



Curso de

Matemáticas para Data Science: Probabilidad

Francisco Camacho

[C1] ¿Qué es la probabilidad?

Incertidumbre y probabilidad

Incertidumbre

Tomar decisiones
con información
incompleta.



“

**El azar no es más que la
medida de nuestra ignorancia.
Los fenómenos fortuitos
son, por definición, aquellos
cuyas leyes o causas
simplemente ignoramos**

”

Henri Poincaré

Probabilidad

Es un lenguaje
que nos da
herramientas
para cuantificar
la incertidumbre.

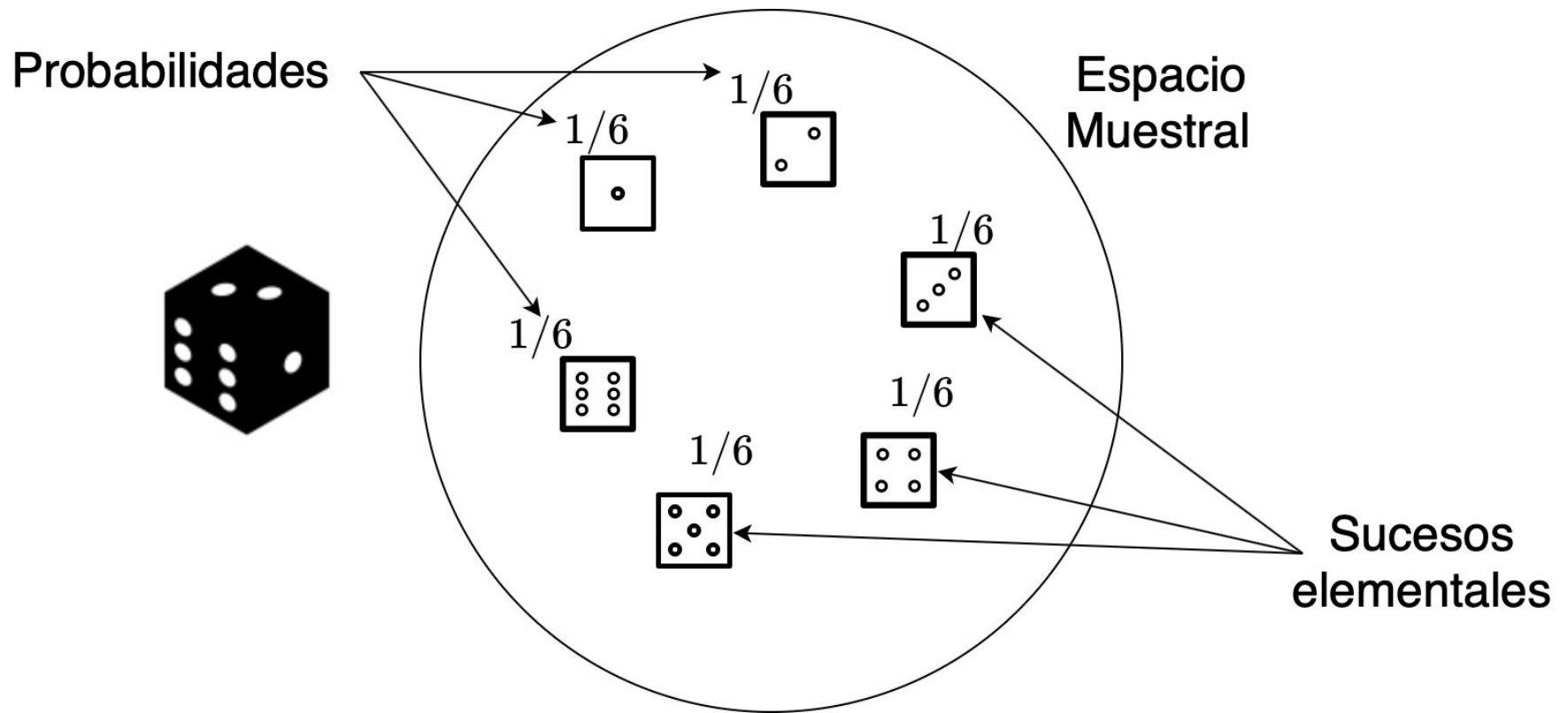


Axiomas de la probabilidad

- **Suceso elemental:**
“El resultado de lanzar un dado es 4”.
- **Suceso:**
“El resultado de lanzar un dado es par”.

$$P = \frac{\text{Nº sucesos exitosos}}{\underbrace{\text{Nº sucesos totales}}_{\text{creencia del total}}}$$

Axiomas de la probabilidad



Axiomas de la probabilidad

$$0 \leq P \leq 1$$

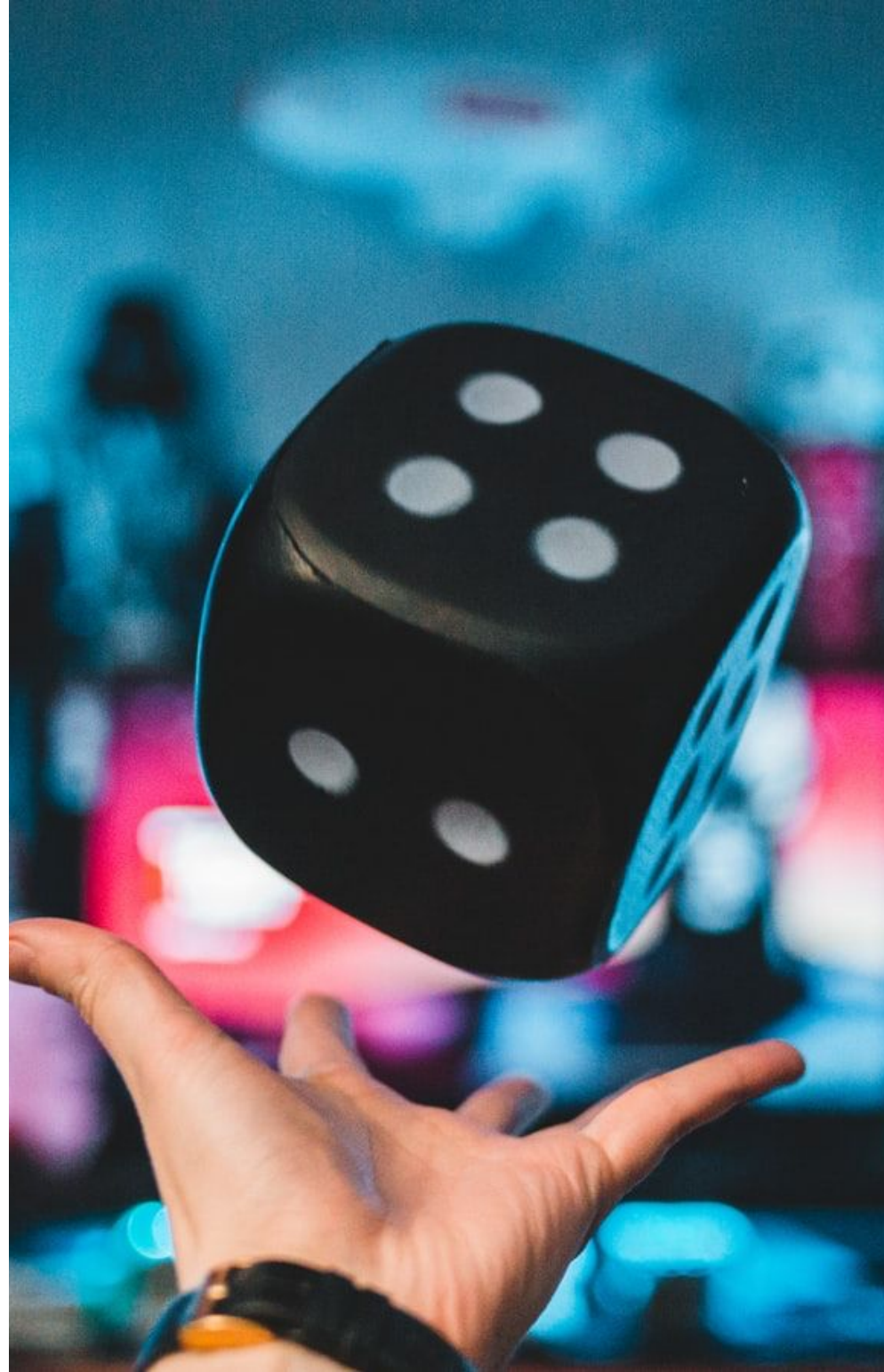
$$\text{certeza} \rightarrow P = 1$$

$$\text{imposibilidad} \rightarrow P = 0$$

$$\text{disjuntos} \rightarrow P(A \cup B) = P(A) + P(B)$$

¿Qué es realmente la probabilidad?

Es una creencia que tenemos sobre la ocurrencia de eventos elementales.



[C2] Probabilidad en Machine Learning

Incertidumbre y probabilidad

Fuentes de incertidumbre

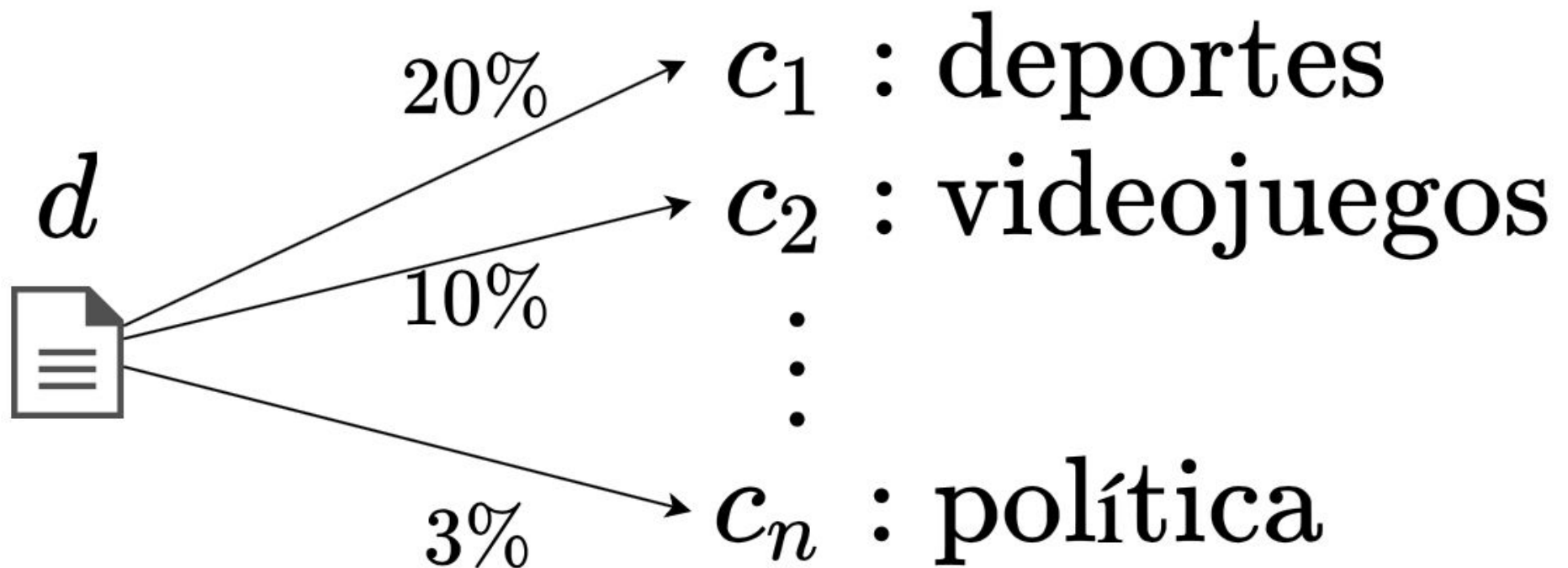
- Datos
- Atributos del modelo
- Arquitectura del modelo



A photograph of a chalkboard with handwritten mathematical equations. The equations are written in white chalk on a dark, textured surface. The equations involve variables like $\frac{cM}{dt}$, $\frac{dS}{dt}$, $\frac{S}{P_r}$, and S , along with constants and parameters like q_{vact} , $g_0(N-N_0)$, $(1-\epsilon S)$, $\frac{M_e}{T_n}$, $\frac{M}{T_p}$, T_0 , q_0 , $\frac{P_0 N}{T_n}$, $\frac{S}{T_p}$, $T_{cp} \lambda_0$, $T_{\text{act}} \eta_{\text{nc}}$, and ϵ . The equations are grouped by large curly braces on the left and right sides. A boxed equation $S \leq \frac{1}{\epsilon}$ is also visible on the right side.

$$\left\{ \begin{array}{l} \frac{cM}{dt} = \frac{1}{q_{\text{vact}}} - g_0(N-N_0)(1-\epsilon S)S + \frac{M_e}{T_n} - \frac{M}{T_p} \\ \frac{dS}{dt} = T_0 q_0(N-N_0)(1-\epsilon S)S + \frac{P_0 N}{T_n} - \frac{S}{T_p} \\ \frac{S}{P_r} = \frac{T_{cp} \lambda_0}{T_{\text{act}} \eta_{\text{nc}}} = 0 \end{array} \right. \quad \left\{ \begin{array}{l} S \leq \frac{1}{\epsilon} \end{array} \right.$$

Modelo de clasificación



Modelo de clasificación

Entrenamiento

Etiquetas



Extractor
atributos

Atributos



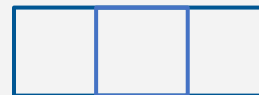
Algoritmo
ML

Predicción



Extractor
atributos

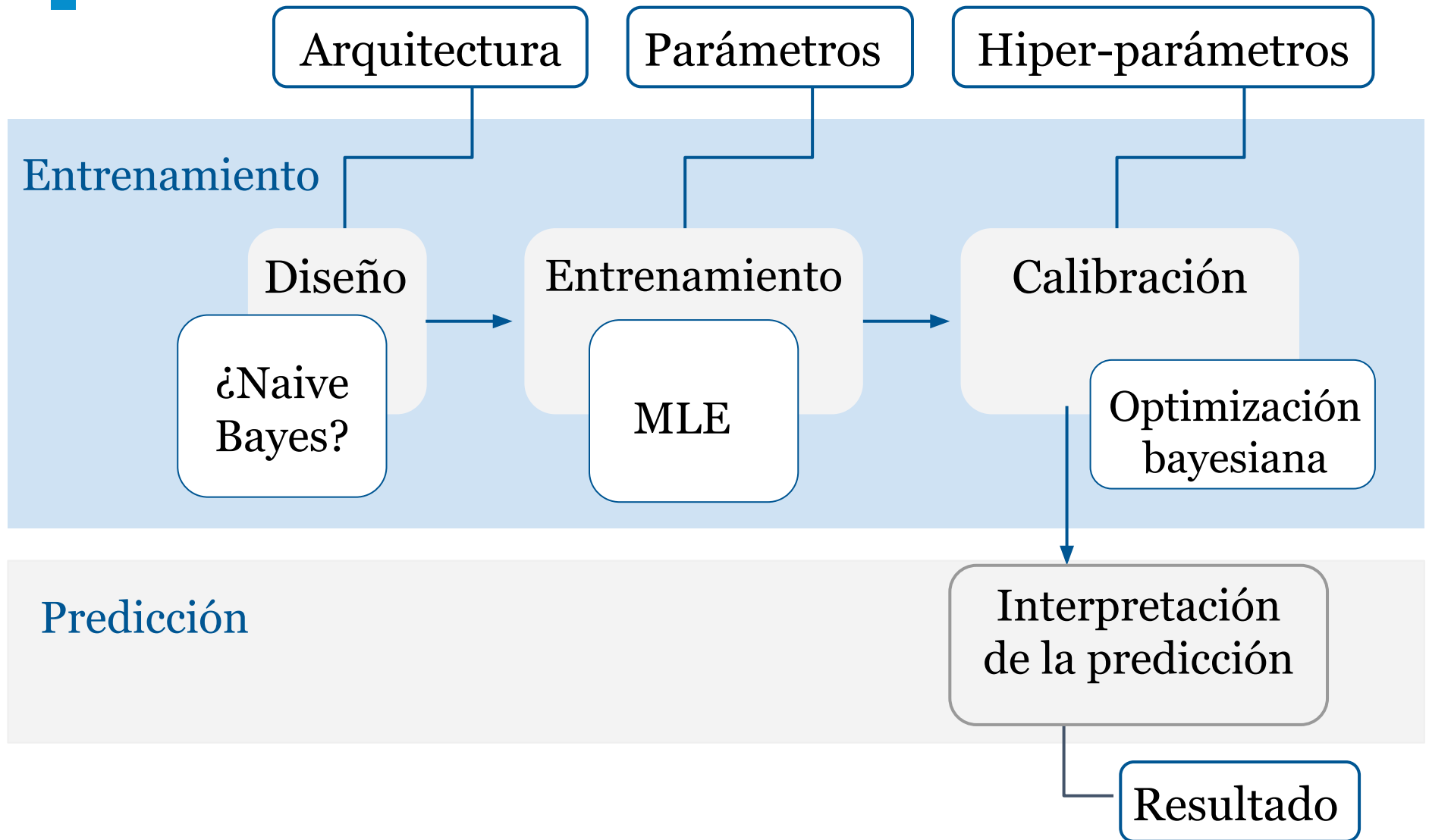
Atributos



Modelo de
clasificación

Etiqueta

Todas las etapas del modelo



[C3] Tipos de probabilidad

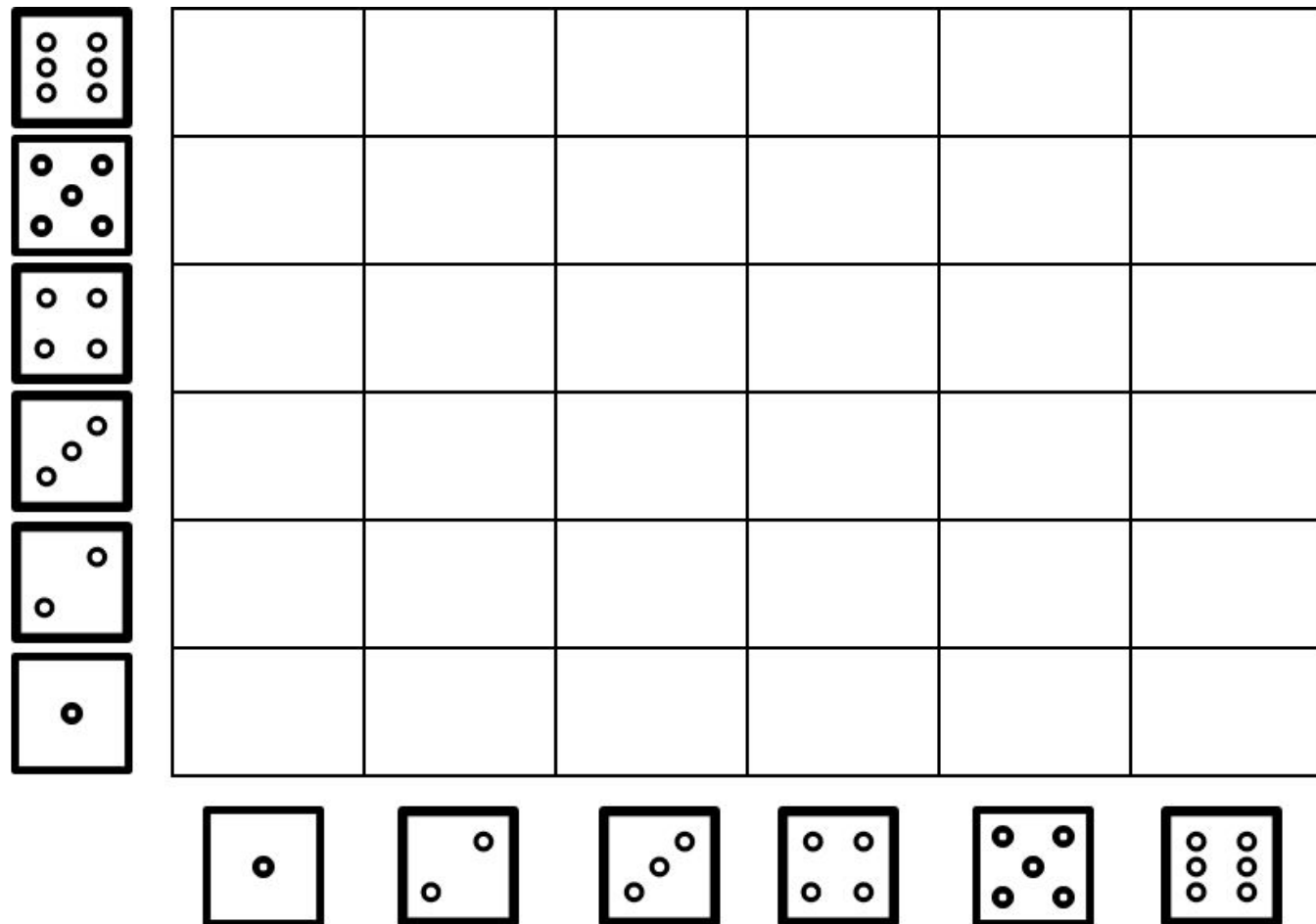
Fundamentos de probabilidad

Tipos de probabilidad

- Conjunta (joint)
- Marginal
- Condicional



Juego de dos dados



Probabilidad conjunta

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

Probabilidad condicional

$$P(A \text{ dado } B) = P(A|B)$$

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

Regla del
producto


[C4] Ejemplos de cálculo de probabilidad


Fundamentos de probabilidad

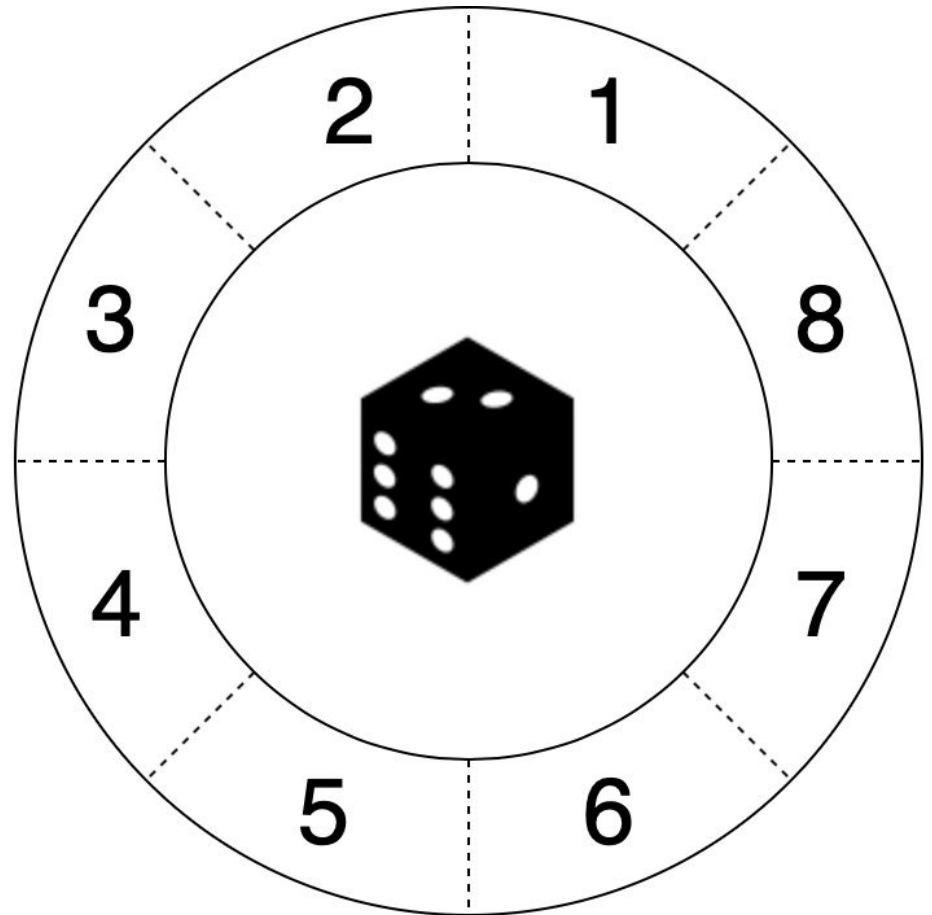
Correlaciones de eventos

- $A = \{\text{el resultado de lanzar un dado es } 4\}$
- $B = \{\text{el resultado de lanzar un dado es par}\}$
- $C = \{\text{el resultado de lanzar un dado es impar}\}$

Juego de ruleta

 → {1, 2, 3, 4}
Jugador 1

 → {5, 6, 7, 8}
Jugador 2



[C5] Ejemplos avanzados con probabilidad

Fundamentos de probabilidad



Paradoja ¿niño o niña?

- Una mujer tiene dos bebés donde el mayor es un varón.
- Una mujer tiene dos bebés donde uno de ellos es varón.

El problema de Monthly Hall



[C6] ¿Qué es una distribución?

Distribuciones de probabilidad

Volvemos al cálculo

X aleatoria \rightarrow $\underbrace{P(X = x)}_{\text{probabilidad de ocurrencia}}$

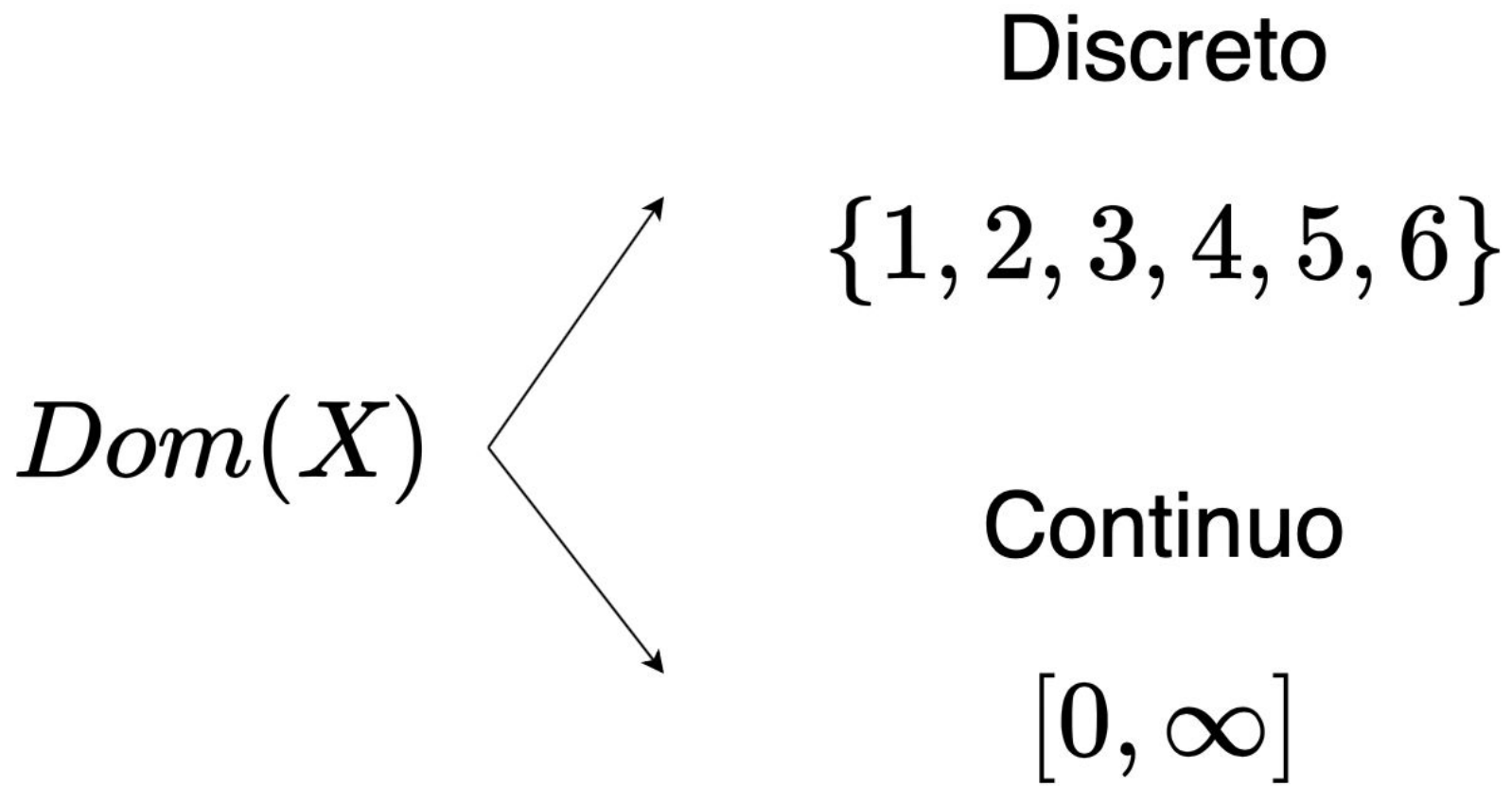
$$P = f(X)$$

Volvemos al cálculo

$X \rightarrow$ variable aleatoria

$x \rightarrow$ valores posibles en el espacio muestral

Volvemos al cálculo



[C7] Distribuciones discretas

Distribuciones de probabilidad

Distribución de Bernoulli

Variables con ocurrencias binarias

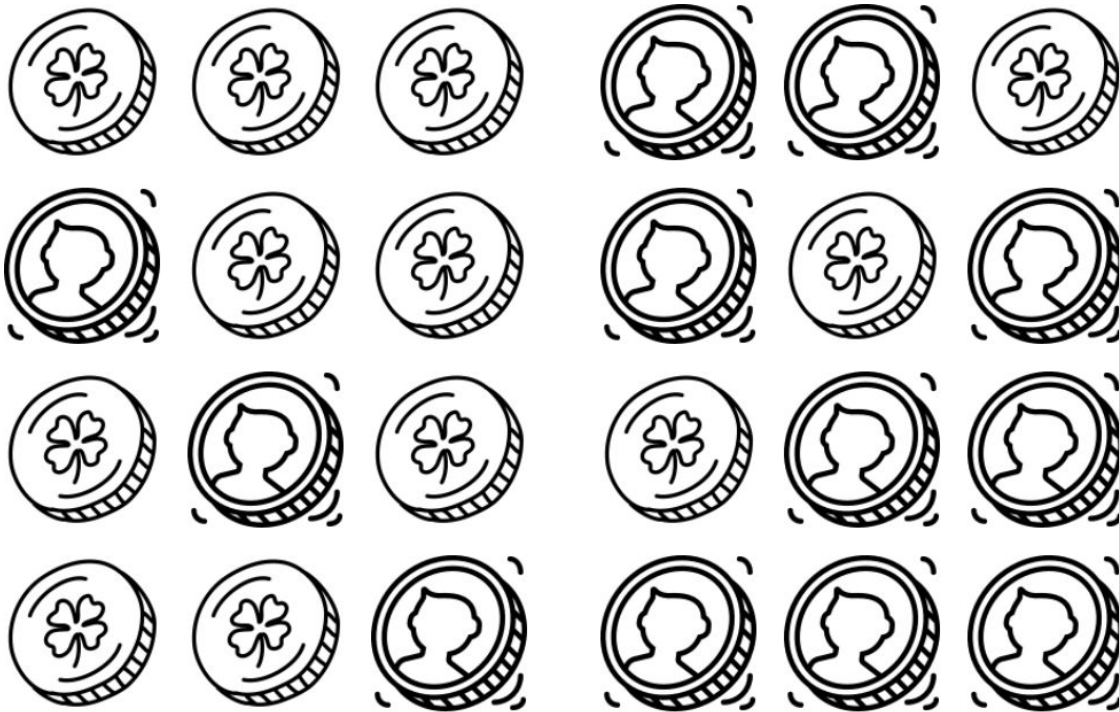


$$P(X = 1) = p$$

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

Distribución binomial

Secuencia repetitiva de eventos tipo Bernoulli

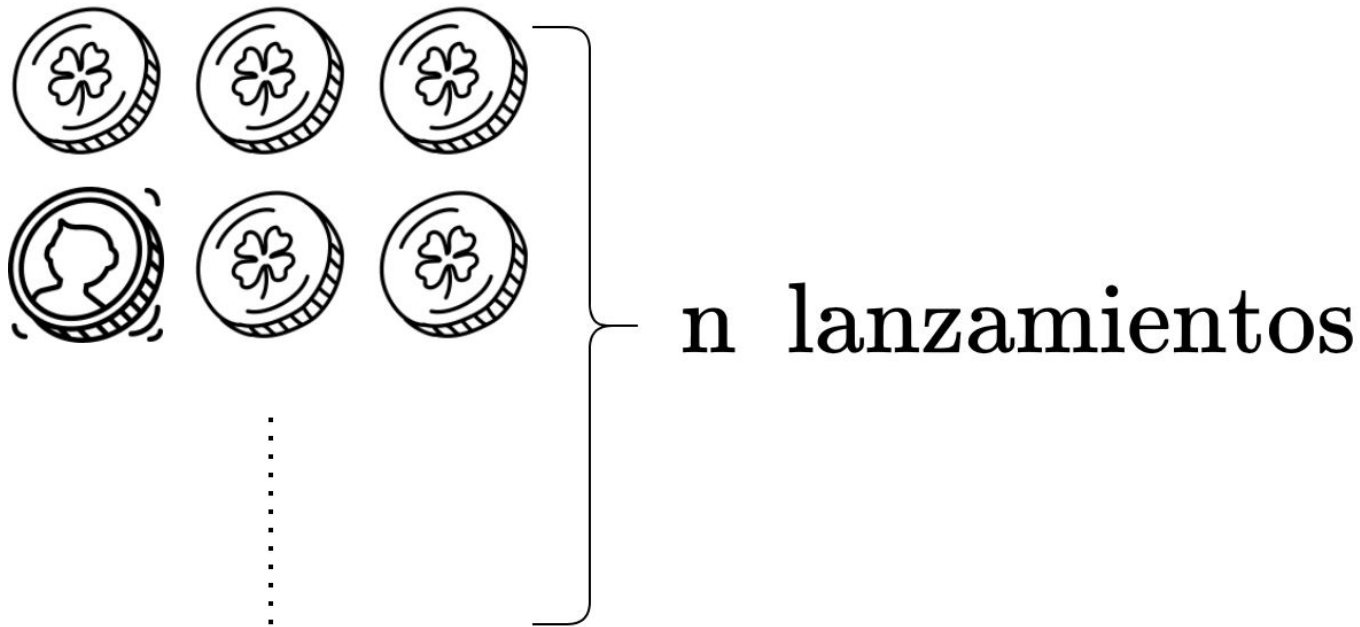


todos los
eventos
igualmente
probables

$$P(2 \text{ caras} | 3 \text{ lanzamientos}) = 3/8$$

Distribución binomial

Secuencia repetitiva de eventos tipo Bernoulli



$$P(k \text{ caras} | n \text{ lanzamientos}) = ?$$

Distribución binomial

$$P(\text{k caras} | \text{n lanzamientos}) = P(k; n, p)$$

$$P(k; n, p) = \binom{n}{k} p^k (1 - p)^{(n-k)}$$

Distribución multinomial

$$P(X_1, \dots, X_n) = \frac{n!}{k_1! \dots k_n!} p_1^{k_1} \dots p_n^{k_n}$$



Otras distribuciones

- Poisson
- Geométrica
- Hipergeométrica
- Binomial negativa
- ...



[C8] Usando la distribución binomial

Distribuciones de probabilidad

[C9] Distribuciones continuas

Distribuciones de probabilidad

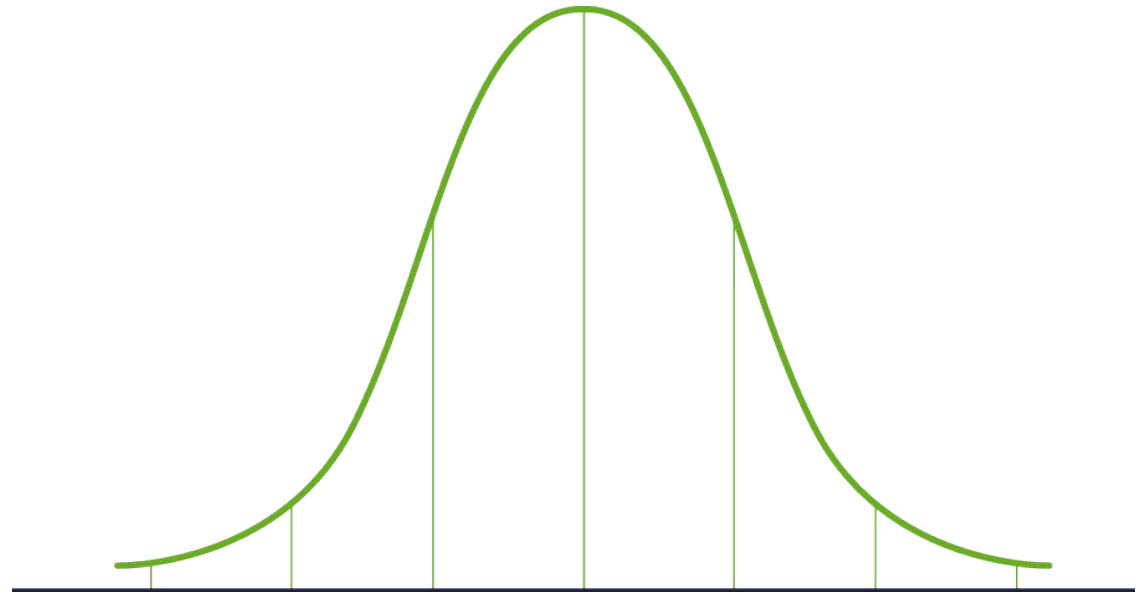
Distribución normal (gaussiana)

Un ejemplo numérico con el siguiente dataset:

<https://seattlecentral.edu/qelp/sets/057/057.html>

Distribución normal (gaussiana)

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{X - \mu}{\sigma} \right)^2$$



Otras distribuciones

- Exponencial
- Pareto
- ...



[C10] ¿Cómo estimar una distribución?

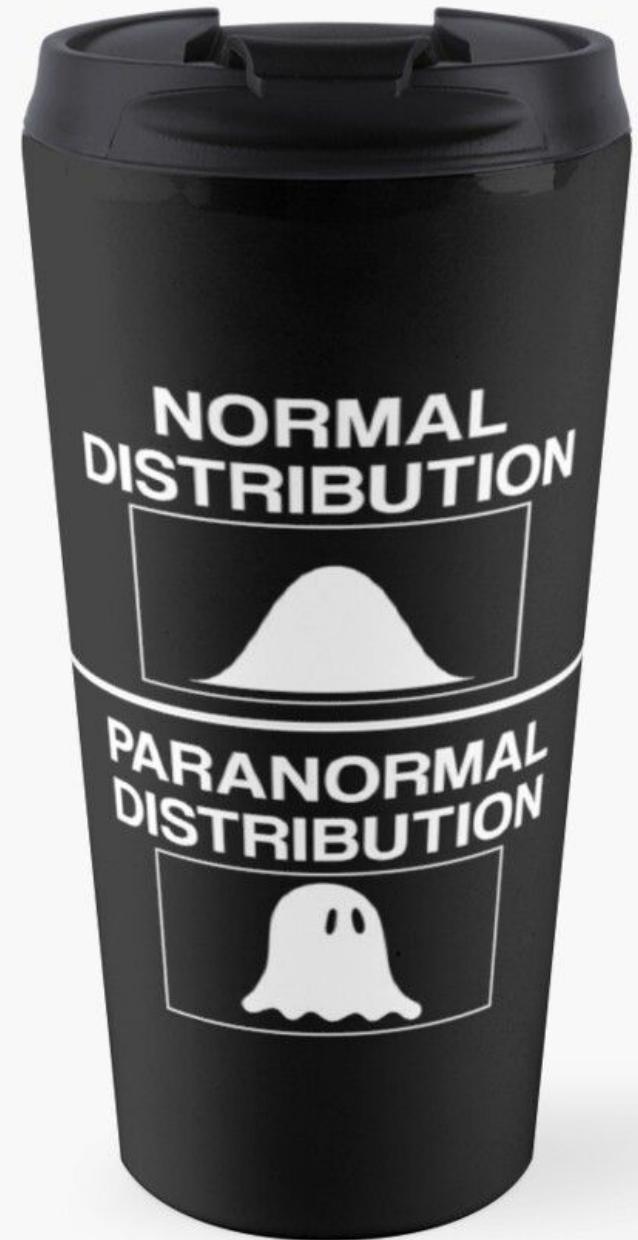
Distribuciones de probabilidad

[C11] ¿Qué es MLE?

Estimación de máxima verosimilitud

MLE

*Un framework
para estimación
de densidades
de probabilidad.*





Elementos de MLE

- **Escoger la distribución:**
Teniendo solo una muestra de los datos.
- **Escoger los parámetros de la distribución:**
Que mejor ajustan la distribución a los datos.

Un problema de optimización

$$P(X; \theta) = L(X; \theta)$$

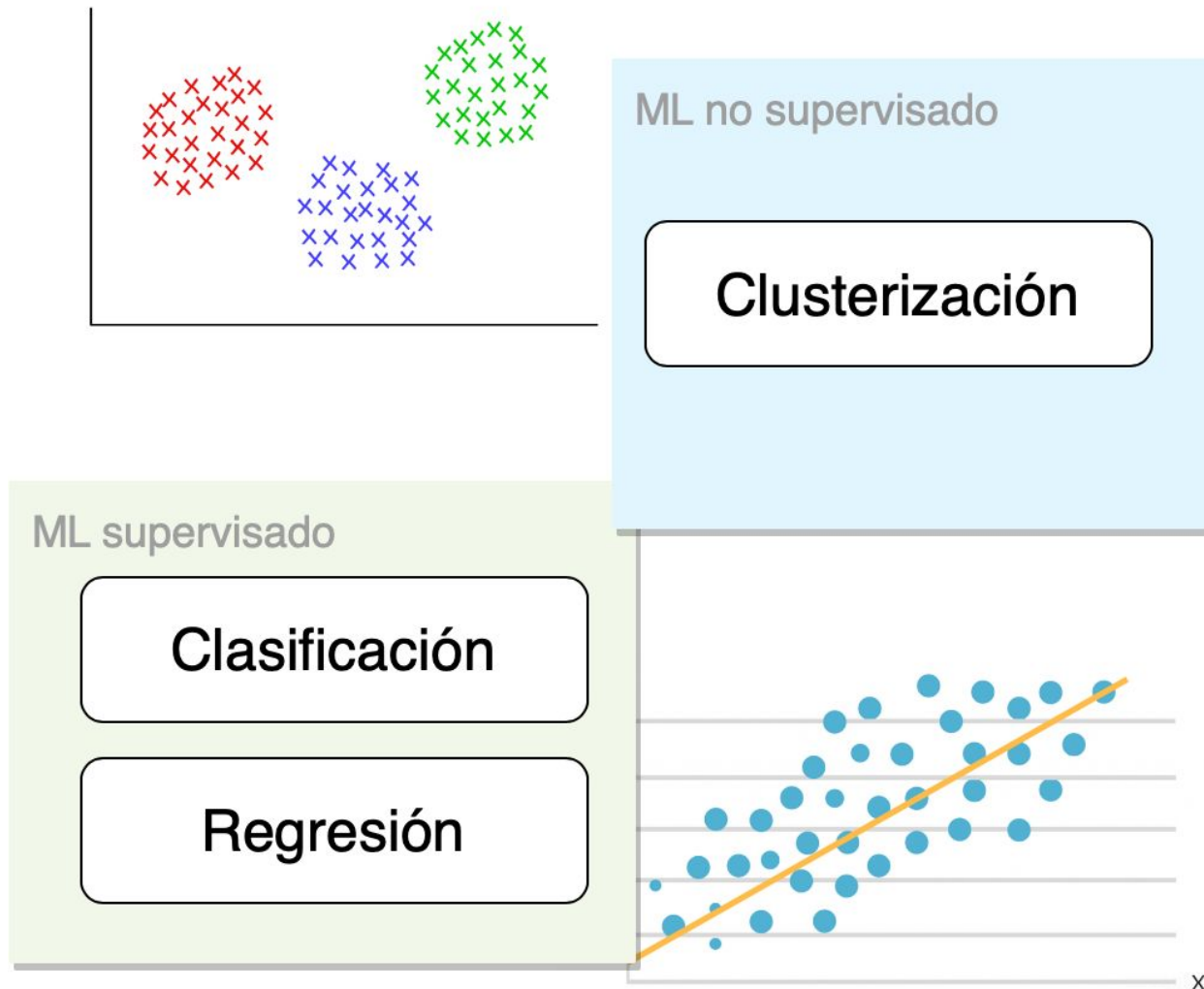
$$\max L(X; \theta) \rightarrow \max \prod_i P(X_i; \theta)$$

$$\max \log L(X; \theta) \rightarrow \max \sum_i \log P(X_i; \theta)$$

[C12] MLE en Machine Learning

Estimación de máxima verosimilitud

ML: ajustar densidades a datos



Regresión lineal con MLE

$$y = \underbrace{m}_{\text{pendiente}} x + \underbrace{b}_{\text{intercepto}} = \underbrace{b_0}_{\text{weight}} x + \underbrace{b_1}_{\text{bias}}$$

$$P(y|x) \rightarrow \max \sum_i \log P(y_i | x_i; \underbrace{h}_{\text{modelo}})$$

Regresión lineal con MLE

$$P(y|x) \rightarrow \max \sum_i \log P(y_i|x_i; \underbrace{h}_{\text{modelo}})$$

$$h \rightarrow y = b_0 x + b_1$$

$$P \rightarrow \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{X - \mu}{\sigma} \right)^2$$

[C13] Regresión logística

Estimación de máxima verosimilitud

[C14] Aplicación de Regresión Logística

Estimación de máxima verosimilitud

[C15] Teorema de Bayes

Inferencia bayesiana

Dos escuelas de probabilidad

Frecuentistas | Bayesianos

SI TAN SOLO CONOCIERAS EL PODER

DEL TEOREMA DE BAYES



Teorema de Bayes

$$\underbrace{P(A|B)}_{\text{posteriori}} = \frac{\overbrace{P(B|A)}^{\text{verosimilitud}} \overbrace{P(A)}^{\text{priori}}}{\underbrace{P(B)}_{\text{evidencia}}}$$

[C16] Bayes en Machine Learning

Inferencia bayesiana

MAP vs MLE

$$P(\overbrace{h}^{\text{hipótesis}} \mid \underbrace{D}_{\text{datos}}) = \frac{P(D|h)P(h)}{P(D)}$$

$$\max P(h|D) \rightarrow \max P(D|h)P(h)$$

Clasificador óptimo de Bayes

