Statistical Data Analysis for HEP

PART-1 of the course (*core* part + *in-depth* part)

Prof. Alexis Pompili (University of Bari Aldo Moro)*

Erasmus+ Teaching Mobility Program / 16-20 October 2023 @ Sofia Physics Faculty

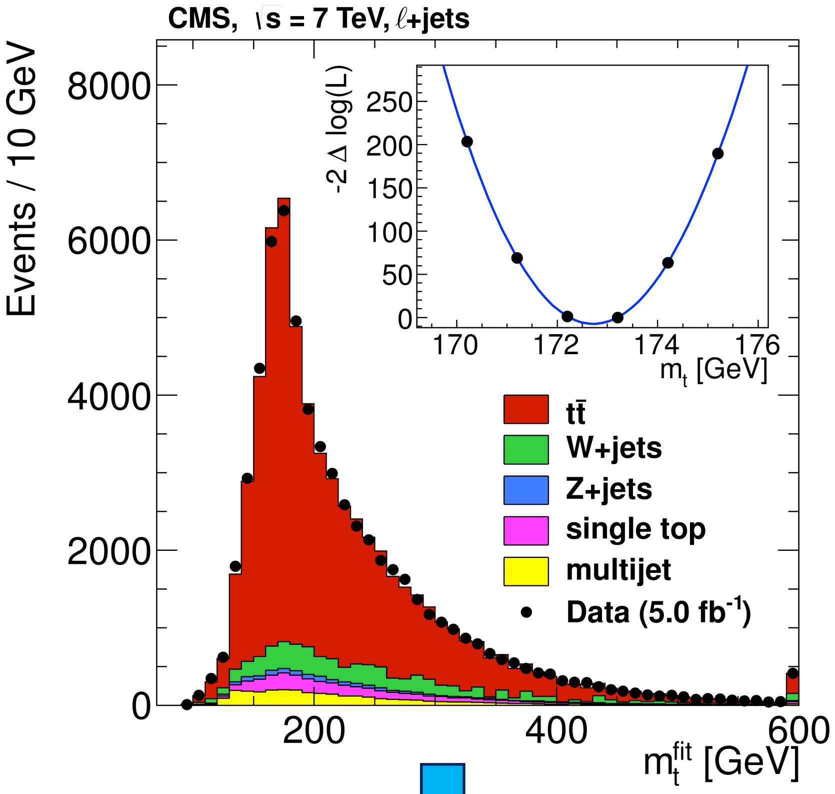
* alexis.pompili@ba.infn.it (or alexis.pompili@cern.ch)

PART 1A - CORE

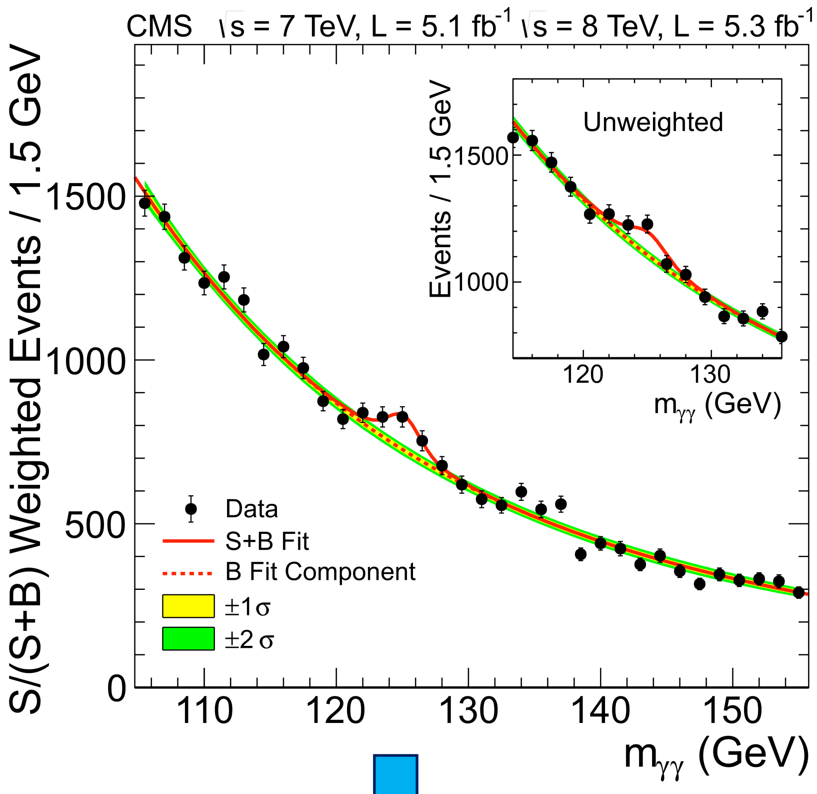
INTRODUCTION: Theory of Probability

MEASUREMENTS

DISCOVERIES



$m_t = 173.49 \pm 1.07$



HIGGS BOSON

The goals of a (experimental) Particle Physicist - II

In modern particle physics experiments, **event data** are recorded by a - usually complex - system of detectors.

Measurements of particle position, particle momentum/energy, time, decay angles etc… are recorded in the **event data**

and are characterized by fluctuations (due to randomness & dilution effects).

Event data are all ***different*** from each other because of:

- **Intrinsic randomness** of the physics process(es) (Quantum Mechanics: $\mathcal{P} \propto |\mathcal{A}|^2$)
- **Detector response** is somewhat random (fluctuations, resolutions, efficiencies, ….)

Some sort of Condition Database records the experimental conditions (alignment, cooling, dead channels, DQM info, ...) which are taken into account when reconstructing the event data.

Tipically, a large number of events are collected by an experiment, each event usually containing large amounts of data → what we study are **distributions** of physical observables (e.g., the mass of a particle, the lifetime, etc.)

The goals of a (experimental) Particle Physicist - III

Distributions of measured quantities in data:

are predicted by a theory model,

depend on some theory parameters,

e.g.: particle mass, cross section, etc.

Given our data sample, we want to:

- measure theory parameters,

  e.g.: m_t= 173.49 ± 1.07 GeV, mH =125.38 GeV

- answer questions about the nature of data

  Is there a Higgs boson? ➔ Yes! (strong evidence? Quantify!)

  Is there a Dark Matter? ➔ No evidence, so far…

  If not, what is the range of theory parameters compatible with the observed data? What parameter range can we exclude?

We should use probability theory on our data and our theory model in order to extract information that will address our questions, that is to say we use statistics for data analysis.

Relation between Probability & Inference - I

Data fluctuate according to the randomness of the physical process governed by a (underlying) physical law
(that the Theory Model should represent)

THEORETICAL MODEL → **PROBABILITY**[*] → DATA (pseudo-data[**])

Known (or assumed correct) the physical process of generation of data (probabilistic model)
… we are able to evaluate the probability of the different outcomes of an experiment

[*] because of the randomness of the process/law ... the calculation of probabilites is involved

[**] when we generate Data according to a model (***Monte Carlo*** generators) we speak about ***pseudo-data***

Relation between Probability & Inference - I

Data fluctuate according to the randomness of the physical process governed by a (underlying) physical law (that the Theory Model should represent)

THEORETICAL MODEL → **PROBABILITY** → DATA

THEORETICAL MODEL ← **INFERENCE** ← DATA

Model parameters in the Theory Model can be estimated with an uncertainty due to fluctuations in the **finite** data sample

In the **statistical inference** the approach is somehow reverted w.r.t. the **theory of probability**:
the physical process or law is under investigation and the statistical methods & techiques try to induce the characteristics of the process on the basis of the (finite) experimental observations

(the word ***induction*** here implies **both** inductive and deductive mental approach along the analysis procedure)

Many processes in nature have uncertain outcomes (their result cannot be predicted in advance).

It is useful to introduce the concept of **random variable**: it represents the outcome of a ***repeatable*** **experiment** whose result is uncertain. Then an **event** consists of the occurrence of a certain particular condition about the value of the random variable resulting from an experiment (in simple words: it is a possible outcome of an experiment).

Note: often in physics : an ***event*** is meant as an ***elementary event***, i.e. it represents a *single outcome*;
on the countrary, in statistics : an ***event*** can represent - in general - a ***subset of possible outcomes***.

Classical probability : if N is the total number of possible outcomes ("cases") of a random variable,
if n is the number of favourable cases for which an event A is realized,
the **probability of an event A** is: $\mathbf{P}(\boldsymbol{A}) = \frac{\boldsymbol{n}}{\boldsymbol{N}}$

(P.S.Laplace, 1749-1827)

Concept of Probability - II

Most experiments in Physics can be ***repeated*** under the same - or at least very similar - conditions.
Such experiments are examples of *random processes* in the sense that, at every repetition, a different outcome is observed.
The result of an experiment may be used to address questions about natural phenomena, …
... for instance about the knowledge of an unknown physical quantity, or the existence or not of some new phenomena.
Statements that answer those questions can be assessed by assigning them a *probability*.
Different definitions of probability apply to cases in which statements refer to repeatable experiments or not:

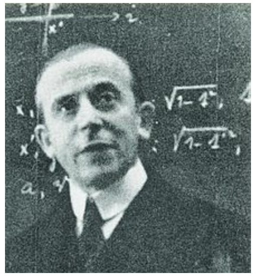
- ***Frequentist probability*** only applies to processes that can be repeated over a reasonably long period of time:

Concept of Probability - II

Most experiments in Physics can be ***repeated*** under the same - or at least very similar - conditions.
Such experiments are examples of *random processes* in the sense that, at every repetition, a different outcome is observed.
The result of an experiment may be used to address questions about natural phenomena, …
... for instance about the knowledge of an unknown physical quantity, or the existence or not of some new phenomena.
Statements that answer those questions can be assessed by assigning them a *probability*.
Different definitions of probability apply to cases in which statements refer to repeatable experiments or not:

➪ ***Frequentist probability*** only applies to processes that can be repeated over a reasonably long period of time:

(R.Von Mises, 1883-1953)



Frequentist probability : is the fraction of the number (N_i) of possible occurrences of an event E_i over the total number of events (N) in a repeatable experiment, in the limit of a very large number of experiments:

$$\mathbf{P}(E_i) = \lim_{N \to \infty} \frac{N_i}{N}$$

Note: - this limit must be intended in an *experimental* (non mathematical!) *sense*
- the true value of the probability would be found only repeating ∞ times the (repeatable) experiment
- in many cases, experience shows that the frequentist probability tends to the classical one (thanks to the ***Law of large numbers***) [ex.: roll a not-loaded dice & execute a large number of rolls] (see 2 slides later)

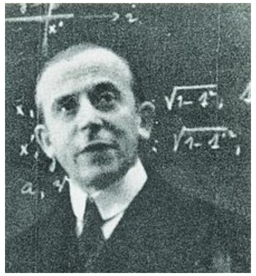
➪

Concept of Probability - II

Most experiments in Physics can be ***repeated*** under the same - or at least very similar - conditions.
Such experiments are examples of *random processes* in the sense that, at every repetition, a different outcome is observed.
The result of an experiment may be used to address questions about natural phenomena, …
... for instance about the knowledge of an unknown physical quantity, or the existence or not of some new phenomena.
Statements that answer those questions can be assessed by assigning them a *probability*.
Different definitions of probability apply to cases in which statements refer to repeatable experiments or not:

➪ ***Frequentist probability*** only applies to processes that can be repeated over a reasonably long period of time:

(R.Von Mises, 1883-1953)



Frequentist probability : is the fraction of the number (N_i) of possible occurrences of an event E_i over the total number of events (N) in a repeatable experiment, in the limit of a very large number of experiments:

$$\mathbf{P}(E_i) = \lim_{N \to \infty} \frac{N_i}{N}$$

Note: - this limit must be intended in an experimental (non mathematical!) sense
- the true value of the probability would be found only repeating ∞ times the (repeatable) experiment
- in many cases, experience shows that the frequentist probability tends to the classical one (thanks to the ***Law of large numbers***) [ex.: roll a not-loaded dice & execute a large number of rolls] (see next slide)

➪ ***Bayesian probability*** applies also to an hypothesis or statement that can be true (or false): the probability of a certain hypothesis (or theory) is represented by the ***degree-of-belief*** *(subjective)* that the hypothesis is true (or false).

- Assuming to repeat N times an experiment whose outcome is a RV x having a given probability distribution, the ***average*** (it will be called *sampling mean* later) - that is itself a RV - is expressed as: $\bar{x} = \frac{x_1 + \cdots + x_N}{N}$

An elementary example is provided by rolling the dice:

the average of N rolls of the type x_i that can assume the values (1,2,3,4,5,6) changes with N and its distribution becomes more and more peacked as N increases; eventually for $N \to \infty$ the distribution becomers a Dirac's δ centered at the value $< x >= \frac{1+2+3+4+5+6}{6} = 3.5$

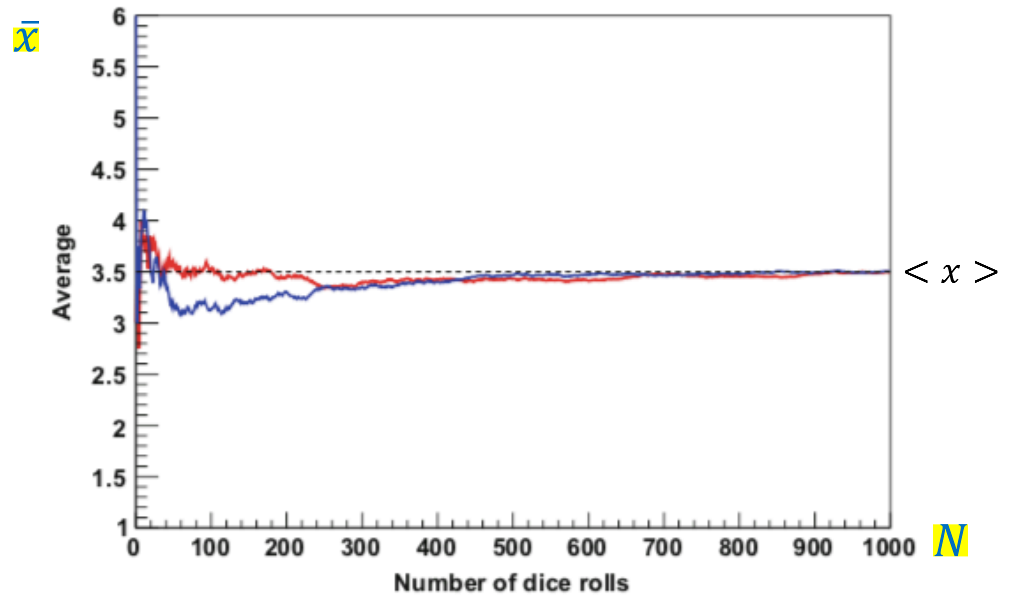


Fig. 1.10 An illustration of the law of large numbers using a computer simulation of die rolls. The average of the first N out of 1000 random extraction is reported as a function of N. 1000 extractions have been repeated twice (*red and blue lines*) with independent random extractions

(borrowed from L.Lista book)

This convergence represents how the law of large numbers "works": large values of N correspond to smaller fluctuations of the result and to a visible convergence towards the expect value of 3.5.
Ideally - when $N \to \infty$ - $\bar{x}$ would be no longer a RV but would take the single possible value 3.5

- The **law of large numbers** has many empirical verifications for the vast majority of random experiments and has a broad validity range.

Interpretation of Probability

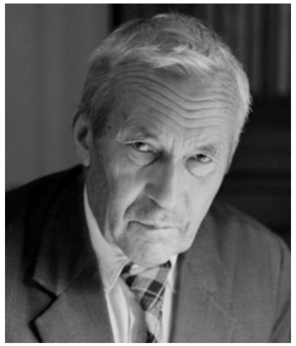
We have just introduced **two different interpretations** of the probability: Frequentist & Bayesian probabilities; note that both are consistent with Kolmogorov axioms.

- **Frequentist probability** refers to a **relative frequency** that can be evaluated for repeatable experiments (for instance when we measure particle scatterings or radioactive decays).
  In this course we will assume/use/refer-to … this concept of probability.

- **Bayesian probability** refers to a **subjective probability** where instead of outcomes we have hypotheses (statements that can be true or false).

In particle physics the frequency interpretation is often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena (for instance the probability that Higgs boson exists, or in handling systematic uncertainties).

In most cases the two approaches give (asymptotically) similar results.



To formalize - in a correct mathematical way - the concept probability, A.N.Kolmogorov (1903-1987) proposed (1933) an **axiomatic approach** (the ***set theory*** can help intuitively to handle axioms and theorems):

- being… Ω the set of possible outcomes, $E \in \Omega$ a certain possible outcome/result/event)

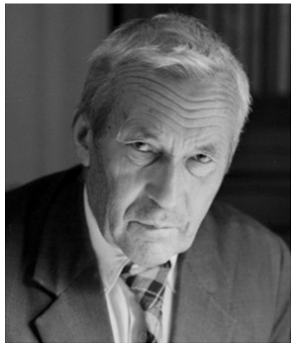
Axiom-1 : $P(\Omega) = 1$ (i.e. the experiment must have a result) [it's the ***normalization condition*** !]

Axiom-2 : $P(E \in \Omega) \geq 0$

union

Axiom-3: property of additivity : $P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$ for ALL E_i being DISJOINT

Axiomatic approach to Probability



To formalize - in a correct mathematical way - the concept probability, A.N.Kolmogorov (1903-1987) proposed (1933) an **axiomatic approach** (the ***set theory*** can help intuitively to handle axioms and theorems):

- being… Ω the set of possible outcomes, $E \in \Omega$ a certain possible outcome/result/event)

Axiom-1 : $P(\Omega) = 1$ (i.e. the experiment must have a result) [it's the ***normalization condition*** !]

Axiom-2 : $P(E \in \Omega) \geq 0$

union

Axiom-3: property of additivity : $P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$ for ALL E_i being DISJOINT

Every concept/definition of probability is required to be compatible with the axiomatic probabiity and with the derived …

… properties: $P(E) = 1 - P(E^*)$, $P(E \in \Omega) \leq 1$, $P(\emptyset) = 0$

intersection

… & theorems: **Additivity theorem** : $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ with $E_1, E_2 \in \Omega$ GENERIC ($\rightarrow$ NOT NECESSARILY DISJOINT)

(it can be easily demonstrated see later)

relative complement

⇨ includes Axiom-3 if E_1, E_2 are disjoint : $P(E_1 \cap E_2) = 0 \Rightarrow P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Joint Probability

Joint probability : $P(A \cap B)$: probability that two events (A & B) happen concurrently

$= 0$ IF A & B DISJOINT ($A \cap B = \emptyset$)

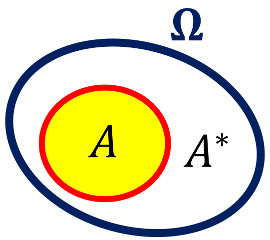
$= P(A) \cdot P(B)$ IF A & B INDEPENDENT

$= P(A) + P(B) - P(A \cup B)$ IF A & B GENERIC ⇦ from the Additivity Theorem! (*)

To deal with **non** independent events we have to introduce the concept of ***conditional probability*** (next slide)

(*) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \Leftrightarrow P(E_1 \cap E_2) = P(E_1) + P(E_2) - P(E_1 \cup E_2)$

Suppose to *restrict* the possible outcomes of an experiment to the subset $A \subset \Omega$ and introduce the …



Conditional probability : $P(E|A)$: probability of event E given the restriction $A \subset \Omega$

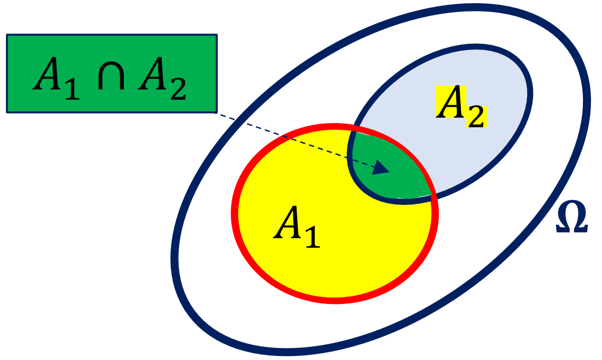
Note: if $A^* \neq \emptyset$ it holds $P(E|A) > P(E)$; this introduces the need to "renormalize" the conditional probability: $P(A|A) \equiv 1$

The following properties hold:

1) $P(A_2|A_1) = P(A_1 \cap A_2|A_1)$ [see figure]

2) ratios of probabilities should not change with the applied restriction:

$$\frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{P(A_1 \cap A_2|A_1)}{P(A_1|A_1)} \quad P(A_1|A_1) \to 1$$



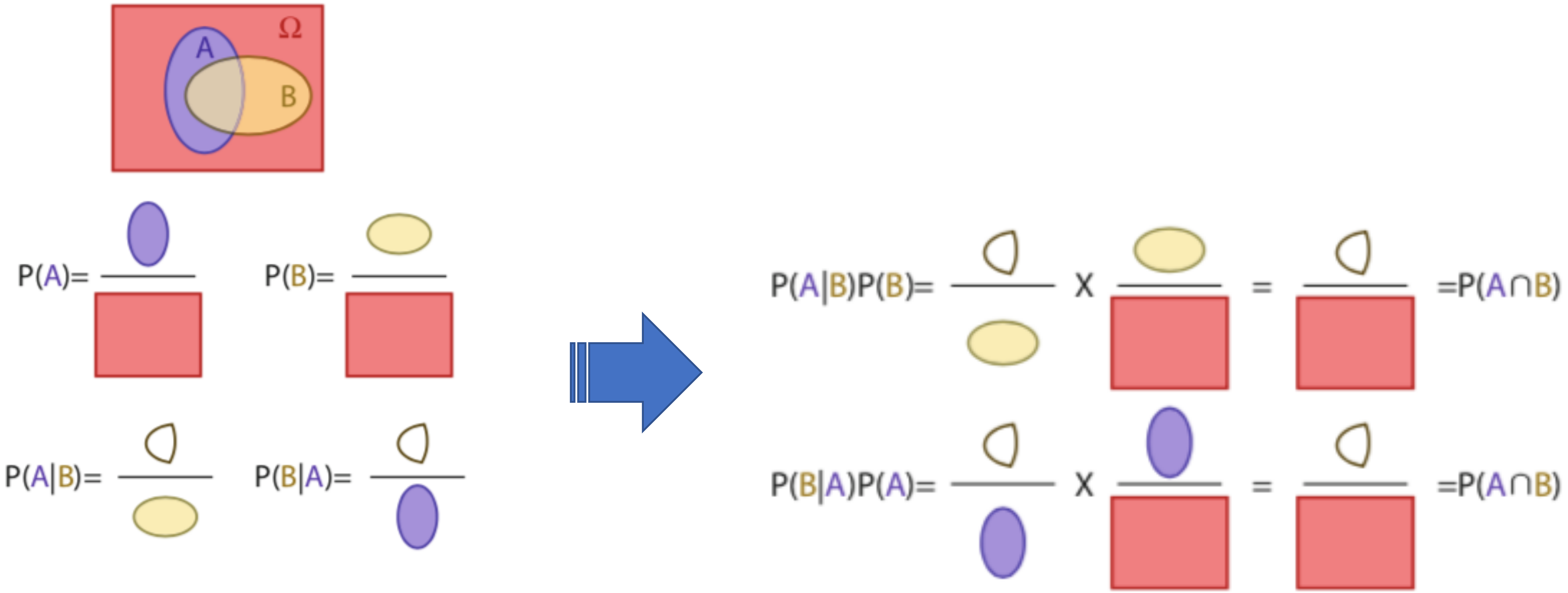
Putting together (1) & (2) : $\frac{P(A_1 \cap A_2)}{P(A_1)} = P(A_2|A_1)$

For completeness (and coherence) we define : $P(A_2|A_1) = 0 \;\; IF \;\; P(A_1) = 0$

We can now formally define the conditional probability: $P(B|A) = \frac{P(B \cap A)}{P(A)}$: probability of event B given the event A already happened

For ***independent*** events: $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B) \cdot P(A)}{P(A)} = P(B)$ (just another way to express independence)

Note: it can be demonstrated that is satisfies the axioms of Kolmogorov



(borrowed from L.Lista book)

This famous theorem by T.Bayes relates the two conditional probabilities $P(B|A)$ with $P(A|B)$ where $A, B \in \Omega$

We've already written $P(B|A) = \dfrac{P(B \cap A)}{P(A)}$ but …

(T.Bayes, 1702-1761)

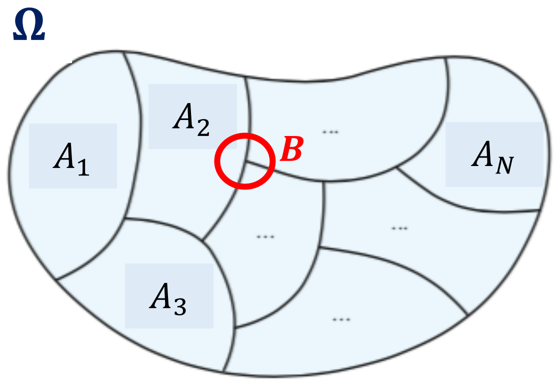
…we can equally write (A, B are exchangeable): $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Putting together: $P(A|B) \cdot P(B) = P(A \cap B) = P(B|A) \cdot P(A)$. Thus : $P(A|B) = P(B|A) \cdot \dfrac{P(A)}{P(B)}$

Bayes theorem - II (extended version)

We derived the basic version of the Bayes' theorem: $P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$

A generalization/extension of the theorem can be obtained by introducing …
… the ***Law of the total probability*** as follows:



if we have sets of events $\{A_i\}_i$ that are **disjoint** and **fully cover Ω** (namely $\Omega = \bigcup_i A_i$)

and if $\boldsymbol{B} \in \boldsymbol{\Omega}$ is a generic event, we can calculate $P(B)$ exploiting the fact that

$B = B \cap \Omega = B \cap \bigcup_i A_i = \bigcup_i (B \cap A_i)$ and $(B \cap A_i)$ are disjoint,

and thus, the total probability can be obtained by the following sum:

Axiom of additivity — definition of conditional probability

$$\boldsymbol{P(B)} = P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i) \cdot P(A_i)$$

representing the so called **Law of total probability**

Now Bayes' theorem can be rewritten in its extended version (with A and B subsets of Ω):

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

Note: nothing forbids A to be one of the A_i, say A_k) :

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_i \boldsymbol{P(B|A_i) \cdot P(A_i)}}$$

Bayes theorem - III (discussion) [note: on textbooks *a-priori* is called *prior*]

This theorem can be discussed in a **frequentist context** (in which a probability cannot be associated to an hypothesis!), [and it can be helpful when designing an experiment] in the following way:

subset of events of interest

detection/reconstruction/selection

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

$P(A)$ → production probability of the events of interest
(tipically estimated in previous experiment(s) or …
suggested by a model used in generating simulated events, etc…
(the so called **a-priori probability)**

$P(B)$ → general detection/reconstruction/selection **efficiency**
(performs the task to ensure normalization)

$P(A|B)$ → probability to detect/reconstruct/select the specific kind of events of interest **given** a positive response by our detector/reconstruction algorithm/selector
(the so called a-posteriori probability, i.e. after having carried out the experiment)

$P(B|A)$ → probability to detect/reconstruct/select the specific kind of events of interest
(it includes the acceptance and the efficiency effects **in relation to** the characteristics of the particular events of interest)

Bayes's theorem - IV (discussion) [note: on textbooks *a-priori*/*a-posteriori* is called *prior*/*posterior*]

Considering also the applications that are discussed in the in-depth part, it is worthy to stress the importance of the following aspect: **the Bayes theorem provides a prescription about how to move from an a-priori to an a-posteriori probability after having carried out an experiment/test**.

The factor most relevant in the previous expression would be the **a-priori probability $P(A)$** because it can represent many concepts in different contexts. For instance, it can refer to:

- a previous empirical result/measurement
- a complex description used in a simulation (Monte Carlo) where all the previous knowledge is implemented
- a prediction from theory or from an assumed model
- a hypothesis (this makes sense in a Bayesian context;
  for the frequentist, a hypothesis is not a RV, and a probability can't be associated to it)

In the bayesian context $P(A)$ can be a degree of belief (subjective) that the hypothesis that a certain theory is correct is true.
Neglecting the normalization term at the denominator, the Bayes theorem assumes the following form:

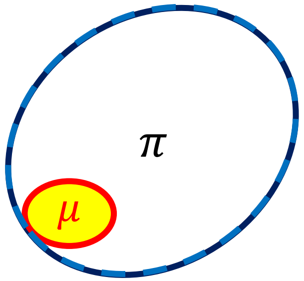
POSTERIOR PROBABILITY — LIKELIHOOD — PRIOR PROBABILITY

$$P(theory|data) \propto P(data|theory) \cdot P(theory)$$

hypothesis (theory) — experimental result (data)

``Bayesian statistics provides no fundamental rule for assigning the prior probabiity to a theory, but once this has been done, it says how one's degree of belief should change in the light of experimental data'' (G. Cowan).
``Bayesian statistics is appropirate only when it is desired (or unavoidable) to put physicist's prior beliefs explicitely into the statistical analysis'' (F. James).

- Assume we are given a particles' beam with 90% of pions (π) and 10% of muons (μ): $P(\pi) = 0.9, P(\mu) = 0.1$
To carry out a test beam we would better need muons, and thus suppose we can use an equipment T (trigger/selector) which filters muons at the expense of pions.
This filtering is characterized by the following performances:



$\boldsymbol{\Omega} = \mu \cup \pi$
[μ, π are disjoint]

  - *efficiency* of the selection for muons: $\varepsilon(\mu) = 0.95$ ⟺ $\boldsymbol{P(T|\mu) = 0.95}$: triggered if $\boldsymbol{\mu}$
  - *efficiency* of the selection for pions: $\varepsilon(\pi) = 0.05$ ⟺ $\boldsymbol{P(T|\pi) = 0.05}$: triggered if $\boldsymbol{\pi}$
    (this is a source of **contamination** for the muons!!)

- Let's calculate: **1)** the total efficiency of the trigger/selector, **2)** the $\boldsymbol{\mu}$ enrichment i.e. $P(\mu|T)$

(1) $$\boldsymbol{\varepsilon_{tot}} = P(T) = P(T|\mu) \cdot P(\mu) + P(T|\pi) \cdot P(\pi) = (0.95 \cdot 0.1) + (0.05 \cdot 0.9) = 0.095 + 0.045 = 0.14 \equiv 14\%$$

↑ total probability law

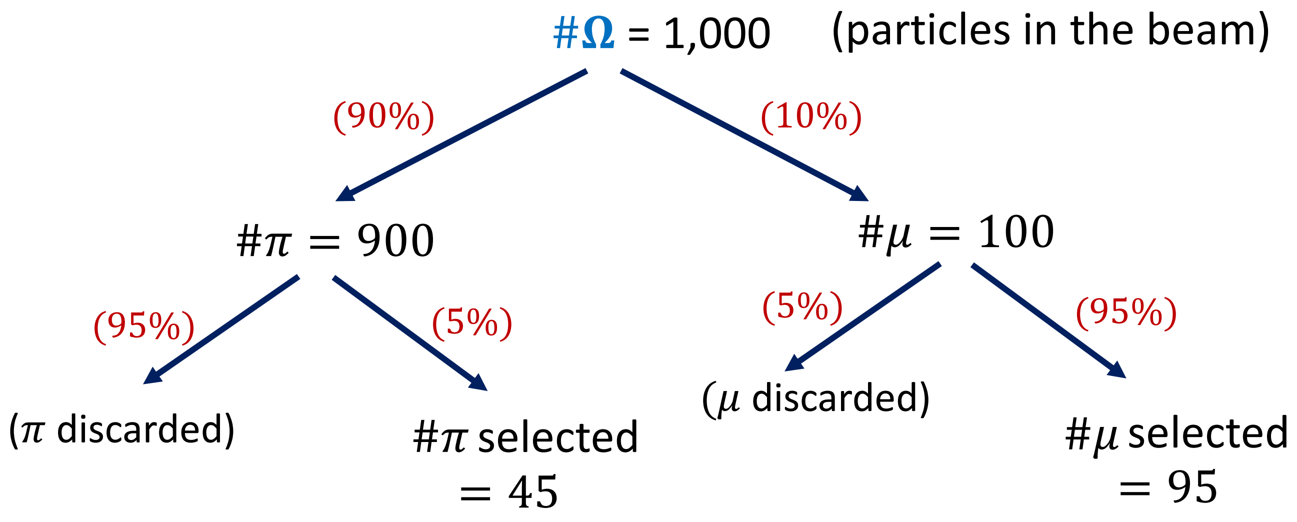
With this trigger only the 14% of the particles in the beam are selected!

(2) $$\boldsymbol{P(\mu|T)} = \frac{P(T|\mu) \cdot P(\mu)}{P(T|\mu) \cdot P(\mu) + P(T|\pi) \cdot P(\pi)} = \frac{P(T|\mu) \cdot P(\mu)}{\boldsymbol{\varepsilon_{tot}}} = \frac{(0.95 \cdot 0.1)}{0.14} = \frac{0.095}{0.14} = 0.679 \equiv 67.9\%$$

↑ Bayes Theorem

By applying this trigger … the pions:muons ratio in the beam goes from 90:10 to ~32:68

- When the problem under study by applying the Bayes theorem is not complicated… this **graphical method** can be useful:



$$\frac{(\#particles)_{selected}}{(\#particles)_{total}} = \frac{95+45}{1000} = \frac{140}{1000} = 14\%$$

$$\frac{\#\mu_{\ selected}}{(\#particles)_{selected}} = \frac{95}{45+95} = \frac{95}{140} \approx 68\%$$

CHARACTERISTICS of MEASUREMENTS

- When we carry out an experimental measurement we must separate the purely **statistical component** from those "non statistical" (called **systematics components**):

$$measure(\text{“}central\ value\text{”})\ \pm\ statistical\ uncertainty\ \pm\ systematic\ uncertainty : m\ ^{+a}_{-b}\ ^{+c}_{-d}$$

A good measurement requires to be able to reduce as much as possible both uncertainties.

IF we have accumulated not much data (low statistics)… we can afford a conservative evaluation of the sources of systematics uncertainties (approximated by excess)

IF we have accumulated a lot of data (high statistics)… the statistical uncertainty will be relatively small and… …we cannot afford a conservative evaluation of systematics uncertainties: we must evaluate the systematics effect with good accuracy with the aim to bring the systematic uncertainties to the same level of the statistical uncertainty !

- Recap: @ "low" statistics : we can afford systematic uncertainty $\lesssim$ statistical uncertainty (relatively large)

  @ "high" statistics : we must work so that systematic uncertainty $\approx$ statistical uncertainty (relatively small)

- If the problem is particular difficult to require the execution - on a computing machine - of the simulation (MC) of your physical system under exam, in order to compare real and simulated data, …

  … it can happen to identify a systematic error ("**bias**") in the real data and to correct the measurement (central value) according to a **correction ("shift")** derived from the data-MC comparison.

  In this circumstance **the statistical uncertainty on the measurement carried out on the simulated data must be considered a systematic uncertainty for the (corrected) measurement in real data**.

  This implies the need to have enough statistics for your simulated data samples (**importance to produce enough simulated data**).

Example: https://arxiv.org/pdf/hep-ex/9902011.pdf (**CLEO** experiment's charmed mesons lifetime measurement) [see next slide]

Phys. Rev. Lett. 82 (1999) 4586

The systematic uncertainties for the D meson lifetimes are listed in Table I and are described below. They can be grouped into three categories:

Reconstruction of the D decay length and proper time. Errors in the measurement of the reconstructed decay length can be due to errors in the measurement of the decay vertex, (1) the global (2) detector scale, and the beam spot. (3) The bias in the decay vertex position is estimated to be $(0.0 \pm 0.9\ \mu m)$ from a "zero-lifetime" sample of $\gamma\gamma \to \pi^+\pi^-\pi^+\pi^-$ events. This corresponds to a measured proper-time uncertainty of ± 1.8 fs. In addition, the vertex reconstruction is checked with events with interactions in the beam pipe with a relative uncertainty of $\pm 0.2\%$. The sums of these uncertainties in quadrature yield the systematic

(1) uncertainties due to the decay vertex measurement. The global detector scale is measured to a precision of $\pm 0.1\%$ in surveys and confirmed in the study of events with interactions in the

(2) beam pipe. The changes in the lifetimes due to the variation ($\pm 2\ \mu m$) in the vertical beam spot position and height are another source of systematic error, since (3) the interaction point is calculated from the beam spot and the reconstructed D momentum and decay vertex. Statistical uncertainties for the D masses [2] and the D momentum measurements lead to systematic errors since these quantities are used to convert the decay length into proper time.

TABLE I. Systematic uncertainties for the D meson lifetimes in fs. The systematic uncertainties for the three D^0 modes are weighted with the same weights as the fitted D^0 lifetimes.

Uncertainty	D^0 $K^-\pi^+$	D^0 $K^-\pi^+\pi^0$	D^0 $K^-\pi^+\pi^-\pi^+$	D^0 combined	D^+ $K^-\pi^+\pi^+$	D_s^+ $\phi\pi^+$	
Decay vertex	± 2.0	± 2.0	± 2.0	± 2.0	± 2.8	± 2.1	(1)
Global detector scale	± 0.1	± 0.1	± 0.1	± 0.1	± 0.1	± 0.1	(2)
Beam spot	$^{+0.3}_{-0.1}$	$^{+2.1}_{-0.0}$	$^{+0.3}_{-0.2}$	$^{+0.8}_{-0.1}$	$^{+1.3}_{-1.1}$	$^{+0.7}_{-1.1}$	(3)
D meson mass	± 0.1	± 0.1	± 0.1	± 0.1	± 0.3	± 0.1	
D meson momentum	$^{+0.2}_{-0.0}$	$^{+0.1}_{-0.2}$	$^{+0.3}_{-0.1}$	$^{+0.2}_{-0.1}$	$^{+0.6}_{-0.0}$	± 0.1	
Signal probability	$^{+0.4}_{-0.1}$	$^{+0.1}_{-0.2}$	$^{+0.1}_{-0.2}$	$^{+0.3}_{-0.1}$	$^{+1.2}_{-8.1}$	$^{+1.3}_{-1.8}$	
$t - M(D)$ correlation	± 0.6	± 0.6	± 1.0	± 0.7	± 1.7	± 1.5	
Large proper times	± 1.2	± 3.4	± 0.2	± 1.5	± 0.3	± 0.5	
Background	± 0.5	± 2.4	± 3.0	± 1.5	± 6.3	± 2.9	
MC statistics	± 0.9	± 2.3	± 2.2	± 1.6	± 6.6	± 2.4	
Total	$^{+2.7}_{-2.6}$	$^{+5.6}_{-5.2}$	± 4.4	$^{+3.5}_{-3.4}$	$^{+9.9}_{-12.7}$	$^{+4.9}_{-5.1}$	

Checking the algorithms with simulated events. Charm meson candidate selection requirements can cause systematic biases in the lifetime measurements. We estimate these biases with simulated events and correct for the biases as described above. We include the statistical uncertainties in the measured lifetimes from the samples of simulated events as systematic uncertainties in the results.

PRECISION & ACCURACY

- **Precision of a measurement:** term that expresses that the result of a measurement can be obtained with great detail (many significative cyphers).

Numerically, it is represented by the random (or "**statistical**") **uncertainty** !

- **Accuracy of a measurement:** term that expresses the maximum possible deviation of the result of a measurement from the result of an ideal measurement; thus it is associated to the maximum systematic error that the experimental instrumentation can introduce in the measurement.

Numerically, it's represented by the maximum "**systematic**" **uncertainty** that the used instrumentation/method can introduce!

Wrapping up:

A precise measurement is a measurement affected by a very small statistical uncertainty;
The systematic uncertainties cannot be eliminated but enough (hopefully strongly) reduceable.

An accurate measurement is a measurement affected by a minimized systematic uncertainty (or anyway, lower than the statistical uncertainty);
The systematic uncertainties cannot be eliminated but hopefully can be *minimized*.

PART 1B - IN-DEPTH SLIDES

Demonstration of the additivity theorem

Let us consider the general case in which are not disjoint and thus have a not null intersection $A \cap B \neq \emptyset$.

From the theory set it is easy to be convinced that the following two expressions hold:

(1) $A \cup B = A \cup [B - (A \cap B)]$

(2) $B = [B - (A \cap B)] \cup (A \cap B)$

We can now apply the Kolmogorov axioms (property of additivity) for the disjoint (sets of) events:

– A and $[B - (A \cap B)]$ are disjoint: $P(A \cup B) = P(A) + P([B - (A \cap B)])$ **(3)**

– $(A \cap B)$ and $[B - (A \cap B)]$ are disjoint: $P(B) = P([B - (A \cap B)]) + P(A \cap B)$ **(4)**

Subtracting member by member the (4) from (3) one gets:

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

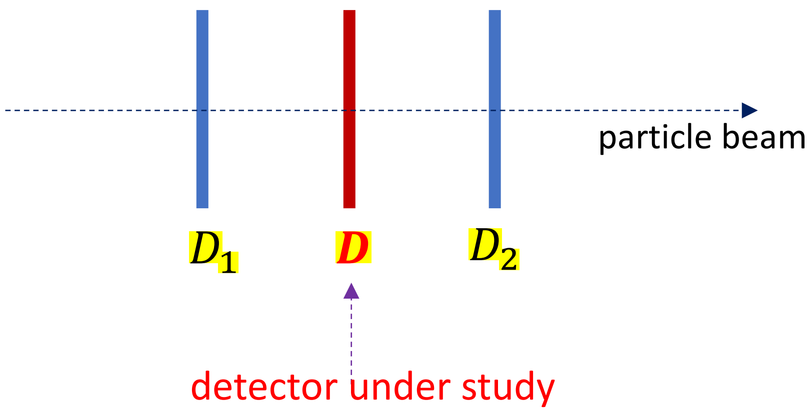
$$\Longleftrightarrow \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \text{QED}$$

Role of the Bayes's Theorem

From a Scientific American article of 2006 (*):

More generally, Bayes's theorem is used in any calculation in which a "marginal" probability is calculated (e.g., p(+), the probability of testing positive in the example) from likelihoods (e.g., p(+|s) and p(+|h), the probability of testing positive given being sick or healthy) and prior probabilities (p(s) and p(h)): p(+)=p(+|s)p(s)+p(+|h)p(h). Such a calculation is so general that almost every application of probability or statistics must invoke Bayes's theorem at some point. In that sense Bayes's theorem is at the heart of everything from genetics to Google, from health insurance to hedge funds. It is a central relationship for thinking concretely about uncertainty, and--given quantitative data, which is sadly not always a given--for using mathematics as a tool for thinking clearly about the world.

(*) https://www.scientificamerican.com/article/what-is-bayess-theorem-an/



Detection efficiencies are probabilities ! (*)

To measure the detection efficiency of the detector under test we need to select ***all and only*** the particles that cross the system and are detected by both "telescope" detectors D_1 & D_2 (that are read in ***time coincidence***).
The intersection expresses the time coincidence in the sense that the probability to have a particle of the beam detected by both of them is given by $P(D_1 \cap D_2)$ [reminder: intersection is a logical-AND]!

Of course, $P(D_1 \cap D_2)$ is a joint probability but note that the two "telescope" detectors work independently, thus:

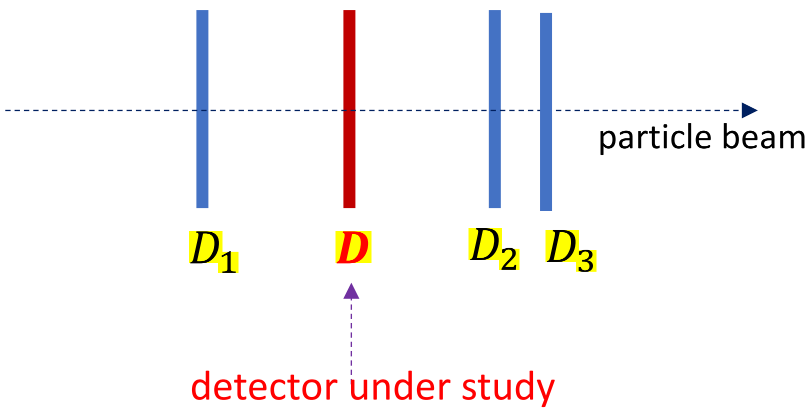
$$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2) \quad (\#)$$

As seen in previous slide, $P(D_1 \cap D_2)$ can also be expressed in terms of conditional probability as follows:

$P(D_1 \cap D_2) = P(D_2|D_1) \cdot P(D_1)$ and since the detectors work independently it holds $P(D_2|D_1) = P(D_2)$ and thus the (#).

The ***total efficiency*** of the telescope can be calculated to know the useful particle flux to study the detector under test.
Obviously … for a telescope with 2 similar detectors (i.e. identical efficiencies $\varepsilon_1 = \varepsilon_2 \equiv \varepsilon_D$) we get (*):

$$\varepsilon_{tot,2} \underset{(*)}{\equiv} P(D_1 \cap D_2) \underset{(\#)}{=} P(D_1) \cdot P(D_2) \equiv \varepsilon_1 \cdot \varepsilon_2 = \varepsilon_D^2$$



Adding a third detector as in the figure implies …
to have detector D_1 in coincidence with **any** of one between D_2 & D_3 !

The involved joint probability is now: $P(D_1 \cap (D_2 \cup D_3))$

[reminder : intersection is a logical-AND, union is a logical-OR]

Now we get : $P(D_1 \cap (D_2 \cup D_3)) = P(D_1) \cdot P(D_2 \cup D_3) = P(D_1) \cdot [P(D_2) + P(D_3) - P(D_2 \cap D_3)]$

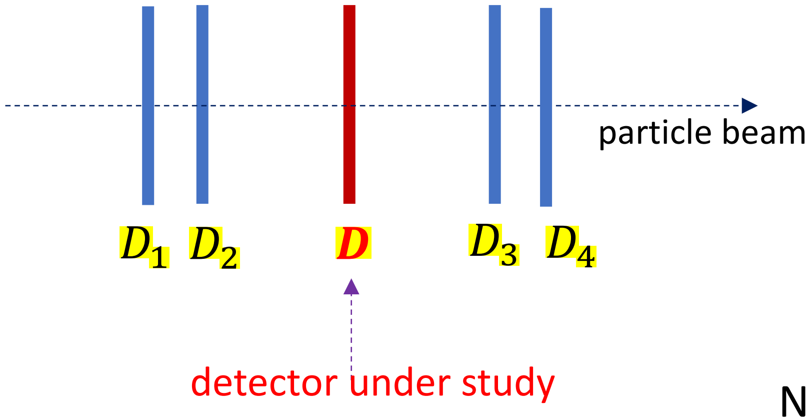
additivity theorem

Since also the detectors D_2 & D_3 work independently it holds: $P(D_2 \cap D_3) = P(D_2) \cdot P(D_3)$

Overall : $P(D_1 \cap (D_2 \cup D_3)) = P(D_1) \cdot [P(D_2) + P(D_3) - P(D_2 \cap D_3)] = P(D_1) \cdot [P(D_2) + P(D_3) - P(D_2) \cdot P(D_3)]$

In this way the ***total efficiency*** of the telescope can be easily calculated as :

$$\varepsilon_{tot,3} \equiv P(D_1 \cap (D_2 \cup D_3)) = P(D_1) \cdot [P(D_2) + P(D_3) - P(D_2) \cdot P(D_3)] = \varepsilon_D \cdot [2\varepsilon_D - \varepsilon_D^2] = \varepsilon_D^2 \cdot [2 - \varepsilon_D] \equiv \varepsilon_{tot,2} \cdot [2 - \varepsilon_D]$$



Adding a fourth detector as in the figure implies that …
… we can have detector D_1 in coincidence with **any** of one between D_3 & D_4
… and the same holds for D_2 !

The involved joint probability is now: $P((D_1 \cup D_2) \cap (D_3 \cup D_4))$

Now we get : $P((D_1 \cup D_2) \cap (D_2 \cup D_3)) = P(D_1 \cup D_2) \cdot P(D_3 \cup D_4) =$

additivity theorem

$$= [P(D_1) + P(D_2) - P(D_1 \cap D_2)] \cdot [P(D_3) + P(D_4) - P(D_3 \cap D_4)]$$

$$\equiv [P(D_1) + P(D_2) - P(D_1 \cap D_2)] \cdot [P(D_3) + P(D_4) - P(D_3 \cap D_4)]$$

$$= [P(D_1) + P(D_2) - P(D_1) \cdot P(D_2)] \cdot [P(D_3) + P(D_4) - P(D_3) \cdot P(D_4)]$$

Since the detectors $D_{1(3)}$ & $D_{2(4)}$ work independently it holds:

$$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2)$$
$$P(D_3 \cap D_4) = P(D_3) \cdot P(D_4)$$

Overall : $\varepsilon_{tot,4} = P\big((D_1 \cup D_2) \cap (D_3 \cup D_4)\big) = [2\varepsilon_D - \varepsilon_D^2] \cdot [2\varepsilon_D - \varepsilon_D^2] = \varepsilon_D \cdot [2 - \varepsilon_D] \cdot \varepsilon_D \cdot [2 - \varepsilon_D] =$

$$= \varepsilon_D^2 \cdot [2 - \varepsilon_D]^2$$

$$= \varepsilon_{tot,2} \cdot [2 - \varepsilon_D]^2$$

Total efficiency calculations for the «test beam»

Summarizing up:

2 detectors: $\varepsilon_{tot,2} = \varepsilon_D^2$

3 detectors: $\varepsilon_{tot,3} = \varepsilon_{tot,2} \cdot [2 - \varepsilon_D]$

4 detectors: $\varepsilon_{tot,4} = \varepsilon_{tot,2} \cdot [2 - \varepsilon_D]^2 = \varepsilon_{tot,3} \cdot [2 - \varepsilon_D]$

… which shows that adding one more detector, either upstream or downstream, introduces a multiplicative factor $[2 - \varepsilon_D]$>1 which therefore increases the overall efficiency!

Numerically, let's assume $\varepsilon_D = 0.8$ … then $\varepsilon_{tot,2} = 0.64$

$\varepsilon_{tot,3} = (2 - 0.8) \cdot 0.64 \approx 0.77$

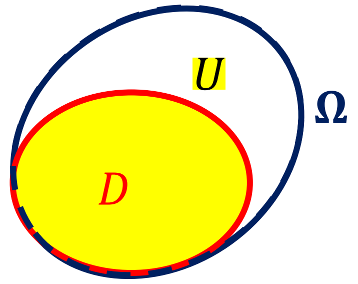
$\varepsilon_{tot,4} = (2 - 0.8)^2 \cdot 0.77 \approx 0.92$

A telescope is usually built by 4 identical detectors (2 upstream and 2 downstream) and this ensures to be using, during the test beam, most of the MIPs of the beam inpinging the detector (i.e. more than 90%).

- Assume to know that if a person is really ill the probability that the medical test gives a positive result is very high (say 98%); of course the test will have also some small probability (say 2%) to give a false positive result on an healthy person.
  If a random person is tested positive and diagnosed with an ilness … what is the probability that he/she is effectively ill?
  A common mistake (for who doesn't know the Bayes theorem!) is to think that this probability would be simply (98-2)%=96%.
  We will discuss that in the next slides.

- Preliminarily let's discuss an example of application of the Law of total probability in the epidemiological context:

An illness M hits in 1 year 10% of men and 5% of women.
If the population Ω of 10,000 people is composed by 45% of men and 55% of women, …
… find the expected number of ill persons in a year.



In this case we have $\Omega = U \cup D$ with $A_1 \equiv D$ and $A_2 \equiv U$ that are *disjoint*!

The law of total probability is:

$$\boldsymbol{P(M)} = \sum_{i=1,2} P(B|A_i) \cdot P(A_i) \equiv P(M|D) \cdot P(D) + P(M|U) \cdot P(U) = [(0.05) \cdot (0.55)] + [(0.1) \cdot (0.45)] = 0.0725$$

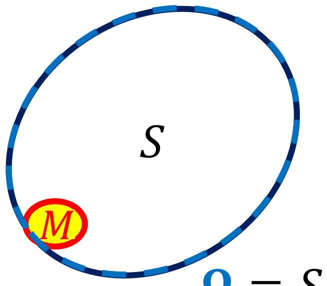
i.e. the expected ill population is 7.25% of the total one, and the expected number of ill persons is: $10{,}000 \cdot 0.0725 = 725$

Bayes Theorem application to an epidemiological use case - II

Consider an illness M hitting the 0.1% of the population.
A medical test (to be used in population screening) is characterized by the following performances:

- it gives positive result at 98% for an ill person
- it has the 3% of probability to give positive result for an healthy person

Question: **what is the probability of being ill if the test gives a positive result?**



$\Omega = S \cup M$
[S, M are disjoint]
ill
healthy

The data:
- probability of being ill: $P(M) = 0.001$
- probability of being healthy: $P(S) = 0.999$
- *efficiency* of the test (positive if ill): $\boldsymbol{P(+|M) = 0.98}$ ...this implies also: $P(-|M) = 0.02$
- *contamination* ("*fake rate*") of the test (positive in spite of being heathy): $\boldsymbol{P(+|S) = 0.03}$ ...this implies also: $P(-|S) = 0.97$

To answer the question, I need to evaluate: $\boldsymbol{P(M|+)}$. This can be done just by applying the (extended) Bayes Theorem:

$$P(M|+) = \frac{P(+|M)\cdot P(M)}{P(+|M)\cdot P(M) + P(+|S)\cdot P(S)} = \frac{0.98\cdot 0.001}{(0.98\cdot 0.001)+(0.03\cdot 0.999)} \cong 0.0317 \approx 3.2\%$$

$$= \frac{P(+\cap M)}{P(+\cap M) + P(+\cap S)} \approx \frac{0.001}{(0.001)+(0.03)} \approx \frac{0.001}{0.03} = \frac{1}{30}$$

The result seems counter-intuitive!
Too low!?! Still it is correct !

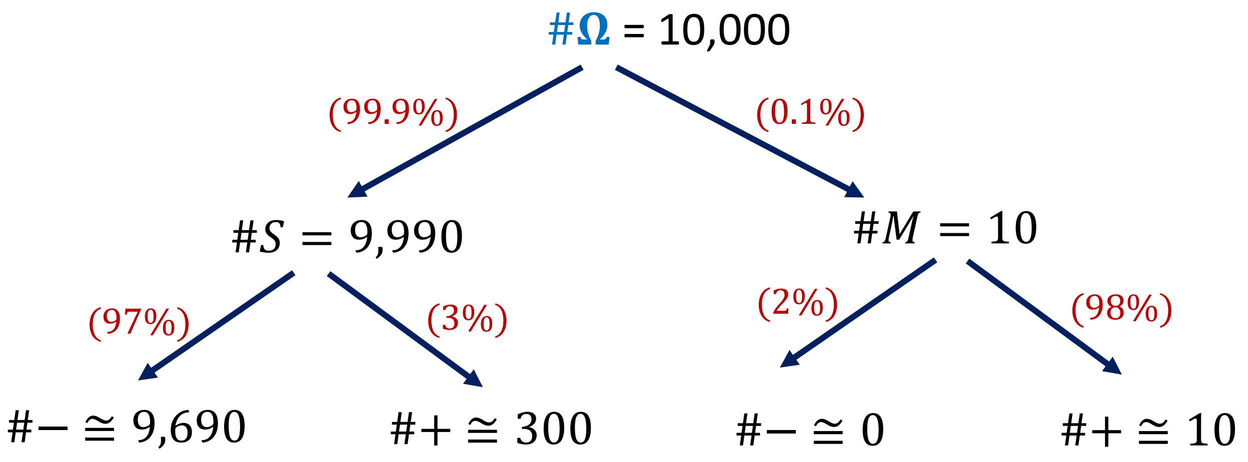
Note: $P(M)$ plays a crucial role together with the "fake rate" $P(+|S)$.
Considering $P(M)$ cannot be changed what matters is $P(+|S)$; a better quality test would have $P(+|S) = 0.003$ so that $P(M|+) \approx 25\%$

Bayes Theorem application to an epidemiological use case - III

The correct interpretation (and usefulness) of the screening of the population relies on the difference between
- prior probability (of being ill): 0.1%
- posterior probability (of being ill): 3.2% ,

…namely how changes the probability to be ill (because belonging to a certain population) in light of a positive resulting test !
Of course, the lowest fake rate of the test the better !

When the problem under study by applying the Bayes theorem is not complicated… this **graphical method** can be useful:



$$\text{probability to be ill if positive} = \frac{\#M}{(\#+)_{tot}} \cong \frac{10}{300+10} \approx 3.2\%$$

Bayes Theorem application to an epidemiological use case : double screening - I

Let us immagine to be able to repeat the medical test ("double screening"): the two consecutive tests are labelled $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$.

Earlier we evaluated $P(M|T)$ and in particular $P(M|+)$; now we are interested in $\boldsymbol{P(M|T_1,T_2)}$ and in particular $\boldsymbol{P(M|+,+)}$!

Step-1) For the **Bayes theorem** we can write: $\boldsymbol{P(M|T_1,T_2)} = P(T_1,T_2|M) \cdot \frac{P(M)}{P(T_1,T_2)}$

Step-2) Reasonably assuming **the two tests are independent** it is possible to express the joint probabilty as a factorization :

$P(T_1,T_2) = P(T_1) \cdot P(T_2)$ … as well as (applying the restriction) … $P(T_1,T_2|M) = P(T_1|M) \cdot P(T_2|M)$

Step-3) Putting all together we get: $\boldsymbol{P(M|T_1,T_2)} = P\ (T_1|M) \cdot P(T_2|M) \cdot \frac{P(M)}{P(T_1) \cdot P(T_2)}$

Step-4) For the **total probability law** we can write: $P\left(T_{1,2}\right) = P\left(T_{1,2}|M\right) \cdot P(M) + P\left(T_{1,2}|S\right) \cdot P(S)$

…thus obtaining $P(T_1) \cdot P(T_2) = [P(T_1|M) \cdot P(M) + P(T_1|S) \cdot P(S)] \cdot [P(T_2|M) \cdot P(M) + P(T_2|S) \cdot P(S)] =$ *(next slde)*

$$= [P(T_1|M) \cdot P^2(M) \cdot P(T_2|M)] + [P(T_1|M) \cdot P(M) \cdot P(T_2|S) \cdot P(S)] + [P(T_1|S) \cdot P(S) \cdot P(T_2|M) \cdot P(M)] + [P(T_1|S) \cdot P^2(S) \cdot P(T_2|S)] =$$

$$\equiv x_1 x_2 x^2 + x_1 y_2 xy + x_2 y_1 xy + y_1 y_2 y^2 = x_1 x_2 x(1-y) + x_1 y_2 xy + x_2 y_1 xy + y_1 y_2 y(1-x) =$$

$x + y = 1 \Leftrightarrow P(M) + P(S) = 1$

$$= x_1 x_2 x - x_1 x_2 xy + x_1 y_2 xy + x_2 y_1 xy + y_1 y_2 y - y_1 y_2 xy) = x_1 x_2 x + y_1 y_2 y + xy[-x_1 x_2 + x_1 y_2 + x_2 y_1 - y_1 y_2]$$

$= \mathbf{0}$ (it can be demonstrated with further algebra)

$$= P(T_1|M) \cdot P(T_2|M) \cdot P(M) + P(T_1|S) \cdot P(T_2|S) \cdot P(S)$$

Therefore, wrapping up: $$\boldsymbol{P(M|T_1,T_2)} = \frac{P(T_1|M) \cdot P(T_2|M) \cdot P(M)}{P(T_1|M) \cdot P(T_2|M) \cdot P(M) + P(T_1|S) \cdot P(T_2|S) \cdot P(S)}$$

For our specific case: $$\boldsymbol{P(M|+,+)} = \frac{P(+|M) \cdot P(+|M) \cdot P(M)}{P(+|M) \cdot P(+|M) \cdot P(M) + P(+|S) \cdot P(+|S) \cdot P(S)} = \frac{0.98 \cdot 0.98 \cdot 0.001}{(0.98 \cdot 0.98 \cdot 0.001) + (0.03 \cdot 0.03 \cdot 0.999)} \cong$$

$$\cong \frac{0.00096}{0.00186} \cong 0.516 \approx 52\%$$

Conclusion: Given a test with an efficiency of 98% and a false positives rate (fake rate) of 3% , the a-priori (=pre-test) probability to be ill is 0.1%; after a positive test it grows to 3.2%; by doing a second test giving again a positive result the illness probabilty rises to 51.6%. These percentages further rise if the fake rate lowers.

Bayes Theorem application to an epidemiological use case : double screening - II

There is a **simpler** way to proceed (**more compliant to the bayesian approach**) to obtain the same result :

let's reconsider $P(M|T_1,T_2) = P\,(T_1|M) \cdot P(T_2|M) \cdot \dfrac{P(M)}{P(T_1) \cdot P(T_2)}$

… and reshuffle the factors as follows: $P(M|T_1,T_2) = \dfrac{P(T_2|M)}{P(T_2)} \cdot \dfrac{P\,(T_1|M) \cdot P(M)}{P(T_1)}$

$P(M|T_1)$: a-posteriori probability after the 1st test

In other words the a-posteriori probability of the 1st test becomes the a-priori (prior) probability for the 2nd test !

In our specific case the calculation is much straightforward :
[note that $P'(+)$ and thus $P'(M)$ and $P'(S) = 1 - P'(M)$ … refer now to the situation of the population after the 1st screening]

$$P(M|+,+) = \frac{P(+|M)}{P'(+)} \cdot P(M|+) = \frac{P(+|M)P(M|+)}{P(+|M) \cdot P'(M) + P(+|S) \cdot P'(S)} = \frac{0.98 \cdot 0.0317}{(0.98 \cdot 0.0317) + (0.03 \cdot (1 - 0.0317))}$$

$$= \frac{0.031066}{0.031066 + 0.029049} = \frac{0.031066}{0.060115} \cong 0.5168 \approx 52\%$$

What we have just seen can be formalized in the Bayesian approach as follows:

in the iteration n the a-priori probability is taken as the a-posteriori probability of the iteration $n-1$:

$$\boldsymbol{P_n(A_k|B)} = \frac{P(B|A_k)\cdot P_{n-1}\,(A_k)}{\sum_i P(B|A_i)\cdot P_{n-1}\,(A_i)} \quad \text{with } i = 1, \dots, k, \dots, N$$