Introducción

José Antonio Perusquía Cortés

Inferencia Estadística - Semestre 2025-II

¿ Qué es la estadística ?

Statistics is about gathering data and working out what the numbers can tell us. From the earliest farmer estimating whether he had enough grain to last the winter to the scientists of the Large Hadron Collider confirming the probable existence of new particles, people have always been making inferences from data. Statistical tools like the mean or average summarise data, and standard deviations measure how much variation there is within a set of numbers. Frequency distributions - the patterns within the numbers or the shapes they make when drawn on a graph - can help predict future events. Knowing how sure or how uncertain your estimates are is a key part of statistics.

Today vast amounts of digital data are transforming the world and the way we live in it. Statistical methods and theories are used everywhere, from health, science and business to managing traffic and studying sustainability and climate change. No sensible decision is made without analysing the data. The way we handle that data and draw conclusions from it uses methods whose origins and progress are charted here.

Julian Champkin
Significance magazine

"Statistics is the science of information gathering, especially when the information arrives in little pieces instead of big ones." Bradley Efron is good at putting things simply. He talked to **Julian Champkin**.

statistics

science

Print | Cite | Share | Feedback

Written by David R. Anderson, Thomas A. Williams | All
Fact-checked by The Editors of Encyclopaedia Britannica
Article History

statistics, the science of collecting, analyzing, presenting, and interpreting data.

Introduction to Statistics

Sheldon M. Ross, in Introduction to Probability and Statistics for Engineers and Scientists (Fifth Edition), 2014

1.1 Introduction

It has become accepted in today's world that in order to learn about something, you must first collect data. *Statistics* is the art of learning from data. It is concerned with the collection of data, its subsequent description, and its analysis, which often leads to the drawing of conclusions.

Introduction

Andrew F. Siegel, in Practical Business Statistics (Seventh Edition), 2016

1.2 What is Statistics?

Statistics is the art and science of collecting and understanding data. Since *data* refers to any kind of recorded information, statistics plays an important role in many human endeavors.

Orígenes

AD 2 Chinese census under the Han dynasty finds 57.67 million people in 12.36 million households – the first census from which data survives, and still considered by scholars to have been accurate.

431 BC Attackers besieging Plataea in the Peloponnesian war calculate the height of the wall by counting the number of bricks. The count was repeated several times by different soldiers. The most frequent value (the mode) was taken to be the most likely. Multiplying it by the height of one brick allowed them to calculate the length of the ladders needed to scale the walls.

1303 A Chinese diagram entitled "The Old Method Chart of the Seven Multiplying Squares" shows the binomial coefficients up to the eighth power – the numbers that are fundamental to the mathematics of probability, and that appeared five hundred years later in the west as Pascal's triangle.

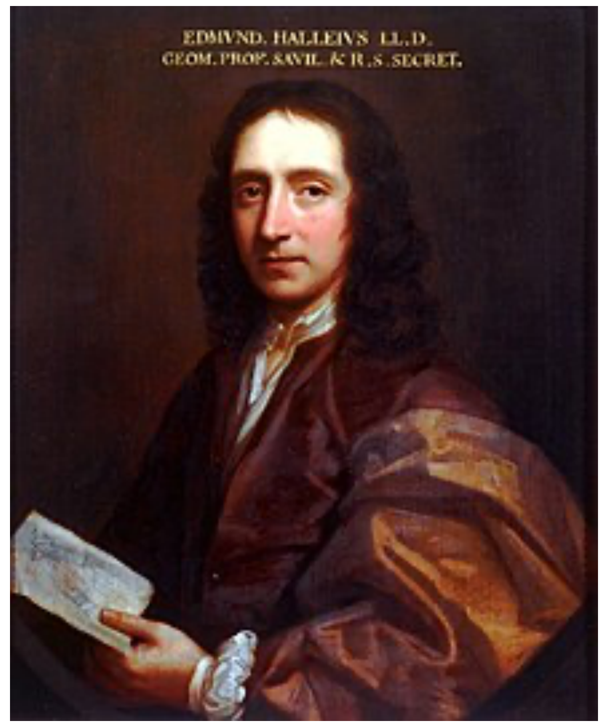
400 BC In the Indian epic the *Mahabharata*, King Rtuparna estimates the number of fruit and leaves (2095 fruit and 50 000 000 leaves) on two great branches of a vibhitaka tree by counting the number on a single twig, then multiplying by the number of twigs. The estimate is found to be very close to the actual number. This is the first recorded example of sampling – "but this knowledge is kept secret", says the account.

1150 Trial of the Pyx, an annual test of the purity of coins from the Royal Mint, begins. Coins are drawn at random, in fixed proportions to the number minted. It continues to this day.

Cimientos

1560 Gerolamo Cardano calculates probabilities of different dice throws for gamblers.

1654 Pascal and Fermat correspond about dividing stakes in gambling games and together create the mathematical theory of probability.



1693 Edmund Halley prepares the first mortality tables statistically relating death rates to age – the foundation of life insurance. He also drew a stylised map of the path of a solar eclipse over England – one of the first data visualisation maps.

1713 Jacob Bernoulli's *Ars conjectandi* derives the law of large numbers – the more often you repeat an experiment, the more accurately you can predict the result.

1761 The Rev. Thomas Bayes proves Bayes' theorem – the cornerstone of conditional probability and the testing of beliefs and hypotheses.

1805 Adrien-Marie Legendre introduces the method of least squares for fitting a curve to a given set of observations.

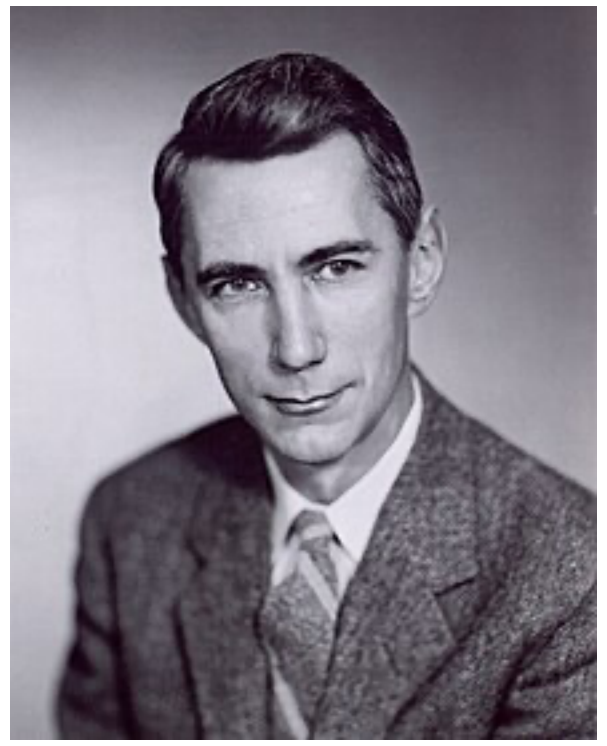
1808 Gauss, with contributions from Laplace, derives the normal distribution – the bell-shaped curve fundamental to the study of variation and error.

1894 Karl Pearson introduces the term "standard deviation". If errors are normally distributed, 68% of samples will lie within one standard deviation of the mean. Later he develops chi-squared tests for whether two variables are independent of each other.

Era moderna

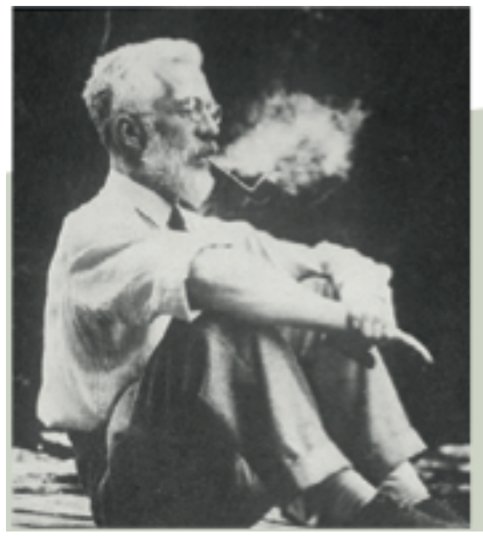
1908 William Sealy Gosset, chief brewer for Guinness in Dublin, describes the *t*-test. It uses a small number of samples to ensure that every brew tastes equally good.

1948 Claude Shannon introduces information theory and the "bit" – fundamental to the digital age.

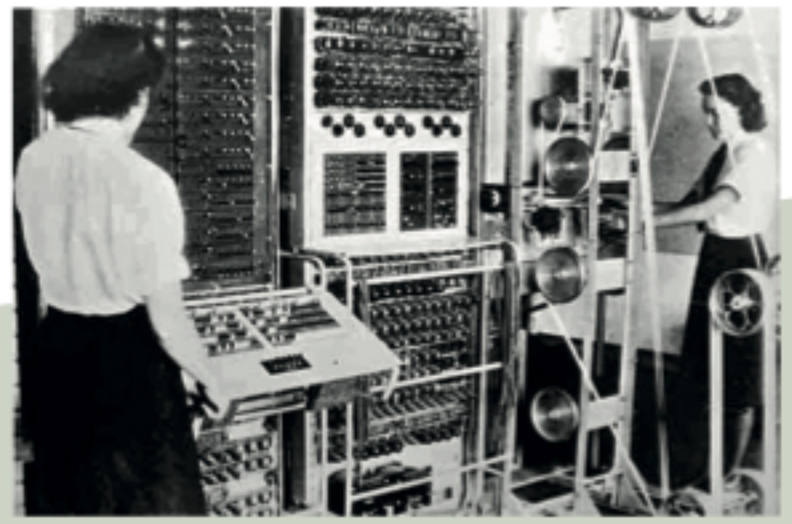


1937 Jerzy Neyman introduces confidence intervals in statistical testing. His work leads to modern scientific sampling.

1958 The Kaplan–Meier estimator gives doctors a simple statistical way of judging which treatments work best. It has saved millions of lives.

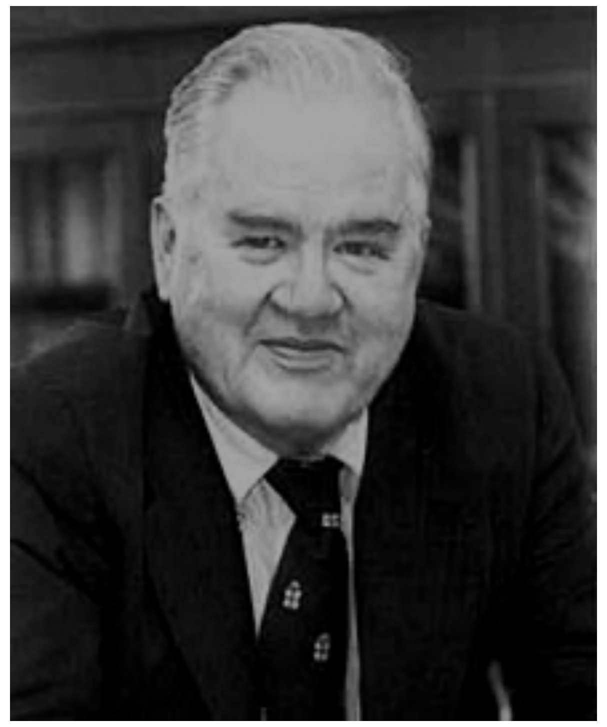


1935 R. A. Fisher revolutionises modern statistics. His *Design of Experiments* gives ways of deciding which results of scientific experiments are significant and which are not.

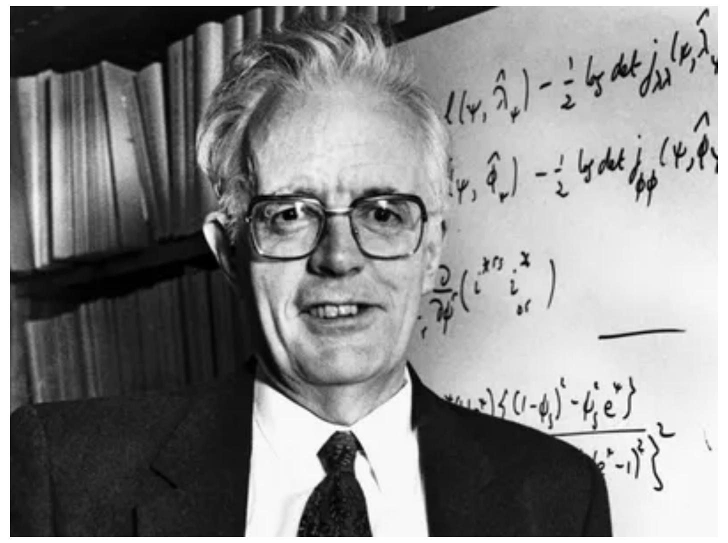


1940-45 Alan Turing at Bletchley Park cracks the German wartime Enigma code, using advanced Bayesian statistics and Colossus, the first programmable electronic computer.

1977 John Tukey introduces the box-plot or box-and-whisker diagram, which shows the quartiles, medians and spread of data in a single image.



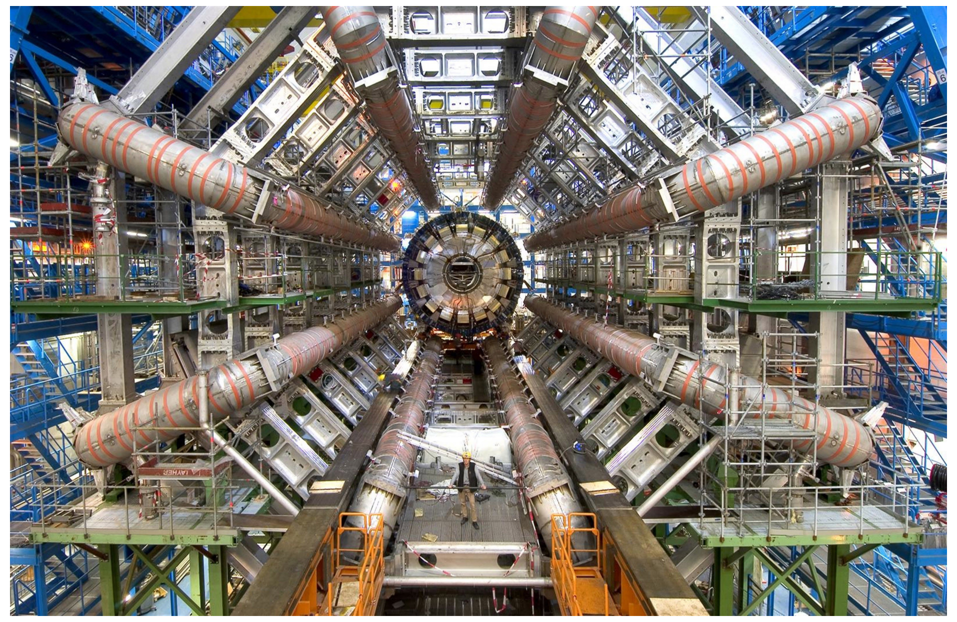
Era moderna



1972 David Cox's proportional hazard model and the concept of partial likelihood.

1993 The statistical programming language "R" is released, now a standard statistical tool.

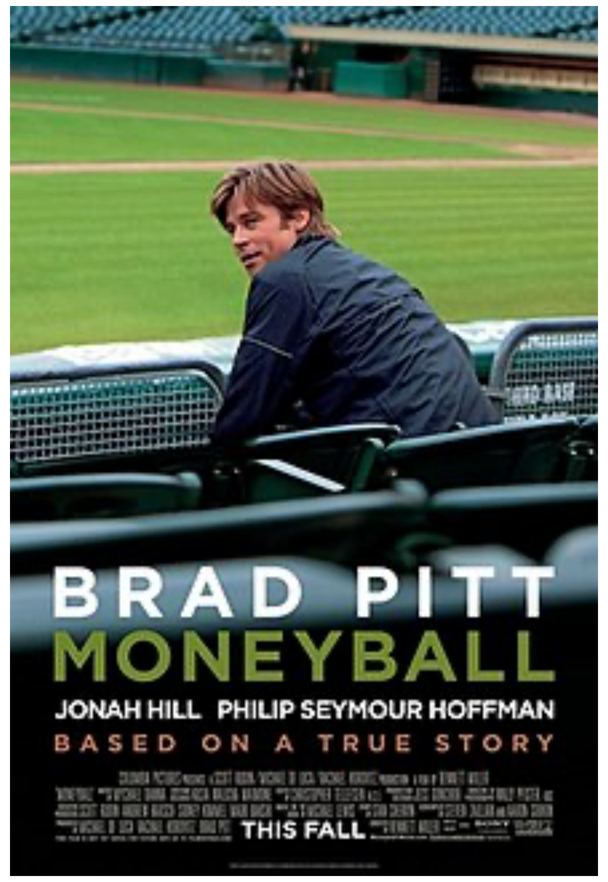
1997 The term "Big Data" first appears in print.



2012 The Large Hadron Collider confirms existence of a Higgs boson-like particle with probability of five standard deviations – around one chance in 3.5 million that all they are seeing is coincidence.

1979 Bradley Efron introduces bootstrapping, a simple way to estimate the distribution of almost any sample of data.

2002 Paul DePodesta uses statistics – "sabermetrics" – to transform the fortunes of the Oakland Athletics baseball team; the film *Moneyball* tells the story.

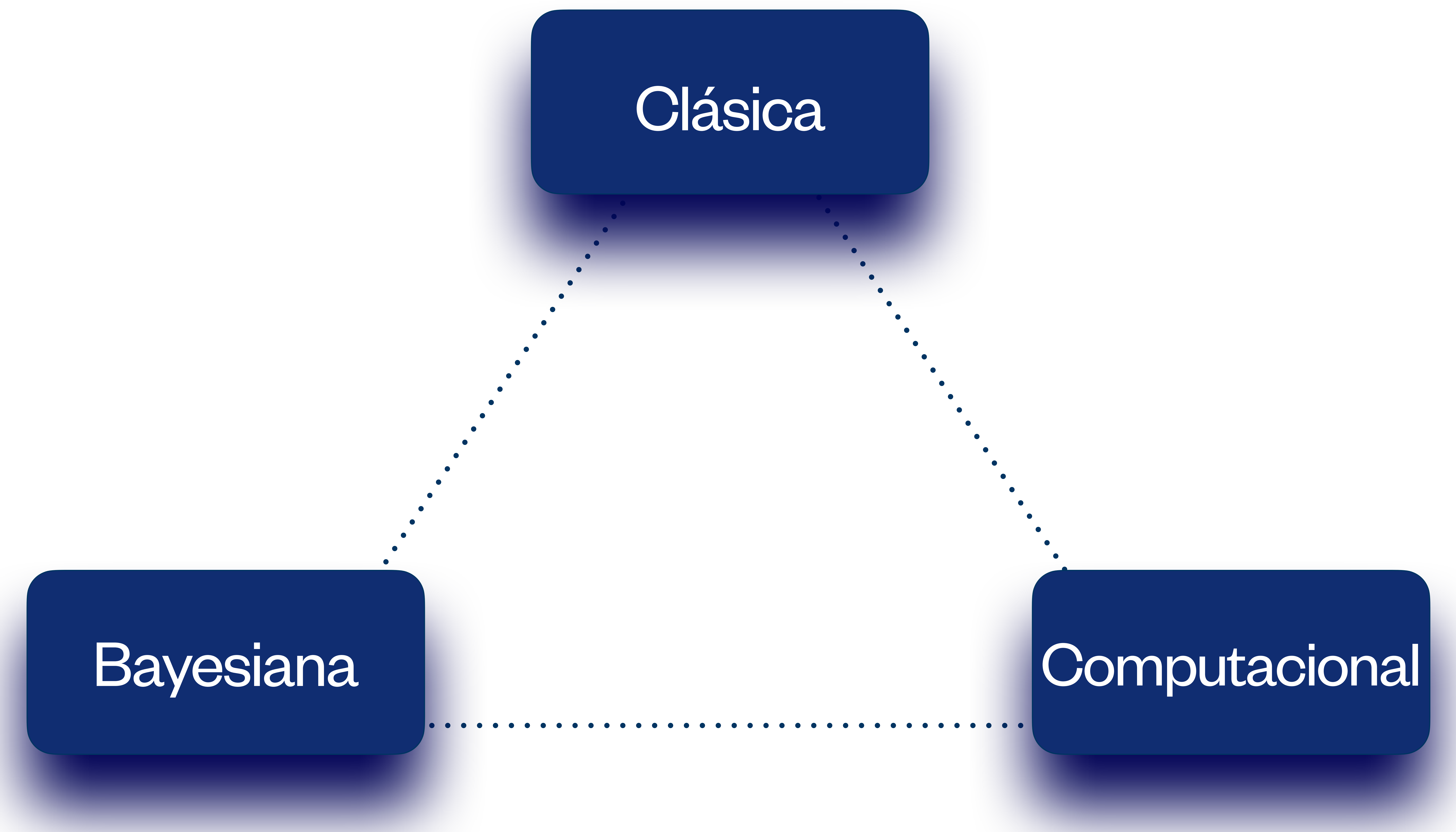


Otros eventos

- El teorema de representación de Bruno de Finetti
- Desarrollo de los métodos MCMC (e.g. Metropolis - Hastings, Gibbs, Hamiltoniano, filtro de partículas, entre muchos otros)
- Inicios de la estadística bayesiana no paramétrica (Ferguson, 1973 y Doksum 1974) y auge a partir del S. XXI
- Y muchos más ...

Enfoques de la estadística

La estadística inferencial moderna



Conceptos iniciales

Población

Muestra

- Aleatorio
- Conglomerados
- Estratificado
- Otros

Recordatorio

Definición 1 (Espacio de probabilidad)

Un espacio de probabilidad es una terna $(\Omega, \mathscr{F}, \mathbb{P})$, donde Ω es un conjunto arbitrario, $\mathscr{F}$ es una σ-álgebra de conjuntos de Ω y $\mathbb{P}$ es una medida de probabilidad.

Definición 2 (Variable aleatoria)

Una variable aleatoria real es una función $X : \Omega \to \mathbb{R}$ tal que para cualquier Boreliano B, se cumple que $X^{-1}(B) \in \mathscr{F}$.

Definición 3 (Espacio parametral)

Al conjunto de valores que el parámetro θ puede tomar se le llama espacio paramétrico (o parametral) y se le denota por Θ.

Muestra aleatoria

Definición 4

Si $X_1, \ldots, X_n$ es un conjunto de **variables aleatorias, independientes e idénticamente distribuidas (v.a.i.i.d.)**, entonces se dice que $X_1, \ldots, X_n$ es una **muestra aleatoria (m.a.)**. De esta forma su densidad conjunta estará dada por

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$