laSalle
Universitat Ramon Llull

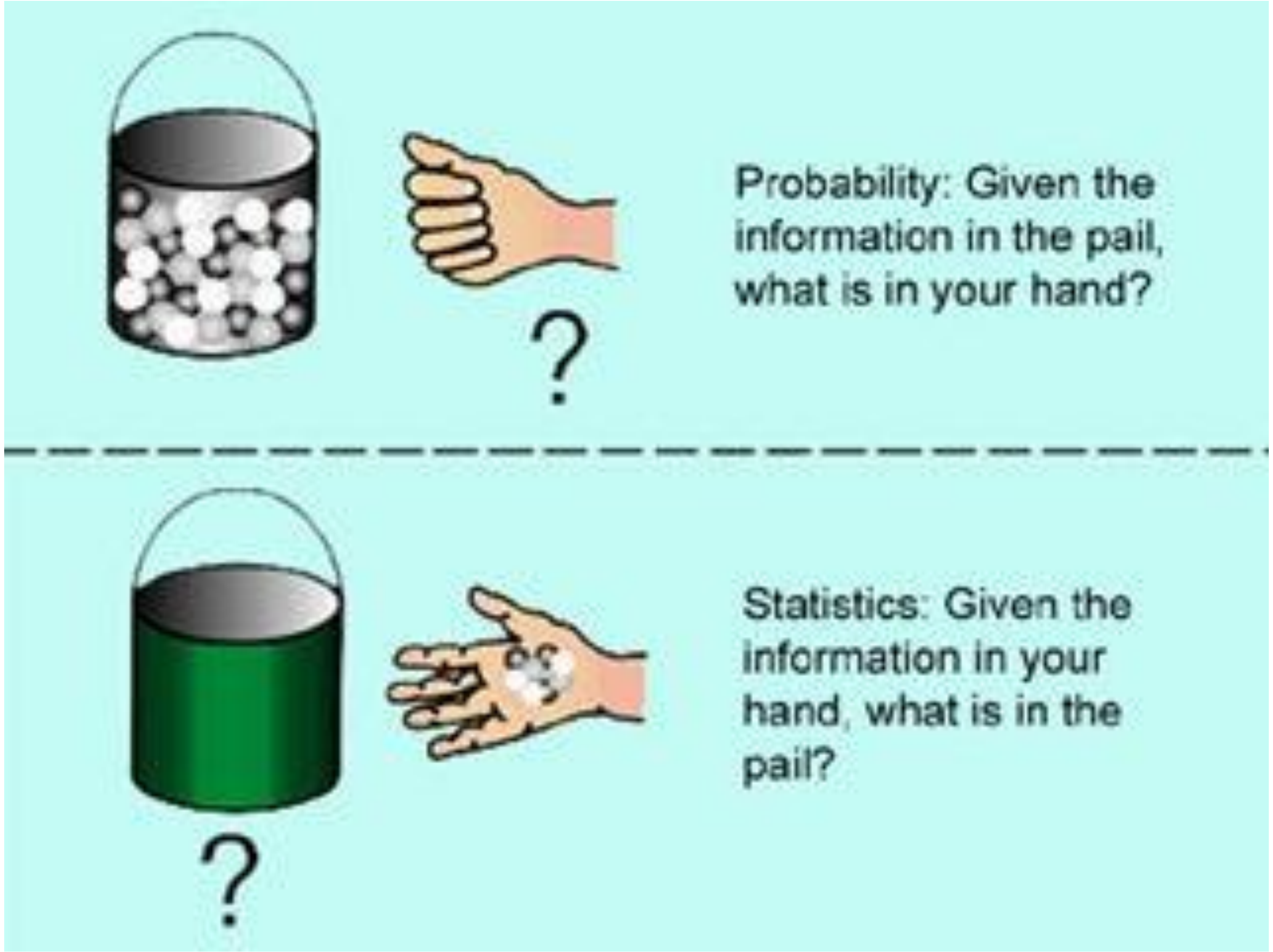
Repaso de Probabilidad

Xavier Vilasís

Master Universitario en Ciencia de los Datos/Data Science

- Diferencia entre probabilidad y estadística
- Definición de probabilidad
- Probabilidad condicionada
- Teorema de Bayes
- Definición, tipos y caracterización de variables aleatorias.
- Valor esperado, media y desviación estándar.
- Ley de los grandes números.
- Teorema del límite central.

Probabilidad vs Estadística



Definición de Probabilidad

$$p(A) = \lim_{\mathcal{N}\to\infty} f_{\mathcal{N}}(A) = \lim_{\mathcal{N}\to\infty} \frac{\mathcal{N}_A}{\mathcal{N}}.$$

$$p(A) = \frac{\text{Number of cases for } A}{\text{All possible cases}}$$

Let's consider a **Probability Space**. A **Probability** is a mapping from the field of events to real numbers,

$$p: \quad \begin{array}{c} \mathcal{F} \to \Re \\ A \mapsto p(A) \end{array}$$

which fulfills the three following axioms

Axiom 1

$$\forall A \in \mathcal{F}, \qquad p(A) \geq 0.$$

Axiom 2

$$p(\Omega) = 1.$$

Axiom 3

$$\forall A, B \in \mathcal{F} / A \cap B = \emptyset$$

$$p(A \cup B) = p(A) + p(B).$$

Probabilidad Condicionada

$$\boxed{p(A \mid B) = \frac{p(A \cap B)}{p(B)}.}$$

Two events A and B are **INDEPENDENT** if

$$p(A \cap B) = p(A)p(B).$$

Total Probability Theorem

Let A_i, $i = 0, \ldots, n$ a partition of Ω, that is

$$\Omega = \bigcup_{i=1}^{n} A_i, \qquad A_i \cap A_j = \emptyset, \quad \forall i, j.$$

For any event B,

$$p(B) = \sum_{i=1}^{n} p(B \mid A_i)p(A_i).$$

Teorema de Bayes

Let A_i, $i = 0, \ldots, n$ be a partition of Ω,
For any event B and any element of the partition A_i,

$$p(A_i \mid B) = \frac{p(B \mid A_i)p(A_i)}{\sum_{i=1}^{n} p(B \mid A_i)p(A_i)}.$$

Variables Aleatorias

A **Random Variable** X
is a mapping from a
Probability Space
onto real numbers,

$$X: \quad \Omega \to \Re$$
$$A \mapsto X(A)$$

such that the experiment outcome
must be a finite number
and all intervals of the type
$]-\infty, x]$
are events.

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

- Discretas

$$f_X(x) = P(X = x)$$

- Continuas

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

- Mixtas

...

http://zoonek2.free.fr/UNIX/48_R/07.html

Valor Esperado, Media y Desviación Estándard

- Valor Esperado

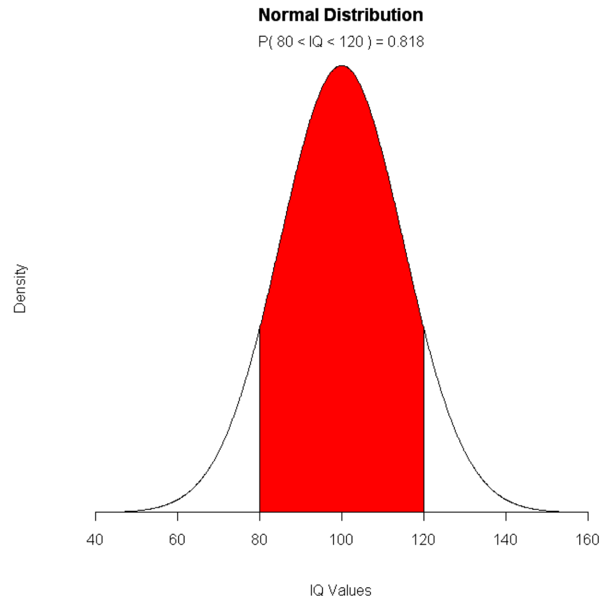
$$< g(X) >= \int_{-\infty}^{\infty} dx \;\, f(x)g(x).$$

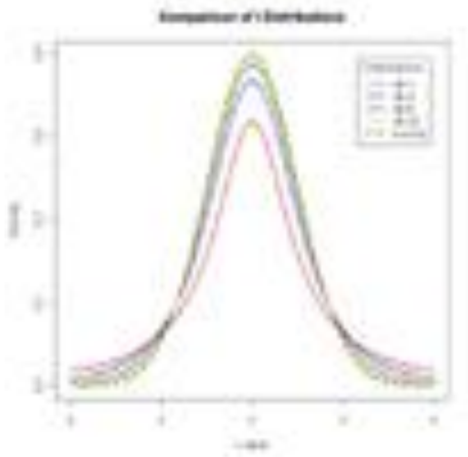
- Media

$$\mu =< x >= \int_{-\infty}^{\infty} dx \;\, f(x)x,$$

- Varianza, desviación estándard

$$\sigma^2 =< (x-\mu)^2 >= \int_{-\infty}^{\infty} dx \;\, f(x)(x-\mu)^2$$





[Graphs from https://www.statmethods.net/advgraphs/probability.html](https://www.statmethods.net/advgraphs/probability.html)

Ley de los Grandes Números

Let X be a random variable with probability density $f(x)$.
X has mean μ and variance σ^2.
We perform n **independent** repetitions of this experiment.
Each repetition will be described by a random variable, X_i,
$i = 1, \ldots, n$, with probability density $f(x)$.
We define the *average* random variable,

$$m_n = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

The **Law of Large Numbers** states that

$$\lim_{n \to \infty} P\left[\mid m_n - \mu \mid < \epsilon \right] = 1$$

Teorema del Límite Central

Consider n independent random variables, x_i $i = 1, \ldots, n$, eventually with different probability densities, average μ_i and variance σ_i^2.
We define x to be the sum of the n random variables.

$$x = x_1 + x_2 + \cdots + x_n.$$

The average of x is

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n.$$

and its variance

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2.$$

In the limit $n \to \infty$, x is a Gaussian random variable.
In other words. If we define,

$$z = \frac{1}{\sigma}(x - \mu),$$

$$\lim_{n\to\infty} f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$