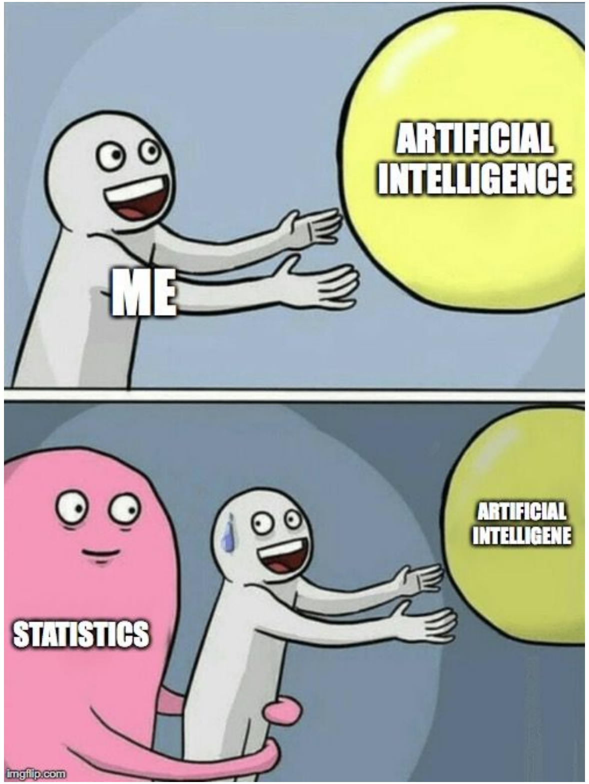
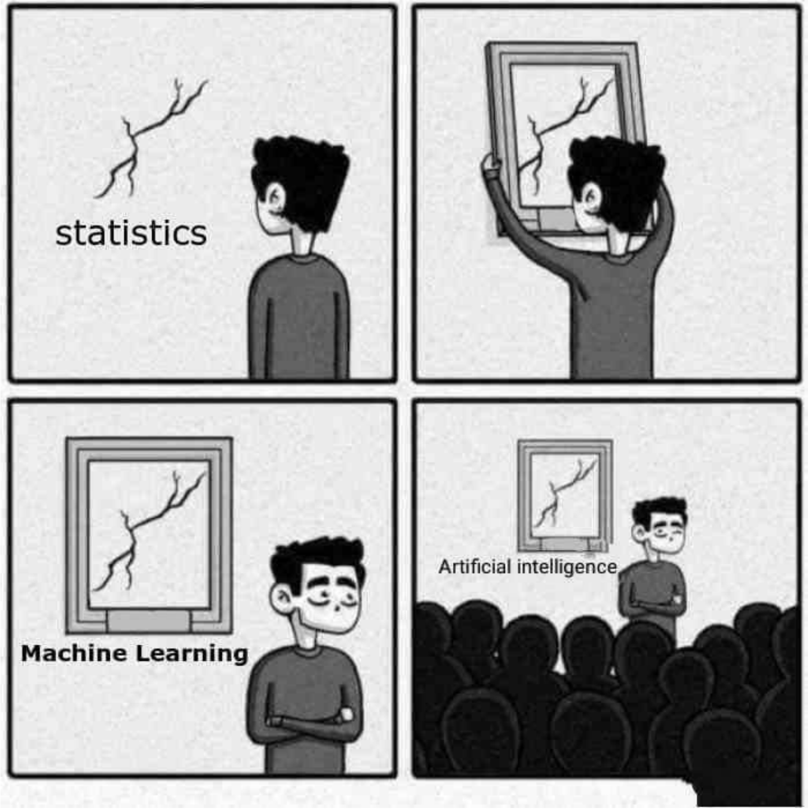
Probabilidad y Estadística

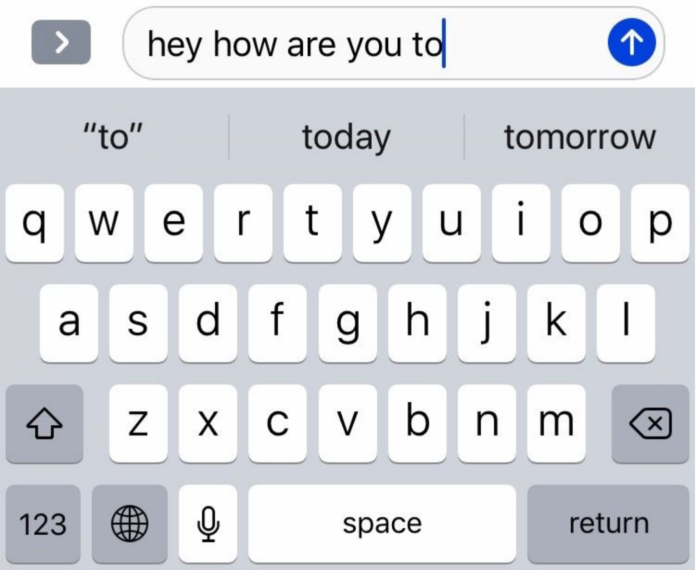
Clase 1

Ud. se encuentra aquí:



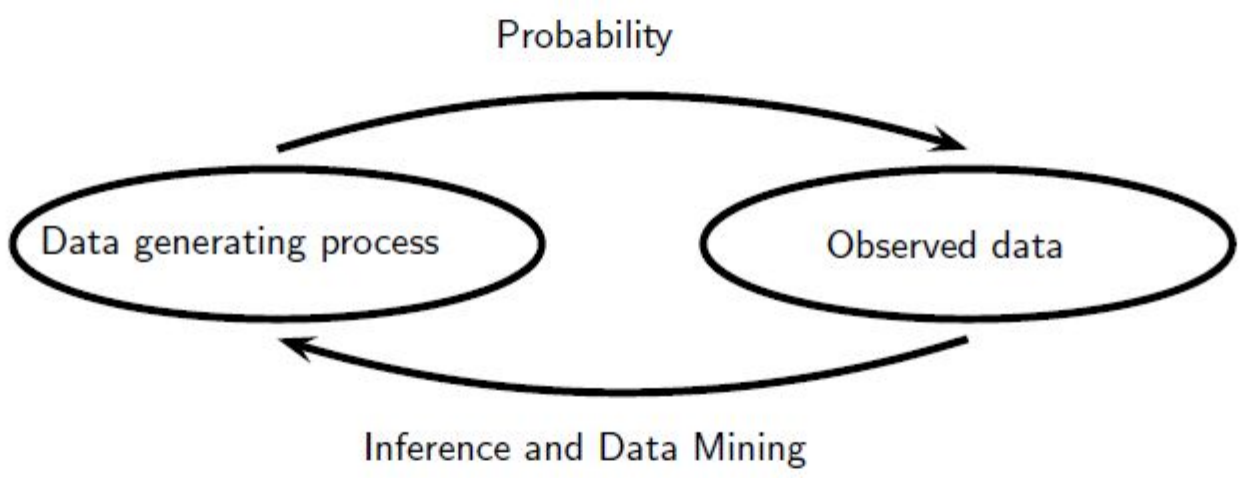
para poder estar aquí:





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	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

Repaso

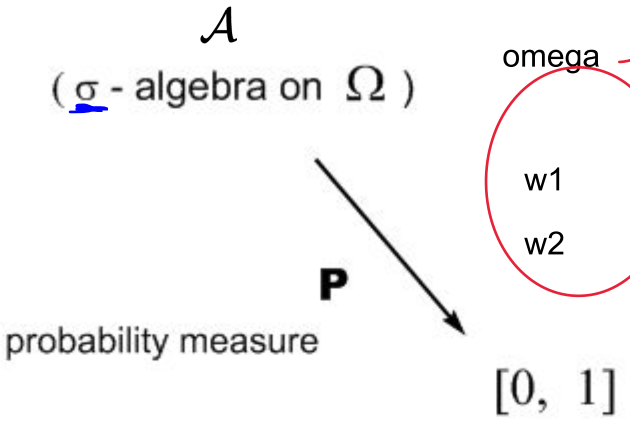
Espacio de Probabilidad $(\Omega, \mathcal{A}, P)$ y variables aleatorias

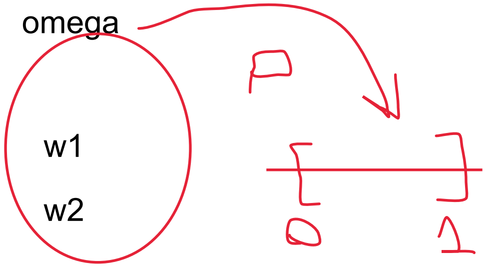
E= lanzar un dado equilibrado y observar qué n° sale
omega = {1,2,3,4,5,6}

A=que salga n° par= {2,4,6}
B=que salga n° impar ={1,3,5}
C=que salga el n°3= {3}

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

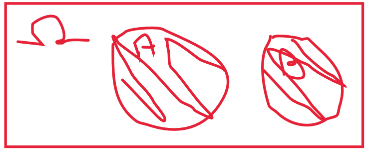
B1, B2, B3,...... entonces $\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$

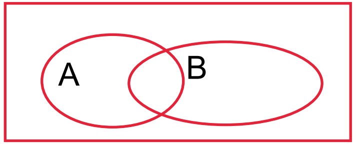




$$0 \le \mathbb{P}(A) \le 1, \ \forall A \in \mathcal{A}$$
$$\mathbb{P}(\Omega) = 1$$
$$A \cap B = \emptyset, \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

A y B no disjuntos (no m.e.)





P(AUB) = P(A) +P(B)-P(A I B)

Regla aditiva (o regla de la suma)

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

Probabilidades condicionales y proba. total

A=que salga el n°1={1}
ocurrió B=que salga un n° impar ={1,3,5}

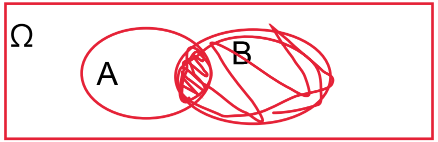
P(A) = 1/6

P(A/B) = 1/3

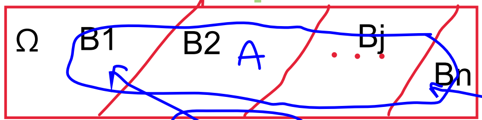
B ocurrió

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

Regla del producto

$$\mathbb{P}(A) = \underbrace{\mathbb{P}(A \cap B_1) + \ldots + \mathbb{P}(A \cap B_n)}_{\text{Regla de la suma para eventos m.e.}} = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \ldots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

Regla de la suma para eventos m.e.

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:

P(Bi/A) = P(A I Bi)
P(A)

probabilidades a priori

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

probabilidad a posteriori de Bi (habiendo ocurrido A)

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

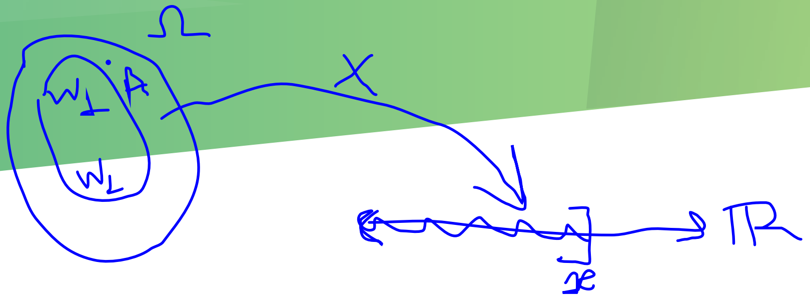
P(A/B) = P(A) o P(B/A) = P(B)

P(A I B) = P(A/B).P(B)
P(A I B) = P(A).P(B) siempre que A y B sean independientes

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

Variables aleatorias

Variables aleatorias



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

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

$X^{-1}[-\infty, x] \in \mathcal{A}$

$\forall x \in \mathbb{R}$

X tiene asociada una función de distribución, definida como

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

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

$F_X(x)$ es monótona no decreciente

$F_X(x)$ es continua por derecha

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

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.

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

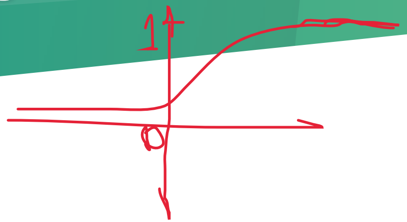
$$p_X(x) = \mathbb{P}(X = \underline{x}) \qquad \sum_{x \in R_X} p_X(x) = 1 \qquad 0 \leq p_X(x) \leq 1$$

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

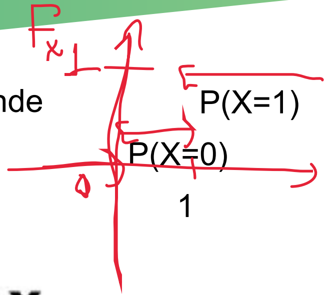
$$f_X(x) = \frac{dF_X(x)}{dx} \qquad \int_{-\infty}^{+\infty} f_X(x)\,dx = 1 \qquad f_X(x) \geq 0$$

¿Qué propiedades debían cumplir cada una de estas funciones?

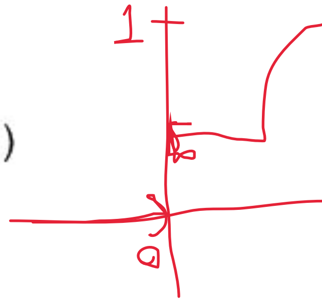
Espacio de Probabilidad (Ω,A,P) y variables aleatorias



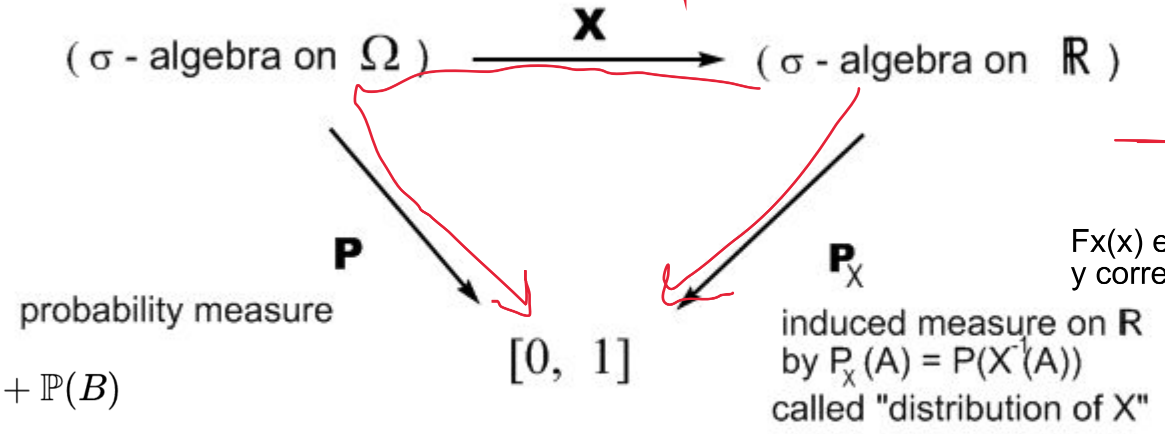
Fx(x) es f. de distribución y corresponde a una v.a. absolutamente continua



Fx(x) es función de distribución y corresponde a una v.a. discreta



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



Fx(x) es f. de distribución y corresponde a una v.a. mixta

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

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

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

Vectores aleatorios

Distribución conjunta y marginales

función de probabilidad conjunta de X e Y

pxy(x,y)	$Y=0$	$Y=1$	px(x)
X=0	$1/10$	$2/10$	$3/10$
X=1	$3/10$	$4/10$	$7/10$
py(y)	$4/10$	$6/10$	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:

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

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

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

Independencia de v.a.

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

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

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

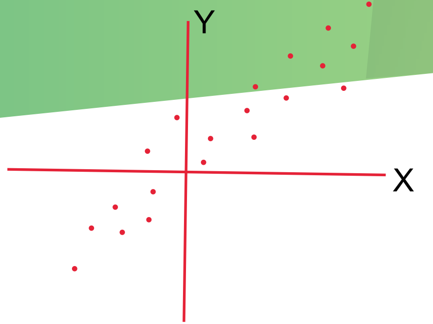
Momentos

$\sum_{x} x \, p_x(x)$

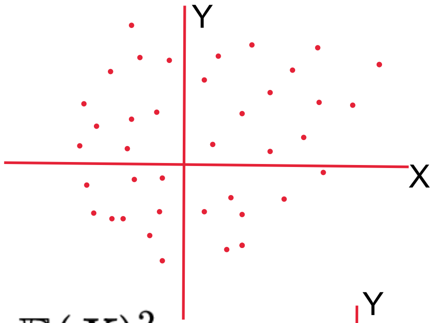
E(X) = $\int_{x}$ x. fx(x) dx

Esperanza (o media):

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

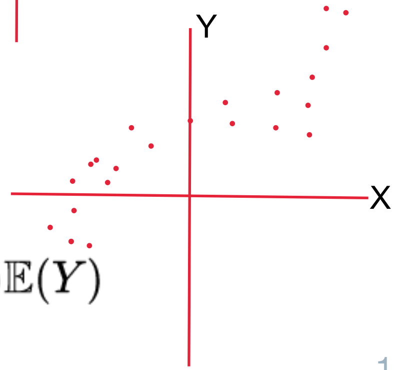


Varianza:



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

Covarianza:



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

Algunas distribuciones útiles

Variables discretas

p=probabilidad de éxito   px(x)=p^x.(1-p)^(1-x)   x=0,1

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

Y=sum(i=1,n) Xi   py(y) = Cn,y . p^(n-y).(1-p)^(n-y)   y=0,1,....n

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

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

pz(z) = (1-p)^(z-1).p

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

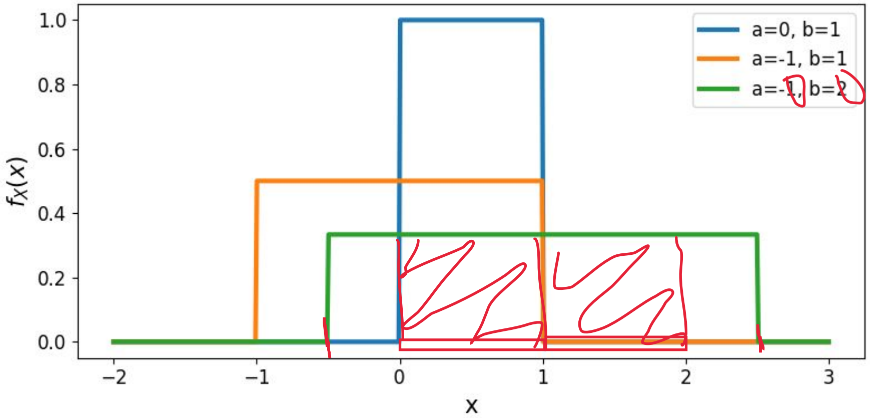
Variables continuas

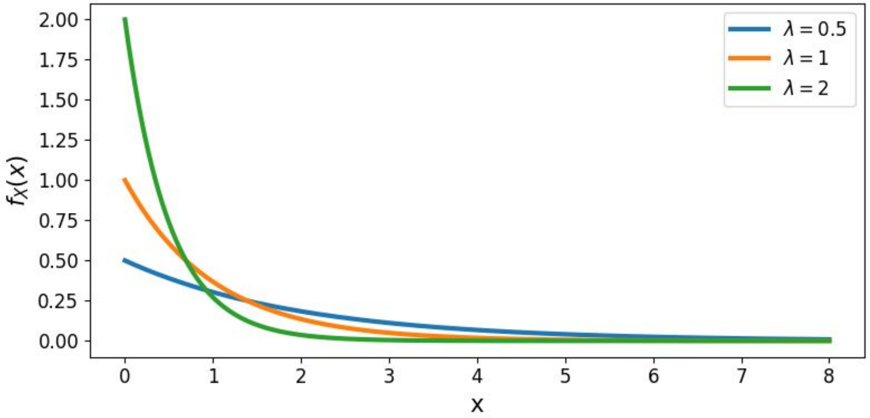
- **Uniforme:** todos los puntos son equiprobables. $X \sim \mathcal{U}(a, b)$

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

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

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



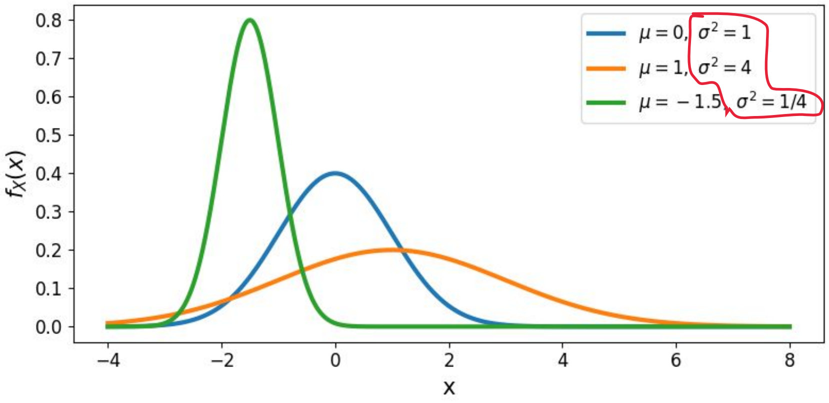


Variables continuas

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

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

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



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

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

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

(combinación lineal de normales es normal)

Ejercicio 1

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

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

Sea además $Y \sim N(2,9)$

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

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

$cov(x_1, x_2) = \mathbb{E}[(X_1-\mu_1)(X_2-\mu_2)]$ $\sim \frac{1}{n}\sum_{i=1}$

$\}+$

$\}-$

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

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

, Σ es definida positiva.

media

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

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

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

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

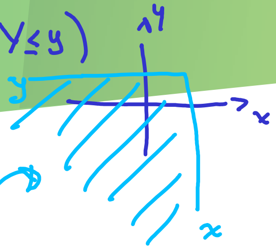
$$\Sigma = \begin{bmatrix} \sigma_1^2 & cov(X_1, X_2) & \ldots & cov(X_1, X_n) \\ cov(X_2, X_1) & \sigma_2^2 & \ldots & cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n, X_1) & cov(X_n, X_2) & \ldots & \sigma_n^2 \end{bmatrix}$$

Ejercicio 3

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

$f_X(x) = \frac{d}{dx} F_X(x)$

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



$(\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})$

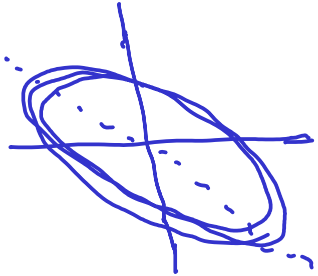
$P\left(\binom{X}{Y} \in \right)$

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

$$f_{X,Y}(x,y) = \frac{1}{2\pi 0.6} e^{-\frac{1}{2}[x \quad y] \underbrace{\begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}^{-1}}_{\text{cuadrática}} \begin{bmatrix} x \\ y \end{bmatrix}} \quad ; \ x, y \in \mathbb{R}$$

$=0 \quad =0 \quad =1 \quad =1 \quad =-0{,}8$

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)



$X \sim N(0,1)$

$Y \sim N(0,1)$

$$= \int_{-\infty}^{-1} \int_{-\infty}^{2} f_{XY}(x,y)\, dx\, dy$$

$$= F_{XY}(2,-1) \leftarrow \dots \text{python}$$

$P(X = 2 \,;\, Y = -1) = 0$

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.