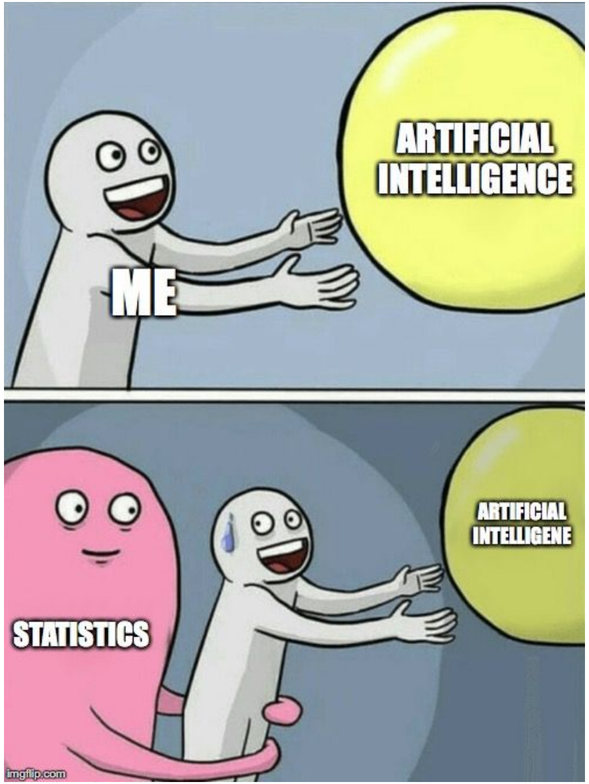
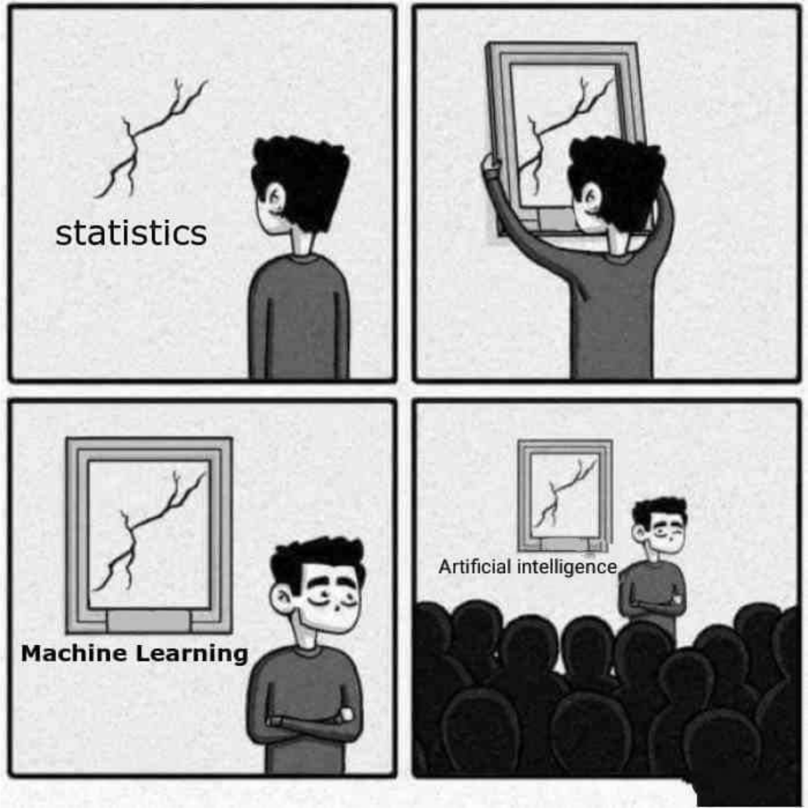
Probabilidad y Estadística

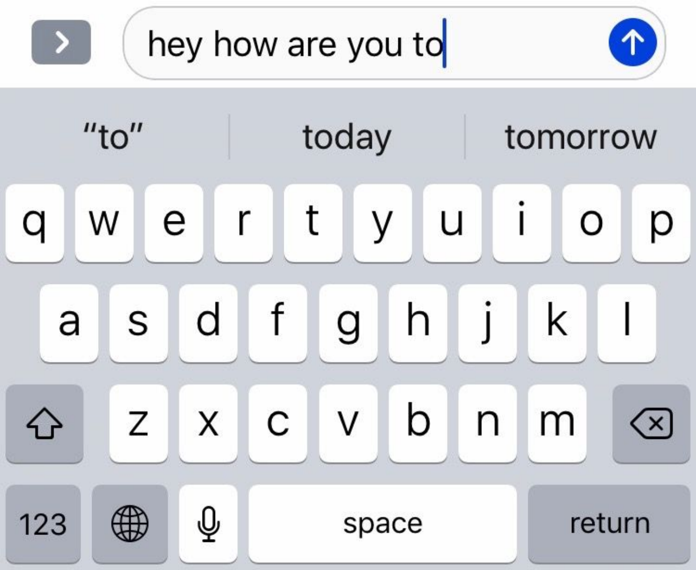
Clase 1

Ud. se encuentra aquí:



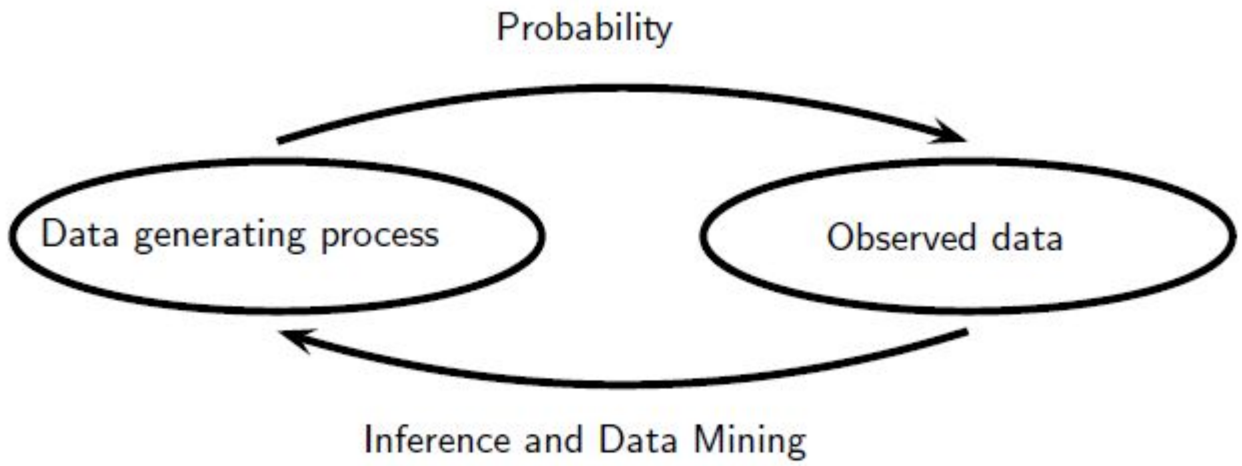
para poder estar aquí:





Ejemplos de probabilidad en IA?

Probabilidad vs. Estadística



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

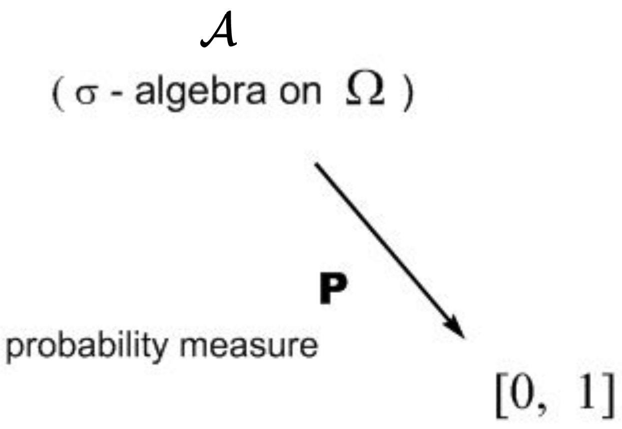
Cronograma

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

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



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

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

Probabilidades condicionales y proba. total

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

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

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

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

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:

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

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

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

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

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

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

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.

- 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

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

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

$$f_X(x) = \frac{dF_X(x)}{dx}$$

¿Qué propiedades debían cumplir cada una de estas funciones?

$$P(Inf \mid \oplus) = \frac{P(\oplus \mid Inf)^{*}\, P(Inf)^{*}}{P(\oplus \mid Inf)P(Inf) + P(\oplus \mid I^c)P(I^c)} \sim 0{,}19$$

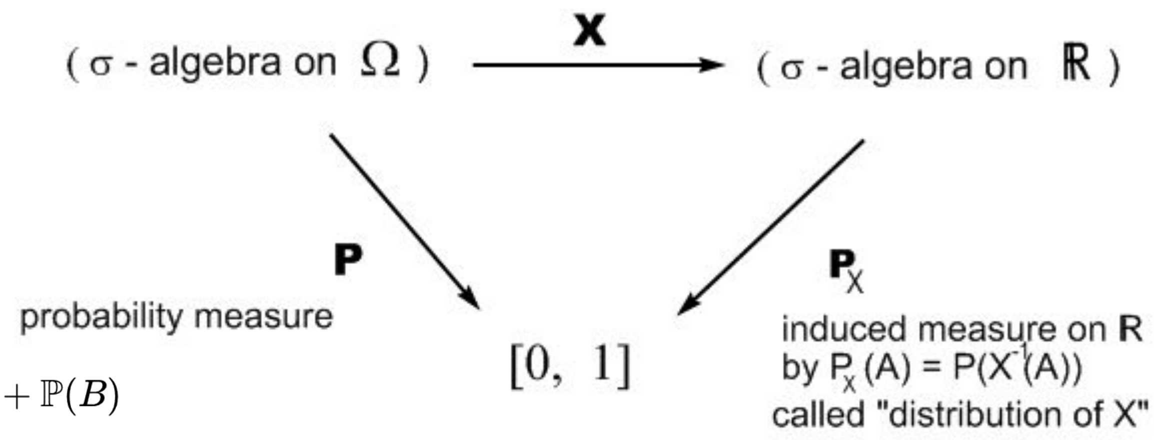
0,99

$P(Inf) = 0{,}005$, $P(No\ Inf) = 0{,}995$

$P(\oplus \mid No\ Inf) = 1 - 0{,}99$

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

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



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

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

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

Vectores aleatorios

Distribución conjunta y marginales

$F_X(b) =$

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

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:

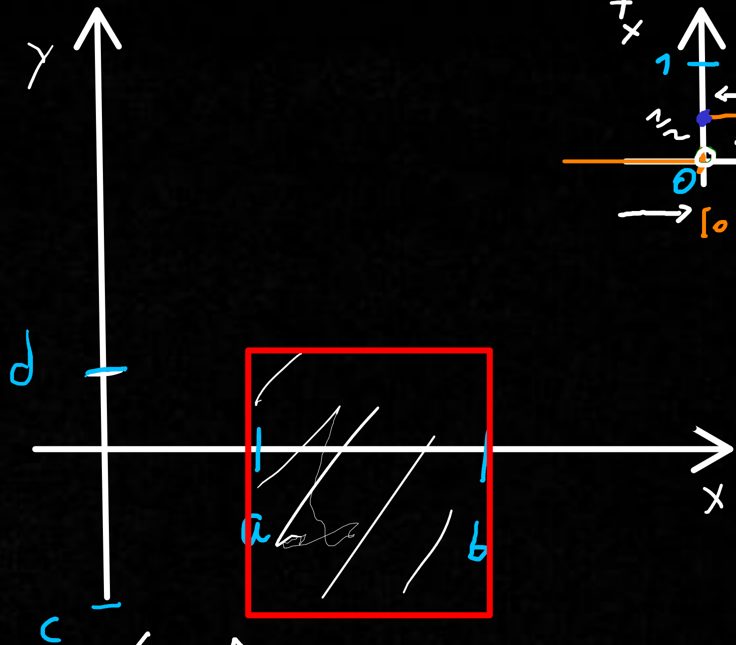
$$\mathbb{P}(\underbrace{a < X \leq b}, c < Y \leq d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

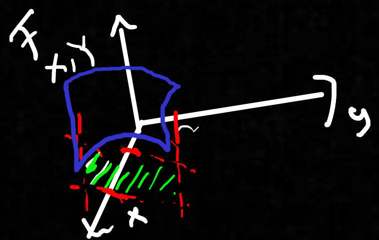
Caso continuo: $f_{X,Y}(x,y)$ es la función de densidad conjunta y se definen las funciones de densidad marginales como $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$ y $2x, \; \frac{1}{2} I\{a \leq x \leq b\}$

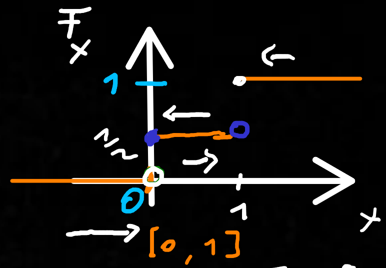
$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$$

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

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$







- $F_X \in [0,1]$
- Monótona No De
- Cont por Derec
- $\lim_{x\to-\infty} F_X(x) = 0$

$$F_X(x) \xrightarrow[x\to+\infty]{} 1$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Dada una moneda equilibrada

$Y =$ *" # Numero de tiros hasta observar la primera cara"*

$\Omega = \{\{c\}, \{x\}\}$

$Y : \Omega \longrightarrow \mathbb{N} \subset \mathbb{R}$

$Y = \mathbb{N}$

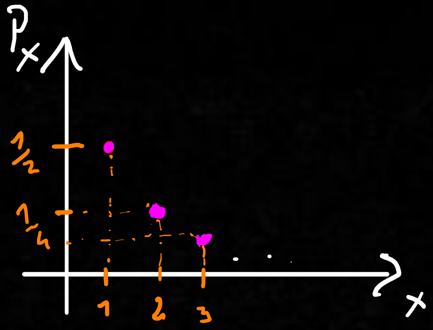
$P(B|A) = P(B)$

$P(A \cap B) = P(A)P(B)$

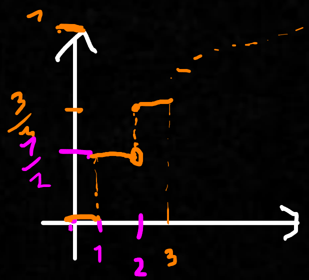
$$P_Y(1) = P(Y=1) = \frac{1}{2} = 0.5$$

$$P_Y(2) = P(Y=2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

$$P_Y(3) = P(Y=3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$



- $P_X(x) \geq 0$
- $\sum_{x \in A} P_X(x) = 1$
- $\sum_{x \in B} P_X(x) = P_X(X \in B)$



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:

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

Momentos

x

Momentos

$f_X(x)$ $\begin{cases} f_X \\ p_X \end{cases}$

$E[X] = \int_{\mathbb{R}} x f_X(x)\,dx$; $\sum_x x\,p_X(x)$

X, Y

$E[aX + bY] = aE[X] + bE[Y]$

Esperanza (o media):

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

Varianza:

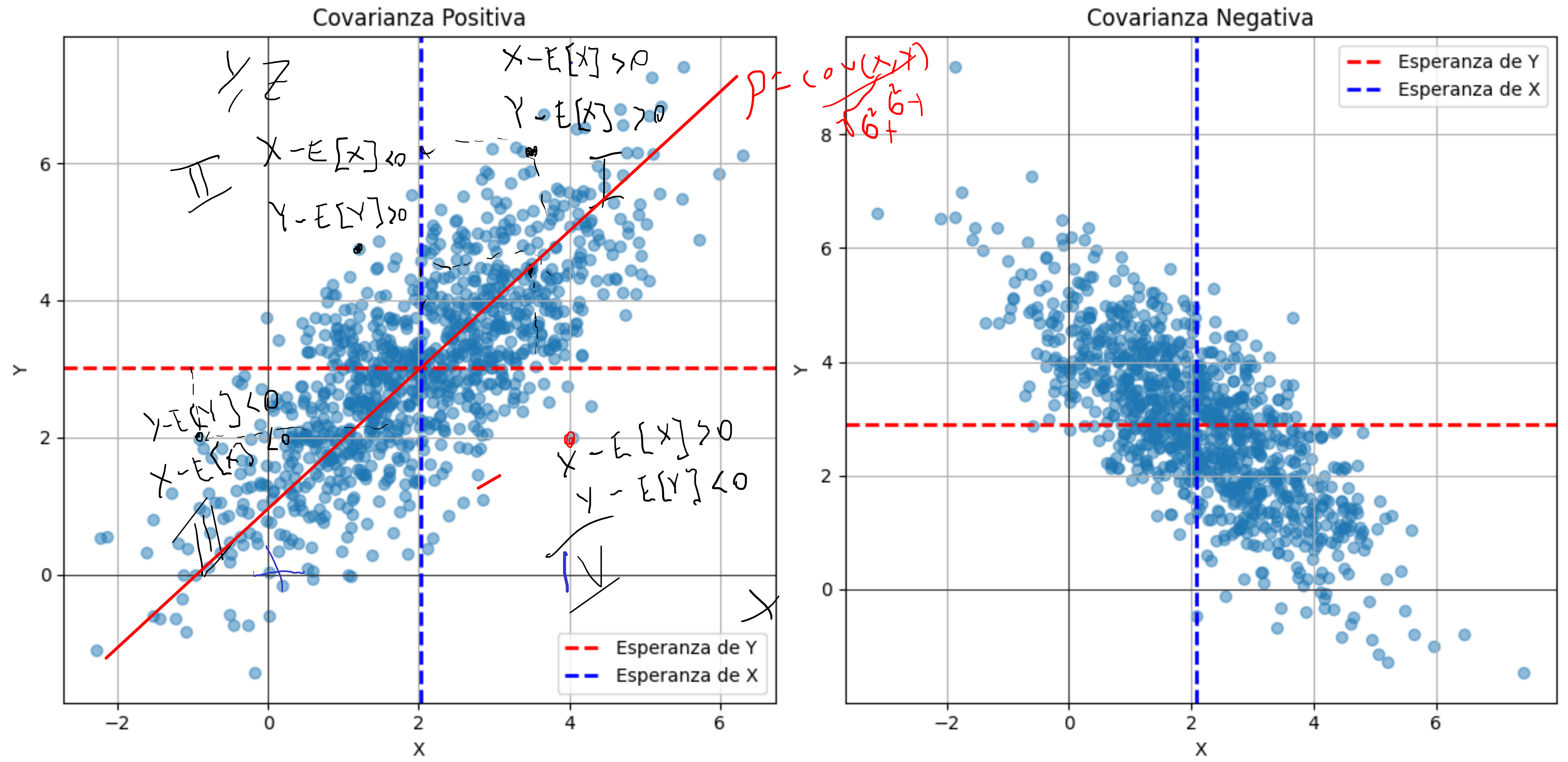
$E[X^2 - 2X\underbrace{E[X]}_{\mu_X} + (E(X))^2] = E[X^2] - 2\mu_X E[X] + E^2[X]$

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

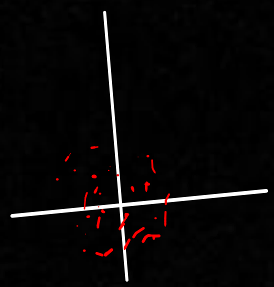
$= E[X^2] - 2\mu_X^2 + (\mu_X)^2 = E[X^2] - \mu_X^2 = E[X^2] - (E[X])^2$

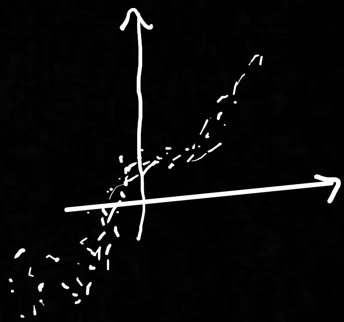
Covarianza:

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



Covarianza Positiva
Esperanza de Y
Esperanza de X
X
Y
Covarianza Negativa
Esperanza de Y
Esperanza de X
X
Y





Algunas distribuciones útiles

Variables discretas

$X = \begin{cases} 1 & \text{Exito} \\ 0 & \text{Fracaso} \end{cases}$ $X \sim Ber(p)$

$P_X = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$

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

EFEE

EEEE FF

$n = 6$, EEFEFE

$x = 4$

$n - x = 2$

- **Binomial(n, p)**: cantidad de éxitos en n ensayos.

$Y \sim Bin(n, p)$ $P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$;

- **Geométrica(p)**: cantidad de ensayos que debo realizar hasta observar el primer éxito

$P(Z=3)$

$Z \sim Geo(1/2)$ $P_Z(z) = (1-p)^{z-1} p$

$\left(\frac{1}{2}\right)^{3-1} \cdot \frac{1}{2}$

Ejercicio 0

$P_X(X=3) =$

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

$X \sim Geo(1/6)$

$n=5,\ p = 1/6$

$Z \sim Bin(5, 1/6)$

$P_Z(1) = \binom{5}{1} p (1-p)^4$

Variables continuas

$I\{a \le x < b\} = \begin{cases} 1 & a \le x < b \\ 0 & \text{en otro caso} \end{cases}$

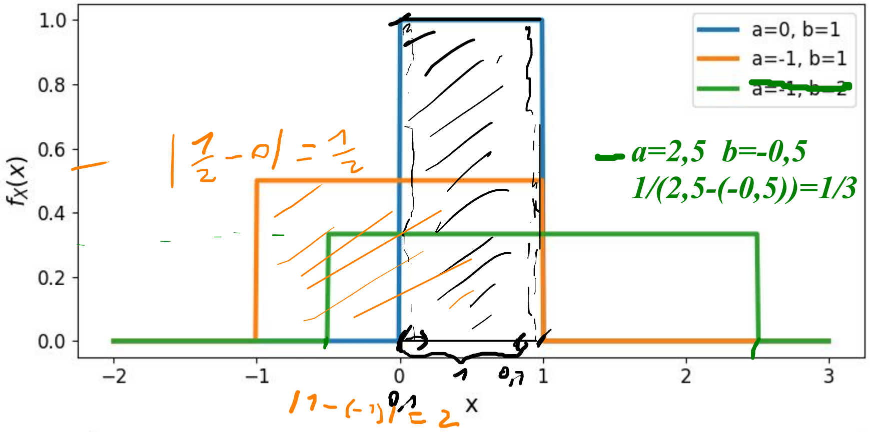
$|a - b| = \varepsilon$ $\quad \int_{\mathbb{R}} f_x = 1$

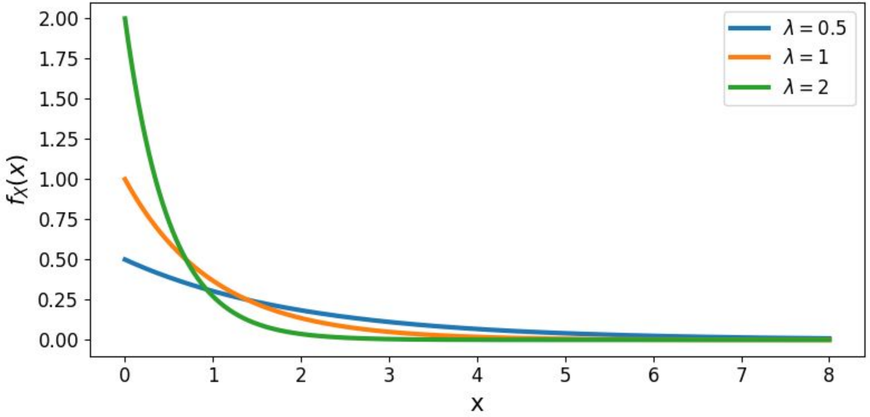
- **Uniforme:** todos los puntos son equiprobables. $X \sim \mathcal{U}(a, b)$

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

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

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



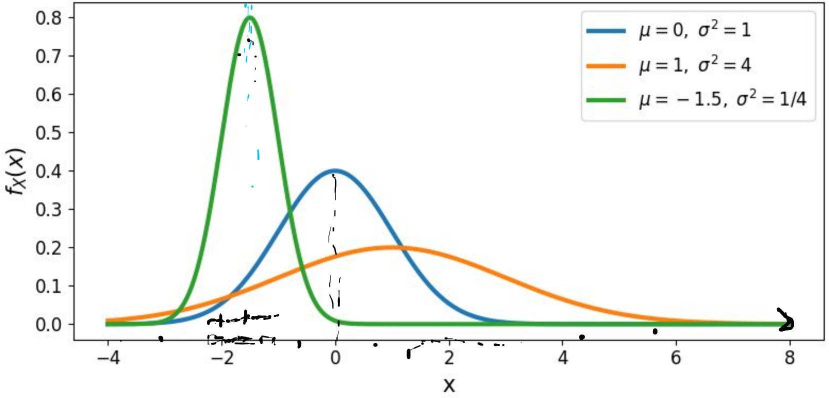


Variables continuas

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

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

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



$X \sim \mathcal{N}(0,1)$, $f_X(x) =$

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

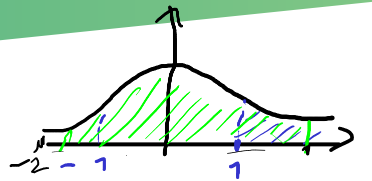
$$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow \overset{Z \sim}{\frac{X-\mu}{\sigma}} \sim \mathcal{N}(0,1) \qquad \text{(estandarización)}$$

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

(combinación lineal de normales es normal)

Ejercicio 1

$X \sim N(0,1)$



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

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

Sea además Y ~ N(2,9) $\quad X + Y \sim N(2 \cdot \mu_X + \mu_Y, 4\sigma^2 + \sigma^2)$

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

$\sim N(2, 13)$

Ejercicio 2

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

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

media

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

matriz de covarianza .

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

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

Distribuciones marginales

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \ldots$

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

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

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

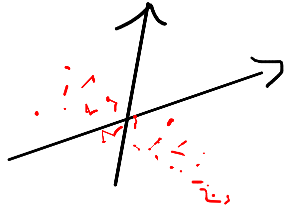
Ejercicio 3

$x - \underline{0} = x$

Sean X,Y dos v.a. Con función de densidad conjunta $\sigma_1^2 = 1, \sigma_2^2 = 2$

$[\underline{x} - \underline{\mu}]$

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



$=0 \quad =0 \quad =1 \quad =1$

1. Calcular E[X], E[Y], var(X), var(Y), y cov(X,Y) $= -0,8 < 0$
2. Hallar las densidades marginales de X e Y
3. Calcular P(X<2, Y<-1)

$X \sim N(\mu_X = 0, \sigma^2 = 1) \qquad Y \sim N(0,1)$

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.