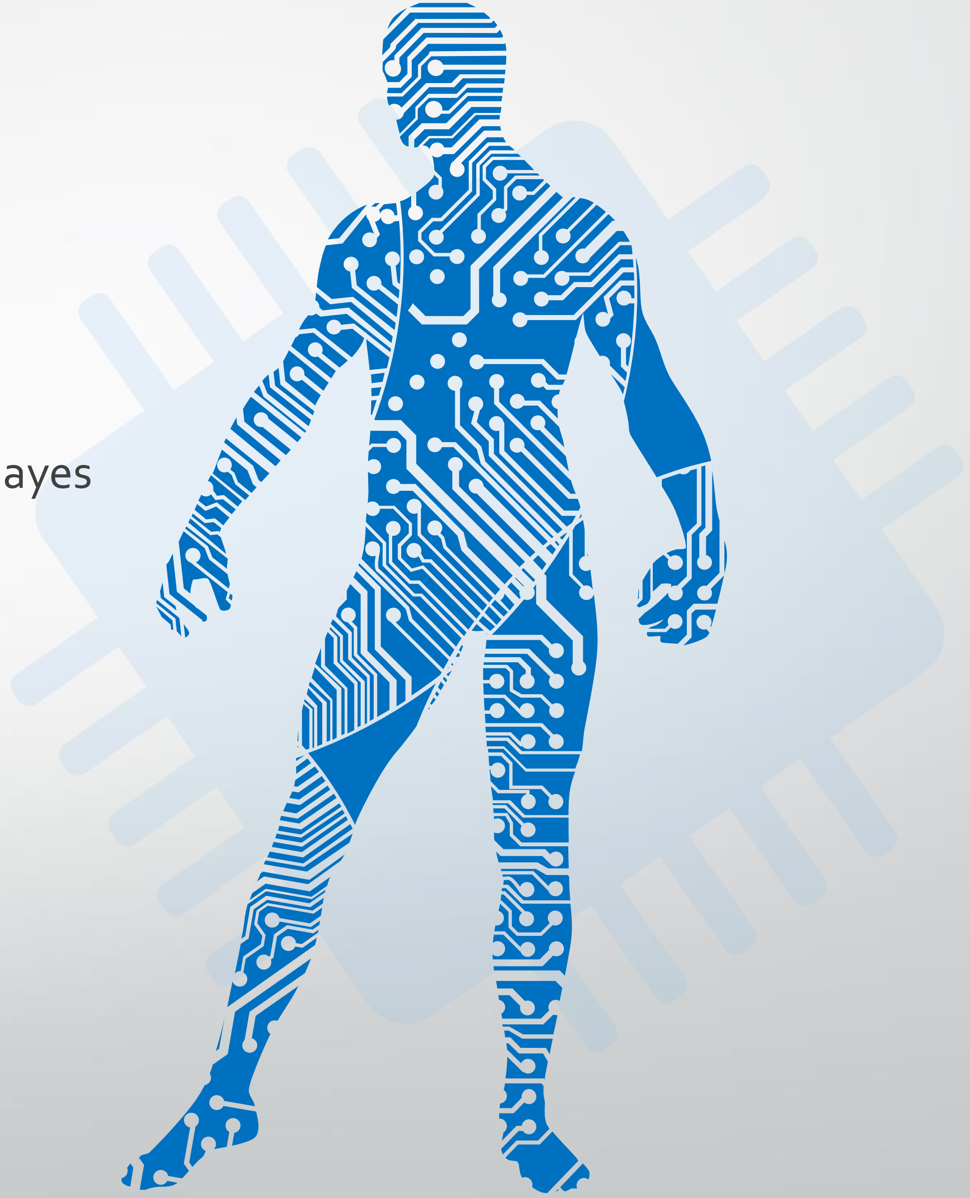
APRENDIZAJE DE MÁQUINA

PROBABILIDAD

AGENDA

01 Repaso Probabilidad

Introducción, variables aleatorias, Teorema de Bayes



PROBABILIDAD

DEFINICIONES

PROBABILIDAD CONDICIONADA

PROBABILIDAD

DEFINICIONES

Espacio muestral $\boldsymbol{\Omega}$: conjunto de todas los resultados $\boldsymbol{\omega}$ de un experimento aleatorio. Donde cada $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ se define como una descripción completa del estado real del mundo después del experimento.

Espacio de eventos $\mathcal{F}$: conjunto cuyos elementos $A \in \mathcal{F}$ (denominados eventos) son subconjuntos de $\boldsymbol{\Omega}$.

Medida probabilística: es una función $P: \mathcal{F} \rightarrow \mathbb{R}$ que satisface los **3 axiomas de probabilidad**.

1) $P(A) \geq 0, \forall A \in \mathcal{F}$
2) $P(\Omega) = 1$
3) Sí $A_1, A_2, \ldots$ son eventos disjuntos por lo tanto:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

PROBABILIDAD

DEFINICIONES

Ejemplo: realizamos el experimento de lanzar 2 monedas al aire.

- **Espacio muestral** Ω:
- **Evento** $\boldsymbol{E}$: que en la 1° moneda salga cara
- **Medida probabilística** P: la probabilidad del evento E (suponiendo que cada resultado es igualmente probable).

PROBABILIDAD
DEFINICIONES

Respuesta: realizamos el experimento de lanzar 2 monedas al aire.

- Espacio muestral Ω:

$$S = \{(H,H),(H,T),(T,H),(T,T)\}$$

- Evento E: que en la 1° moneda salga cara

$$E = \{(H,H),(H,T)\}$$

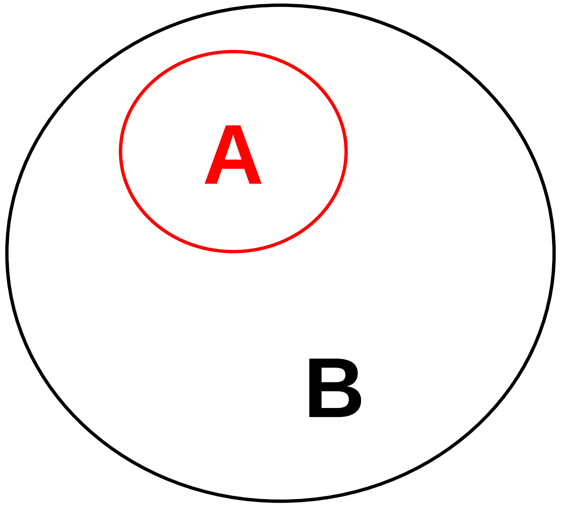
- Medida probabilística P: la probabilidad del evento E sería.

$$\mathrm{P}(E) = \frac{2}{4} = \frac{1}{2}$$

PROBABILIDAD

PROPIEDADES

1 If $A \subseteq B \Longrightarrow P(A) \leq P(B)$.

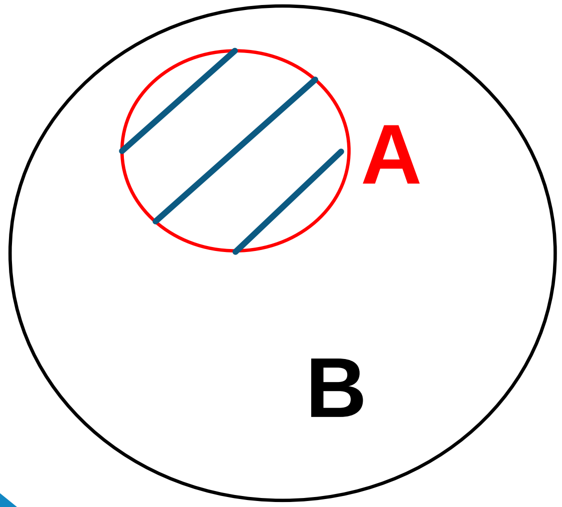


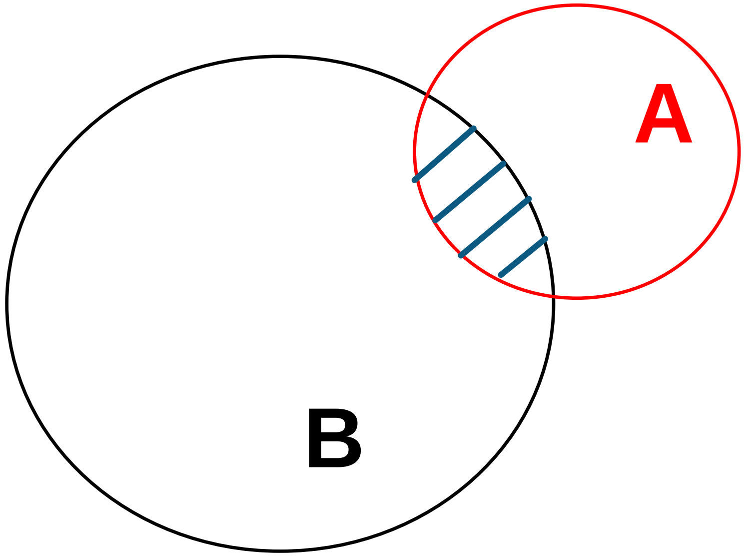
PROBABILIDAD

PROPIEDADES

2 $P(A \cap B) \leq \min(P(A), P(B)).$

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

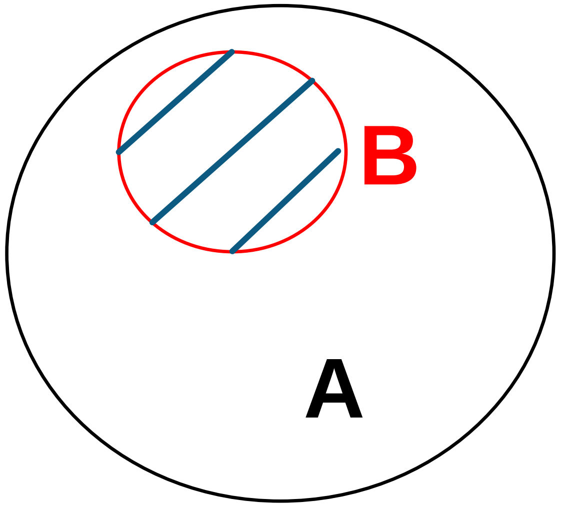




$P(A \cap B) < P(A)$

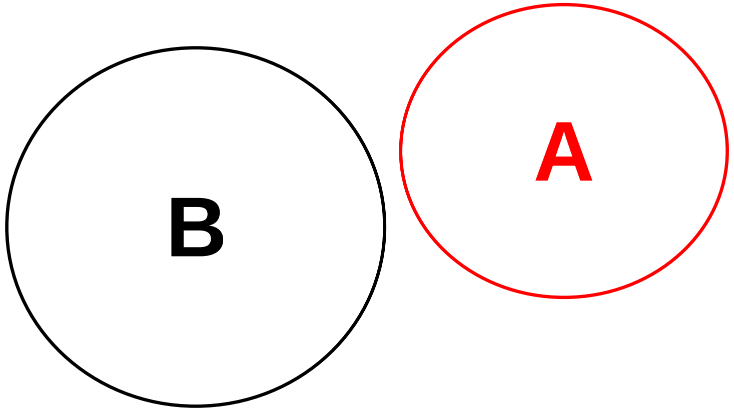
$P(A \cap B) < P(B)$

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

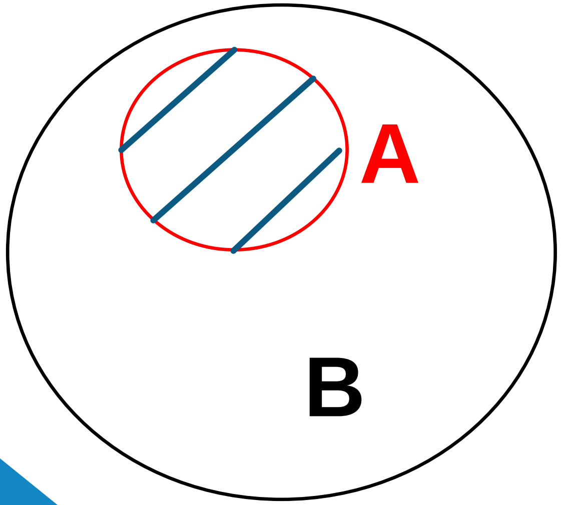


PROBABILIDAD
PROPIEDADES

3 (Union Bound) $P(A \cup B) \leq P(A) + P(B)$.

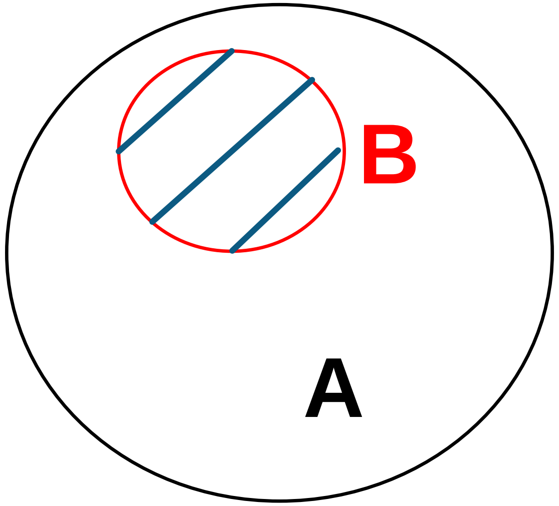


$\boldsymbol{P(A \cup B) = P(A)}$



$\boldsymbol{P(A \cup B) = P(A) + P(B)}$

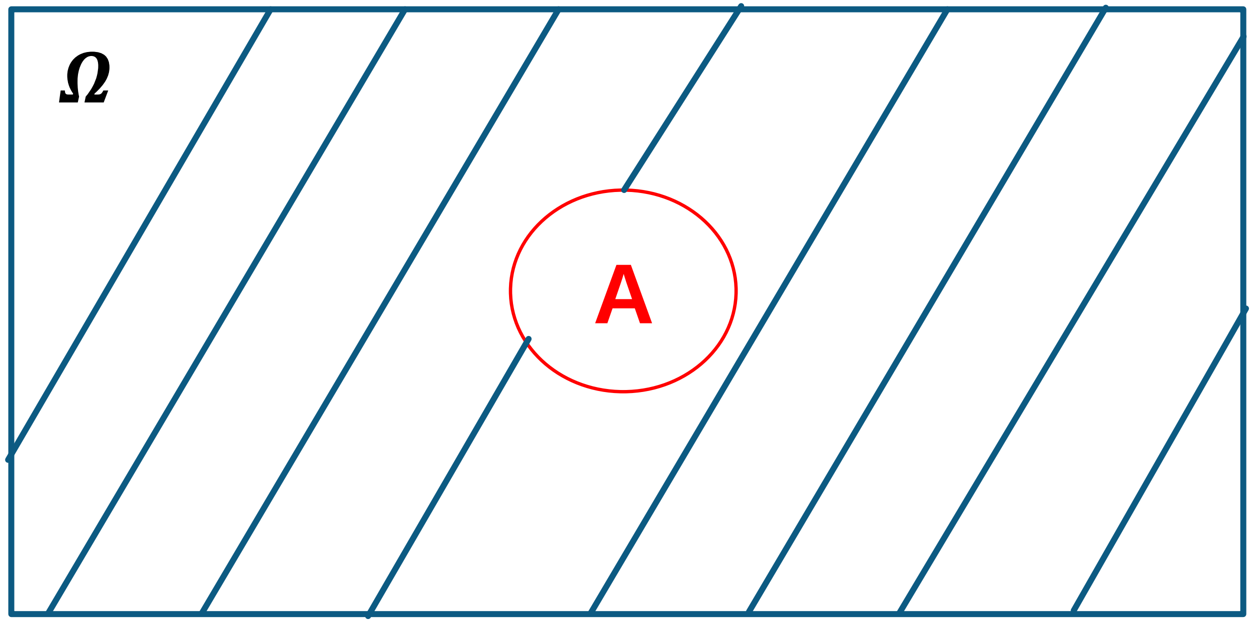
$\boldsymbol{P(A \cup B) = P(B)}$



PROBABILIDAD

PROPIEDADES

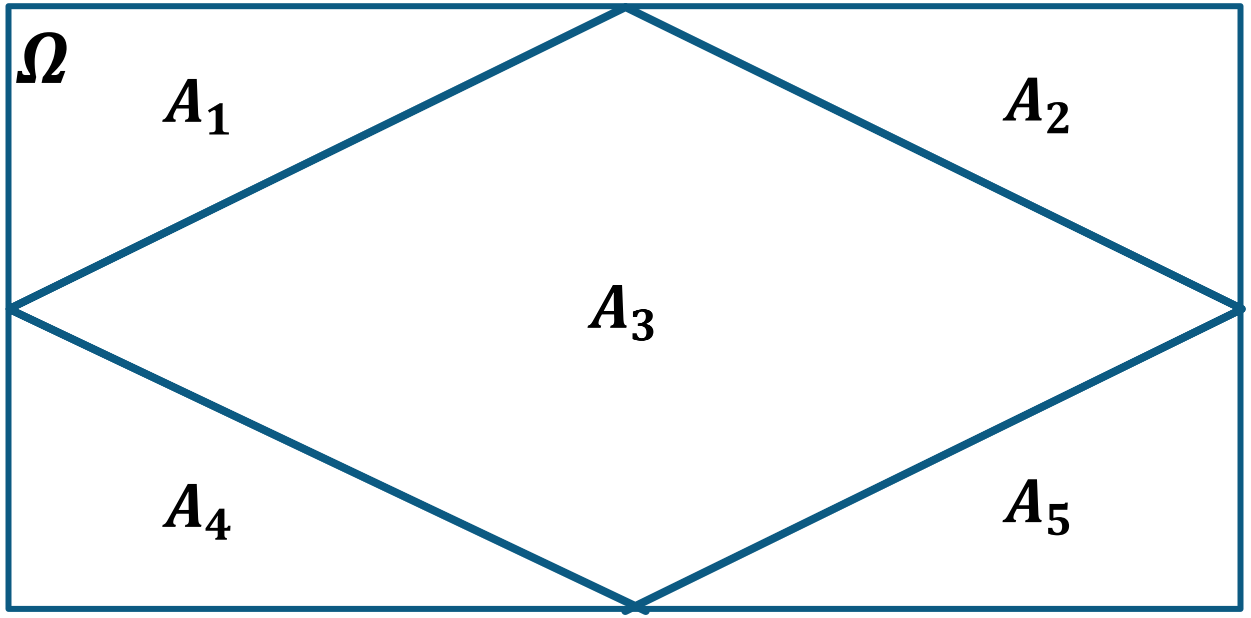
4 $P(\Omega \setminus A) = 1 - P(A).$



PROBABILIDAD

PROPIEDADES

5 (Law of Total Probability) If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^{k} A_i = \Omega$, then $\sum_{i=1}^{k} P(A_k) = 1$.



PROBABILIDAD

PROBABILIDAD CONDICIONADA

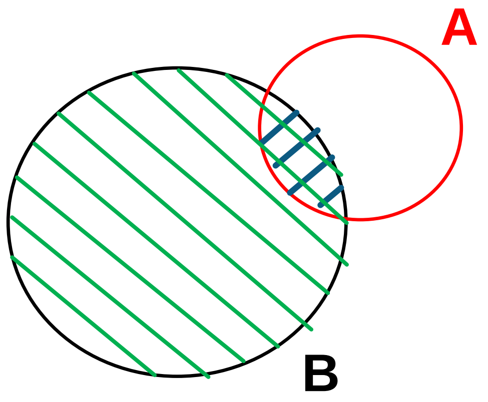
Probabilidad condicionada: la medida de probabilidad $P(A/B)$ define la probabilidad de un evento A después de observar la ocurrencia del evento B con una probabilidad $P(B) \neq 0$.

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

Independencia: se dice que 2 eventos AyB son independientes si la observación del evento de B no afecta a la probabilidad del evento A.

$$\boldsymbol{P(A/B) = P(A)}$$

$$\therefore \boldsymbol{P(A \cap B) = P(A)P(B)}$$



PROBABILIDAD

PROBABILIDAD CONDICIONADA

Ejemplo: una moneda es lanzada al aire 2 veces. Suponiendo que todos los resultados son equi-probables, ¿cuál es la probabilidad de que ambos lanzamientos hayan resultado en cara (H) dado que

a) la 1° moneda resultó en H?
b) al menos un lanzamiento resultó en H?

PROBABILIDAD

PROBABILIDAD CONDICIONADA

Respuestas:

a) $P(ambos\ H\ /\ 1°\ H) = \frac{1}{2}$

b) $P(ambos\ H\ /\ al\ menos\ 1\ H) = \frac{1}{3}$

PROBABILIDAD

VARIABLES ALEATORIAS

PROBABILIDAD
VARIABLES ALEATORIAS

Muchas veces, **NO** nos importan los **resultados** $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ de un experimento. Lo que nos **importa** son **funciones** de estos resultados $\boldsymbol{X}(\boldsymbol{\varpi})$**.**

Formalmente una **variable aleatoria** es una función:

$$X: \Omega \to \mathbb{R}$$

Variables aleatorias discretas	Variables aleatorias continuas
$\boldsymbol{X}(\boldsymbol{\omega})$ puede tomar un valor **finito** de **valores**.	$\boldsymbol{X}(\boldsymbol{\omega})$ puede tomar un valor **infinito** de **valores**.
$P(X = k) \coloneqq P(\{\varpi: X(\omega) = k\})$	$P(a \leq X \leq b) \coloneqq P(\{\varpi: (a \leq X(\omega) \leq b)\})$
Ejemplo: $\boldsymbol{X}(\boldsymbol{\omega})$ es el número de caras que ocurren en una secuencia de lanzamientos ω.	**Ejemplo:** $\boldsymbol{X}(\boldsymbol{\omega})$ es el tiempo de decaimiento radioactivo de una partícula.

PROBABILIDAD

FUNCIONES DE PROBABILIDAD

Se necesitan diferentes **medidas** de **probabilidad** cuando se plantean **variables aleatorias.** Para esto se definen **funciones** de **probabilidad**:

1. Función de Distribución Acumulada (discreta y continua) → CDF.

2. Función de Masa de Probabilidad (discretas)→ PMF.

3. Función de Densidad de Probabilidad (continuas) → PDF.

NOTA: cuando la variable aleatoria $\boldsymbol{X}$ toma un valor específico, se denota con minúscula $\boldsymbol{x}$.

PROBABILIDAD

FUNCIÓN DE DISTRIBUCIÓN ACUMULADA

La Función de Distribución Acumulada es una función $F_X : \mathbb{R} \to [0,1]$ que especifica la medida de probabilidad como:

$$F_X(x) \triangleq P(X \leq x).$$

Intuitivamente se puede decir que está función define las probabilidades de todos los eventos $A_i \in \Omega$ cuando $x \to \infty$

Propiedades:

$$0 \leq F_X(x) \leq 1.$$

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

$$\lim_{x \to \infty} F_X(x) = 1.$$

$$x \leq y \Longrightarrow F_X(x) \leq F_X(y).$$

PROBABILIDAD

FUNCIÓN DE MASA DE PROBABILIDAD

La Función de Masa de Probabilidad **asigna una medida** de **probabilidad** a cada **valor** que puede **tomar** la **variable** aleatoria $\boldsymbol{X}$.

$$p_X(x) \triangleq P(X = x).$$

Propiedades:

$$0 \leq p_X(x) \leq 1.$$

$$\sum_{x \in Val(X)} p_X(x) = 1.$$

$$\sum_{x \in A} p_X(x) = P(X \in A).$$

Donde $Val(X)$ representa todos los posibles valores que puede tomar

PROBABILIDAD

FUNCIÓN DE DENSIDAD DE PROBABILIDAD

La Función de Densidad de Probabilidad **asigna una medida** de **probabilidad** a cada **valor** que puede **tomar** la **variable** aleatoria X en un **intervalo continuo**.

Formalmente se define como la derivada de la Función de Distribución Acumulada:

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

Dicha función **puede no existir**.

Propiedades:

$$f_X(x) \geq 0\,.$$

$$\int_{-\infty}^{\infty} f_X(x) = 1.$$

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

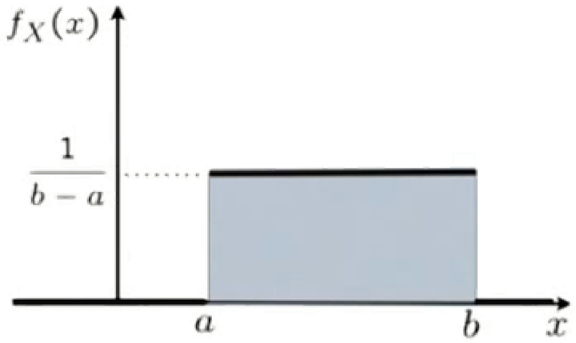
PROBABILIDAD

E J E M P L O S   P R Á C T I C O S

Ejemplo:

Suponiendo que X es una variable aleatoria continua y su PDF está descrita por:

$$f_X(t) = \frac{1}{b-a}, \forall t \in [a, b]$$
$$f_X(t) = 0 \text{ de otra forma}$$



Calcule la gráfica de su CDF.

PROBABILIDAD

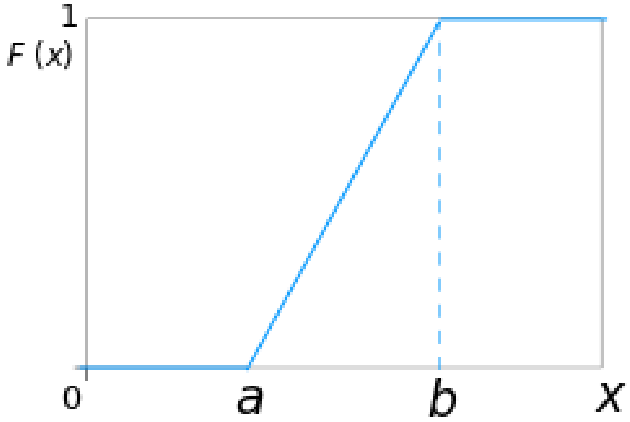
EJEMPLOS PRÁCTICOS

Respuesta:

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

$$F_X(x) = \int_{-\infty}^{a} \frac{1}{b-a}dx + \int_{a}^{x} \frac{1}{b-a}dx + \int_{x}^{b} \frac{1}{b-a}dx$$

$$F_X(x) = 0 + \frac{x}{b-a} + 0$$



PROBABILIDAD

VALOR ESPERADO

Suponemos que $\boldsymbol{X}$ es una **variable aleatoria discreta** con una **PMF** $\boldsymbol{p_X(x)}$ y $\boldsymbol{g}: \mathbb{R} \rightarrow \mathbb{R}$ una función arbitraria. En este caso, $\boldsymbol{g(X)}$ se considera una **variable aleatoria**, por lo que se define al **valor esperado** de $\boldsymbol{g(X)}$ como:

$$E[g(X)] \triangleq \sum_{x \in Val(X)} g(x) p_X(x).$$

PROBABILIDAD

VALOR ESPERADO

Suponemos que X es una **variable aleatoria continua** con una **PDF $f_X(x)$,** por lo tanto el **valor esperado** de $g(X)$ sería:

$$E[g(X)] \triangleq \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

PROBABILIDAD

VALOR ESPERADO

Intuitivamente lo que se calcula en el **valor esperado** es un **“promedio ponderado”** de los **valores de** $\boldsymbol{g(x)}$ donde los **pesos** están dados ya sea **por** $\boldsymbol{f_X(x)}$ **o** $\boldsymbol{p_X(x)}$**.**

NOTA: cuando $\boldsymbol{g(x) = x}$ el valor esperado de la variable aleatoria $\boldsymbol{X}$ sería la **media aritmética**.

Propiedades:

$E[a] = a$ for any constant $a \in \mathbb{R}$.

$E[af(X)] = aE[f(X)]$ for any constant $a \in \mathbb{R}$.

(Linearity of Expectation) $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.

For a discrete random variable X, $E[1\{X = k\}] = P(X = k)$.

PROBABILIDAD

V A R I A N Z A

La **varianza** de una **variable aleatoria** X es la **medida** de como está **concentrada** la **distribución** de la **variable aleatoria** X **alrededor** de la **media**.

$$Var[X] \triangleq E[(X - E(X))^2]$$

PROBABILIDAD

V A R I A N Z A

TAREA:

Demostrar la siguiente igualdad.

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

PROBABILIDAD

V A R I A N Z A

Propiedades:

$Var[a] = 0$ for any constant $a \in \mathbb{R}$.

$Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Ejemplo:

Calcular la media y la varianza de una variable aleatoria X con PDF $f_X(x) = 1, \forall\ x \in [0,1], 0$ en otro rango.

PROBABILIDAD

VALOR ESPERADO

Respuesta:

$$\boldsymbol{E[X] = \frac{1}{2}}$$

$$\boldsymbol{Var[X] = \frac{1}{12}}$$

PROBABILIDAD

EJEMPLOS DE VARIABLES ALEATORIAS

PROBABILIDAD

EJEMPLOS DE VARIABLES ALEATORIAS

Discretas:

BERNOULLI

$$p(x) = \begin{cases} p & \text{if } p = 1 \\ 1 - p & \text{if } p = 0 \end{cases}$$

BINOMIAL

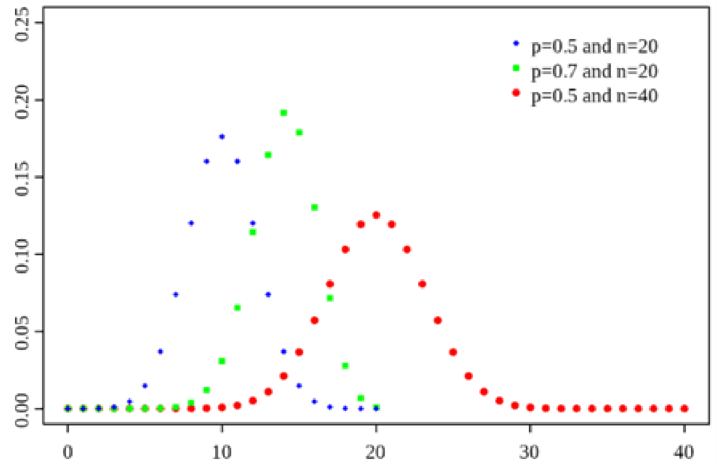
$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

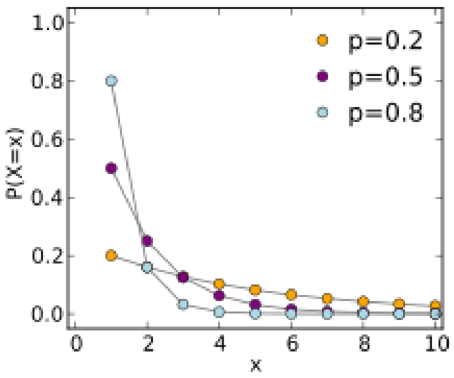
GEOMÉTRICA

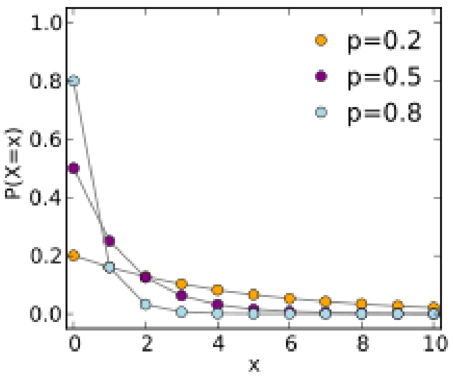
$$p(x) = p(1-p)^{x-1}$$

POISSON

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$







PROBABILIDAD

EJEMPLOS DE VARIABLES ALEATORIAS

Continuas:

UNIFORME

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

EXPONENCIAL

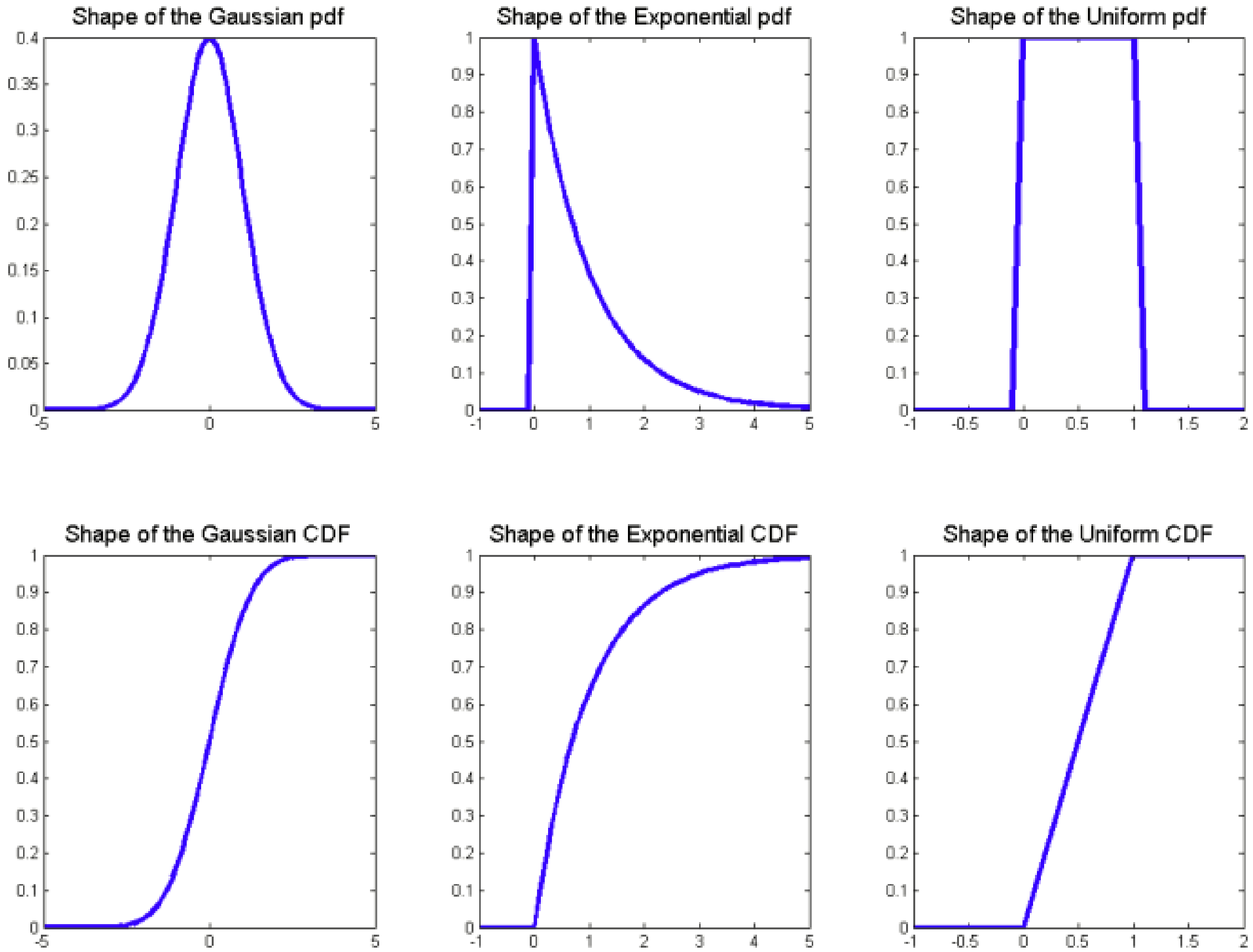
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

NORMAL

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

PROBABILIDAD

EJEMPLOS DE VARIABLES ALEATORIAS



PROBABILIDAD

EJEMPLOS DE VARIABLES ALEATORIAS

TAREA:

Obtener la media y varianza de las 4 distribuciones discretas.

Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0. \end{cases}$	p	$p(1-p)$
$Binomial(n,p)$	$\binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq n$	np	npq
$Geometric(p)$	$p(1-p)^{k-1}$ for $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$e^{-\lambda}\lambda^x/x!$ for $k = 1, 2, \ldots$	λ	λ

PROBABILIDAD

DOS VARIABLES ALEATORIAS

PROBABILIDAD

DISTRIBUCIONES CONJUNTAS Y MARGINALES ACUMULADAS

Si se quieren conocer los **valores** de **dos variables** aleatorias $\boldsymbol{X}$ y $\boldsymbol{Y}$ de manera **simultánea** , se necesita la **distribución** de **conjunta acumulada** de $\boldsymbol{X}$ y $\boldsymbol{Y}$:

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

Las **funciones** de **distribución** $\boldsymbol{F_X(x)}$ y $\boldsymbol{F_Y(y)}$ se denominan funciones de **distribución acumulada marginal**

$$\begin{aligned} F_X(x) &= \lim_{y\to\infty} F_{XY}(x,y)dy \\ F_Y(y) &= \lim_{x\to\infty} F_{XY}(x,y)dx. \end{aligned}$$

PROBABILIDAD

DISTRIBUCIONES CONJUNTAS Y MARGINALES ACUMULADAS

Propiedades:

$$0 \leq F_{XY}(x, y) \leq 1.$$

$$\lim_{x,y \to \infty} F_{XY}(x, y) = 1.$$

$$\lim_{x,y \to -\infty} F_{XY}(x, y) = 0.$$

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y).$$

PROBABILIDAD

FUNCIONES DE MASA CONJUNTAS Y MARGINALES

Si $\boldsymbol{X}$ y $\boldsymbol{Y}$ son **variables aleatorias discretas**, la **función** de **masa** de **probabilidad conjunta** $p_{XY}: \mathbb{R}x\mathbb{R} \rightarrow [0,1]$ está definida por:

$$p_{XY}(x, y) = P(X = x, Y = y).$$

Propiedades:

$$0 \leq P_{XY}(x, y) \leq 1 \text{ for all } x, y.$$

$$\sum_{x \in Val(X)} \sum_{y \in Val(Y)} P_{XY}(x, y) = 1.$$

PROBABILIDAD

FUNCIONES DE MASA CONJUNTAS Y MARGINALES

La **funciones** de **masa** de **probabilidad marginales** de $\boldsymbol{X}$ y $\boldsymbol{Y}$ está definidas por:

$$p_X(x) = \sum_y p_{XY}(x, y).$$

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

PROBABILIDAD

FUNCIONES DE DENSIDAD CONJUNTAS Y MARGINALES

Si $\boldsymbol{X}$ y $\boldsymbol{Y}$ son **variables aleatorias continuas**, con una **distribución conjunta** F_{XY} **diferenciable** en **todo** el espacio, se define la **función** de **densidad** de **probabilidad** como:

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}.$$

$$\iint_{x \in A} f_{XY}(x,y)\,dx\,dy = P((X,Y) \in A).$$

PROBABILIDAD

FUNCIONES DE DENSIDAD CONJUNTAS Y MARGINALES

Las **funciones** de **densidad** de **probabilidad marginales** de ***X*** y ***Y*** serían:

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

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

PROBABILIDAD

FUNCIONES DE DENSIDAD CONJUNTAS Y MARGINALES

Ejemplo:

La **función** de **densidad conjunta** de $\boldsymbol{X}$ y $\boldsymbol{Y}$ está dada por:

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, \quad 0 < y < \infty \\ 0 \text{ de otra forma} \end{cases}$$

Calcular $\boldsymbol{P(X > 1, Y < 1)}$

PROBABILIDAD

FUNCIONES DE DENSIDAD CONJUNTAS Y MARGINALES

Respuesta:

$$P(X > 1, Y < 1) = e^{-1}(1 - e^{-2})$$

PROBABILIDAD

DISTRIBUCIONES CONDICIONADAS

La **función** de **masa** de **probabilidad condicionada** en el caso discreto:

$$p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$$

La **función** de **densidad** de **probabilidad condicionada** en el caso continuo:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

PROBABILIDAD

VALOR ESPERADO

Variables discretas:

$$E[g(X,Y)] \triangleq \sum_{x \in Val(X)} \sum_{y \in Val(Y)} g(x,y) p_{XY}(x,y).$$

Variables continuas:

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dx dy.$$

PROBABILIDAD

C O V A R I A N Z A

Para estudiar la relación de 2 variables aleatorias se utiliza la covarianza.

$$Cov[X,Y] \triangleq E[(X - E[X])(Y - E[Y])]$$

Cuando $\boldsymbol{Cov}[\boldsymbol{X},\boldsymbol{Y}] = \mathbf{0}$ se dice que $\mathbf{X}$ y $\mathbf{Y}$ no están correlacionadas.

TAREA:

Demostrar que:

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

PROBABILIDAD

MÚLTIPLES VARIABLES

ALEATORIAS

PROBABILIDAD

DISTRIBUCIONES

Suponiendo que se tienen $\boldsymbol{n}$ **variables aleatorias continuas** $X_1, X_2, \ldots, X_n$ se tienen las siguientes **distribuciones**:

Función de probabilidad conjunta acumulada

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) \quad = \quad P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

Función de densidad de probabilidad conjunta

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) \quad = \quad \frac{\partial^n F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n)}{\partial x_1 \ldots \partial x_n}$$

PROBABILIDAD

D I S T R I B U C I O N E S

Suponiendo que se tienen $\boldsymbol{n}$ **variables aleatorias continuas** $X_1, X_2, \dots, X_n$ se tienen las siguientes **distribuciones**:

Función de densidad de probabilidad marginal de $\boldsymbol{X_1}$

$$f_{X_1}(X_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots x_n) dx_2 \dots dx_n$$

Función de densidad de probabilidad condicionada

$$f_{X_1 | X_2, \dots, X_n}(x_1 | x_2, \dots x_n) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots x_n)}{f_{X_2, \dots, X_n}(x_1, x_2, \dots x_n)}$$

PROBABILIDAD

REGLA DEL PRODUCTO

La función de **densidad** de **probabilidad conjunta** se puede expresar como el **producto** de las **probabilidades condicionadas**:

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n) &= f(x_n | x_1, x_2 \ldots, x_{n-1}) f(x_1, x_2 \ldots, x_{n-1}) \\
&= f(x_n | x_1, x_2 \ldots, x_{n-1}) f(x_{n-1} | x_1, x_2 \ldots, x_{n-2}) f(x_1, x_2 \ldots, x_{n-2}) \\
&= \ldots = f(x_1) \prod_{i=2}^{n} f(x_i | x_1, \ldots, x_{i-1}).
\end{aligned}
$$

PROBABILIDAD

INDEPENDENCIA

La **propiedad** de **independencia** se **generaliza** para $\boldsymbol{n}$ variables aleatorias $X_1, X_2, \dots, X_n$:

$$f(x_1, \dots, x_n) = f(x_1) f(x_2) \cdots f(x_n).$$

Se dice que $\boldsymbol{k}$ **eventos** $A_1, A_2, \dots, A_k$ son **mutuamente independientes** si para **cualquier subconjunto** $S \subseteq \{1, 2, \dots, k\}$ se tiene que:

$$P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i).$$

PROBABILIDAD

VECTORES ALEATORIOS

PROBABILIDAD

VECTORES ALEATORIOS

Cuando se trabaja con $\boldsymbol{n}$ **variables aleatorias** es conveniente **representarlas** utilizando un **vector**, denominado **vector aleatorio**, que realiza un **mapeo de $\boldsymbol{\Omega} \to \mathbb{R}^{\boldsymbol{n}}$:**

$$X = [X_1\ X_2\ \ldots\ X_n]^T$$

PROBABILIDAD

VALOR ESPERADO

Se presenta el **cálculo** del **valor esperado** para $\boldsymbol{n}$ **variables** aleatorias **continuas** en donde se tiene una **función** de **ponderación** $\boldsymbol{g}(\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$y una **función** de **densidad** de **probabilidad** $\boldsymbol{f_{X_1 X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}$**.**

$$E[g(X)] = \int_{\mathbb{R}^n} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots x_n) dx_1 dx_2 \ldots dx_n,$$

PROBABILIDAD

TEOREMA DE BAYES

PROBABILIDAD

DERIVACIÓN DEL TEOREMA

Suponiendo que se tienen **2 variables aleatorias** $\boldsymbol{X}$y $\boldsymbol{Y}$ **discretas** se puede escribir la probabilidad condicionada como:

$$p(Y/X) = \frac{p(X,Y)}{p(X)}$$

Aplicando la **propiedad simétrica** $\boldsymbol{p(Y,X) = p(X,Y)}$ y la **regla** del **producto** de probabilidad $\boldsymbol{p(X,Y) = p(X/Y)p(Y)}$ se tiene que:

$$p(Y/X) = \frac{p(X/Y)p(Y)}{p(X)}$$

PROBABILIDAD

DERIVACIÓN DEL TEOREMA

Además se sabe por la **regla** se la **suma** de **probabilidades**, que la **probabilidad** de **un evento** es igual a la suma de las intersecciones de ese **evento** con todos los demás **eventos**:

$$p(X) = \sum_{Y} p(X,Y) = \sum_{Y} p(X/Y)\, p(Y)$$

$$p(Y/X) = \frac{p(X/Y)p(Y)}{\sum_{Y} p(X/Y)\, p(Y)}$$

PROBABILIDAD

A P L I C A C I Ó N D E L T E O R E M A

Ejemplo:

2. *Question:* A diagnostic test has a probability 0.95 of giving a positive result when applied to a person suffering from a certain disease, and a probability 0.10 of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.5 % of the population are sufferers. Suppose that the test is now administered to a person about whom we have no relevant information relating to the disease (apart from the fact that he/she comes from this population). Calculate the following probabilities:

(a) that the test result will be positive;

(b) that, given a positive result, the person is a sufferer;

PROBABILIDAD

APLICACIÓN DEL TEOREMA

Respuesta:

$$\text{(a)}\quad \mathbf{P}(T) = \mathbf{P}(T|S)\mathbf{P}(S) + \mathbf{P}(T|S')\mathbf{P}(S') = (0.95 \times 0.005) + (0.1 \times 0.995) = 0.10425.$$

$$\text{(b)}\quad \mathbf{P}(S|T) = \frac{\mathbf{P}(T|S)\mathbf{P}(S)}{\mathbf{P}(T|S)\mathbf{P}(S) + \mathbf{P}(T|S')\mathbf{P}(S')} = \frac{0.95 \times 0.005}{(0.95 \times 0.005) + (0.1 \times 0.995)} = 0.0455.$$

PROBABILIDAD

EXPLICANDO EL TEOREMA

Interpretando el **Teorema** de **Bayes**:

$$\overset{\text{Posterior}}{P(A|B)} = \frac{\overset{\text{Likelihood}}{P(B|A)} * \overset{\text{Prior}}{P(A)}}{\underset{\text{Evidence}}{P(B)}}$$

1. **Prior**: son las "**creencias**" que tenemos sobre como se distribuye la variable aleatoria $\boldsymbol{A}$ **antes** de obtener **evidencia** $\boldsymbol{B}$**.**

2. **Posterior:** captura la **distribución** de la **variable aleatoria A después** de haber recopilado la **evidencia** $\boldsymbol{B}$.

3. **Likelihood**: expresa la **probabilidad** de que nuestras **creencias** (distribución de $\boldsymbol{A}$) sean ciertas acorde a la evidencia B.

PROBABILIDAD

EXPLICANDO EL TEOREMA

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Leer páginas **21 – 23** del libro "***Pattern Recognition and Machine Learning"*** de Christopher Bishop, 1° Edición.