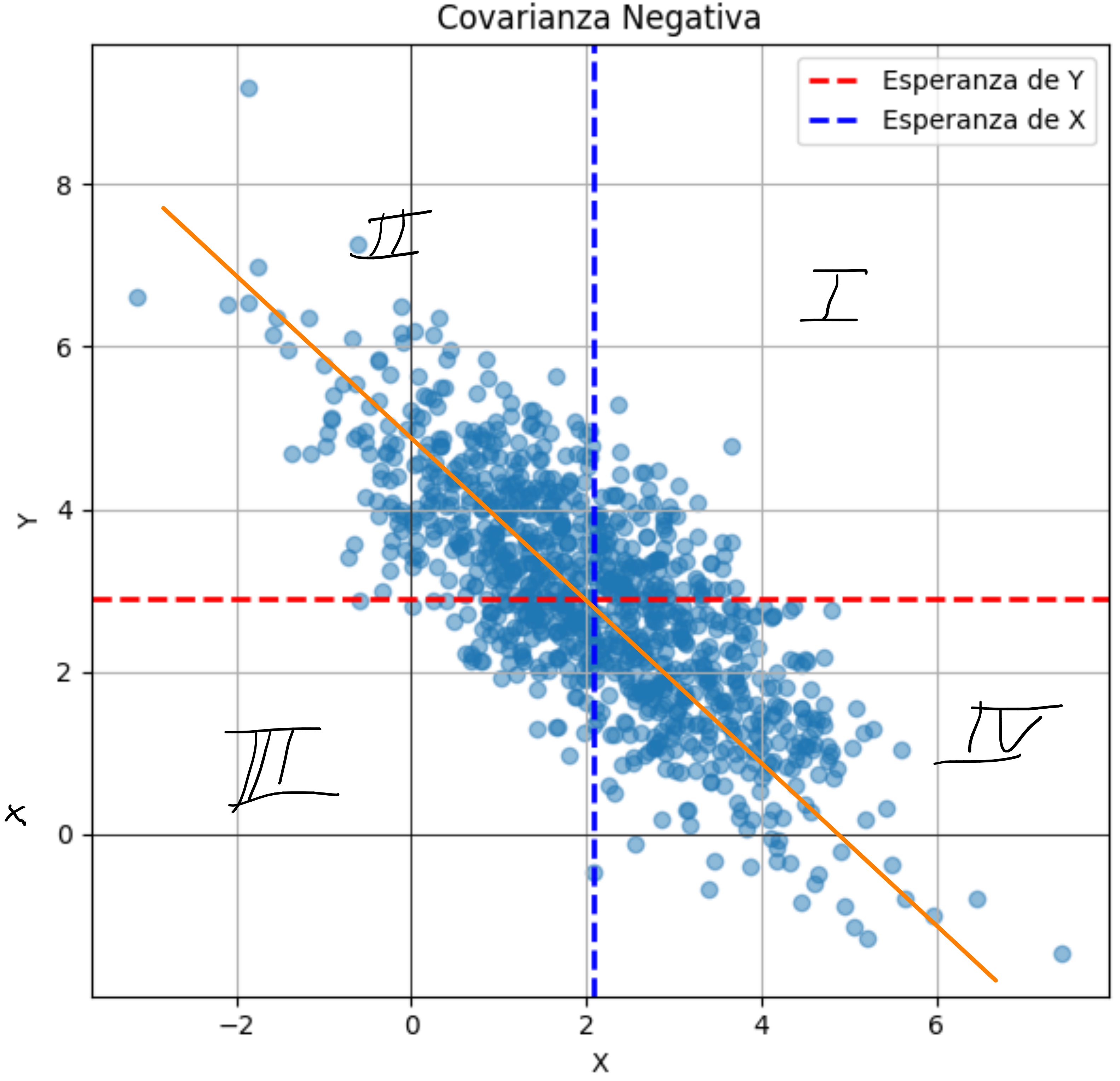
Covarianza: X, Y

$cov(X,Y) > 0$

$cov(X,Y) < 0$

$$cov(X,Y) = \mathbb{E}[\underbrace{(X - \mathbb{E}[X])}\underbrace{(Y - \mathbb{E}[Y])}] = \mathbb{E}[\underbrace{X \cdot Y}] - \underbrace{\mathbb{E}(X)}\underbrace{\mathbb{E}(Y)}$$


Covarianza Positiva
Y
$X - E(X) > 0$
I
II
III
IV
$E(Y)$
0,1
(0,0)
$Y - E(Y) < 0$
x
Esperanza de Y
Esperanza de X
E (X)
−2
0
2
4
6


Covarianza Negativa
Esperanza de Y
Esperanza de X
II
I
III
IV
Y
X
8
6
4
2
0
−2

Algunas distribuciones útiles

$X \sim F_X$

Variables discretas

$$P(X=x) = \underbrace{p^x (1-p)}_{1}{}^{1-x}$$

$$P(X=1) = p \quad \wedge \quad P(X=0) = (1-p)$$

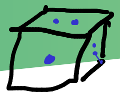
- $X \sim$ **Bernoulli(p)**: X = {0,1}. Asociada a la ocurrencia o no de un éxito

$n \geq x$

- $Y \sim$ **Binomial(n, p)**: cantidad de éxitos en n ensayos. $\binom{n}{x} p^x (1-p)^{n-x}$
- $Z \sim$ **Geométrica(p)**: cantidad de ensayos que debo realizar hasta observar el primer éxito

$$p(1-p)^{x-1}$$

Ejercicio 0



$p = \frac{1}{6}$

$P(x=1) + P(x=2)$

Se tira sucesivamente un dado

1. Probabilidad de necesitar menos de 3 tiros hasta ver el primer 2
2. Probabilidad de ver exactamente un 2 en 5 tiros

$Z \sim$ • **Geométrica(p)**: cantidad de ensayos que debo realizar hasta observar el primer éxito

$1-\ X \sim G(\frac{1}{6})$

$p(1-p)^{x-1}$

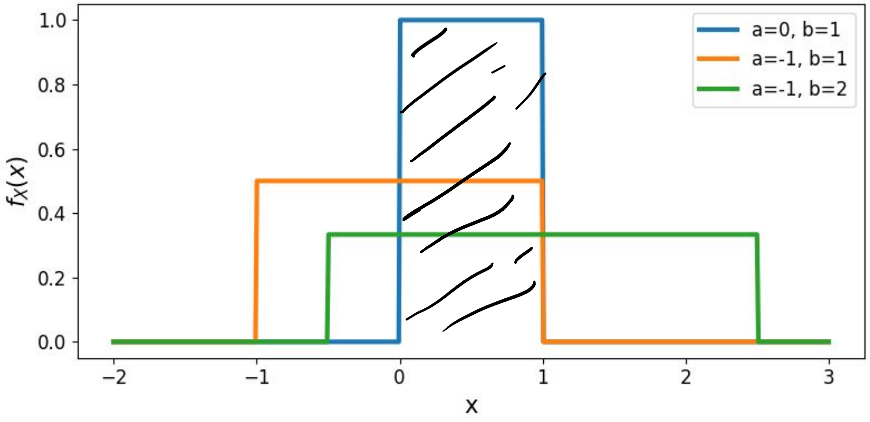
*1/6 + 5/6 * 1/6*

$2-\ Y \sim Binomial(5, \frac{1}{6})$ ¿ P

Variables continuas

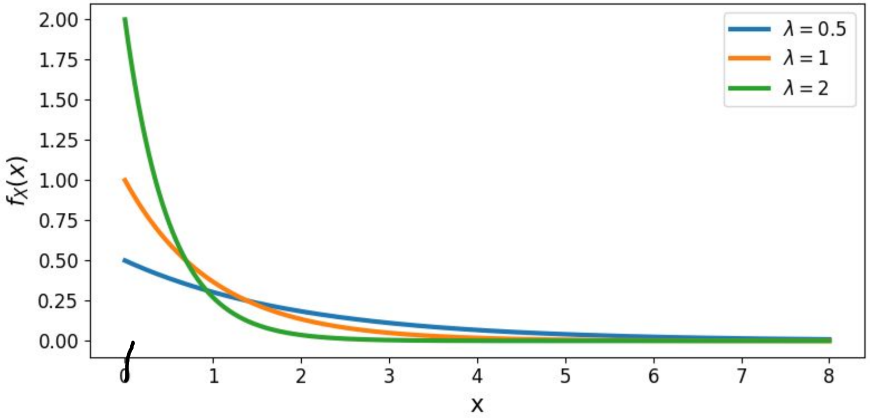
- **Uniforme:** todos los puntos son equiprobables. $X \sim \mathcal{U}(a,b)$

$$f_X(x) = \frac{1}{b-a}\mathbf{I}\{a < x < b\}$$



- **Exponencial:** sirve para modelar tiempos hasta eventos que no tienen memoria. $X \sim \mathcal{E}(\lambda)$

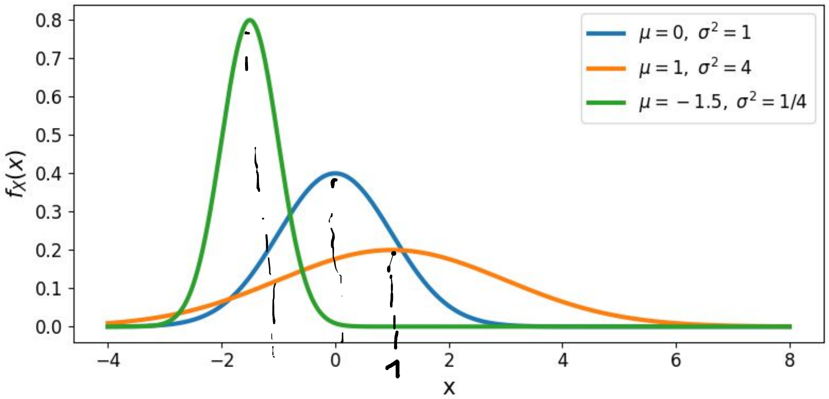
$$f_X(x) = \lambda e^{-\lambda x}\mathbf{I}\{x > 0\}$$



Variables continuas

- **Normal (gaussiana).** $X \sim \mathcal{N}(\mu, \sigma^2)$

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



μ es la media
$\boldsymbol{\sigma}^2$ es la
varianza

Propiedades: Sean X, Y dos v.a. Independientes

$$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow \boxed{\frac{X-\mu}{\sigma}} \sim \mathcal{N}(0, 1)$$ (estandarización)

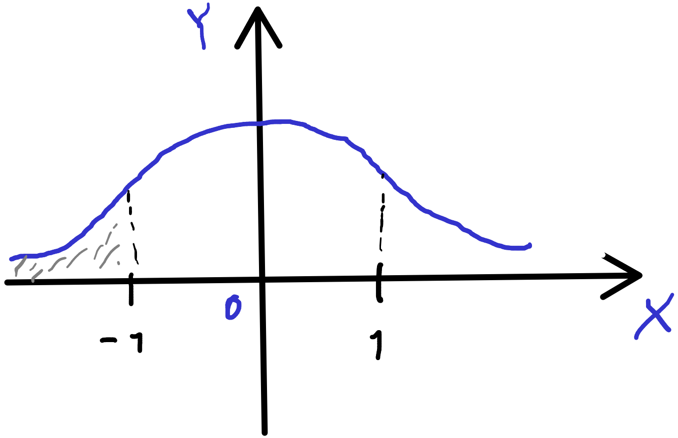
$$\underbrace{X \sim \mathcal{N}(\mu_X, \sigma_X^2)}, \underbrace{Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)} \rightarrow aX + bY \sim \mathcal{N}(\underbrace{a\mu_X + b\mu_Y}_{\mu}, \overbrace{a^2\sigma_X^2 + b^2\sigma_Y^2}^{\sigma^2})$$

(combinación lineal de normales es normal)

Ejercicio 1

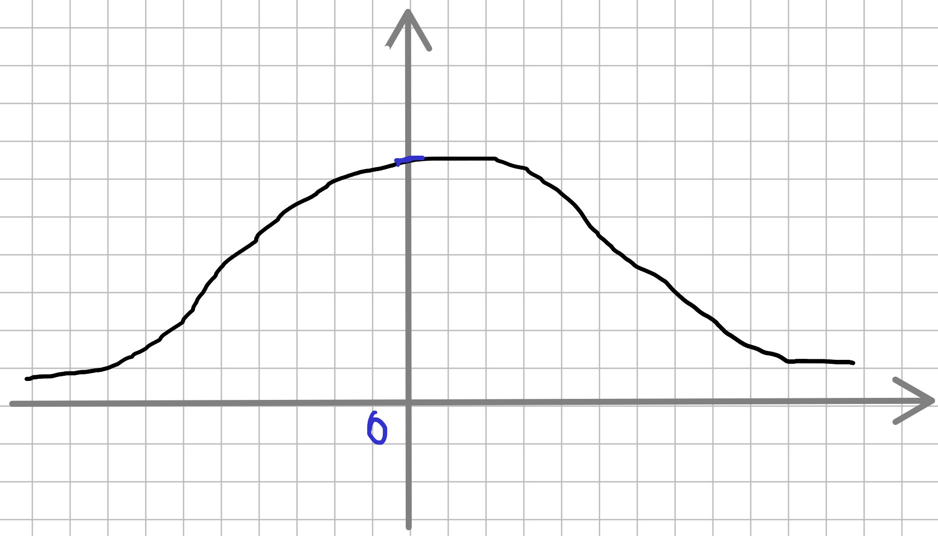
Sea X una v.a con distribución normal estándar. Hallar la probabilidad de que

1. $P(X>1) = 1 - P(X \leq 1) = 1 - F_X(1)$
2. X<-1
3. |X| <2
4. Hallar los cuantiles 0.1 y 0.9



Sea además Y ~ N(2,9)

1. Hallar P(2X+Y <5)


0

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
```

```
1-stats.norm.cdf(1)
```

```
0.15865525393145707
```

Ejercicio 2

X es V.A.C.

$X \sim \mathcal{E}(\lambda = 1/5)$

El tiempo (en minutos) entre llamadas a un call center tiene una distribución exponencial de parámetro 1/5.

1. Calcular la probabilidad de que la primer llamada llegue después de 2 minutos
2. Calcular la probabilidad de que la primer llamada llegue después de los 5 minutos, si se sabe que en los primeros 3 minutos no se recibieron llamados

Distribución normal multivariada

Función de densidad conjunta

Sea $\underset{n}{\underline{X}} = [X_1, \ldots, X_n]^T$ un vector aleatorio continuo, diremos que X tiene distribución normal multivariada, si su función de densidad conjunta es de la forma

$$\underbrace{f_{\underline{X}}(\underline{x})} = \frac{1}{(2\pi)^{n/2}\underbrace{|\Sigma|}^{1/2}} e^{\underbrace{-\frac{1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})}}$$

$cov(X_2,X_1) = cov(X_1,X_2)$

media $\quad \pi = 3{,}14\ldots$ matriz de covarianza .

$$\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & cov(X_1,X_2) & \ldots & cov(X_1,X_n) \\ cov(X_2,X_1) & \sigma_2^2 & \ldots & cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n,X_1) & cov(X_n,X_2) & \ldots & \sigma_n^2 \end{bmatrix}$$

Notación: $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$

Distribuciones marginales

$f_{X,Y}$

$f_X \quad f_Y$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & cov(X_1, X_2) & \ldots & cov(X_1, X_n) \\ cov(X_2, X_1) & \sigma_2^2 & \ldots & cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n, X_1) & cov(X_n, X_2) & \ldots & \sigma_n^2 \end{bmatrix}$$

Ejercicio 3

Sean X,Y dos v.a. Con función de densidad conjunta

$$f_{X,Y}(x,y) = \frac{1}{2\pi 0.6} e^{-\frac{1}{2}\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}}$$

1. Calcular E[X], E[Y], var(X), var(Y), y cov(X,Y)
2. Hallar las densidades marginales de X e Y
3. Calcular P(X<2, Y<-1)

Bibliografía

Bibliografía

"Mathematical Statistics with Applications", Dennis D. Wackerly, William Mendenhall III, Richard L. Scheaffer.

"All of Statistics: A Concise Course in Statistical Inference", Larry Wasserman.