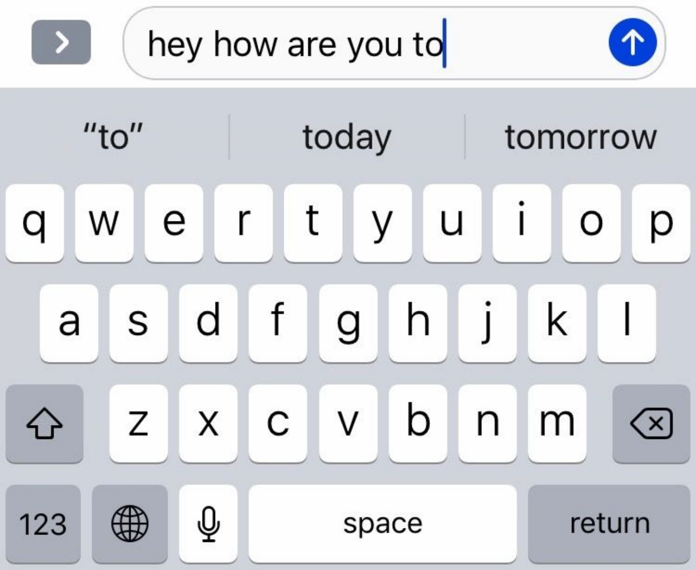
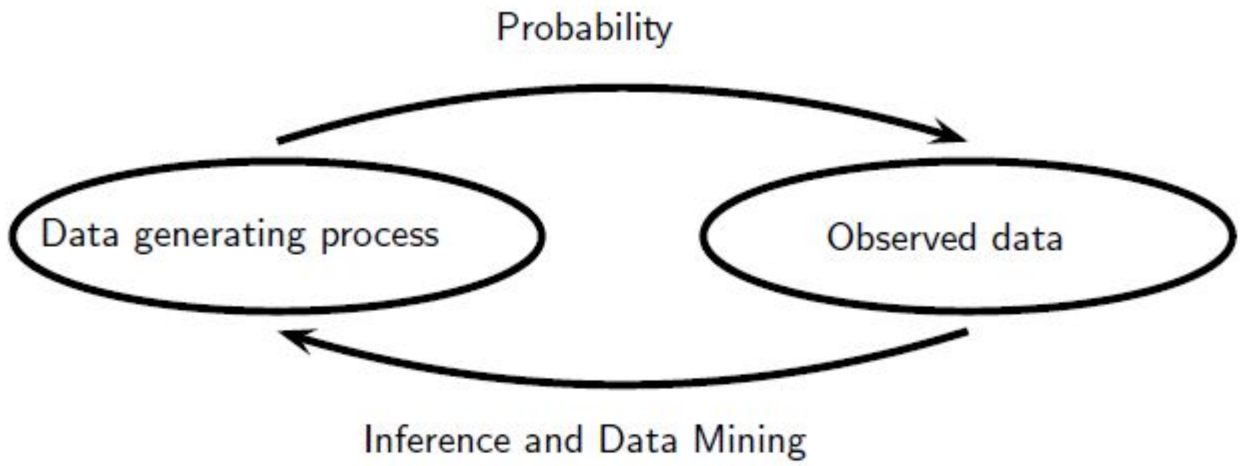
Probabilidad y Estadística

Clase 1



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

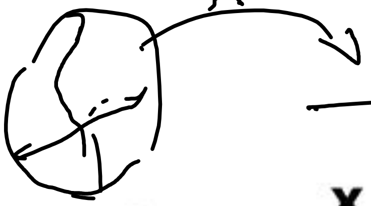
Repaso

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

$\mathcal{A}$ es un conjunto de partes

si es discreto $\{w_0, w_1, w_2\}$ → $\{\{\emptyset\}, \{w_0\}, \dots, \{w_0, w_1\}, \dots, \{w_0, w_1, w_2\}.\}$

Ω $\xrightarrow{X}$ $\mathbb{R}$



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

$\mathcal{A}$,

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

$P(X \leq 0)$

P probability measure

[0, 1]

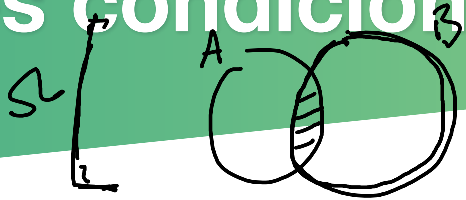
P_X induced measure on **R** by $P_X(A) = P(X^{-1}(A))$ called "distribution of X"

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

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

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

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

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

→ Normaliza

$P(B|B) = 1$ ($P(\Omega) = 1$)

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

1

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

$P(B_i|A) . P(A) = P(A \cap B_i)$

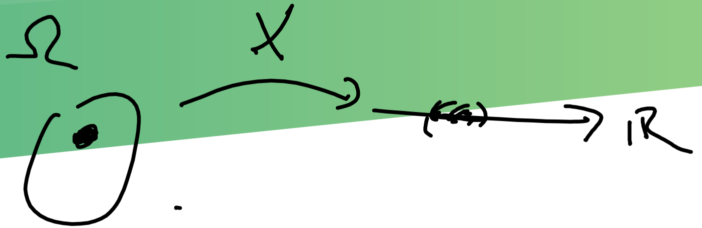
$$\boxed{\mathbb{P}(B_i|A)} = \frac{\boxed{\mathbb{P}(A|B_i)}\mathbb{P}(B_i)}{\underbrace{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}_{= P(A)}}$$

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

$\underbrace{P(B|A)}_{P(B)} . P(A) = \underbrace{P(A|B)}_{P(A)} . P(B) = \boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)} \equiv P(A|B) = P(A)$

Variables aleatorias

Variables aleatorias

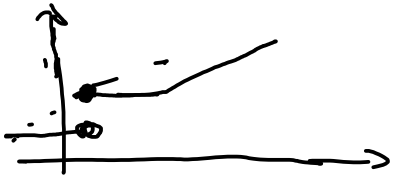


Una va. X es una función que mapea eventos a números.

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

X tiene asociada una función de distribución, definida como

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



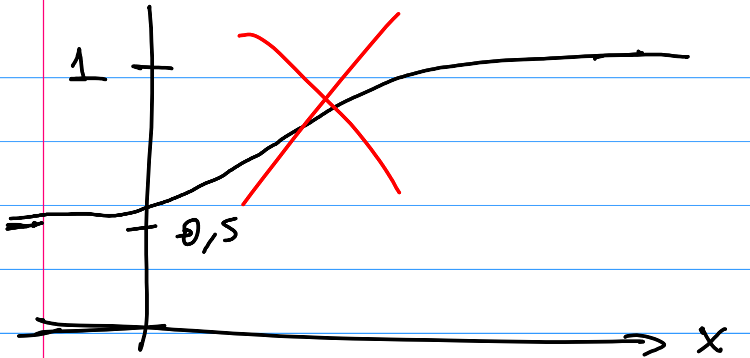
$F_X(x) \in [0,1] \; \forall \, x \in \mathbb{R}$

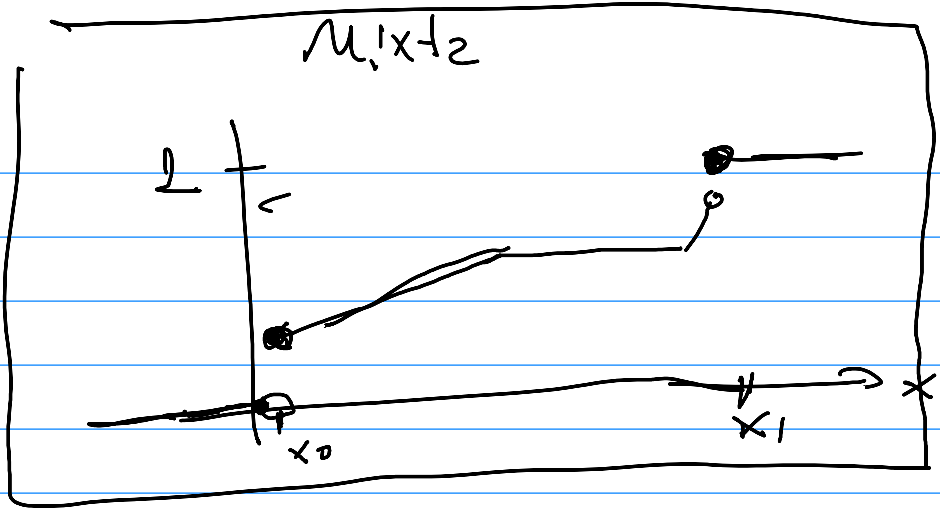
$F_X(x)$ es monótona no decreciente

$F_X(x)$ es continua por derecha

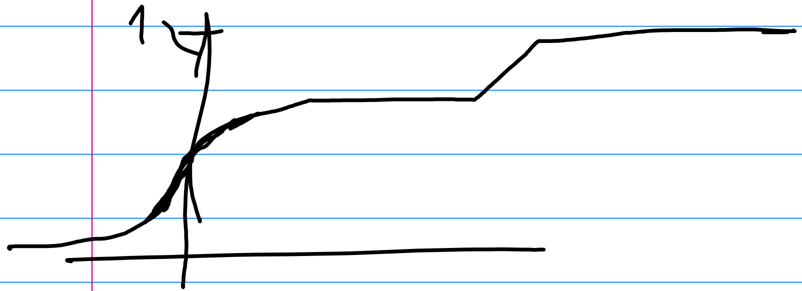
$$\lim_{x \to x_0,\; x > x_0} F_X(x) = F_X(x_0)$$

$\lim_{x \to -\infty} F_X(x) = 0$ y $\lim_{x \to \infty} F_X(x) = 1$

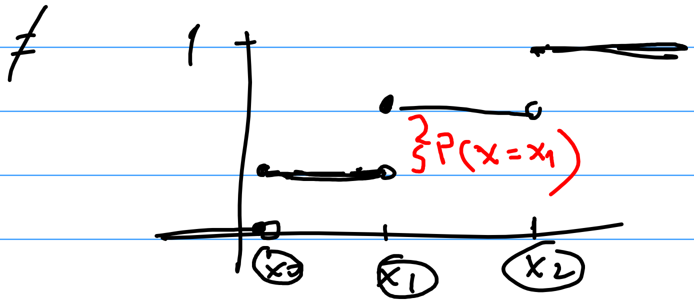




continue



discrète



Tipos de v.a.

$A \xrightarrow{X} \mathbb{R}$ cont

discreta $\to \mathbb{Z}$

$x_0 \quad x_1$

- Discretas (v.a.d): toman valores en un conjunto discreto o enumerable 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

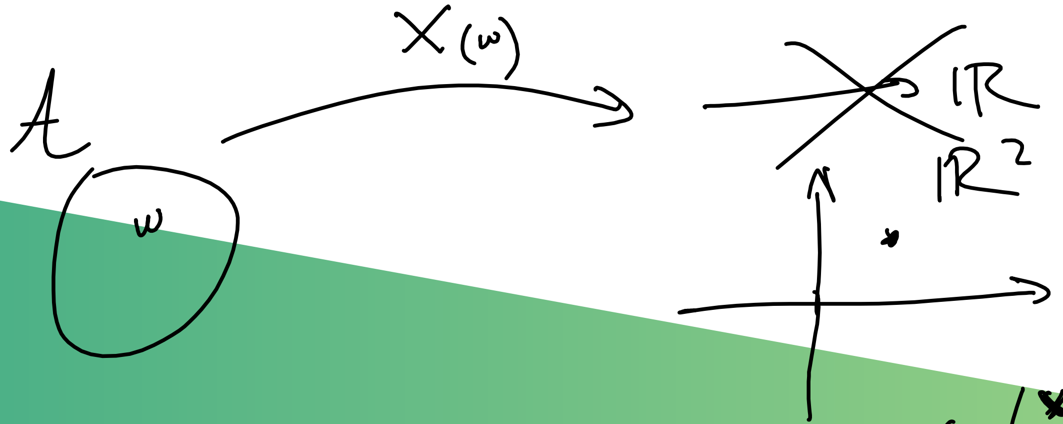
$$f_X(x) = \frac{dF_X(x)}{dx} \geq 0 \qquad F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = \mathbb{P}(X \leq x)$$

$A \subset \mathbb{R}$,

$$\mathbb{P}(X \in A) = \int_{x \in A} f_X(x)\,dx$$

¿Qué propiedades debían cumplir cada una de estas funciones?

$$\int_{x \in \mathbb{R}} f_X(x)\,dx = 1$$



$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Vectores aleatorios

Distribución conjunta y marginales

$\mathbb{R}^2$ $\begin{bmatrix} X \\ Y \end{bmatrix}$

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

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:

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

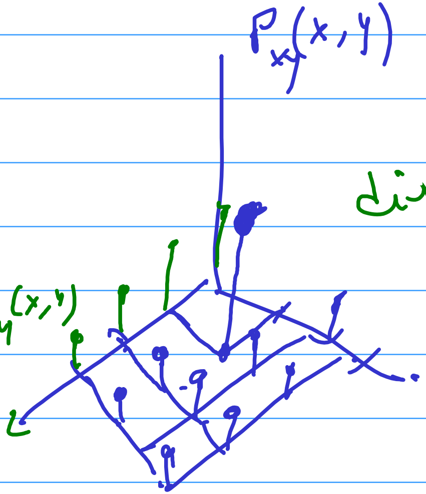
Caso continuo: $f_{X,Y}(x,y)$ es la función de densidad conjunta y se definen las funciones de densidad marginales como $\boxed{f_X(x)} = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$ y $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$

$f_{XY}(x,y) = \frac{d^2}{dx\,dy} F_{XY}(x,y)$

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

continua

$$f_X(x) = \int_{\mathbb{R}} \underline{f_{XY}(x,y)\,dy}$$



discreta

$$P_X(x) = \sum_y P_{xy}(x,y)$$

Independencia de v.a.

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

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

$A \perp\!\!\!\perp B$

$P(A \cap B) \overset{\Updownarrow}{=} P(A)P(B)$

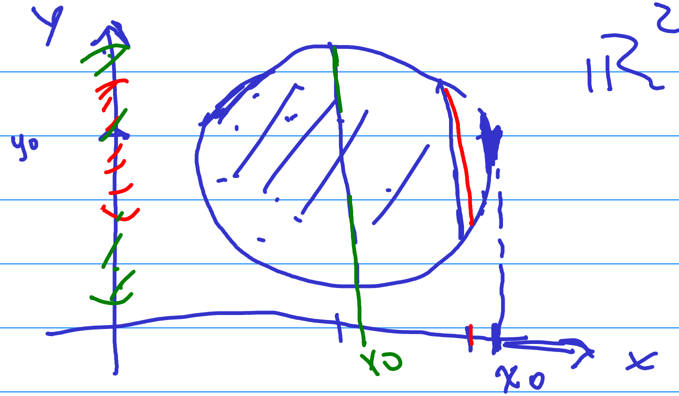
Caso discreto:

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

Caso continuo:

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

$(x,y) \sim \mathcal{U} \quad \longleftrightarrow \quad P_{xy}$ o f_{xy} es cte. $\forall x,y$

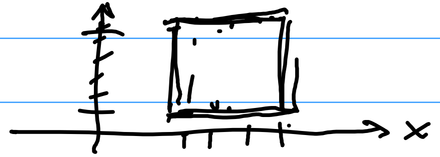


$x, y \in$ 

¿Pueden ser $x \perp\!\!\!\perp y$? No

$$\boxed{f(x,y) = f(x)\, f(y)}$$

si son $\perp\!\!\!\perp$ $\Rightarrow$ el dominio de $f_{xy}(x,y)$ es un producto tensorial (incluye $\mathbb{R}^2$)



Momentos

Momentos

sin cambio de vars — cambio de vars

$g(x)=X$ $\quad X \sim f_X$

$Y = g(X) \sim f_Y$

Esperanza (o media):
(vale para vectores)

$$\mathbb{E}[g(X)] = \int g(x) \overset{\text{densidad}}{f_X(x)} dx = \int y \, f_Y(y)\, dy$$
$$= \sum g(x) p_X(x)$$

Varianza:

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

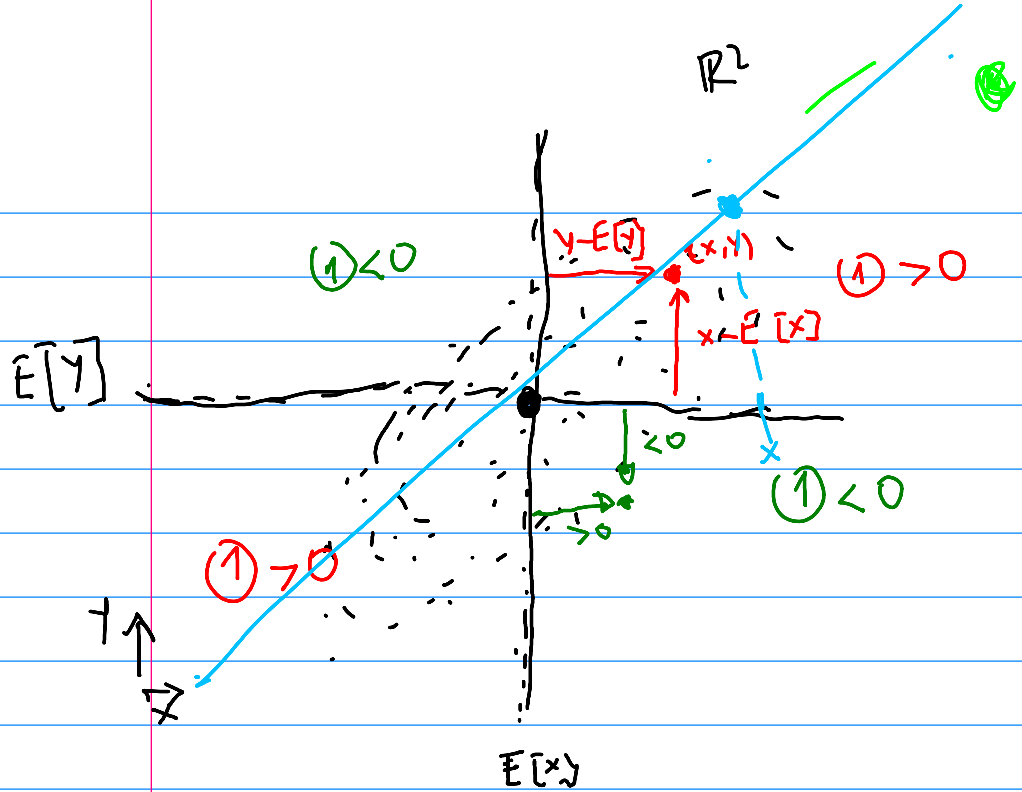
cte

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

Covarianza:

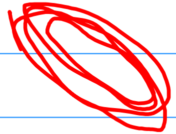
$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}(X)\mathbb{E}(Y)$$

$\begin{pmatrix} X \\ Y \end{pmatrix}$

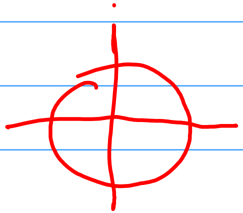


$$\text{cov}(x,y) = E\Big[\underbrace{(x - E[x])(y - E[y])}_{①}\Big]$$

> 0



$\text{cov} \leq 0$



$\text{cov} = 0$

Algunas distribuciones útiles

Variables discretas

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

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

  $P(X=1)$, $X \sim Bin(5, 1/6)$

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

Ejercicio 0

Se tira sucesivamente un dado $\rightarrow P(x) = \frac{1}{6}\ , \ x \in \{1, \dots, 6\}$

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

1)

- $X_1 = 2$.
- $X_1 \neq 2$ y $X_2 = 2$

$$P(A) = \underbrace{\frac{1}{6}}_{P(x=2)} + \underbrace{\frac{5}{6}}_{P(x\neq 2)} \cdot \underbrace{\frac{1}{6}}_{P(x=2)} \approx 0,3$$

$X \sim$ geométrica$(1/6)$

$$P(A) = P(X < 3) = P(X \leq 2) = F_X(2)$$

2) $A =$ "veure exact. un 2 en 5 tirs" $\qquad P(X_i = x) = \frac{1}{6} \quad x \in \{1, 2, 3, \ldots 6\}$

X_1	X_2	X_3	X_4	X_5
2	$\neq 2$	$\neq 2$	$\neq 2$	$\neq 2$
	2		$\neq 2$	
		2		
			2	
				2

5 esdeveniments incompatibles

$\leftarrow \boxed{\frac{1}{6}} \cdot \boxed{\left(\frac{5}{6}\right)^4} \cdot 5 = P(A)$

$\frac{1}{6} = P(X_1 = 2)$

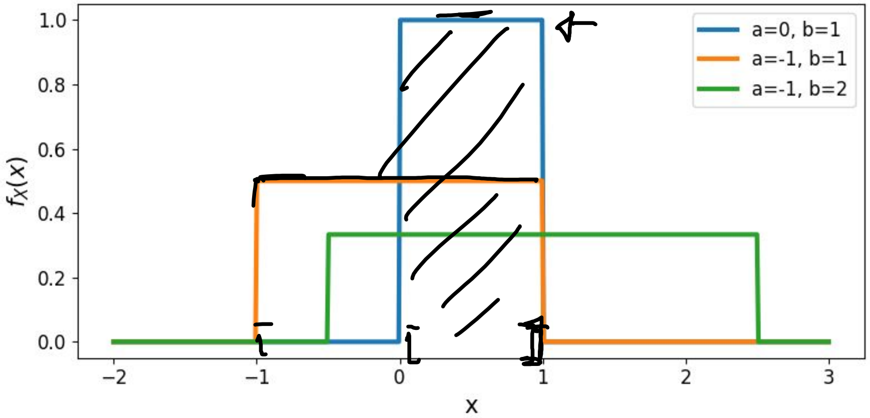
$\left(\frac{5}{6}\right)^4 = P\left(\bigcap_{i=2}^{5} X_i \neq 2\right)$

Variables continuas

$(b-a)\cdot f = 1$

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

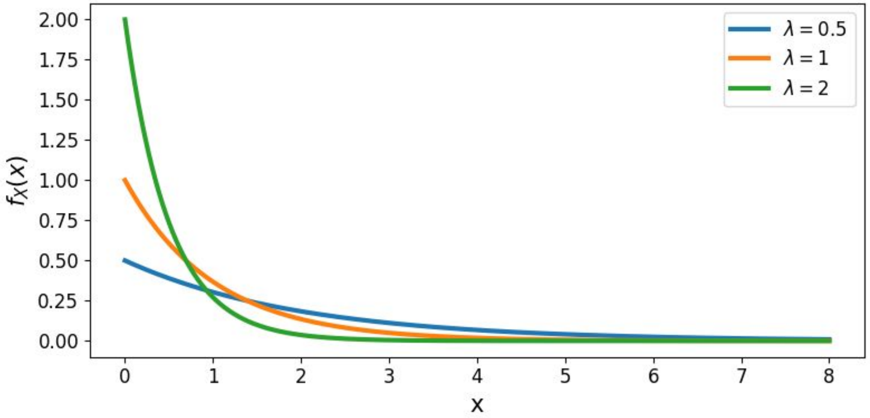
$$f_X(x) = \underbrace{\frac{1}{b-a}}_{\text{Normaliza}}\mathbf{I}\{\underbrace{a < x < b}_{\text{soporte}}\}$$



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

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

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

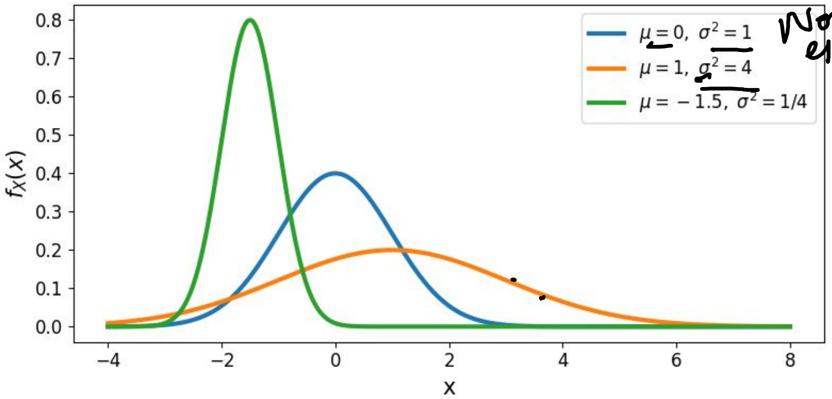


Variables continuas

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

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

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



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

estandarización

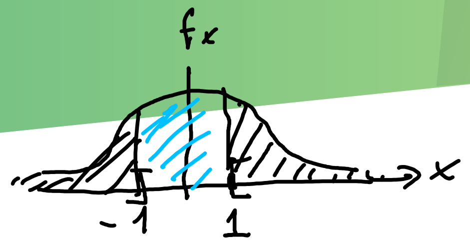
$$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow \boxed{\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)}$$ (estandarización)

Var = E[(X-EX)²]

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

(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
2. X<-1
3. |X| < 1
4. Hallar los cuantiles 0.1 y 0.9

Sea además Y ~ N(2,9)

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

$$P(X>1) = \int_1^{\infty} f_X(x)\,dx = 1 - \underbrace{P(X \le 1)}_{F_X(1)} \simeq 0{,}15$$

$$P(X<-1) = F_X(-1) \simeq 0{,}15$$

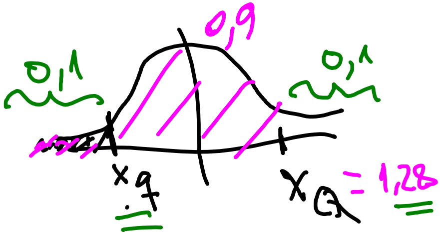
$$P(|X|<1) = 1 - 2 \cdot P(X>1)$$

$$P(-1<X<1) = F_X(1) - F_X(-1)$$

$$x_q \,/\, \underbrace{P(X<x_q)}_{F_X(x_q)} = 0{,}1$$

$$F_X(x_q) = 0{,}1$$

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



$$W = 2X + Y \quad , \quad X \sim N(0,1)$$

$$Y \sim N(2,9)$$

$$W \sim N\left(\overbrace{2 \cdot 0 + 2}^{2} \quad \overbrace{+\ 2^2\, 1 + 9}^{13}\right)$$

$$P(W<5) = P\left(\underbrace{\frac{W-2}{\sqrt{13}}}_{Z \sim N(0,1)} < \frac{5-2}{\sqrt{13}}\right)$$

$$= P\left(Z < \frac{5-2}{\sqrt{13}}\right) = F_Z\left(\frac{5-2}{\sqrt{13}}\right)$$

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 $P(T>5 \mid T>3) = P(T>2) = 0{,}67$

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

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

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

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

media

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

matriz de covarianza .

es simétrica

$$\Sigma = \begin{bmatrix} \sigma_1^2 & cov(X_1, X_2) & \ldots & cov(X_1, X_n) \\ cov(X_2, X_1) & \sigma_2^2 & \ldots & cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n, X_1) & cov(X_n, X_2) & \ldots & \sigma_n^2 \end{bmatrix} \succ 0 \text{ (def+)}$$

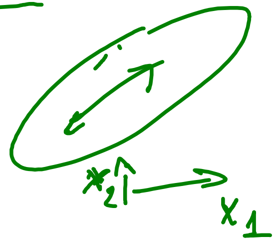
Notación: $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$

Distribuciones marginales

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

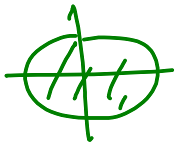
cov(x_1, x_2)

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



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

si cov $= 0$



$$f_{xy}(x,y) := \left(\begin{pmatrix}\mu_x \\ \mu_y\end{pmatrix}, \begin{bmatrix}\sigma_x^2 & 0 \\ 0 & \sigma_y^2\end{bmatrix}\right) = i$$

general $\left\{ \text{si } X \perp\!\!\!\perp Y \implies \mathrm{cov}(X,Y) = 0 \right.$

para la Normal vale también

$$\boxed{\text{si } \underline{\mathrm{cov}(X,Y)=0} \quad y \quad (X,Y) \text{ es } \underline{\text{Normal}} \\ \Downarrow \\ X \perp\!\!\!\perp Y}$$

Dem

$$\Sigma = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$$

$$f_{XY}(x,y) = \underbrace{f_X(x)}_{N(\mu_X, \sigma_X^2)} \cdot \underbrace{f_Y(y)}_{N(\mu_Y, \sigma_Y^2)} \implies X \perp\!\!\!\perp Y$$

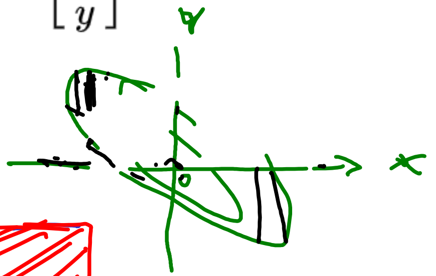
Ejercicio 3

$F_X = P(X \le x)$ ← en $\mathbb{R}$

$\boxed{F_{XY} = P(X \le x, Y \le y)}$ ← en $\mathbb{R}^2$

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

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



=0 =0 =1 =1 =−0,8

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) $= F_{XY}(2,-1) \approx 0{,}13$

$P = \int f_{XY}\, dxy$. complicado

$P(X, Y \in$ ▨ $)$

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.