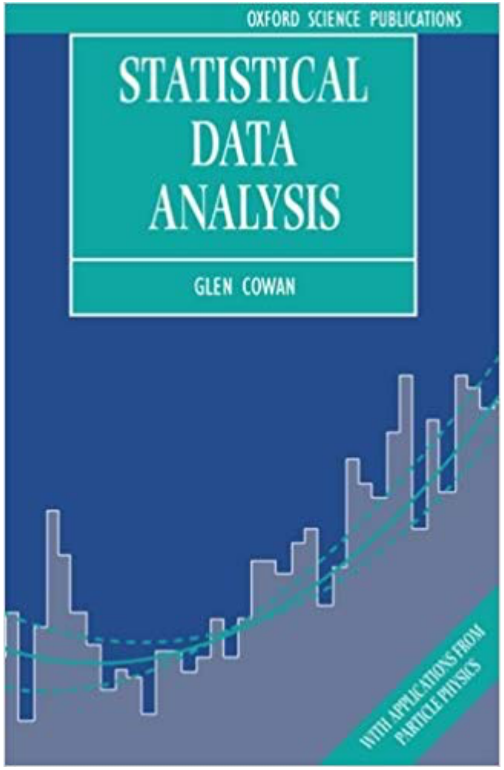
The language of random variables

Reference Textbook

G. Cowan, Statistical Data Analysis,
Oxford Science Publications (1998)

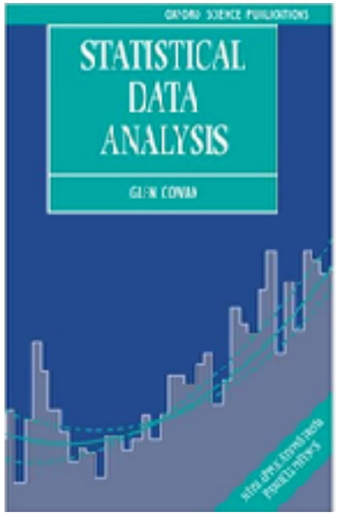


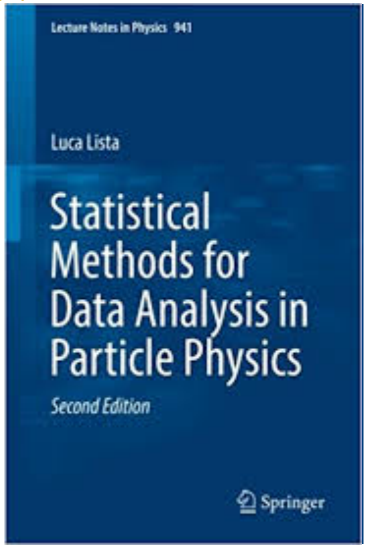
Support Material

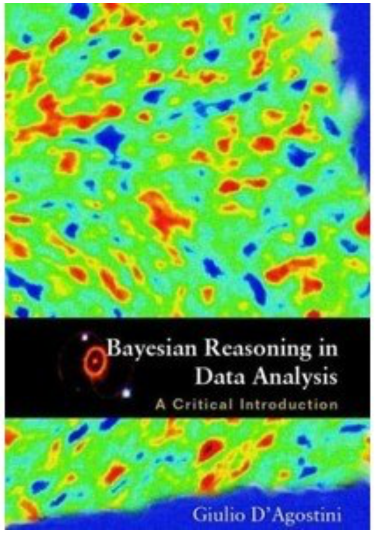
G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford, 1998.
PDG

L. Lista Statistical methods for Data Analysis, 2nd Ed. Springer, 2018

G. Cowan PDG
http://pdg.lbl.gov/2017/reviews/rpp2017-rev-statistics.pdf

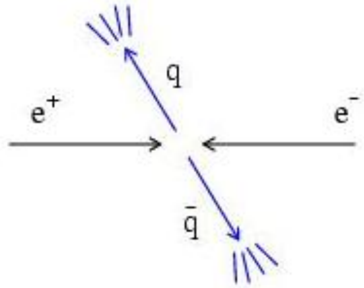






C. Bini, Lezioni di statistica per la fisica sperimentale, Edizioni Nuova Cultura
G. D'Agostini, Bayesian reasoning in data analysis, World Scientific (https://cds.cern.ch/record/395902/files/CERN-99-03.pdf)

Data analysis in particle physics



Observe events of a certain type

Measure characteristics of each event (particle momenta, number of muons, energy of jets,...)

Theories (e.g. SM) predict distributions of these properties up to free parameters, e.g., α, G_F, M_Z, α_s, m_H, ...

Some tasks of data analysis:

Estimate (measure) the parameters;

Quantify the uncertainty of the parameter estimates;

Test the extent to which the predictions of a theory are in agreement with the data.

Dealing with uncertainty

In particle physics there are various elements of uncertainty:

theory is not deterministic

quantum mechanics

random measurement errors

present even without quantum effects

things we could know in principle but don't

e.g. from limitations of cost, time, ...

We can quantify the uncertainty using PROBABILITY

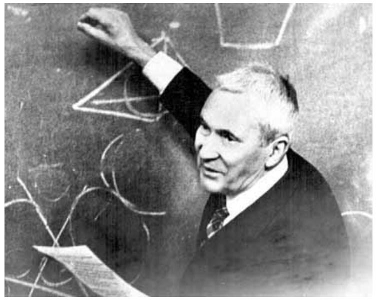
A definition of probability

Consider a set S with subsets A, B, ...

For all $A \subset S, P(A) \geq 0$

$$P(S) = 1$$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$



Kolmogorov axioms (1933)

From these axioms we can derive further properties, e.g.

$$P(\overline{A}) = 1 - P(A)$$

$$P(A \cup \overline{A}) = 1$$

$$P(\emptyset) = 0$$

if $A \subset B$, then $P(A) \leq P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional probability, independence

Also define conditional probability of *A* given *B* (with $P(B) \neq 0$):

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

E.g. rolling dice: $P(n < 3 \,|\, n \text{ even}) = \frac{P((n<3) \cap n \text{ even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$

Subsets *A*, *B* independent if: $P(A \cap B) = P(A)P(B)$

If *A*, *B* independent, $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$

N.B. do not confuse with disjoint subsets, i.e., $A \cap B = \emptyset$

Interpretation of probability

I. Relative frequency

A, B, ... are outcomes of a repeatable experiment

$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n}$$

cf. quantum mechanics, particle scattering, radioactive decay...

II. Subjective probability

A, B, ... are hypotheses (statements that are true or false)

$$P(A) = \text{degree of belief that } A \text{ is true}$$

- Both interpretations consistent with Kolmogorov axioms.
- In particle physics frequency interpretation often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena:

  systematic uncertainties, probability that Higgs boson exists,...

Bayes' theorem

From the definition of conditional probability we have,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

but $P(A \cap B) = P(B \cap A)$, so

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

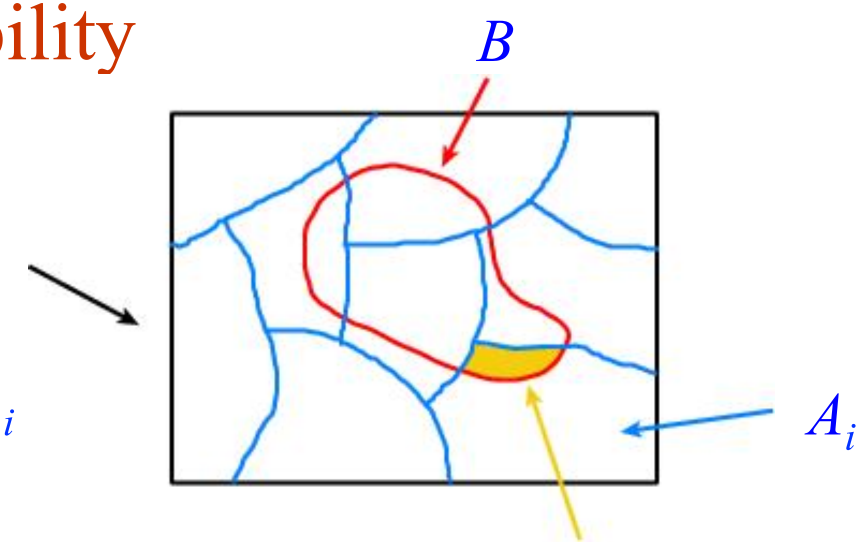
Bayes' theorem

First published (posthumously) by the Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370; reprinted in Biometrika, **45** (1958) 293.

The law of total probability

Consider a subset B of the sample space S,



divided into disjoint subsets A_i such that $\cup_i A_i = S$,

$\rightarrow \quad B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i),$

$\rightarrow \quad P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$

$\rightarrow \quad P(B) = \sum_i P(B|A_i) P(A_i)$ law of total probability

Bayes' theorem becomes

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

An example using Bayes' theorem

Suppose the probability (for anyone) to have a disease D is:

$$P(\text{D}) = 0.001$$
$$P(\text{no D}) = 0.999$$

← prior probabilities, i.e., before any test carried out

Consider a test for the disease: result is + or −

$$P(+|\text{D}) = 0.98$$
$$P(-|\text{D}) = 0.02$$

← probabilities to (in)correctly identify a person with the disease

$$P(+|\text{no D}) = 0.03$$
$$P(-|\text{no D}) = 0.97$$

← probabilities to (in)correctly identify a healthy person

Suppose your result is +. How worried should you be?

An example using Bayes' theorem

Suppose the probability (for anyone) to have a disease D is:

$$P(\mathrm{D}) = 0.001$$
$$P(\text{no D}) = 0.999$$

← prior probabilities, i.e., before any test carried out

Consider a test for the disease: result is + or −

$$P(+|\mathrm{D}) = 0.98$$
$$P(-|\mathrm{D}) = 0.02$$

← probabilities to (in)correctly identify a person with the disease

$$P(+|\text{no D}) = 0.03$$
$$P(-|\text{no D}) = 0.97$$

← probabilities to (in)correctly identify a healthy person

Suppose your result is +. How worried should you be?

Bayes' theorem example (cont.)

The probability to have the disease given a + result is

$$p(\mathrm{D}|+) = \frac{P(+|\mathrm{D})P(\mathrm{D})}{P(+|\mathrm{D})P(\mathrm{D}) + P(+|\text{no D})P(\text{no D})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$$= 0.032$$ ← posterior probability

i.e. you're probably OK!

Your viewpoint: my degree of belief that I have the disease is 3.2%.

P(D | +) and P(+ | D) can be very different (depends on the prior!)

Frequentist Statistics − general philosophy

In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations (shorthand: $\vec{x}$).

Probability = limiting frequency

Probabilities such as

P (Higgs boson exists),
$P\,(0.117 < \alpha_s < 0.121)$,

etc. are either 0 or 1, but we don't know which.

The tools of frequentist statistics tell us what to expect, under the assumption of certain probabilities, about hypothetical repeated observations.

The preferred theories (models, hypotheses, ...) are those for which our observations would be considered 'usual'.

Probability definition (freqentist)

- A bit more formal definition of probability:
- Law of large numbers:

$$P(E) = p \quad \text{if} \quad \frac{N(E)}{N} \xrightarrow{P} p$$

- i.e.: $\forall \varepsilon \quad \lim_{N\to\infty} P\left(\left|\frac{N(E)}{N} - p\right| < \varepsilon\right) = 1$

… isn’t it a circular definition?

$$\langle x \rangle = \frac{1+2+3+4+5+6}{6} = 3.5\,.$$

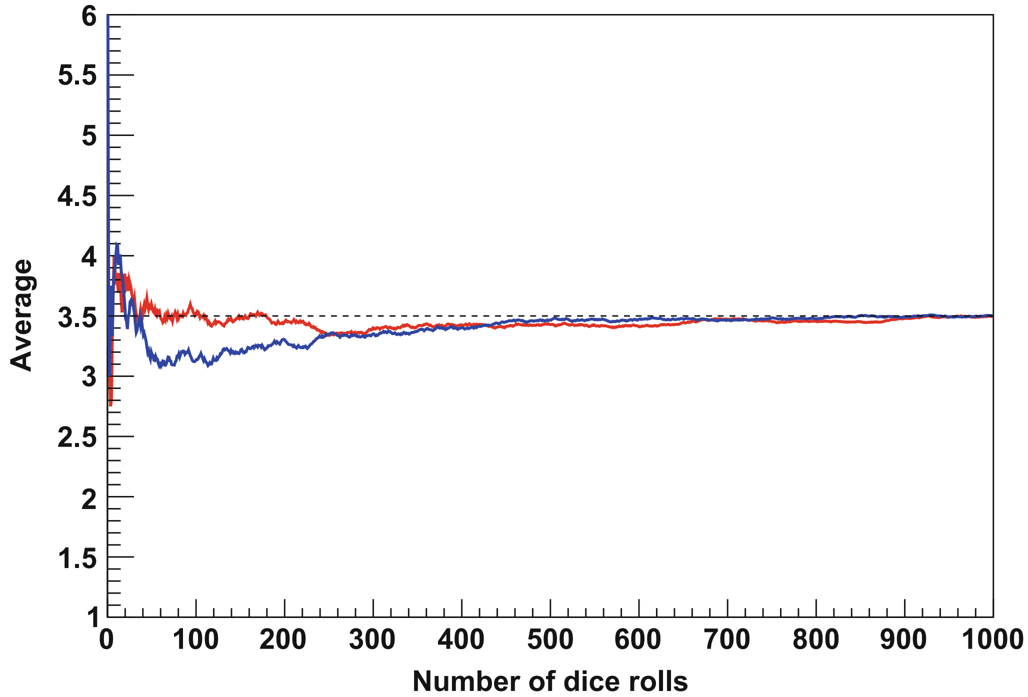


Fig. 1.10 An illustration of the law of large numbers using a computer simulation of die rolls. The average of the first N out of 1000 random extraction is reported as a function of N. 1000 extractions have been repeated twice (*red and blue lines*) with independent random extractions

Frequentist probability definition:

$$P(E) = p \quad \text{if} \quad \forall\, \varepsilon \lim_{N\to\infty} P\left(\left|\frac{N(E)}{N} - p\right| < \varepsilon\right)$$

(somewhat a circular definition: a probability in terms of another probability)

Bayesian Statistics − general philosophy

In Bayesian statistics, use subjective probability for hypotheses:

probability of the data assuming hypothesis H (the likelihood)

prior probability, i.e., before seeing the data

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

posterior probability, i.e., after seeing the data

normalization involves sum over all possible hypotheses

Bayes' theorem has an "if-then" character: If your prior probabilities were $\pi(H)$, then it says how these probabilities should change in the light of the data.

No general prescription for priors (subjective!)

Definition of probability

- There are two main different definitions of the concept of probability
- Frequentist
  - **Probability** is the ratio of the number of occurrences of an event to the total number of experiments, in the *limit* of very large number of **repeatable** experiments.
  - Can only be applied to a specific classes of events (repeatable experiments)
  - Meaningless to state: "*probability that the lightest SuSy particle's mass is less tha 1 TeV*"
- Bayesian
  - **Probability** measures someone's the degree of belief that something is or will be true: would you bet?
  - Can be applied to most of unknown events (past, present, future):
    - "*Probability that Velociraptors hunted in groups*"
    - *"Probability that S.S.C Naples will win next championship"*

or Rome, Juventus etc..

Problems with probability definitions

- Frequentist probability is, to some extent, circularly defined
  - A phenomenon can be proven to be random (i.e.: obeying laws of statistics) only if we observe infinite cases
  - F.James et al.: *"this definition is not very appealing to a mathematician, since it is based on experimentation, and, in fact, implies unrealizable experiments ($N \rightarrow \infty$)"*. But a physicist can take this with some pragmatism
  - A frequentist model can be justified by details of poorly predictable underlying physical phenomena
    - Deterministic dynamic with instability (chaos theory, …)
    - Quantum Mechanics is intrinsically probabilistic…!
  - A school of statisticians state that Bayesian statistics is a more natural and fundamental concept, and frequentist statistic is just a special sub-case
- On the other hand, Bayesian statistics is subjectivity by definition, which is unpleasant for scientific applications.
  - Bayesian reply that it is actually inter-subjective, i.e.: the real essence of learning and knowing physical laws…
- Frequentist approach is preferred by the large fraction of physicists (probably the majority, but Bayesian statistics is getting more and more popular in many application, also thanks to its easier application in many cases

Frequentist vs Bayesian

- The Bayesian infers from the data using priors

posterior $P(H|x) \approx P(x|H) \cdot P(H)$

- Priors is a science on its own. Are they objective? Are they subjective?

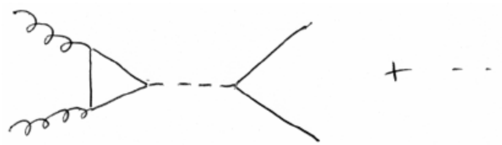
- The Frequentist calculates the probability of an hypothesis to be inferred from the data based on a large set of hypothetical experiments Ideally, the frequentist does not need priors, or any degree of belief while the Baseian posterior based inference is a "Degree of Belief".

- However, NPs (Systematic) inject a Bayesian flavour to any Frequentist analysis

Theory ↔ Statistics ↔ Experiment

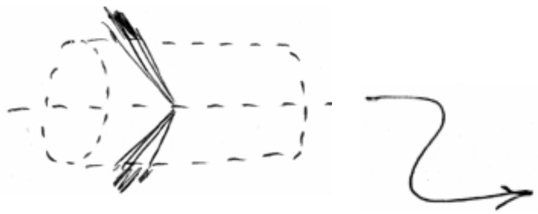
Theory (model, hypothesis):

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \bar{\psi} \not{D} \psi + \cdots$$

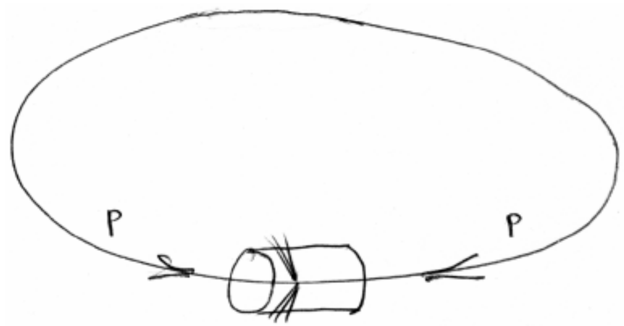


$$\sigma = \frac{G_F \alpha_s^2 m_H^2}{288 \sqrt{2\pi}} \times \sim$$

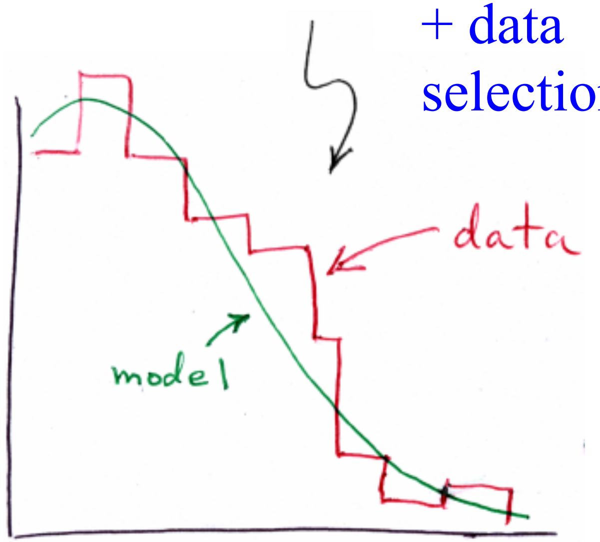
+ simulation
of detector
and cuts



Experiment:



+ data
selection



Data analysis in particle physics

Observe events (e.g., pp collisions) and for each, measure a set of characteristics:

particle momenta, number of muons, energy of jets,...

Compare observed distributions of these characteristics to predictions of theory. From this, we want to:

Estimate the free parameters of the theory: $m_H = 125.4$

Quantify the uncertainty in the estimates: $\pm\ 0.4$ GeV

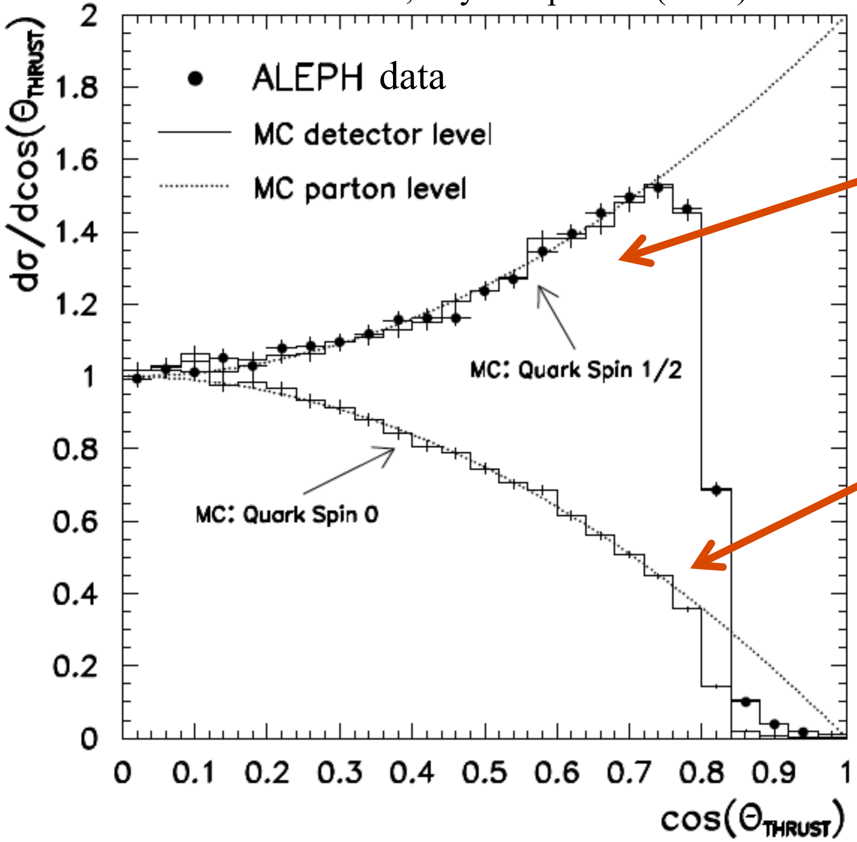
Assess how well a given theory stands in agreement with the observed data: 0^+ good, 2^+ bad

To do this we need a clear definition of PROBABILITY

Data analysis in particle physics: testing hypotheses

Test the extent to which a given model agrees with the data:

ALEPH, Phys. Rept. 294 (1998) 1-165



spin-1/2 quark model "good"

spin-0 quark model "bad"

In general need tests with well-defined properties and quantitative results.