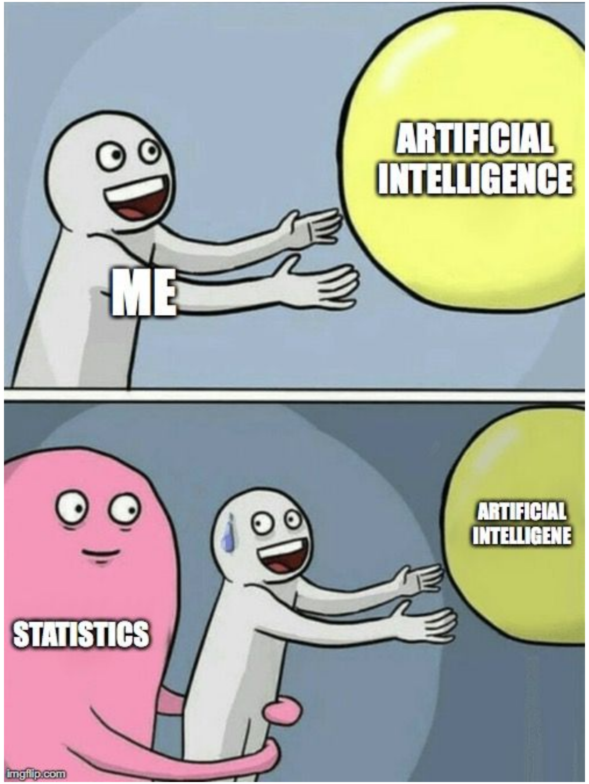
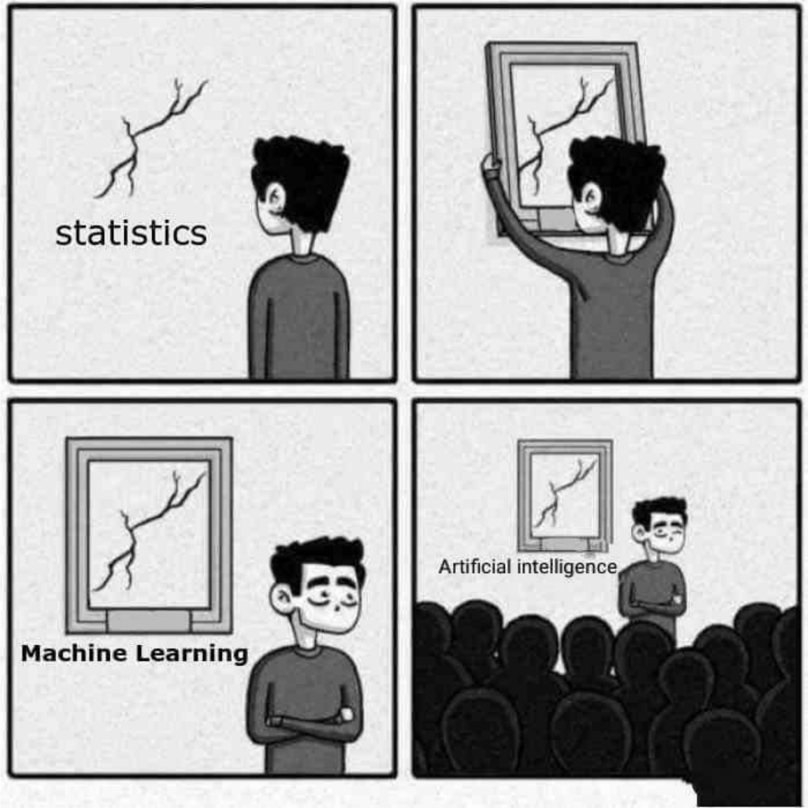
Probabilidad y Estadística

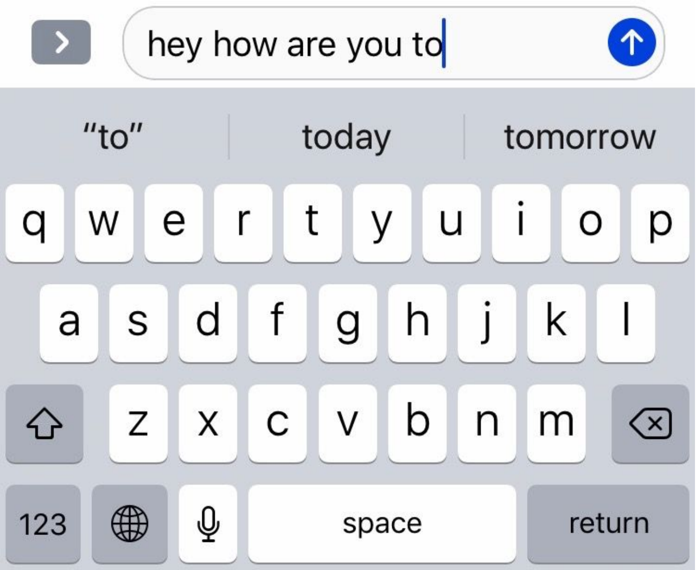
Clase 1

Ud. se encuentra aquí:



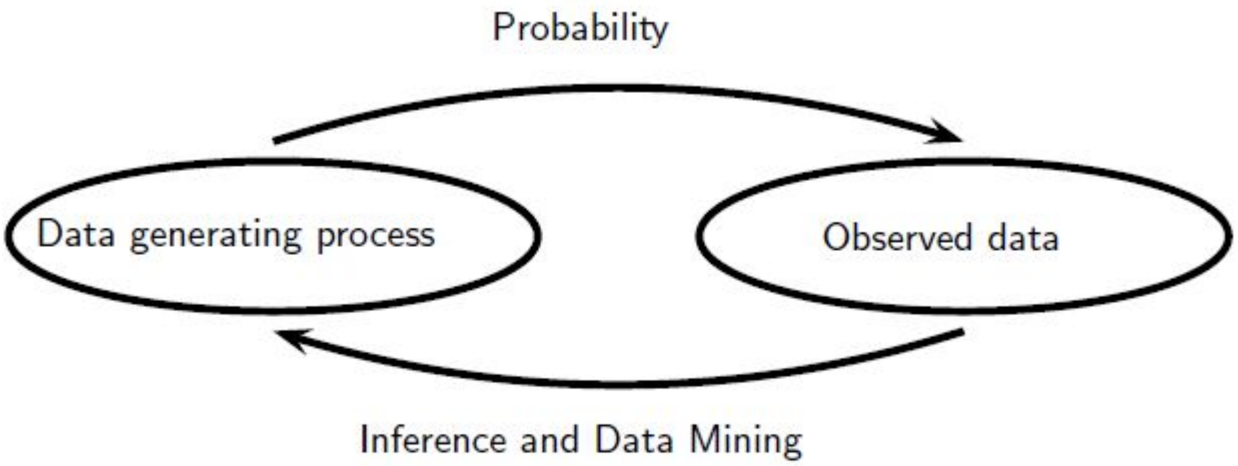
para poder estar aquí:





Ejemplos de probabilidad en IA?

Probabilidad vs. Estadística



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

Cronograma

1) Matricularse al campus.

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

2 semanas

Examen
Cuatrimestral

Repaso

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

$\Omega = \{w_1, w_2\}$, $\mathcal{A} = \{\emptyset, \{w_1\}, \{w_2\}, \{w_1, w_2\}\}$

Espacio Muestral — Un conjunto

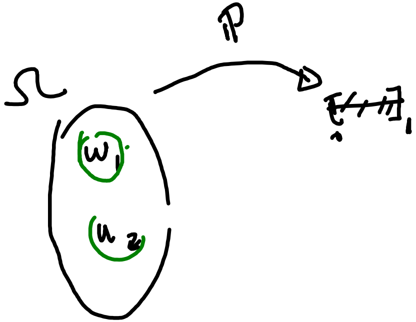
σ-Álgebra

medida de prob.

$$\left\{ \begin{array}{c} \Omega \in \mathcal{A} \\ B \in \mathcal{A} \Rightarrow B^c \in \mathcal{A} \\ B, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A} \end{array} \right\}$$

$\mathcal{A}$

(σ - algebra on Ω)



P

probability measure

[0, 1]

$P(\emptyset) = 0$

$P(w_1) = p_1$

$P(w_2) = p_2$

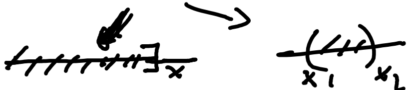
$P(\Omega) = 1$

$$\left\{ \begin{array}{c} 0 \le \mathbb{P}(A) \le 1, \ \forall A \in \mathcal{A} \\ \mathbb{P}(\Omega) = 1 \\ \underbrace{A \cap B = \emptyset}_{\text{disjuntos}}, \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \end{array} \right.$$

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

A∩B

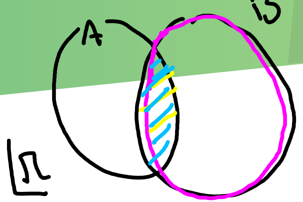
$\Omega \in \mathbb{R}$



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

Probabilidades condicionales y proba. total

$\rightarrow \quad P(A\cap B) = P(A|B)P(B)$



B → es como Ω para A|B

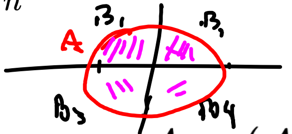
Def: Se llama **probabilidad condicional** de A dado B ($\mathbb{P}(A|B)$) a la probabilidad de que ocurra el evento A sabiendo que B ha ocurrido, y está definida por

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

→ Normaliza

satisface $P(B|B)=1$

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

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

$$= P(A\cap\Omega) = P\Big(A\cap \bigcup_j B_j\Big) = P\Big(\bigcup_j (A\cap B_j)\Big) = \sum_j P(A\cap B_j)$$

Fórmula de probabilidad total

Teorema de Bayes e independencia

$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)\,P(B_i)}{\sum P(A|B_i) \cdot P(B_i)}$

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

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

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

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

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

Variables aleatorias

Variables aleatorias

$\Omega \ni \omega_x \xrightarrow{X} \{1, \dots, 4\}$

3) $\lim\limits_{x \to x_0} F_X(x) = F_X(x_0)$

$x > x_0$

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

semirrecta

x

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

$F_X(x_Q) = 0{,}9$

$\mathbb{P}(X \leq x_q) = F_X(x_q) = 0{,}1$ "el cuantil 0,1" es x_q

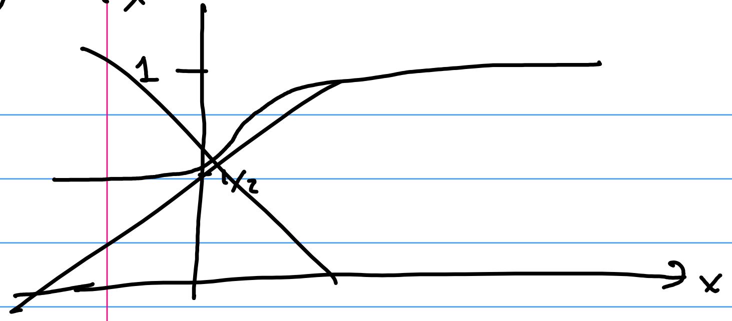
- $F_X(x) \in [0,1] \; \forall x \in \mathbb{R}$ ✓
- 2) $F_X(x)$ es monótona no decreciente
- 3) $F_X(x)$ es continua por derecha ,
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ y $\lim\limits_{x \to \infty} F_X(x) = 1$.

además se define la función de distribución inversa o función cuantil como

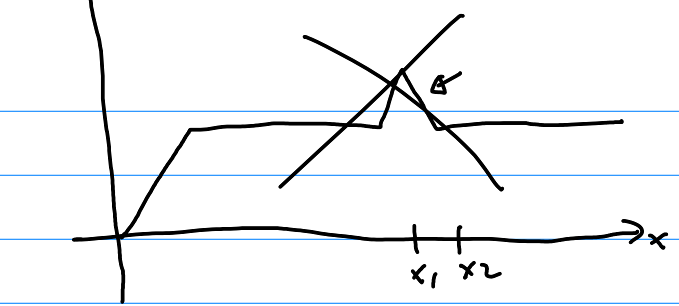
$$F_X^{-1}(q) = \inf\{x : F(x) > q\}$$

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

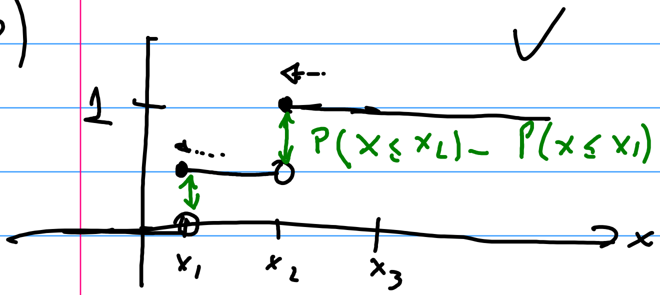
1) $F_X(x) = P(X \leq x)$



2) $P(X \leq x_1) \leq P(X \leq x_2)$

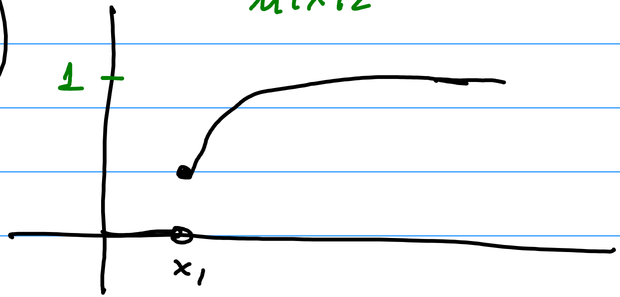


3)



$P(X \leq x_2) - P(X \leq x_1) = P(X = x_2)$

4) mixta

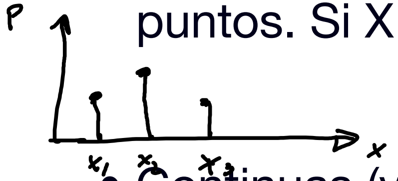


$\sum_i \alpha_i F_{X_i}(x)$ con $\Sigma \alpha_i = 1$, $\alpha_i \geq 0$

es una fun de distrib.

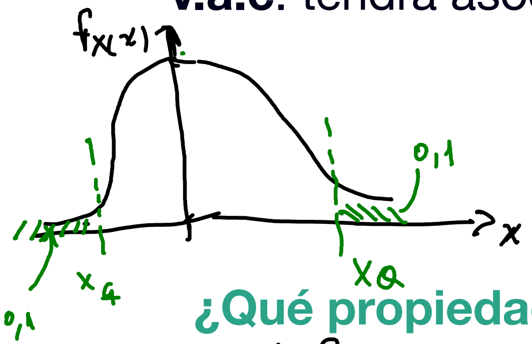
Tipos de v.a.

- 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



$$\boxed{p_X(x) = \mathbb{P}(X = x)} = F_X(x_k) = p_X(x_1) + \dots + p_X(x_n)$$

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



$$f_X(x) = \frac{dF_X(x)}{dx} \overset{T.F.C.}{\equiv} F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$$

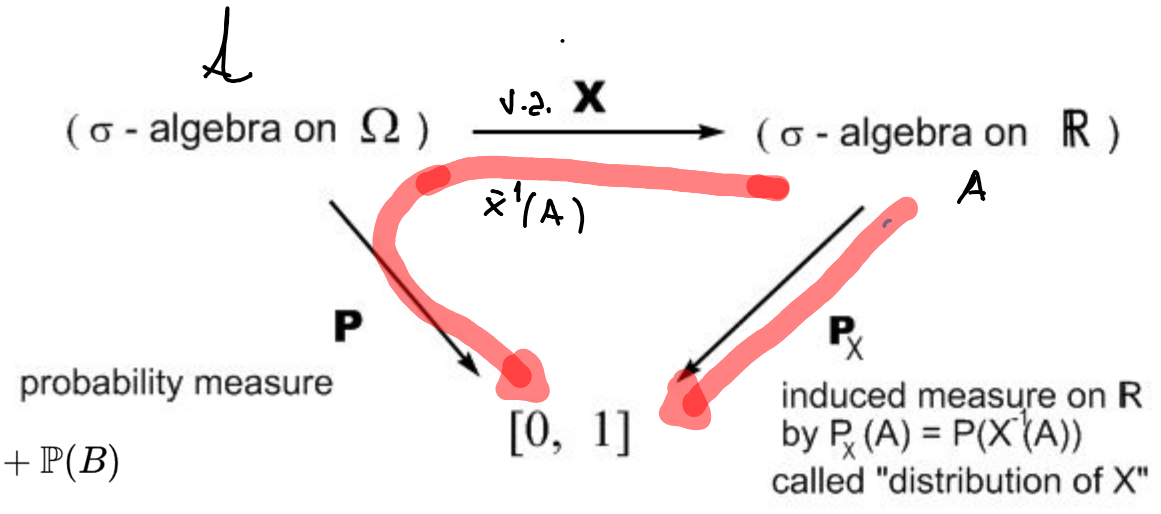
¿Qué propiedades debían cumplir cada una de estas funciones?

1) $f_X(x) \geq 0$, $p_X(x) \geq 0$ $\forall x$.

2) $\int_{-\infty}^{\infty} f_X(x) = 1$, $\sum_{x \in \Omega} p_X(x) = 1$.

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

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



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

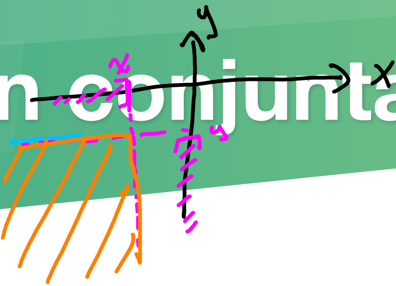
$$F_X(x) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}$$

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

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

Vectores aleatorios

Distribución conjunta y marginales



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

$P_{XY}(x,y)$

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:

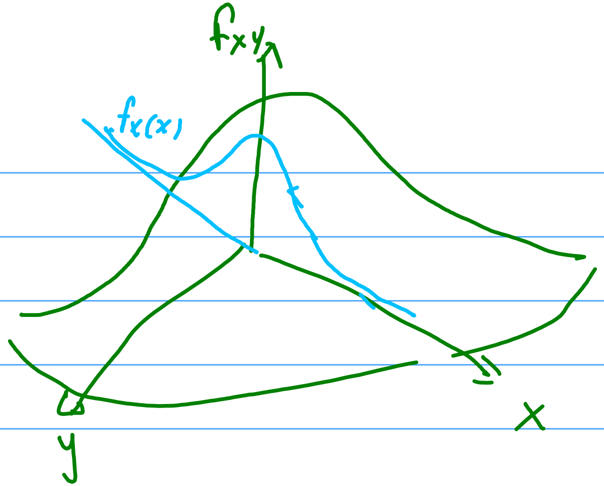
- $\mathbb{P}(a < X \leq b, c < Y \leq d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$ ✓

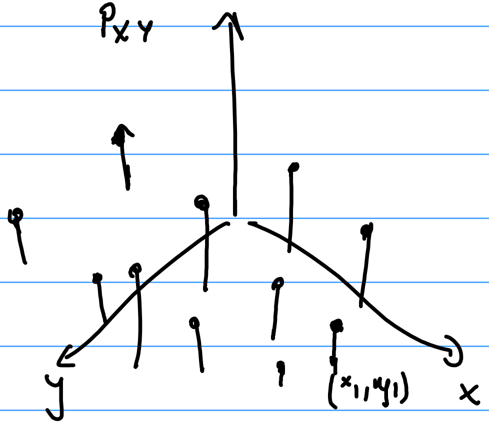
Caso continuo: $f_{X,Y}(x,y)$ es la función de densidad conjunta y se definen las funciones de densidad marginales como $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$ y
$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$

$\neq F_X(x) = \int_{-\infty}^{x} f_X(x)dx = P(X \leq x)$

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
$p_Y(y) = \sum_x p_{X,Y}(x,y)$

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(x,y)\,dx\,dy$$

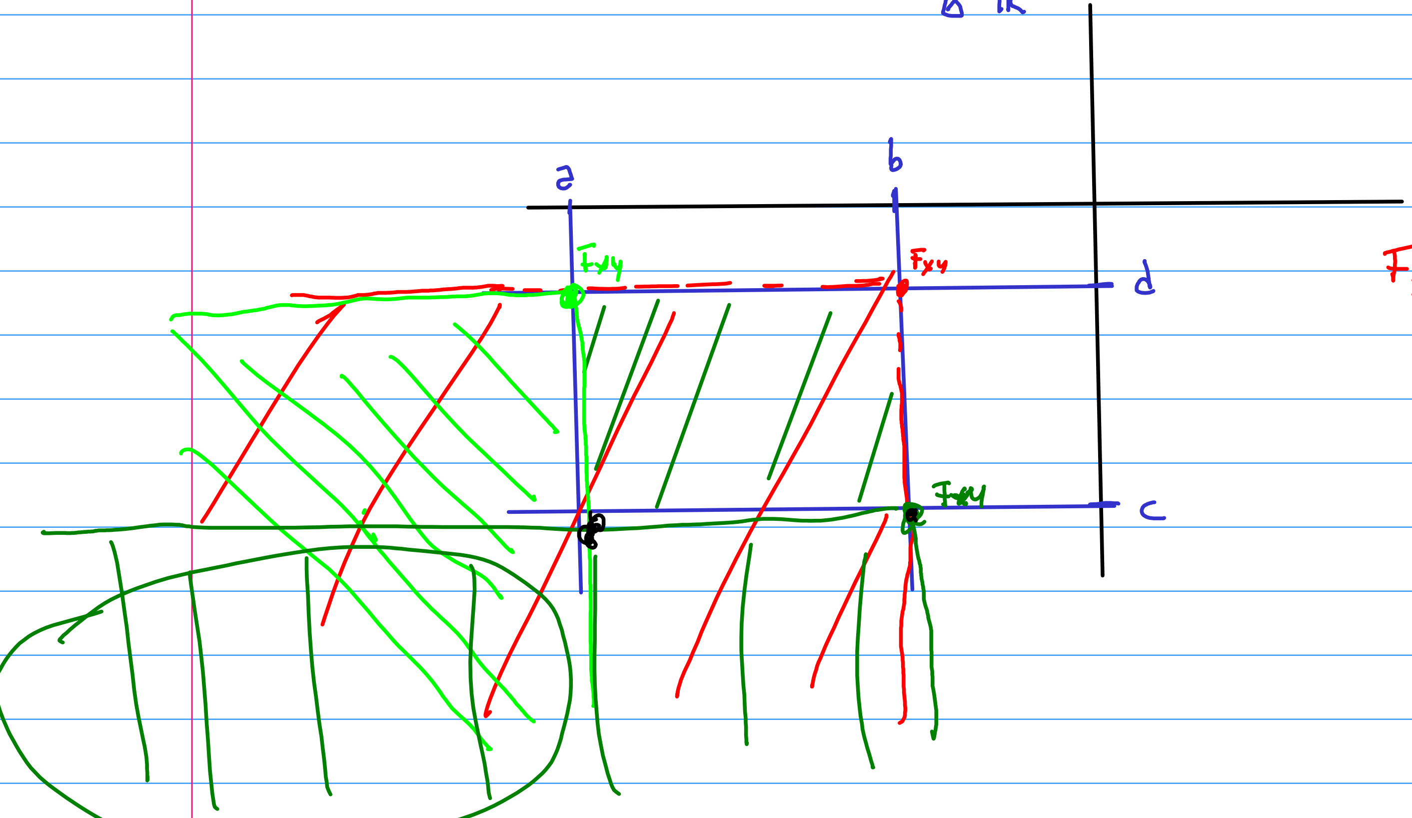




$P(x_1, y_1) = P_1$

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

$\begin{pmatrix} X \\ Y \end{pmatrix}$ $\Omega \to \mathbb{R}^2$



$$F_{X,Y}(b,d) = P(X \le b, Y \le d)$$

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:

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

$$\underbrace{f_{X|Y}}_{+\text{ adelante}}(x) \cdot f_Y(y) = f_{XY}(x,y) \quad \Rightarrow \quad f_{X|Y} = f_X$$

Momentos

Momentos

Esperanza (o media): → es lineal

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

$g(x) = x$

$$\boxed{\begin{aligned}\mathbb{E}[g(X)] &= \int g(x) f_X(x) dx \\ &= \sum g(x) p_X(x)\end{aligned}}$$

Tranf. de variable

$\boxed{Y = g(x)}$

$E[Y] = \int y \, f_Y(y) \, dy$

Varianza:

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

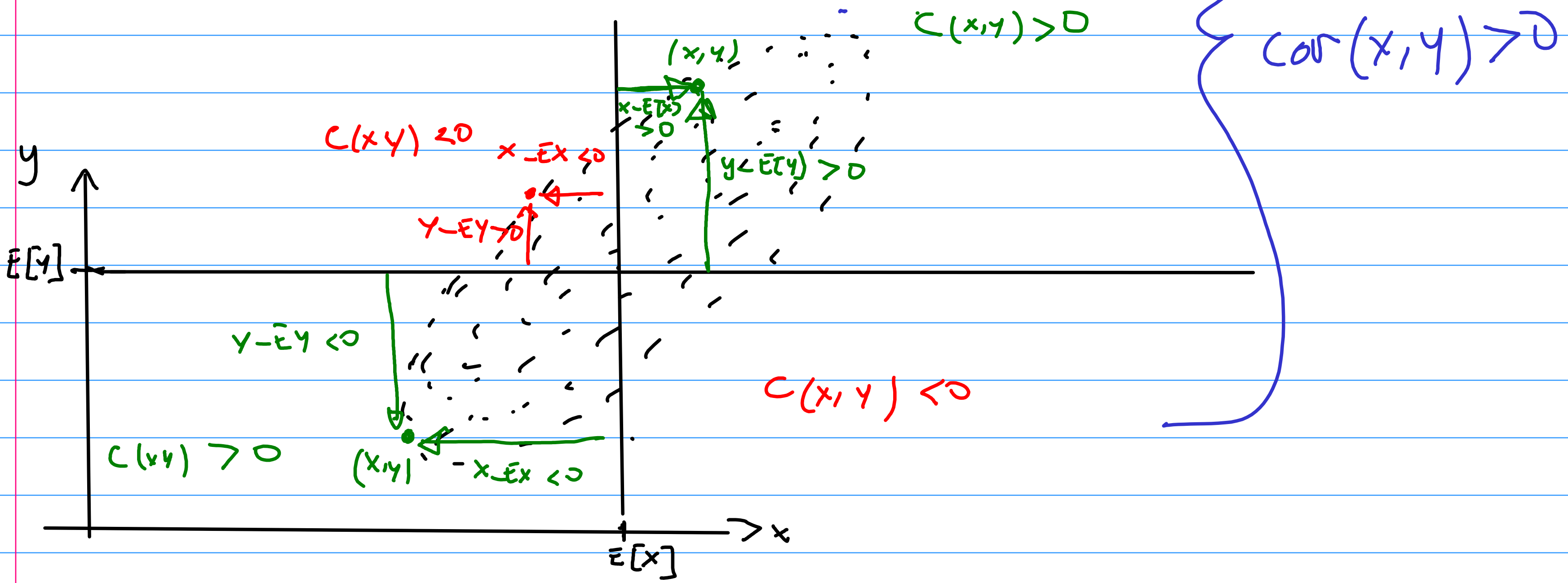
$= cov(X, X)$

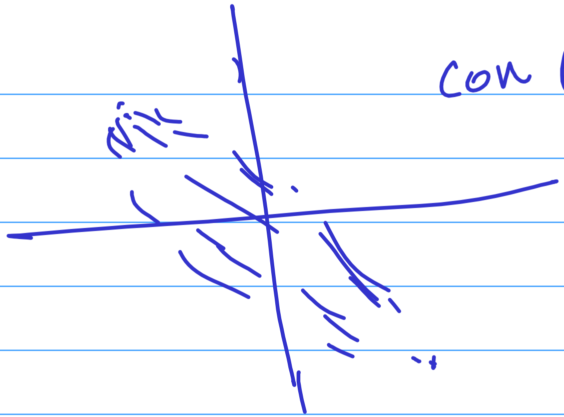
Covarianza:

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

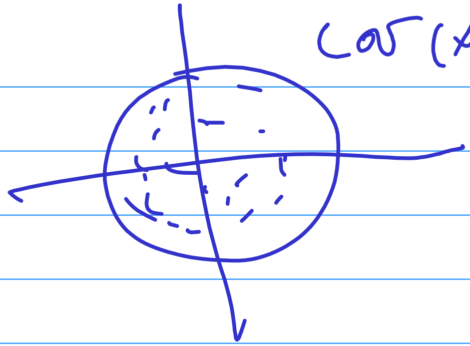
$$\operatorname{cov}(X,Y) = E\Big[\underbrace{(X - E[X])\,(Y - E[Y])}_{C(x,y)}\Big]$$

$$E \to \iint [\quad]\, f_{XY}(x,y)\,dx\,dy \qquad E[X] \to \int x\, f_X(x)\,dx$$





$cov(x,y) < 0$



$cov(x,y) = 0$

Algunas distribuciones útiles

Variables discretas

parámetros

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

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

$\sum_{i=1}^{n} X_i = Y \sim$ **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

$P(Z=3)=p\cdot(1-p)^{3-1}$

Ejercicio 0

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

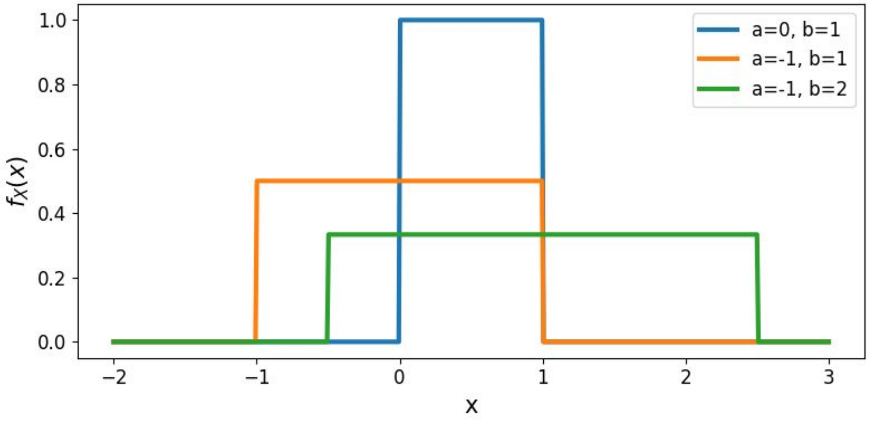
$$\mathbb{1}\{a<x<b\}(x) = \begin{cases} 1 & \text{si } a<x<b \\ 0 & \text{sino} \end{cases}$$

Variables continuas

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

indicadora

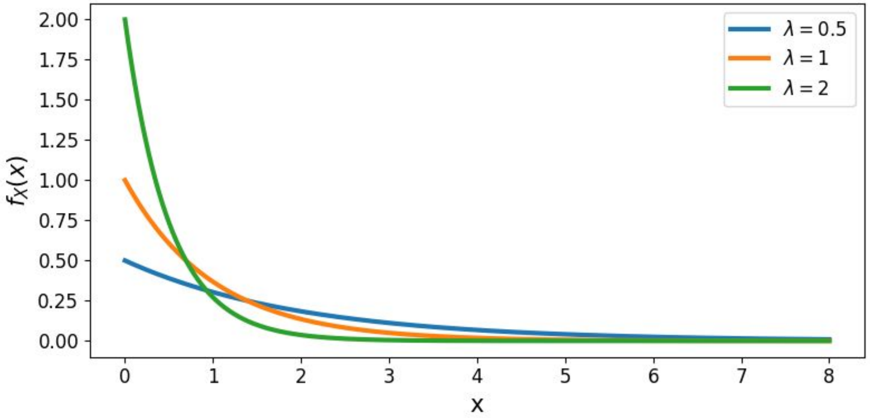
$$f_X(x) = \frac{1}{b-a}\mathbf{I}\{a < x < b\}$$



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

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

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



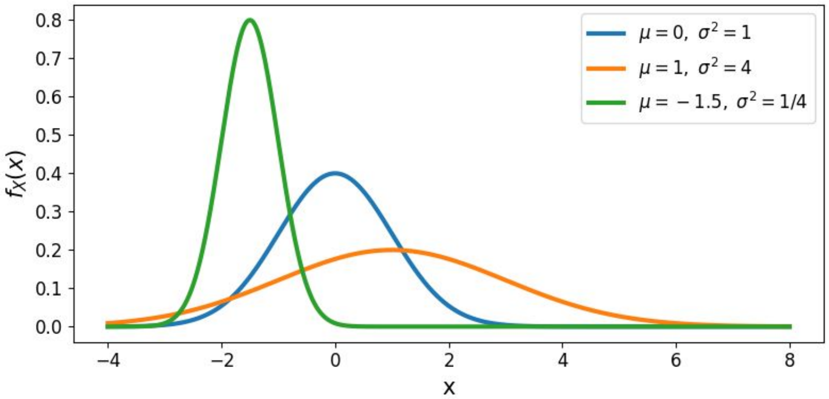
Variables continuas

Media Varianza

- **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
σ^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)

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

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

Sea además Y ~ N(2,9)

1. Hallar P(2X+Y <5) $= P\left(\underbrace{\frac{2X+Y-2}{\sqrt{2^2 \cdot 1 + 1^2 \cdot 9}}}_{Z \sim N(0,1)} < \frac{5-2}{\sqrt{13}} \right) = F_Z\left(\frac{5-2}{\sqrt{13}}\right)$

$2X+Y \sim N(2,13)$

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 $P(X>2) = 1 - P(X \leq 2) = F_X(2)$.
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

$\langle \underline{a}, \underline{b} \rangle = \underline{a} \cdot \underline{b} = \underline{a}^T b \in \mathbb{R}.$

$$\underline{\mu} = E[\underline{X}] = \underbrace{\int \cdots \int}_{p} \underline{x} f_{\underline{X}}(\underline{x})\, d\underline{x}$$

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

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\overbrace{\underbrace{\underset{1\times n}{(\underline{x}-\underline{\mu})^T} \underset{n\times n}{\Sigma^{-1}} \underset{n\times 1}{(\underline{x}-\underline{\mu})}}_{\text{escalar.}}}^{\text{es un p.i.}}} \Longleftrightarrow \frac{(x-\mu)^2}{\sigma^2}$$

media

$$\underline{\mu} = [\mu_1, \ldots, \mu_n]^T \in \mathbb{R}^{n\times 1}$$

matriz de covarianza .

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

simétrica

def +.

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

Distribuciones marginales

p- elementos

marginales

$$\underline{\boldsymbol{X}} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow X_i \sim \mathcal{N}(\underline{\mu_i}, \underline{\sigma_i^2}) \quad \forall\, i = 1, \ldots, p$$

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

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

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 $\underbrace{E[X]}_{=0}$, $\underbrace{E[Y]}_{=0}$, $\underbrace{var(X)}_{=1}$, var(Y), y $\underbrace{cov(X,Y)}_{=-0,8}$ → obs. directa de la param.
2. Hallar las densidades marginales de X e Y ✓
3. Calcular P(X<2, Y<-1) = ... = ~~~

Regl. del Rectángulo

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.