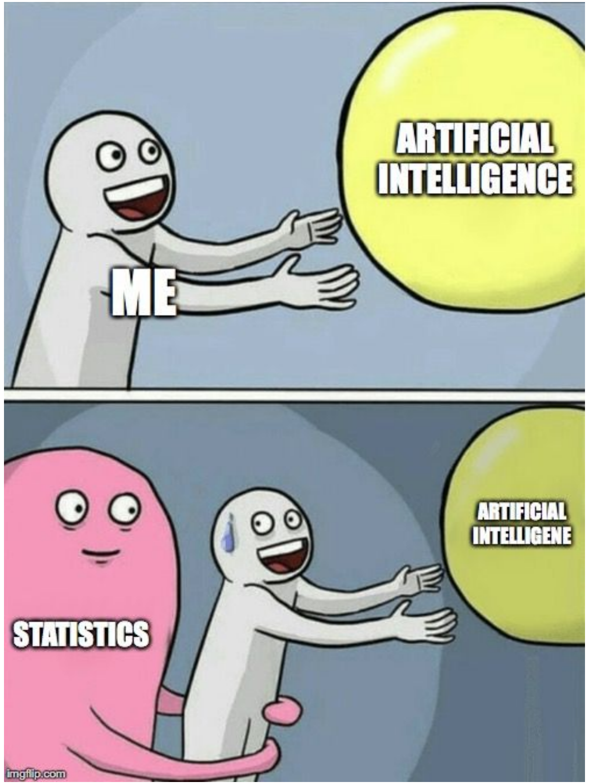
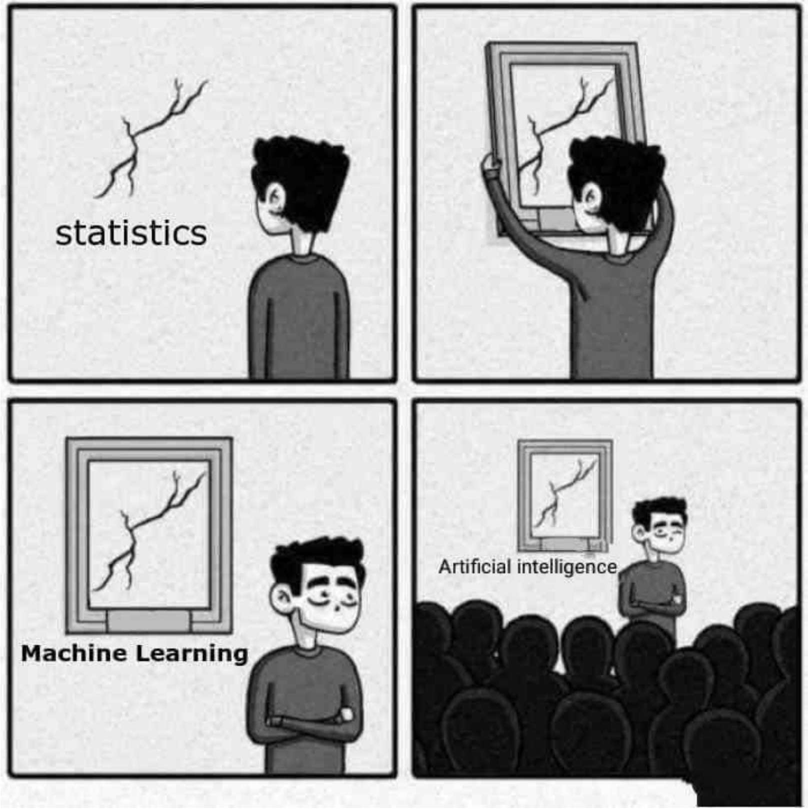
Probabilidad y Estadística

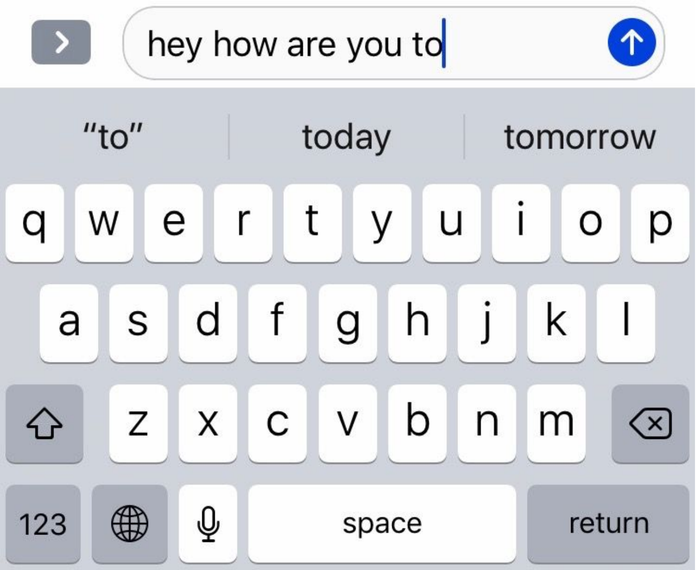
Clase 1

Ud. se encuentra aquí:



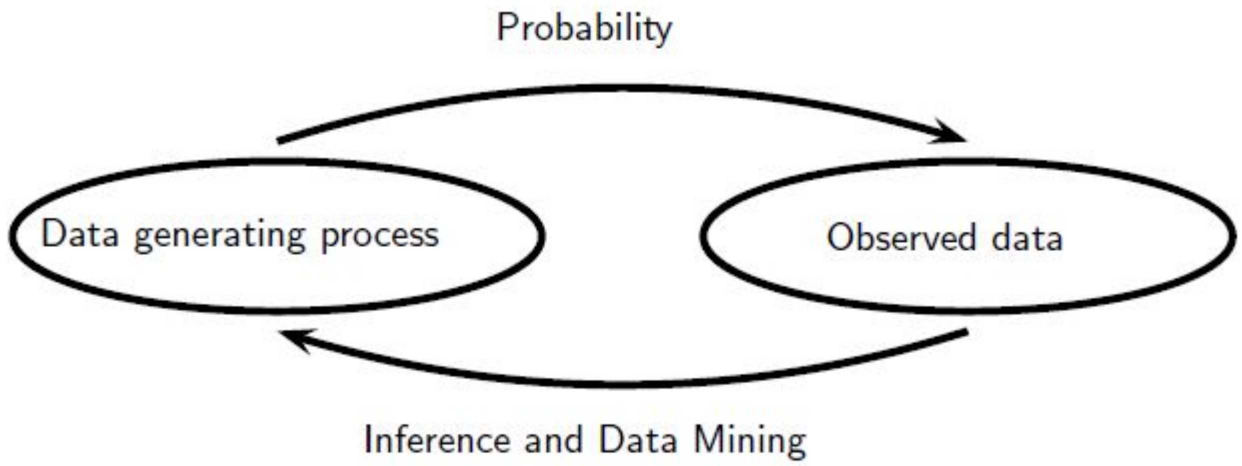
para poder estar aquí:





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.	
TP1	Clase 2	V.a. condicionadas Esperanza condicional → predictivo	P
	Clase 3	ECM y estimadores de cuadrados mínimos Estimador de máxima verosimilitud → aprender modelos	S
TP2	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 $(\Omega, \mathcal{A}, \mathbb{P})$ y variables aleatorias

Espacio Muestral

Algebra de eventos (σ-Algebra)

medida de Pr.

conjunto de eventos

$\Omega = \{\omega_1, \omega_2\}$

$\mathcal{A}$

conjunto de partes de Ω

$\{\emptyset, \{\omega_1\}, \{\omega_2\}, \Omega\}$

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

(σ - algebra on Ω)

$\mathcal{A}$ define los conjuntos medibles (por $\mathbb{P}$)

P

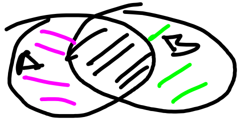
probability measure

[0, 1]

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

medida

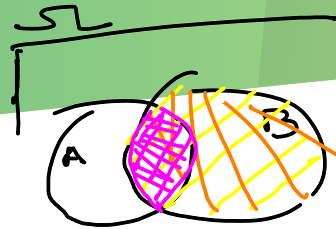
disjuntos



$$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

Probabilidades condicionales y proba. total



$B \approx \Omega$

$P(B|B) = 1$

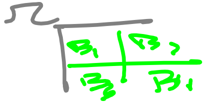
Def: Se llama **probabilidad condicional** de A dado B ($\mathbb{P}(A|B)$) a la probabilidad de que ocurra el evento A sabiendo que B ha ocurrido, y está definida por

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

← Normaliza / $P(B|B)=1$

disjuntos

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



$A = A \cap \Omega = A \cap \left(\bigcup_{j=1} B_j \right)$

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

disjuntos

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

condicional   marginal

Fórmula de probabilidad total

Teorema de Bayes e independencia

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

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

↑ Prob. total

Def: Diremos que dos eventos A y B son **independientes** si y sólo sí vale que (1) = (2) = (3)

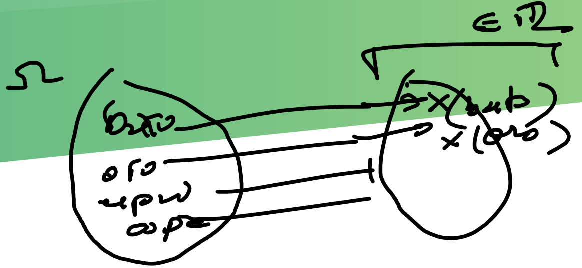
$$\underbrace{P(B|A)}_{= P(B)\ (3)} P(A) = \underbrace{P(A|B)}_{= P(A)\ (2)} \cdot P(B) = \mathbb{P}(A \cap B) \overset{(1)}{=} \mathbb{P}(A)\mathbb{P}(B)$$

$$\Omega \xrightarrow{X} \mathbb{R}$$

σ-Álg

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

define mi σ-Algebra

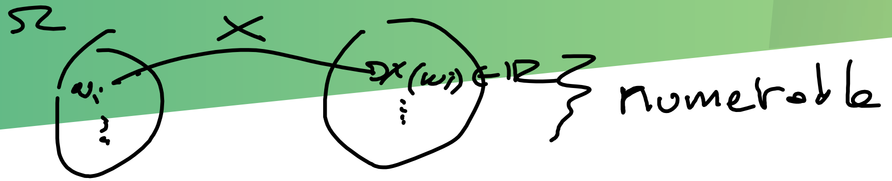
$$F_X(x) = \mathbb{P}(X \leq x)$$

- $F_X(x) \in [0,1] \; \forall x \in \mathbb{R}$
- $F_X(x)$ es monótona no decreciente
- $F_X(x)$ es continua por derecha
- $\lim_{x\to-\infty} F_X(x) = 0$ y $\underbrace{\lim_{x\to\infty} F_X(x)}_{=\mathbb{P}(X\in\mathbb{R})} = 1$

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

$$\text{(1)} \quad p_X(x) = \mathbb{P}(X = x) \quad \forall\, x \text{ en el espacio numerable}$$

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

$$\Omega \xrightarrow{X} \mathbb{R}$$

$$\text{(2)} \quad f_X(x) = \frac{\overbrace{dF_X(x)}^{P(X\leq x)}}{dx} \qquad F_X(x) \overset{\text{TFC}}{=} \int_{-\infty}^{x} f_X(u)\,du$$

¿Qué propiedades debían cumplir cada una de estas funciones?

$$\forall i=1,\dots,n \quad 0 \leq p_X(x_i) \leq 1 \text{ y } \sum_{i=1}^{n} p_X(x_i) = 1 \;; \qquad f_X(x) \geq 0 \;\; \forall x \in \mathbb{R} \text{ y } \int_{-\infty}^{\infty} f_X(u)\,du = 1$$

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

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

$\circledast = P(X^{-1}(A))$

(σ - algebra on Ω) $\xrightarrow{\textbf{X}}$ (σ - algebra on ℝ)

P

probability measure

[0, 1]

→ usamos esta medida por v.a.

$\boxed{P_X(A)} = \circledast$

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

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

$$F_X(x) = \mathbb{P}(\underbrace{X \le x}_{A}), \quad \forall x \in \mathbb{R}$$

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

$$\Omega \xrightarrow{X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}} \mathbb{R}^2$$

Vectores aleatorios

Distribución conjunta y marginales

$\Omega \xrightarrow{\binom{X}{Y}} \mathbb{R}^2$

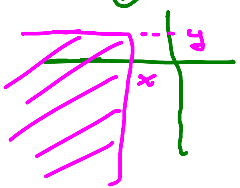
	$Y=0$	$Y=1$	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	1

p_{xy} p_X p_Y

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

$F_X(x) = P(X \leq x)$

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



(*) Caso continuo: $\boxed{f_{X,Y}(x,y)}$ 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

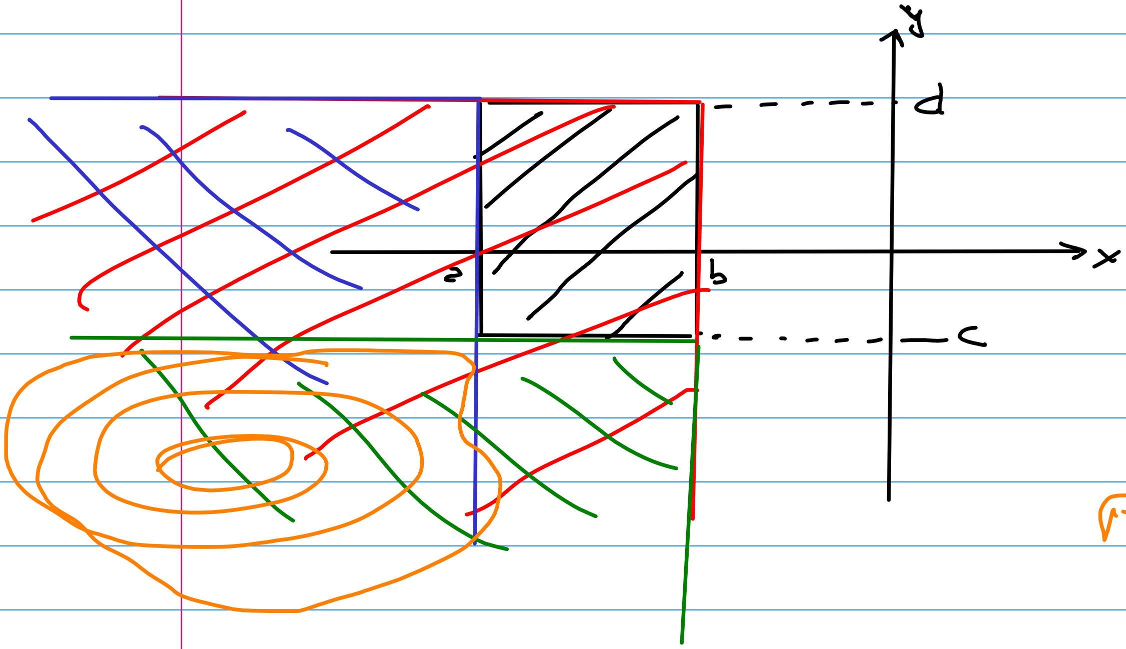
$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$

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

$p_{\cancel{X}Y}(x) = \sum_{\cancel{y}x} p_{X,Y}(x,y)$

(*) $F_{XY}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u,v)\,du\,dv \overset{TFC}{\Longleftrightarrow} f_{XY}(x,y) = \frac{d^2}{dx\,dy} F_{XY}(x,y)$

$$(1) = P(a \le X \le b,\ c \le Y \le d) = \overset{1}{F_{XY}(b,d)} - \overset{1}{F_{XY}(a,d)} - \overset{1}{F_{XY}(b,c)} + F(a,c)$$



$$P((X,Y) \in B) = \int_B f_{XY}(u,v)\,dud$$

Independencia de v.a.

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

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

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

Caso discreto:

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

$\equiv P_{X|Y}(x) = P_X(x)$

Caso continuo:

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

$\equiv f_{X|Y}(x) = f_X(x)$

Momentos

Momentos

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

- Esperanza (o media):

(es lineal)

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

Varianza:

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

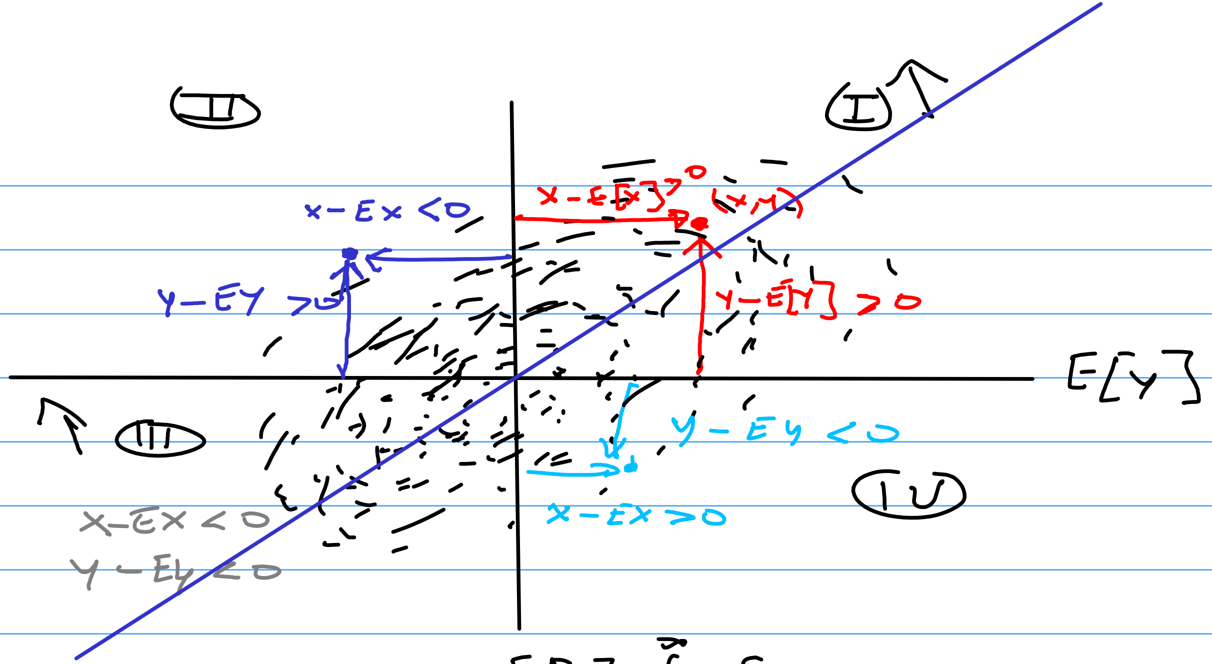
$$\| \quad = E[x^2 - 2xE[x] + E[x]^2] = E[x^2] - \cancel{2} E[x]^2 \cancel{+ E[x]^2}$$

cov(x,x)

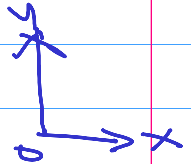
propiedades de E.

Covarianza:

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



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



$$\operatorname{cov}(x,y) = E\left[\left(\underline{X - E[X]}\right)\left(\underline{Y - E[Y]}\right)\right] > 0$$

+	+	} ⊕	(I) y (III)
−	−		
−	+	} ⊖	(II) y (IV)
+	−		

Algunas distribuciones útiles

Variables discretas

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

$P(X=1) = p$ y $P(X=0) = 1-p$.

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

- $Y \sim$ **Binomial(n, p)**: cantidad de éxitos en n ensayos. $P(Y=y) = \binom{n}{y} p^{y} (1-p)^{n-y}$

- $Z \sim$ **Geométrica(p)**: cantidad de ensayos que debo realizar hasta observar el primer éxito $P(Z=3) = p(1-p)^{3-1}$

Ejercicio 0

$\rightarrow P(X=1) = \dots = P(X=6) = \frac{1}{6}$ (es pzoo equiprobable)

Se tira sucesivamente un dado

1. Probabilidad de necesitar menos de 3 tiros hasta ver el primer 2 $\rightarrow$ $Z \sim Geom(1/6)$, $P(Z=3)$
2. Probabilidad de ver exactamente un 2 en 5 tiros

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

$P(Y=1)$

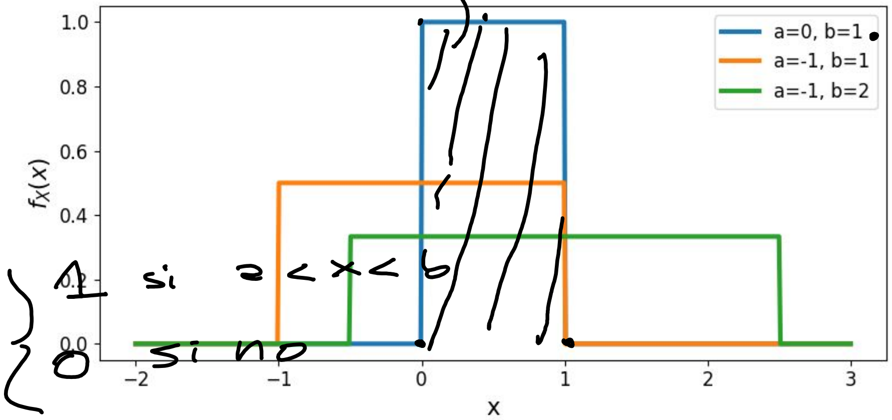
Variables continuas

integra 1

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

función indicadora

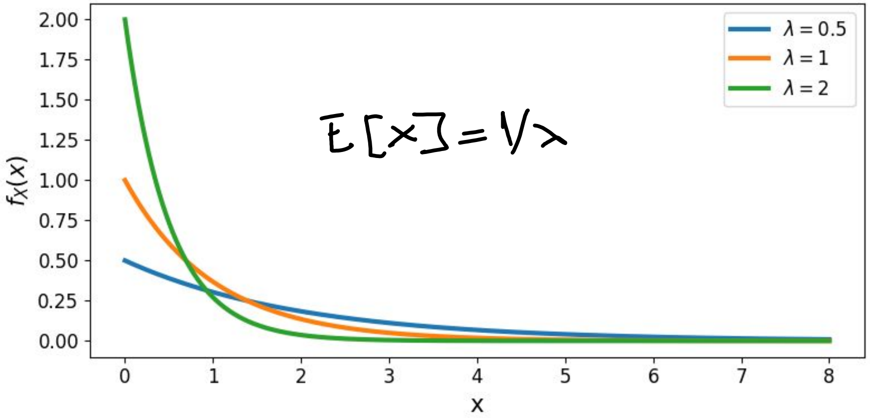
$$f_X(x) = \frac{1}{b-a}\mathbf{I}\{a < x < b\} = \begin{cases} 1 & \text{si } a < x < b \\ 0 & \text{si no} \end{cases}$$



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

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

$\rightarrow P(X > t+s \mid X > s) = P(X > t)$



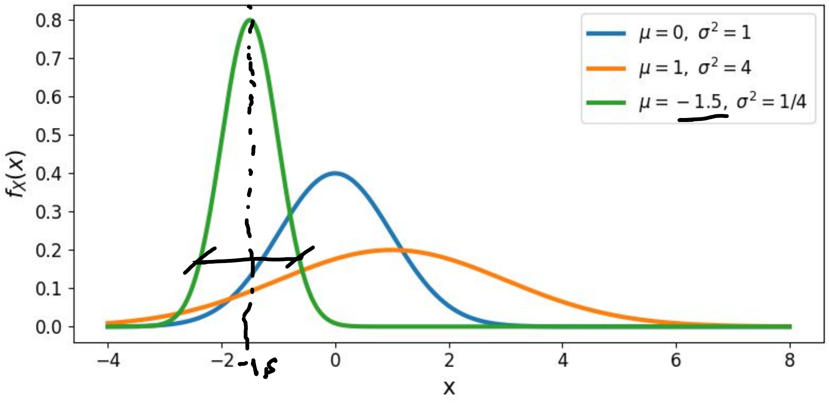
$E[X] = 1/\lambda$

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
σ^2 es la varianza



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

media varianza

"Normal estandarizada"

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

σ ← desvío estándar

indep

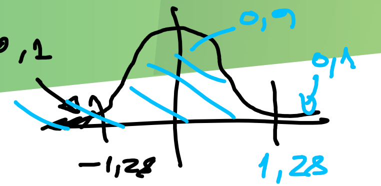
$$X \sim \mathcal{N}(\mu_X, \sigma_X^2), Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \rightarrow aX + bY \sim \mathcal{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

(combinación lineal de normales es normal)

si $X_i \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim \mathcal{N}(\mu, \sigma^2/n)$

Ejercicio 1

$F_X(x) = P(X \le x)$



CA

$E(2X - Y) = 2\overbrace{EX}^{0} - \overbrace{EY}^{2} = 2$

$var(2X - Y) \overset{indep}{=} 4\underbrace{var(X)}_{1} + \underbrace{var(Y)}_{9} = 13$

$\boxed{X \sim N(0,1)}$ → es simétrica.

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

1. $P(X>1) = 1 - P(X \le 1) = 1 - F_X(1) = 0{,}15$
2. $P(X<-1) \overset{vac}{=} P(X \le -1) = F_X(-1) = 0{,}15$ // $P(X < -1) \overset{simetría}{=} P(X > 1)$
3. $P(|X| < 2) = P(-2 < X < 2) = P(X<2) - P(X<-2) = F_X(2) - F_X(-2)$
4. Hallar los cuantiles 0.1 y 0.9

$0{,}1 = P(X \le x_q) = F_X(x_q)$

$x_q = F_X^{-1}(0{,}1) = -1{,}28$

$x_Q = F_X^{-1}(0{,}9) \underset{simetría}{=} -F_X^{-1}(0{,}1) = 1{,}28$

Sea además $\boxed{Y \sim N(2,9)}$

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

$= P\left(\frac{2X - Y - E[2X-Y]}{\underbrace{var(2X-Y)}_{Z \sim N(0,1)}} < \frac{5 - E(2X-Y)}{var(2X-Y)} \right)$

$= P\left(Z < \frac{5 - E(2X-Y)}{var(2X-Y)}\right) = P\left(Z < \frac{5-2}{\sqrt{13}}\right) = F_Z\left(3/\sqrt{13}\right) = 0{,}79$

Ejercicio 2

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

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 $P(X>2) = 1 - P(X \leq 2) = 1 - F_X(2) = 0{,}67$
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

Pérdida de memoria?

$$P(X>5 \mid X>3) \overset{\downarrow}{=} P(X>2)$$

Distribución normal multivariada

Función de densidad conjunta

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

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

$\Sigma^{-1} = \frac{1}{\sigma^2}$

$\leftarrow$ det

media

varianzas

interacciones

matriz de covarianza .

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

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

Distribuciones marginales

diag de Σ

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

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

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

Ejercicio 3

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

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 $\underset{=0}{E[X]}$, $\underset{=0}{E[Y]}$, $\underset{=1}{var(X)}$, $\underset{=1}{var(Y)}$, y $\underset{=-0,8}{cov(X,Y)}$
2. Hallar las densidades marginales de X e Y $y, x \sim N(0, 1)$
3. Calcular P(X≤2, Y≤-1) $= F_{XY}(2,-1) = 0,13$

$F_{XY}(x,y) = P(X \leq x, Y \leq y)$

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.