Introduction to Bayesian Statistics

Context, Machinery of Probability, Interpretations of Probability
(part 1)

Teacher: Matilde Trevisani

DEAMS

A.A. 2025/2026

(aggiornato: 2026-03-02)

Agenda (about 3 lectures)

Introduction to Bayesian Statistics

- Context
- Machinery of Probability
- Interpretations of Probability
- Direct and Inverse problems
- Bayes Theorem
- Modern (classical) Statistics
- Bayesian Statistics and subjective probability
- Prior Distributions
- Current Approaches
- Final Notes

Context

Education in Quantitative Methods

- Most training in quantitative methods progresses within a **frequentist framework**
    - Parameters, statistics, parameter estimates, standard errors, hypothesis tests, confidence intervals, etc.
- Bayesian approaches taught as a “special topic”
- Our context

> Bayesian statistics is ***not a branch*** of statistics. It is a way of looking at the ***whole*** of statistics.
>
> -- *Lindley, as quoted in McGrayne (2011, p. 107)*

Why Bayes?

- Philosofical pov
  - Coherent approach to reasoning that better supports modeling, data analysis, inferences
- Practical pov
  - No arguing that a Bayesian approach seems like a lot of unnecessary work for simple models relative to conventional frequentist approaches
  - But can do more with a Bayesian approach
  - Wanna push the limits of modeling and quantitative methods? Go Bayes!

Starting from simple models (in the first part)

- In order to teach something new, we will do it with models that are familiar to us

You will be tempted:

> "Do I really need to use Bayes to estimate a population mean? C'mon, I've been doing that forever and I haven't needed any of this Bayes business!"

- Focus on simple situations
    - **not** because you should use Bayes every time we have simple models, but because:
    - easier to understand what's new here in the Bayesian approach
    - see similarities with and differences from frequentist methods
    - see Bayesian principles in conventional approaches!

You may be frustrated

> This is hard! I thought this was supposed to be easy. Or easier. Or at least tractable.

However,

> Bayesian statistics is difficult in the sense that thinking is difficult.
> -- *Donald Berry*

From Dogmatism to Ecumenism
Or, What is Our Goal?

Statistical Dogmatism (Sectarianism?)

> The theory of inverse probability [Bayes'inference] is founded upon an error, and must be wholly rejected.
>
> -- *Fisher (1925, p. 10)*

> Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit he might have been wrong.
>
> -- *Lindley, comment to Efron (1986)*

Ecumenism in Statistics

> These days the statistician is often asked such questions as “Are you a Bayesian?” “Are you a frequentist?” “Are you a data analyst?” “Are you a designer of experiments?” I will argue that the appropriate answer to all these questions can be (and preferably should be) “yes" ...
>
> -- *Box (1983, p. 51)*

What Now?

- We still lack a clear foundation of statistical inference which is agreed upon.
- This is not only an abstract issue, it has been argued that it is at the root of practical problems in applications of statistics: the issue of hypotheses testing in applied science (see Nuzzo (2014); Goodman (2016), see also Pauli (2018) for an overview of the issue).
- In next class we will see some modern overviews of the scenario on the foundations of statistics (Senn, Efron and Hastie, Royall).

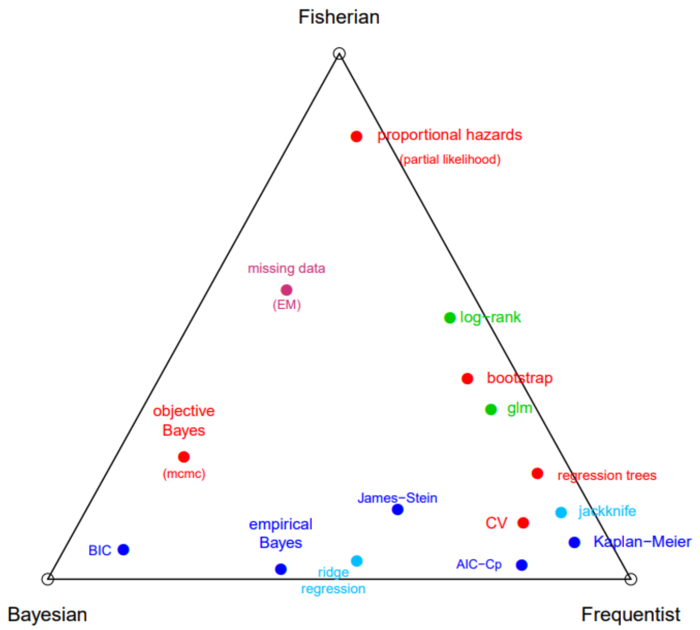
Anticipation: Map of techniques/approaches by Efron and Hastie

Efron and Hastie (2021) map some statistical techniques with respect to their guiding principles and the **relevance of the computational aspect**.

The triangle is not meant to give a complete picture: it shows a **selection of techniques**, as described in the text (*Early Computer-Age Methods*: 15 major topics, **1950s through 1990s**), and the Bayesian, frequentist, and Fisherian influences on them. It shows how these techniques are **based on a mixture of approaches**

Color indicates the **importance of electronic computation** in their development →

It would not be easy to place some XXI century machine learning developments (a philosophically atheistic approach to statistical inference) here (of course they would all be red)



- red, crucial;
- violet, very important;
- green, important;
- light blue, less important;
- blue, negligible.

Some quotes from *Computer age statistical inference, student edition: algorithms, evidence, and data science*

> Statistical inference is an unusually wide-ranging discipline, located as it is at the triple-point of mathematics, empirical science, and philosophy. The discipline can be said to date from 1763, with the publication of Bayes' rule. The most recent quarter of this 250-year history—from the 1950s to the present—is the **"computer age"**, the time when computation, the traditional bottleneck of statistical applications, became faster …. The book is an examination of how statistics has evolved over the past sixty years.
>
> The role of electronic computation is central to our story. This doesn't mean that every advance was computer-related. [But] Almost all topics in twenty-first-century statistics are now computer-dependent.
>
> Very broadly speaking, algorithms are what statisticians do while inference says why they do them.
>
> A particularly energetic brand of the statistical enterprise has flourished in the new century, **data science**, emphasizing algorithmic thinking rather than its inferential justification.

Mixing

The good note is that the factions are no more: to some extent at least, statisticians are willing on taking what is relevant from each approach.

> It seems quite clear that both Bayesian and frequentist methodology are here to stay, and that we should not expect either to disappear in the future. ... Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference.
>
> -- *Bayarri and Berger (2004)*

- In practice this has meant that many Bayesians now consider it reasonable to assess model adequacy (which is incoherent with viewing (posterior distribution on) models as beliefs, as beliefs cannot be wrong)

- Frequentist properties of Bayesian procedures are studied. See hybridization

- Conformal Prediction: a new framework for uncertainty quantification (prediction) with frequentist guarantee, that can be applied to Bayesian Inference $\rightarrow$ Bayesian Conformal Prediction

    - CP can be applied on top of a derived posterior predictive distribution (PPD) to get frequentist coverage guarantees

Today

- On pragmatic grounds, it is reasonable to use whatever approach is best suited for the situation at hand, this is the most common attitude among applied statisticians.

> "Pure" Bayes, "pure" frequentist, "pure" any statistical philosophy, pairs nicely with Port, but when you leave port for the high seas of applications, some degree of impurity is usually necessary. Consequently, statisticians who engage in important studies use their paradigm as an aid to navigation, not as a straightjacket. The goal is to do a good job, and one can't be (too) doctrinaire
>
> -- *Tom Louis, 2019*

- It is also reasonable to interpret Bayesian techniques as modelling techniques rather than a philosophical stance (thus disconnecting it from the subjective interpretation), in this sense the role of the prior can be downplayed, from a source of information to a regularization device (a part of a model).

- Whatever attitude you will take, keep in mind, however, that the Bayesian approach is the only correct one (on a probabilistic reasoning ground) and all other procedures are justified only as approximations of the Bayesian ones.

Machinery of Probability

Axioms of Probability

Axiomatic Definition of Probability (Kolmogorov's Axioms)

Start with a single (discrete) variable X with a finite set of mutually exclusive and exhaustive possible values, $(x_1, \ldots, x_J)$

Probability mass function (**pmf**) specifies the probability for each state, $Prob(X = x_j) = P(X = x_j) = p(x_j)$:

1. $p(x_j) \geq 0$ (non-negativity)
2. $P(X = x_j \, or \, X = x_m) = p(x_j) + p(x_m)$ (additivity)
3. $P(\cup_{j=1}^{J} x_j) = P(\Omega) = 1$ (unit-measure)

Joint, Marginal, and Conditional Distributions

Joint Distributions

- Discrete random variables X and Y
- $P(X = x_j, Y = y_k) = p(x_j, y_k)$ is the ***joint probability*** that X takes on the value x_j ($j = 1, \ldots, J$) and Y takes on the value y_k ($k = 1, \ldots, K$)
- $0 \leq p(x_j, y_k)$ (≤ 1)
- $\sum_j \sum_k p(x_j, y_k) = 1$

Marginal Distributions

- Discrete random variables X and Y
- $P(X = x_j, Y = y_k) = p(x_j, y_k)$ is the joint probability that X takes on the value x_j ($j = 1, \ldots, J$) and Y takes on the value y_k ($k = 1, \ldots, K$)
- The **marginal** probability of X taking on a value x_j is the sum over all possible joint probabilities involving that value of X

$$p(x_j) = \sum_{k=1}^{K} p(x_j, y_k)$$

- The marginal distribution of X is simply the probability distribution of X averaged or marginalized over the information relative to Y
- This can be viewed as a *marginalization* over Y or *integration* of the (noisy variable) Y

Conditional Distributions

- Discrete random variables X and Y
- $P(X = x_j \mid Y = y_k) = p(x_j \mid y_k)$ is the probability that X takes on the value x_j ***given that*** Y takes on the value y_k
  - $p(cancer \mid age, family\,history, smoking\,status)$
  - $p(temperature\,today \mid geographical\,location, temperature\,yesterday)$
- Vitally important in specifying models
- Alternative conception is ***as if***
  - conditional probability is defined as:

$$p(x_j \mid y_k) := \frac{p(x_j, y_k)}{p(y_k)}$$

i.e., the probability of x_j (and y_k) given y_k is the probability that x_j and y_k jointly occur, 'renormalized' under the assumption that y_k occurred, i.e., is the probability of x_j as if y_k has probability 1

Kolmogorov's Axioms for conditional distribuzions

- Hold for conditional probabilities as long as what is on the right of the conditioning bar is held constant

1. $0 \leq p(x_j \mid y_k)$ (≤ 1) for each value of y_k
2. $P(X = x_j \, or \, X = x_m \mid y_k) = p(x_j \mid y_k) + p(x_m \mid y_k)$ for each value of y_k
3. $\sum_{j=1}^{J}(x_j \mid y_k) = 1$ for each value of y_k

Connecting Joint, Marginal, and Conditional Distributions

Joint Distributions From Marginal & Conditional Distributions

- Discrete random variables X and Y
- $p(y_k)$ is the (marginal) probability that $Y = y_k$
- $p(x_j \mid y_k)$ is the (conditional) probability that $X = x_j$ given that $Y = y_k$
- The joint probability of a particular value of X and a particular value of Y is the product of the conditional probability for X given that value of Y and the marginal probability for that value of Y

$$p(x_j, y_k) = p(x_j \mid y_k)p(y_k)$$

(the joint probability follows from the conditional probability definition)

Connecting Joint, Marginal, and Conditional Distributions

Marginal Distributions From Conditional Distributions

- Discrete random variables X and Y
- $p(x_j \mid y_k)$ is the (conditional) probability that $X = x_j$ given that $Y = y_k$
- The marginal probability of X is the sum of its conditional probabilities given all possible values of Y, where each conditional probability is weighted by the associated marginal probability for that value of Y

$$p(x_j) = \sum_k p(x_j, y_k) = \sum_k p(x_j \mid y_k) p(y_k)$$

- Known as the Theorem of Total Probability

Bayes' Theorem

Deriving Bayes' Theorem

$$p(x, y) = p(x \mid y)p(y)$$

The joint probability for x and y is the product of the conditional probability for x given y and the marginal probability for y

$$p(x, y) = p(y \mid x)p(x)$$

Reversing the roles of x and y

$$p(x, y) = p(x, y)$$

Tautology

$$p(y \mid x)p(x) = p(x \mid y)p(y)$$

Substitution

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)}$$

Division *(that is the Bayes' Theorem!)*

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}$$

Works both ways

Probability for Continuous (Random) Variables

Probability for Continuous Variables

- Probability defined over ranges of values
- The full range is referred to as the ***support***
  - E.g., the support for a normal (Gaussian) distribution is $(-\infty, \infty)$
- ***Probability density functions*** (**pdf**) rather than probability mass functions (pmf)
  - E.g., for $x \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

- All the same concepts hold (joint, marginal, independence, conditional independence, Bayes' theorem)
- Mathematics a bit different

Probability for Continuous Variables

$p(X), p(Y)$ defined over ranges of values; $\sum \rightarrow \int$

Discrete	**Continuous**
$\sum_j p(x_j) = 1$	$\int_x p(x_j) = 1$
$\sum_j \sum_k p(x_j, y_k) = 1$	$\int_x \int_y p(x_j, y_k) = 1$
$p(x_j) = \sum_k p(x_j, y_k)$	$p(x_j) = \int_y p(x_j, y_k)$
$p(x_j) = \sum_k p(x_j \mid y_k) p(y_k)$	$p(x_j) = \int_y p(x_j \mid y_k) p(y_k)$

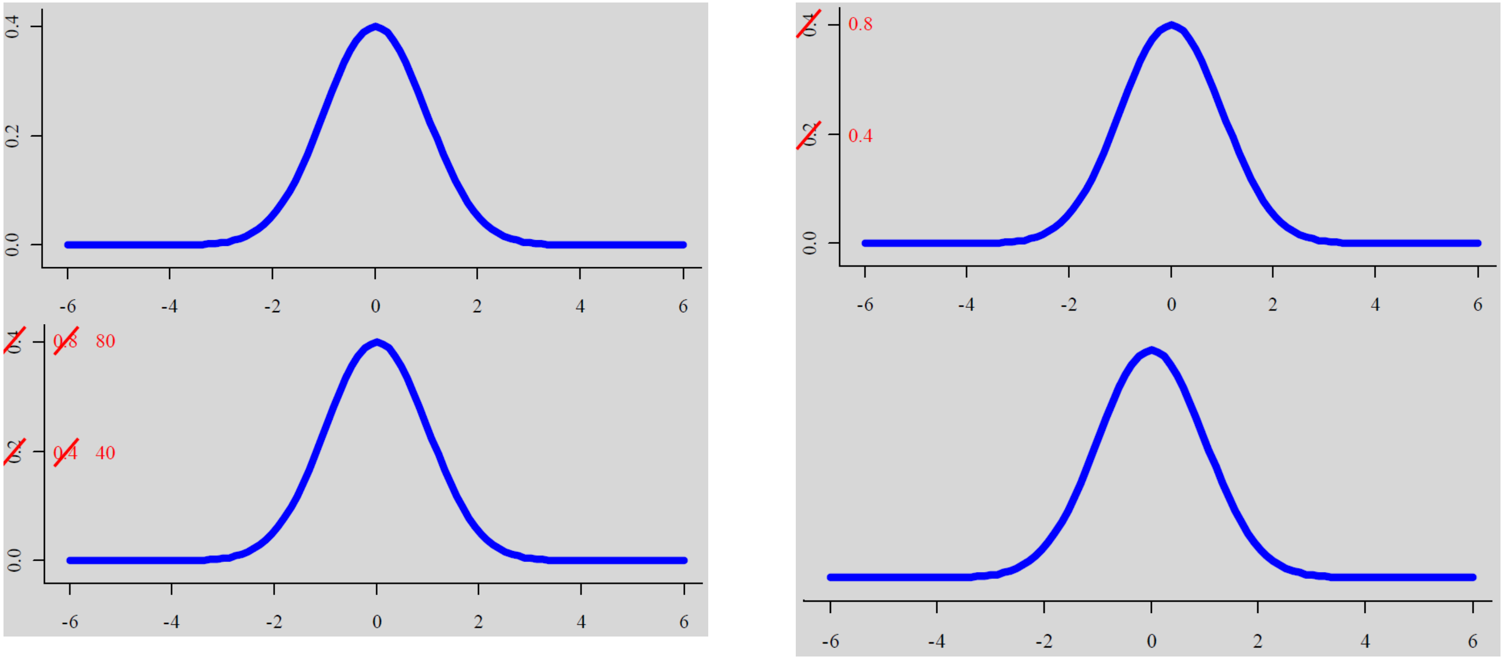
Additional Terminology

Additional Terminology

- A distribution is ***proper*** if it integrates (sums) to a finite quantity
- A distribution is ***normalized*** if it integrates (sums) to 1
- The ***kernel*** of the distribution is the form where any factors that are not functions of any of the variables in the domain are omitted.
- The ***normalizing constant*** serves to normalize the distribution (it is an adjustment)
    - Dropping the normalizing constant serves to render the function ***proportional*** to the probability

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \propto \frac{1}{\sigma}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

Normalized Distribution normalizzata and illustration of proportionality



Normalized pdf (top left) and various proportional kernels

Summary

Summary

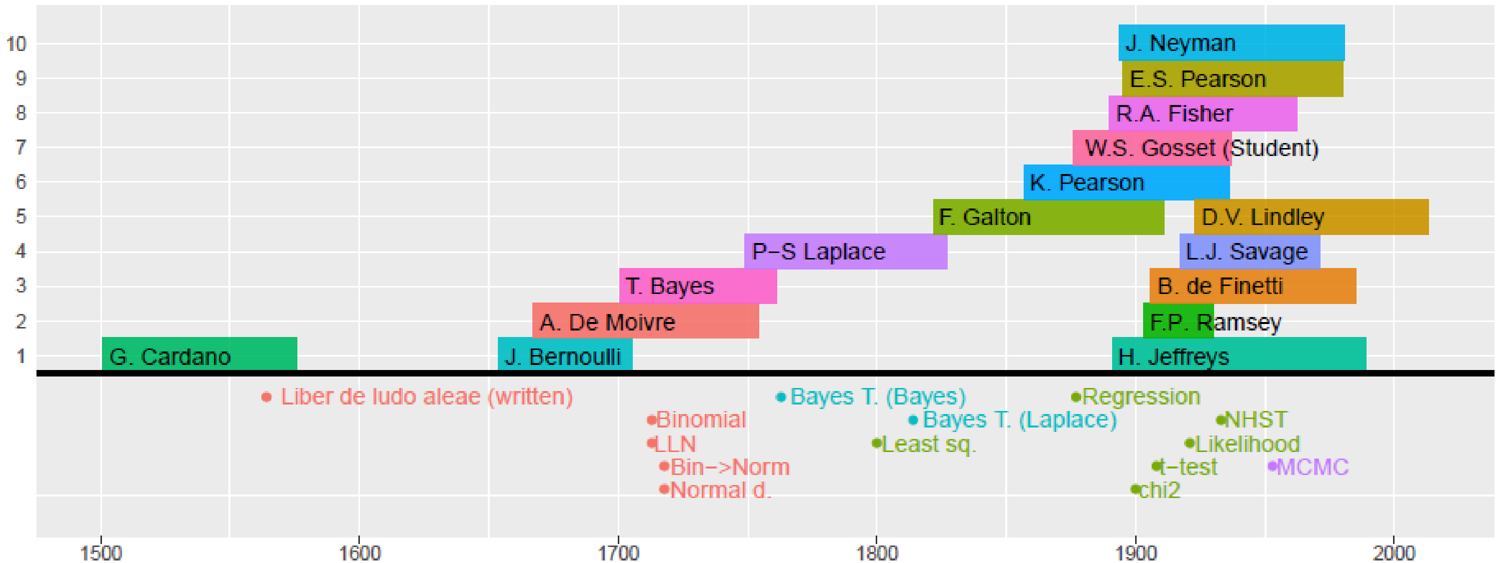
- Axioms of probability
- Joint, marginal, and conditional distributions
- Bayes’ theorem
- Continuous random variables
- Proportionality

Interpretations of Probability

> It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there as seldom been such complete disagreement and breakdown of comunication since Tower of Babel.
>
> -- *Savage (1954)*

Starting from Probability and History

Let's start by introducing Bayesian Statistics from considering its development from probability calculus and its history compared to the one of the concurrent (so called) Classical Statistics.



Historical excursus (Figure courtesy of Francesco Pauli)

Three main definitions of probability

Statistics is the study of **uncertainty**. How is it measured? And how do we make decisions in the presence of it? One way to deal with uncertainty is to think about **probability**.

There are **three main paradigms** according to which we can define probabilities.

- Classical
- Frequentist
- Bayesian

Classical/Frequentist Interpretations of Probability

(Chance, Objectivist, Aleatory)

Scope of Application

> The rational concept of probability, which is the only basis of probability calculus, applies only to problems in which **either the same event repeats itself again and again**, or **a great number of uniform elements are involved at the same time** ...
>
> -- *Von Mises (1957, p. 11)*

Classical approach

Suppose we throw a six-sided die and ask ourselves what is the probability that the die shows a four.

In the classical paradigm, **the possible outcomes** of a *random experiment* **are all equally probable**. So, in the case of a fair die, in which the possible outcomes are six and equally probable, the probability that the die shows a four is one in six (probability of elementary events).

First definition of probability, based on **symmetry**

$$\Pr(\text{event}) = \frac{\#\text{ favourable outcomes}}{\#\text{ possible outcomes}}$$

We could ask a related question, e.g. what is the probability of getting four from the sum of a pair of throws (probability of combinations of elementary events).

$\ldots \rightarrow \Pr(X_1 + X_2 = 4) = \frac{3}{36} = \frac{1}{12}$

Probability born with gambling

Probability calculus is initially developed to study games of chance: developing strategies to win in games was of interest to nobles, who were willing to pay scholars for them.

For example **Galileo** in 1620 wrote a note offering the solution of this issue:

> suppose three dice are thrown and the three numbers obtained added. What is the probability that the total equals 9?

This circumstances help not only because of the money, but also because of the simple structure of the problems involved.

Girolamo **Cardano** (1501-1576)}, wrote the first systematic treatment of probability in 1576: *Liber de ludo aleae*; this, however was not published until 1663. He was a polymath with interests ranging from mathematics to biology. He was also a gambler (and a rumor exists that he did not publish his book on probability because his knowledge gave him an advantage in betting).

Elementary probability

The first examples of probability problems are concerned with simple random mechanisms whose symmetry offered the solution.

> Q: A marble is randomly drawn from an urn containing R red marbles and W white marbles, what is the probability that the marble is red?
> A: By symmetry
>
> $$P(red) = \frac{R}{R+W}$$

Elementary probability and combinatorics

For more "complicated" questions tools were developed to count favourable and unfavourable outcomes.

> Q: We draw a marble from an urn containing R red marbles and W white marbles m times (putting it back in the urn after each draw), what is the probability that r out of m are red?
>
> A: Still by symmetry

$$P(r \text{ red out of } m) = \binom{m}{r}\left(\frac{R}{R+W}\right)^r\left(1-\frac{R}{R+W}\right)^{m-r}$$

Limitations of the classical approach

This approach is appropriate if the outcomes of a random experiment are well-defined and equally probable (e.g., in games of chance).

The probability, calculated according to the classical definition, is also called **a priori**, since it is determined on the basis of a theoretical and non-experimental prediction.

- to be able to apply the formula it is essential to know a priori the values in the numerator and denominator and above all it is necessary to know whether the possible cases are equally possible.

The classical approach is difficult to apply, for example, in these other cases:

- *what is the probability of four coming up if we assume that the die is unfair*,
- *what is the probability of a given adverse event occurring in the administration of a vaccine*, etc.

Frequentist approach

The frequentist definition is based on the observation of an experimental fact and requires imagining the repetition of the random experiment in identical conditions an infinite number of times, so the *limit* value of the relative frequency with which the event occurs in a hypothetical **infinite sequence of repeated trials in the same conditions** corresponds to the probability of the event.

- We can imagine rolling a die an infinite number of times and if the die is fair, about a sixth of the time, we will get a four.
- If out of 10,000 vaccine administrations, a given side effect occurred in only one case, we can define the probability of that adverse event as 1 in 10,000.

From Grounding to Definition

1. Outcomes/results of repetitions vary (differ) in an unpredictable manner
    - *Random variation*
2. *Long-run* regularities emerge
    - Take repeated observations (implicitly assumed independent)
    - S_n = number of times an event (E) occurs out of n observations
    - This relative frequency S_n/n appears to stabilize/converge to some limiting value as $n \to \infty$

- These stable, long-run relative frequencies characterize the situation
- We assign a number, called the ***probability***, $p(E)$, to stand for this long-run frequency

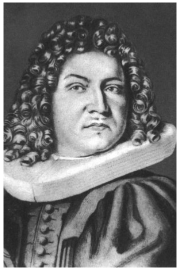
The idea of limiting frequency (in the game of chances)

Still in the context of game of chances it is easy to think of "repeating" events, so the probability of an event materializes as the

> limiting relative frequency of occurrence of the event in a number of repetitions.

This idea was developed and made more precise by Jakob **Bernoulli** (Ars Conjectandi, 1713) and Abraham **De Moivre** (1733) in the **law of large numbers** which links theoretically the probability of an event to the relative frequency in infinite repetitions.

Jakob Bernoulli (Basel 1654-1705), (Jacob, Jacques or James) in *Ars Conjectandi* (1713) discusses the application of probability to gambling. He develops techniques based on combinatorics calculus (and the binomial distribution) and a first version of the law of large numbers.



Abraham De Moivre (1667-1754) in *Laws of Chances* (1718) builds on Bernoulli's (and others) works. One of his main achievements is the formula for the normal distribution and the link between the binomial and the normal distribution.

The law of large numbers

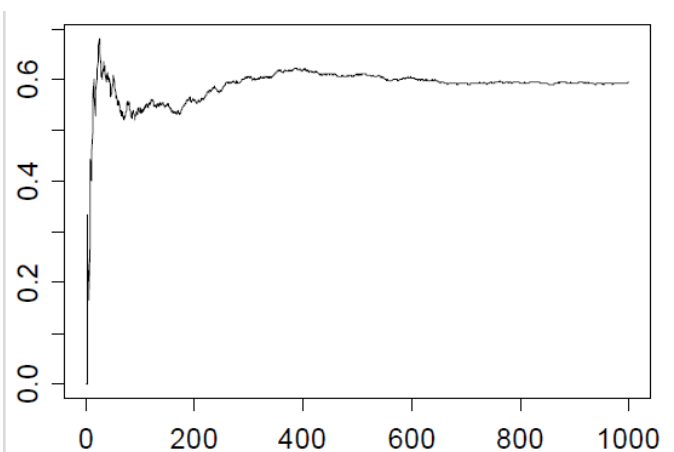
(Strong) Law of large numbers

Let $E_1, \ldots, E_n, \ldots$ be a sequence of independent events such that $P(E_i) = p$ for all i. Let $S_n = \sum_{i=1}^n |E_i|$ be the number of events occurring among the first n. Then

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1.$$

Note that the theorem was already stated, without proof, by Cardano.

Example



Proportion of successes in an unfolding series of Bernoulli variables with $\theta = .6$

Features

- Value of the probability is determined by the (repeatable) situation at hand
    - “Objective”, unique in the sense that it does not depend on the investigator, or any other ancillary circumstances
- Probability is a ***property of the world***
- Defined in situations that are potentially able to be repeated over and over under (essentially) identical conditions
- Only relevant information comes from observing the outcomes of different realizations of the repeatable process
    - Results of flips of the coin
    - Keep checking widgets coming off the line for defects
- Full fruition by Pascal-Fermat correspondence of 1654

When applicable?

Defined in situations that can potentially be repeated over and over again under (essentially) identical conditions.

1) Whether or not a widget coming off a manufacturing assembly line is defective

- Reference to many successive widgets produced by off the line

🆗

2) The life expectancy of university faculty

- Reference to … set of similar university faculty

⚠️

3) Author of *The Federalist papers* - Madison or Hamilton?

- Reference to ?

⛔

Limits of the frequentist approach

Sometimes, therefore, **it is not conceivable** to refer to **a sequence of repeated trials**.

E.g. (other cases),

- If the question is what is the *probability that it will rain tomorrow*, it is not intuitive to imagine an infinite sequence of "tomorrows."
- Or if the question is what is *the probability that the die is rigged*, rolling the die multiple times does not change the result - it is not a repeatable event; even more, there is no random variation, its probability is either 0 or 1.

The frequentist approach aims to be objective in the way it defines probabilities. However, it can encounter deep philosophical issues. Sometimes, **objectivity is merely an illusion**.

Next slides introduce to the **subjective** interpretation of probability.

- It not only overcomes some limits of frequentist interpretation
- It changes the uncertainty nature
  - uncertainty, randomness is no more aleatory (due to the random variation of outcomes, world phenomena)
  - uncertainty is **epistemic** due to the imperfect or limited knowledge of the subject.
  - uncertainty, randomness extends to everything that is not (exactly) known, then, not only world phenomena, though also hypotheses, models, parameters, …
  - uncertainty = *mental uncertainty*, *uncertainty of thought*

A long time is required to arrive at a new conception of uncertainty and a formulation of subjective probability.

Its origins can be traced back to Bayes and Laplace (second half of XVIII century). The new probability fully comes to light only with De Finetti (after 1950).

Subjective Interpretations of Probability

(Degree of Belief, Epistemic)

In Simple Words

> By degree of probability we really mean, or ought to mean, **degree of belief** ....Probability then, refers to and implies belief, more or less, and belief is but another name for imperfect knowledge, or it may be, expresses the **mind in a state of imperfect knowledge**.
> -- *de Morgan (1847, p. 172-173; see Kyburg & Smokler, 1964)*

Quantify Beliefs, Uncertainty

- Want to quantify different beliefs
    - *“I think the coin is just as likely to land on heads as tails.”*
    - *“I am pretty sure that van Buren was the 8th president of the US.”*
    - *“I am almost positive Madison wrote The Federalist papers.”*
    - *“I am not confident it will rain tomorrow.”*
- We are uncertain about something
    - Outcomes/results vary (differ) in an unpredictable manner
    - Events may or may not occur
    - Latent patterns are unobservables
- We assign a number, called the ***probability***, to stand for our belief or amount of uncertainty

Expression of Uncertainty Regarding State of Knowledge

- Probability is determined by the individual's belief
    - Subjective, unique in the sense that they depend on the investigator, or other ancillary circumstances
    - property of an **individual's state of knowledge**, including model of the situation
    - Different individuals in different contexts may have different opinions, backgrounds of information, and degrees of uncertainty.
- An expression of uncertainty
    - $p(Rain) = 0, p(Rain) = .1, p(Rain) = .5, p(Rain) = 1$
- Full fruition by Pascal, 1662

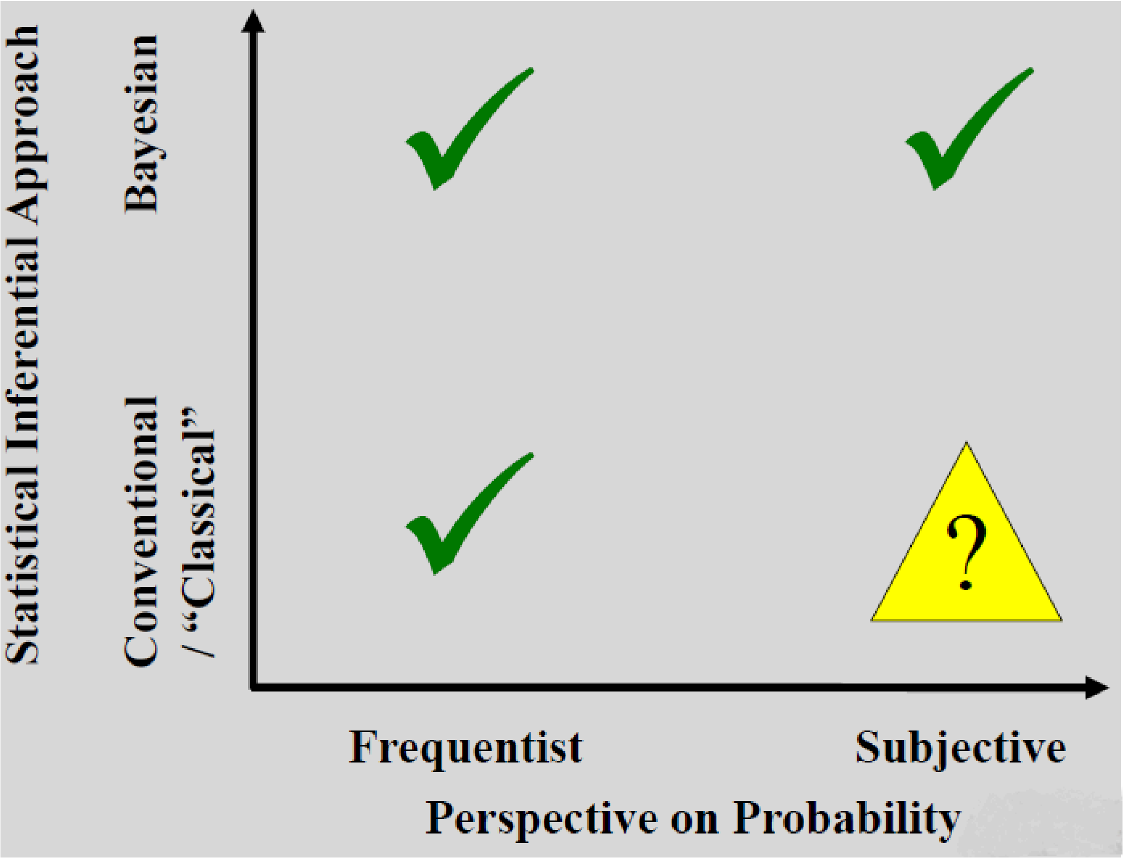
When applicable

- Applies in contexts of repeatable events and nonrepeatable events

1. Whether or not a widget coming off a manufacturing assembly line is defective
   🆗
2. The life expectancy of university faculty
   🆗
3. Author of The *Federalist papers* - Madison or Hamilton?
   🆗

Inferential Frame

Frequentist <-> Subjective
Inferenza Conventional/"Classical" <-> Bayesian



Still on the Bayesian/Subjective Approach

The frequentist approach to inference is based on an objective approach to defining probability. However, it can run into some profound philosophical issues. Sometimes, objectivity is merely an illusion.

The Bayesian approach, on the other hand, is a **subjective approach to probability**. Your probability represents **your personal perspective** - it is your measure of uncertainty and takes into account what you know about a particular problem. Probabilities defined according to the Bayesian approach adhere to the standard rules of probability, thus forming a **coherent system**.

We can quantify probabilities by thinking about what constitutes a **fair bet**.

Subjective definition of probability by de Finetti

The probability is not an objective property of a phenomenon but rather the opinion of a person.

Definition of Subjective probability

> The probability of an event is, for an individual, his degree of belief on the event.

If the probability is a subjective degree of belief, it depends on the information which is subjectively available, and it is also clear that **by random we mean not known for lack of information**.

Bruno de Finetti (1906-1985), Italian probabilist and actuary (for Generali) proposes the subjective definition of probability and the coherence framework, based on the bet interpretation (see *Theory of probability* (1970)).

In *Theory of probability* he wrote

> Probability does not exist

A Fair Bet

For example, we ask ourselves what the probability is that it will rain tomorrow. Therefore, we can ask you what bet you would be willing to make, considering that this bet is fair.

Suppose you are willing to bet that if it rains tomorrow, you win 4 euros, while if it does not rain, you lose 1 euro. Which is equivalent to defining a probability (*odds*) of 4 to 1 that it will not rain tomorrow. If you believe this bet is fair, you should be willing to take it in both directions, which would also mean that if it rains, you lose 4 euros, while if it does not rain, you win 1 euro.

If you consider both bets to be fair, then you are defining the probability of rain as $\frac{1}{1+4} = \frac{1}{5}$ which equals 0.2. We can verify that this framework is correct by calculating your expected return.

In the first bet, your expected return is 0 since you win 4 with a probability of 0.2 and lose 1 with a probability of 0.8. Thus, it is a fair bet. In the second case, you win 1 with a probability of 0.8 and lose 4 with a probability of 0.2, and this expected value is (obviously) also 0. Therefore, the two bets are balanced.

A Fair Bet (cont.)

Therefore, you can use this betting scheme to think about what your personal probability is, based on the bets you would be willing to make.

If you would accept a bet with odds of 1000 to 1 in your favor, but would not accept a bet with odds of 1000 to 1 against you, your personal probability lies somewhere in between.

More formally,

we define the probability of an event $P(E)$ as:

- The price you would pay in exchange for a return of 1 if the event occurs, and 0 otherwise.
- The price you would accept in exchange for having to pay 1 if the event occurs, and 0 otherwise.

In other words, once $P(E)$ is established, you would buy or sell the **random** amount $|E|$ in exchange for $P(E)$.

$$\text{probability} = \text{the amount considered fair to pay for a bet with a value of 1}$$

"To consider fair" essentially means to regard it as indifferent whether you take the role of the bettor or the bookmaker.

A bet is fair if the expected value for each player is 0.

Coherence of Bets

Finally, the concept of **coherence**. The probabilities assigned must satisfy all the axioms of probability; otherwise, one may be inconsistent. That is, suppose you have assigned probabilities to a set of events, but they do not satisfy the axioms of probability. In that case, you can construct a series of bets where the loss (or gain) is certain.

A combination of bets that leads to a certain loss is called a ***Dutch book***. If you follow all the rules of probability, you can be sure that you are consistent.

Dutch book argument

Assume I state the following probabilities

$$P(A) = 0.2; \quad P(B) = 0.3; \quad P(A \cap B) = 0.6$$

Then I have to accept the following bets by an opponent

- he pays 0.2 to receive 1 if A
- he pays 0.3 to receive 1 if B
- he pays 0.4 to receive 1 if $\overline{A \cap B} = \bar{A} \cup \bar{B}$

So I receive 0.9, how much do I pay? It depends on what happens:

- if $AB \rightarrow -2$
- if $A\bar{B} \rightarrow -2$
- if $\bar{A}B \rightarrow -2$
- if $\bar{A}\bar{B} \rightarrow -1$

overall my net outcome is either -0.1 or -1.1, I incur in a sure loss.

Marginal Notes

Frequentist Evaluations in Bayesian Statistics

Bayesian theory considers both epistemic and aleatory probabilities.

"Frequentist" evaluations focus on "frequentist" properties given the model and the random repetition of an observation.

- Asymptotic Consistency
- Correctness:
  - Less crucial in Bayesian inference (as small errors are more important)
- Efficiency:
  - A small quadratic error is desirable.
  - Other utility/cost functions may also be considered.
- Calibration
  - A posterior interval at α% contains the true value in α% of cases.
  - A predictive interval at α% contains future true values in α% of cases.
  - *Approximate* calibration with shorter intervals for likely true values is more important than exact calibration with longer intervals for all possible values.

See Mixing

References

Bayarri, M. J. and J. O. Berger (2004). "The interplay of Bayesian and frequentist analysis".

Box, G. (1983). " An apology for ecumenism in statistics". In: *Scientific inference, data analysis, and robustness*. Ed. by G. Box, T. Leonardi and G. Wu. San Diego: Academic Press, pp. 51-84.

Efron, B. (1986). "Why Isn't Everyone a Bayesian?" In: *The American Statistician* 40.1, pp. 1-5. ISSN: 00031305, 15372731. URL: http://www.jstor.org/stable/2683105 (visited on mar. 03, 2025).

Efron, B. and T. Hastie (2021). *Computer age statistical inference, student edition: algorithms, evidence, and data science*. Vol. 6. Cambridge University Press.

Fisher, R. (1925). *Statistical methods for research workers*. Oliver and Boyd, Edinburgh.

Goodman, S. N. (2016). "Aligning statistical and scientific reasoning". In: *Science* 352 (6290), pp. 1180-1181.

McGrayne, S. B. (2011). *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy*. Yale University Press.

Nuzzo, R. (2014). "Scientific method: Statistical errors". In: *Nature* 506.7487, pp. 150-152. ISSN: 0028-0836.

Pauli, F. (2018). "The p-value Case, a Review of the Debate: Issues and Plausible Remedies". In: *Studies in Theoretical and Applied Statistics*. Ed. by C. Perna, M. Pratesi and A. Ruiz-Gazen. Cham: Springer International Publishing, pp. 95-104. ISBN: 978-3-319-73906-9.

Savage, L. (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

References (continue)

Von Mises, R. (1957). *Probability, Statistics, and Truth*. Allen and Unwin. URL: https://books.google.it/books?id=KSE4AAAAMAAJ.