Statistics for Data Analysis

(an un-exhaustive course on practical Statistics for Physicists)

Luca Stanco

INFN-Padova

luca.stanco@pd.infn.it

https://userswww.pd.infn.it/~stanco/didattica/Stat-An-Dati-Dottorato/

Padova

Dottorato A.A. 2022/23

Three teachers:

1) Luca Stanco (http://www.pd.infn.it/~stanco/CV-2.pdf)
   Particle Physicist (specialized in Neutrinos)
   Slides in: http://www.pd.infn.it/~stanco/didattica/Stat-An-Dati-Dottorato/

2) Tommaso Dorigo (tommaso.dorigo@pd.infn.it)
   Particle Physicist (specialized in Hadron Collider)

3) Denis Bastieri (denis.bastieri@pd.infn.it)
   Astro-Particle Physicist (specialized in Gamma Observation)

One oral examination that deals in two parts:

1) A statistical problem chosen by the student between those illustrated in the course: illustration, critical issues, analysis, results
2) Usual follow up

Generic Layout (much more inside) :

PDF: Probability Density Function(s) 1. Random Variables, Normal Density; Central Limit Theorem; 2. Cumulative Function and Uniform Distribution; 3. Binomial, Poisson, Cauchy and t-Student Functions.	Basic Elements
Probability and Bayes 1. Probability laws; Bayes Theorem for Physicists; Ordering; 2. Posterior probabilities; Credibility Intervals.	Fundamental Concepts
Likelihood and Estimators 1. Chi-Squared and Likelihood functions; Methods of ; 2. Error propagation; Estimators; 3. All on Correlations.	Basic Concepts (T.D.)
Confidence Intervals and Test of Hypothesis 1. Intervals of Confidence, and Statistical Tests 2. H0, H1 and p-values; 3. Cramer-Rao Theorem and Lemma of Neyman-Pearson.	Sophisticated Applications (D.B.)
Applications 1.Feldman & Cousins 2.Monte Carlo's methods and Markov chain 3.Kalman Fiiltering	Applications

Premises, Suggestions, Tricks, Caveats…

- Statistics is a touchy, uncomfortable field
- It is fundamental for the contemporary researchers (physicists et al.)
- It is based on an undefined, circular definition of Probability
  (you will see, I will use quite often the wording "usually")

- Currently, ***Probability*** presents too many Axioms, Postulates and Principles
- Still waiting for a more comprehensive general frame
- Many mistakes in books, articles, internet sites, even recent ones

- But it is applied, and it works usually well in practically all the fields of human rational *(science, finance, work, production …)*

- Statistics has been developed and it is currently under the trust of Mathematics and Economy, i.e. it is currently an "abstract-like" science.
  Physics view in Statistics is taking-up (especially HEP) and it may help a lot !

- Identify the **question** of the problem you are interested to
- Solve the problem underneath, **DO NOT** try to generalize

- Use the up-to-date tool: your computer (!)
  You will NOT understand by reading the slides, but instead following our lessons

11 February 2016: discovery of the Gravitational Waves (and confirmation of Black Holes)

PRL **116,** 061102 (2016) PHYSICAL REVIEW LETTERS week ending 12 FEBRUARY 2016

Observation of Gravitational Waves from a Binary Black Hole Merger

B. P. Abbott *et al.**

(LIGO Scientific Collaboration and Virgo Collaboration)

(Received 21 January 2016; published 11 February 2016)

On September 14, 2015 at 09:50:45 UTC the two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal. The signal sweeps upwards in frequency from 35 to 250 Hz with a peak gravitational-wave strain of 1.0×10^{-21}. It matches the waveform predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than 5.1σ. The source lies at a luminosity distance of 410^{+160}_{-180} Mpc corresponding to a redshift $z = 0.09^{+0.03}_{-0.04}$. In the source frame, the initial black hole masses are $36^{+5}_{-4} M_\odot$ and $29^{+4}_{-4} M_\odot$, and the final black hole mass is $62^{+4}_{-4} M_\odot$, with $3.0^{+0.5}_{-0.5} M_\odot c^2$ radiated in gravitational waves. All uncertainties define 90% credible intervals. These observations demonstrate the existence of binary stellar-mass black hole systems. This is the first direct detection of gravitational waves and the first observation of a binary black hole merger.

* for16 days of running

Solution

The probability of a fake event is: (16/365) * (1/203000) = 2.2 * 10^{-7}

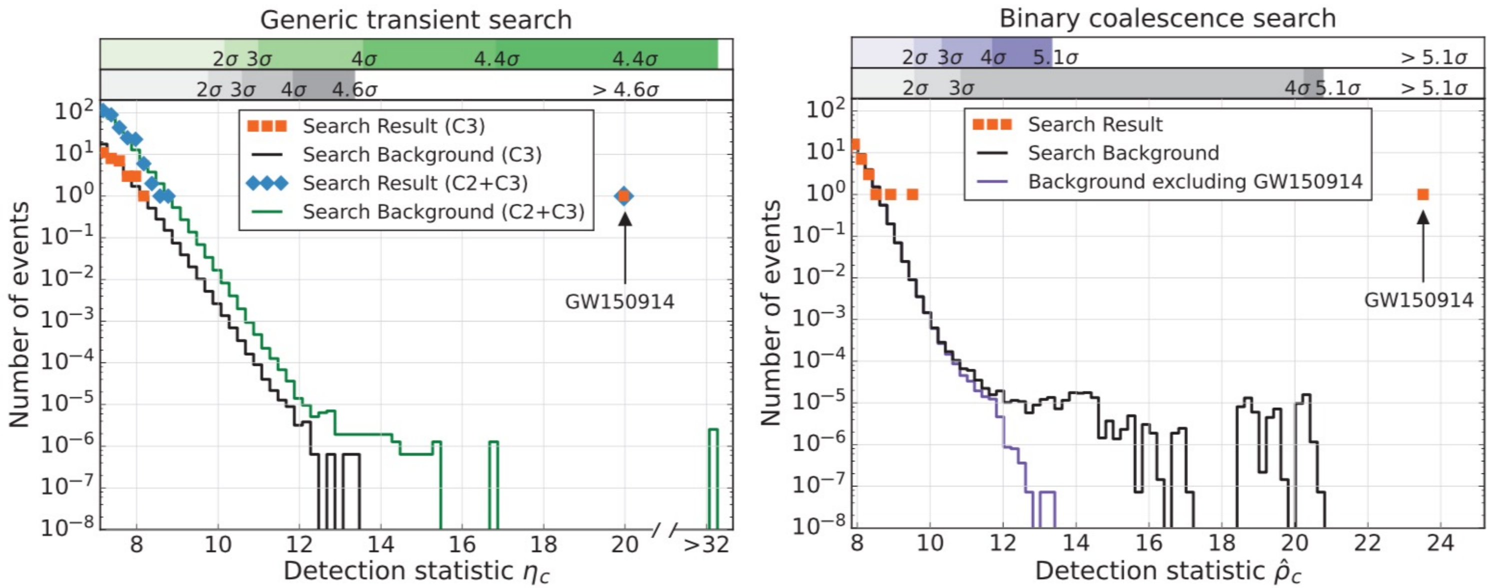
a little less than 5 σ, which corresponds to 5.7 * 10^{-7}

Observation of Gravitational Waves. Its significance is given as
"less than 1 event in 203,000 years (over integration of 16 days of observation), i.e. 5.1 σ"

Binomial with $p = \frac{1}{\frac{16}{365} * \frac{1}{203000}} = 2.16 * 10^{-7}$

(i.e. 5.055 σ at ONE-SIDE !)

Actually, it follows as example from an exact estimation:



50 M hours of CPU…

16 October 2017: discovery of Gravitational Waves from neutron stars collapse, associated to Light

Selected for a Viewpoint in *Physics*

PRL **119,** 161101 (2017) PHYSICAL REVIEW LETTERS week ending 20 OCTOBER 2017

GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral

B. P. Abbott *et al.*[*]

(LIGO Scientific Collaboration and Virgo Collaboration)

(Received 26 September 2017; revised manuscript received 2 October 2017; published 16 October 2017)

On August 17, 2017 at 12:41:04 UTC the Advanced LIGO and Advanced Virgo gravitational-wave detectors made their first observation of a binary neutron star inspiral. The signal, GW170817, was detected with a combined signal-to-noise ratio of 32.4 and a false-alarm-rate estimate of less than one per 8.0×10^4 years. We infer the component masses of the binary to be between 0.86 and 2.26 $M_\odot$, in agreement with masses of known neutron stars. Restricting the component spins to the range inferred in binary neutron stars, we find the component masses to be in the range 1.17–1.60 $M_\odot$, with the total mass of the system $2.74^{+0.04}_{-0.01} M_\odot$. The source was localized within a sky region of 28 $\deg^2$ (90% probability) and had a luminosity distance of 40^{+8}_{-14} Mpc, the closest and most precisely localized gravitational-wave signal yet. The association with the γ-ray burst GRB 170817A, detected by Fermi-GBM 1.7 s after the coalescence, corroborates the hypothesis of a neutron star merger and provides the first direct evidence of a link between these mergers and short γ-ray bursts. Subsequent identification of transient counterparts across the electromagnetic spectrum in the same location further supports the interpretation of this event as a neutron star merger. This unprecedented joint gravitational and electromagnetic observation provides insight into astrophysics, dense matter, gravitation, and cosmology.

DOI: 10.1103/PhysRevLett.119.161101

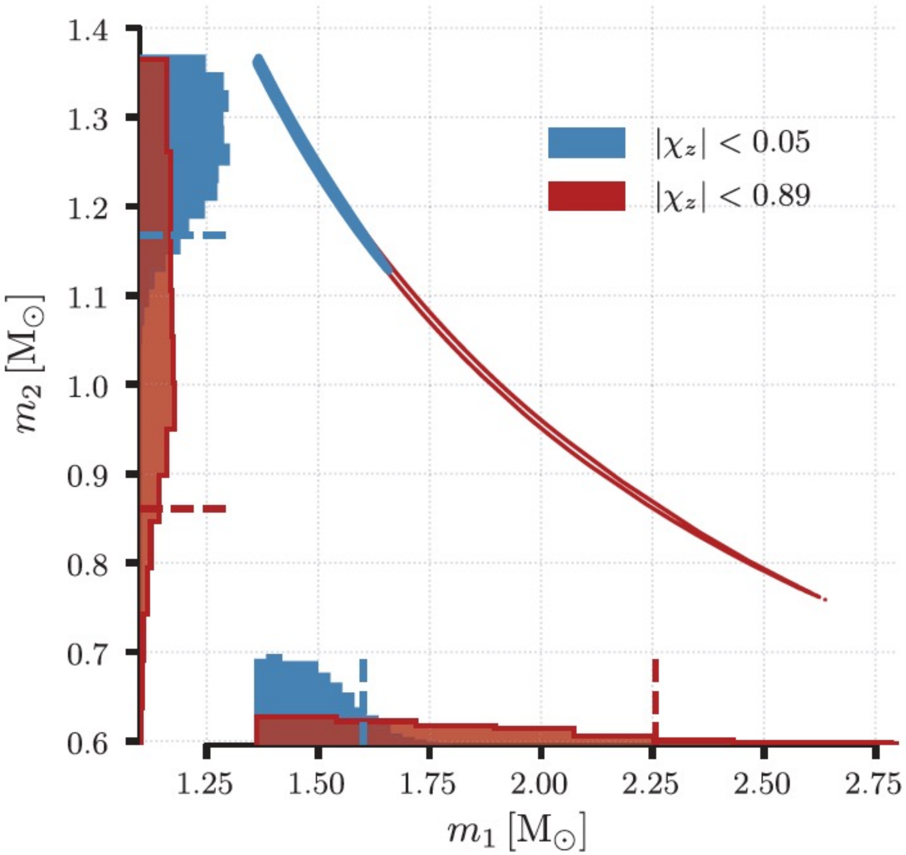


FIG. 4. Two-dimensional posterior distribution for the component masses m_1 and m_2 in the rest frame of the source for the low-spin scenario ($|\chi| < 0.05$, blue) and the high-spin scenario ($|\chi| < 0.89$, red). The colored contours enclose 90% of the probability from the joint posterior probability density function for m_1 and m_2. The shape of the two dimensional posterior is determined by a line of constant $\mathcal{M}$ and its width is determined by the uncertainty in $\mathcal{M}$. The widths of the marginal distributions (shown on axes, dashed lines enclose 90% probability away from equal mass of $1.36 M_\odot$) is strongly affected by the choice of spin priors. The result using the low-spin prior (blue) is consistent with the masses of all known binary neutron star systems.

Masses of the two collapsing objects

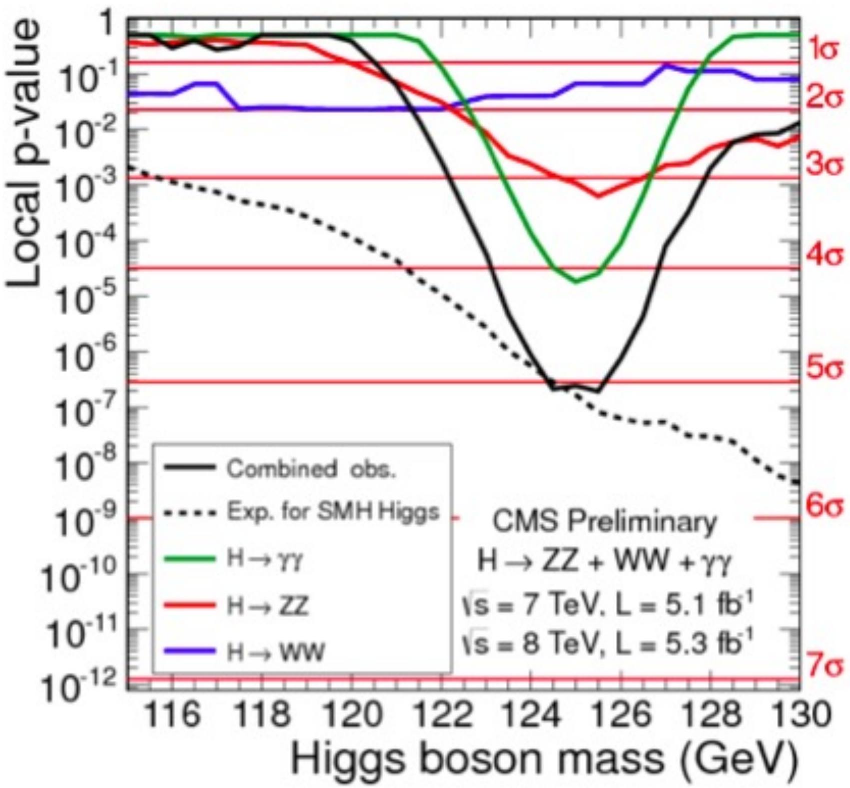
$\mathcal{M}$ is the chirp mass:

$$M = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} = 1.188^{+0.004}_{-0.002}$$

in solar masses

4 July 2012: discovery of the ”Higgs”

Characterization of excess near 125 GeV



adding high sensitivity, but low mass resolution WW

comb. significance: **5.1 σ**

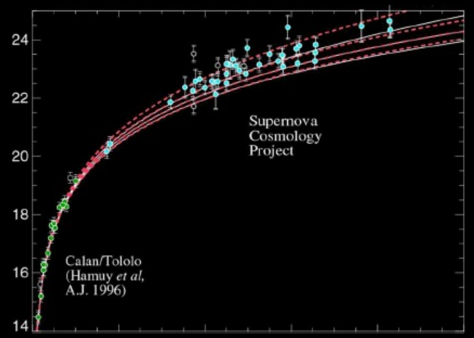
expected significance for SM Higgs: 5.2 σ

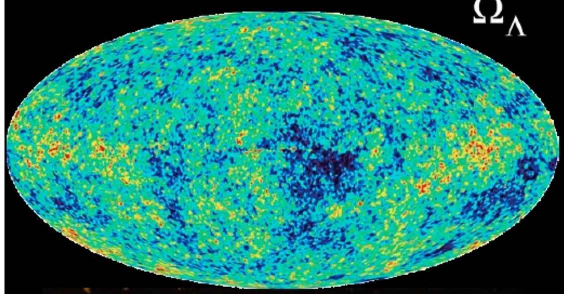
July 4th 2012 The Status of the Higgs Search J. Incandela for the CMS COLLABORATION

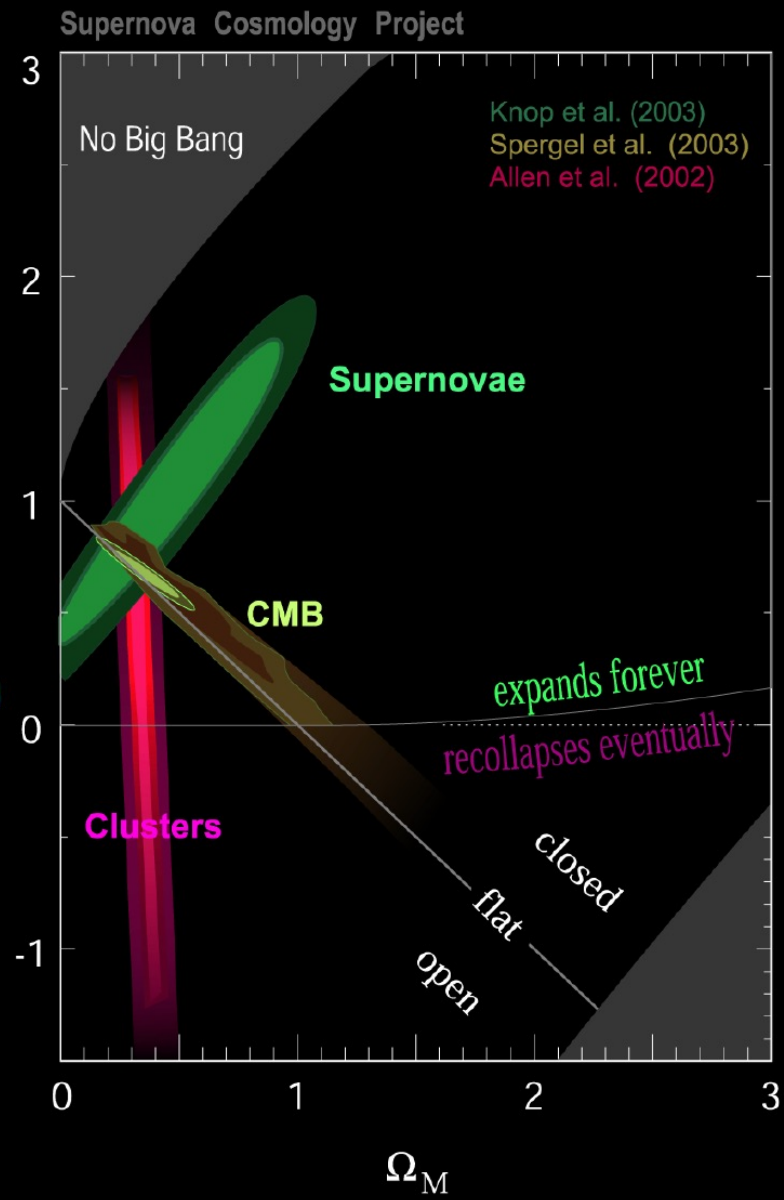
78

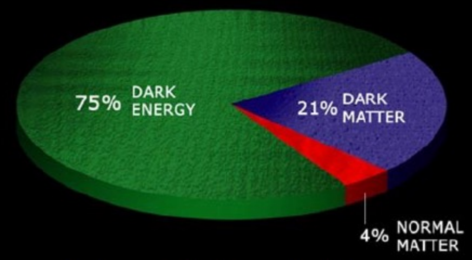
Standard Model for Cosmology

Confirmed by LSS and CMBR fluctuations









75 ± 2% Dark Energy

25 ± 3% Matter

0.5% Bright Stars

Matter (25%):

20% Dark Matter

4.4% Baryons

0.3% νs

We do not understand 96% of the universe!

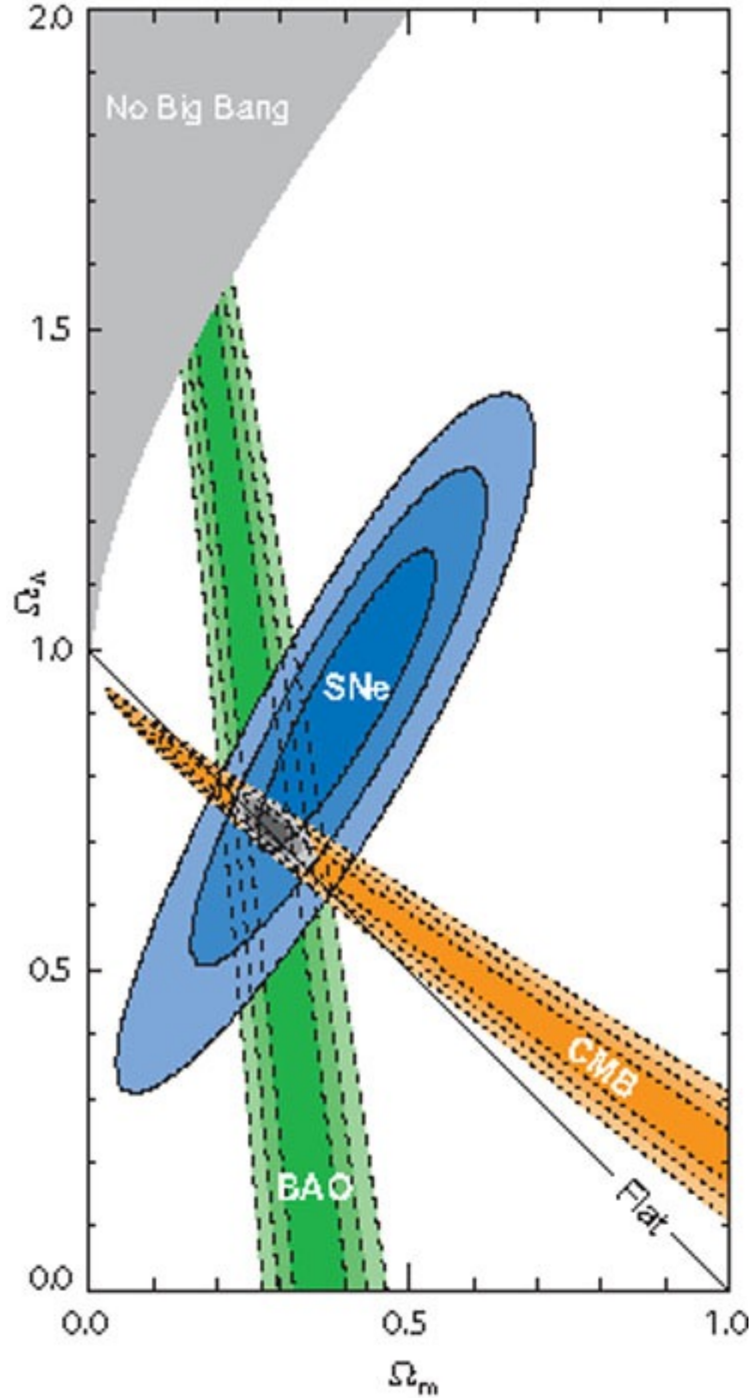
The current major problem in Fundamental Physics

Cosmology Concordance Model:

$\Omega_M+\Omega_\Lambda+\Omega_\kappa \approx \Omega_\Lambda+\Omega_\kappa$ 1 ($\rightarrow$k=0, inflation)

Legend: SN-Ia: Supernove of type Ia
CMB: Cosmics Microwave Background
BAO: Baryon Acoustic Oscillation
Ω_M: density of the mass of the universe
Ω_Λ: density of vacuum (energy
Ω_κ: free parameter

Anomalies drive scientific discoveries !



http://arxiv4.library.cornell.edu/pdf/1004.1711v1

Some suggestions for References

By physicists, for physicists

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

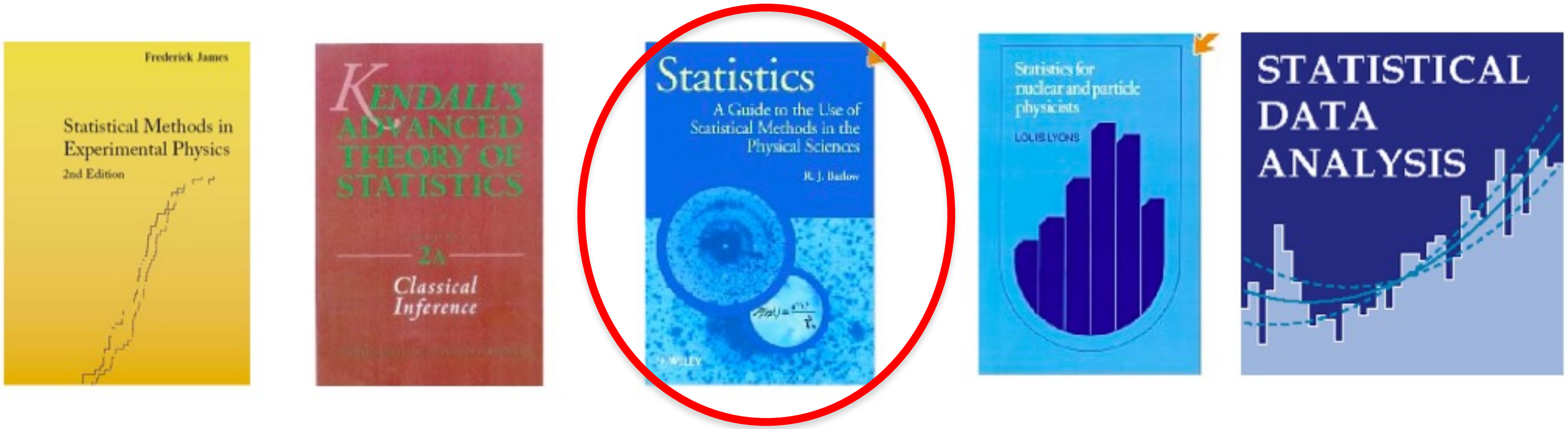
R.J.Barlow, A Guide to the Use of Statistical Methods in the Physical Sciences, John Wiley, 1989;

F. James, Statistical Methods in Experimental Physics, 2nd ed., World Scientific, 2006;

- W.T. Eadie et al., North-Holland, 1971 (1st ed., hard to find);

S.Brandt, Statistical and Computational Methods in Data Analysis, Springer, New York, 1998.

L.Lyons, Statistics for Nuclear and Particle Physics, CUP, 1986.



L. Wasserman, "All of Statistics. A Concise Course in Statistical Inference", Springer, ed. 2004

References *(cont)*

1. M. Loreti (http://wwwcdf.pd.infn.it/labo/INDEX.html)
   in 90ties

1. G. D’Agostini (http://www-zeus.roma1.infn.it/~dagos/perfezionamento.ps
   http://www-zeus.roma1.infn.it/~dagos/prob+stat.html)
   in 90ties

3. G. Bohm, G. Zech (http://www-library.desy.de/preparch/books/vstatmp_engl.pdf)
   Febraury 2010

4. O. Behnke, K. Kroninger, G. Schott, T. Schoner-Sadenius:
   “Data Analysis in High Energy Physics”, 2013

5. I. Narsky, F.C. Porter:
   “Statistical Analysis Techniques in Particle Physics”, 2013

References *(cont)*

Some slides partially developed from:

- Larry Gonick & Woollcott Smith, "The cartoon guide to Statistics"

- PDG, Particle Data Group, C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C, **40**, 100001 (2016)
  (http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-probability.pdf
  http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf
  http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-monte-carlo-techniques.pdf)

-Terascale Statistics School, DESY, Sep. 2008
  http://indico.desy.de/contributionListDisplay.py?confId=1149
-Kyle Cranmer, CERN Academic Training, Feb. 2009
  http://indico.cern.ch/conferenceDisplay.py?confId=48425 and following
- Louis Lyons, Gran Sasso Lectures, Sep. 2010
  http://cfa.lngs.infn.it/seminars/practical-statistics-for-experimental-scientists
- INFN School of Statistics 2013
  https://agenda.infn.it/conferenceDisplay.py?confId=5719
- INFN School of Statistics 2015
  https://agenda.infn.it/conferenceDisplay.py?confId=8095

Data analysis

THE GATHERING, DISPLAY, AND SUMMARY OF DATA;

Probability

THE LAWS OF CHANCE, IN AND OUT OF THE CASINO;

Statistical inference

THE SCIENCE OF DRAWING STATISTICAL CONCLUSIONS FROM SPECIFIC DATA, USING A KNOWLEDGE OF PROBABILITY.

1^0 step
(may we assume that you know what an histogram is ?)

2^0 step
(correct definitons of what you can think already to know)

3^0 step
(all about what you will learned)

PROBABILTY

e.g. dice's roll

Given P(5) = 1/6 what is the P(20 5's out of 100 trials) ?

similarly…

If unbiassed, what is P(n evens out of 100 trials) ?

THEORY→ DATA

STATISTICS

Given 20 5's out of 100, what is P(5) ?

Parameter Determination

Observe 65 evens in 100 trials, is it unbiassed ?

Goodness of fit

Or is P(even) = 2/3 ?

Hypothesis testing (and Inference)

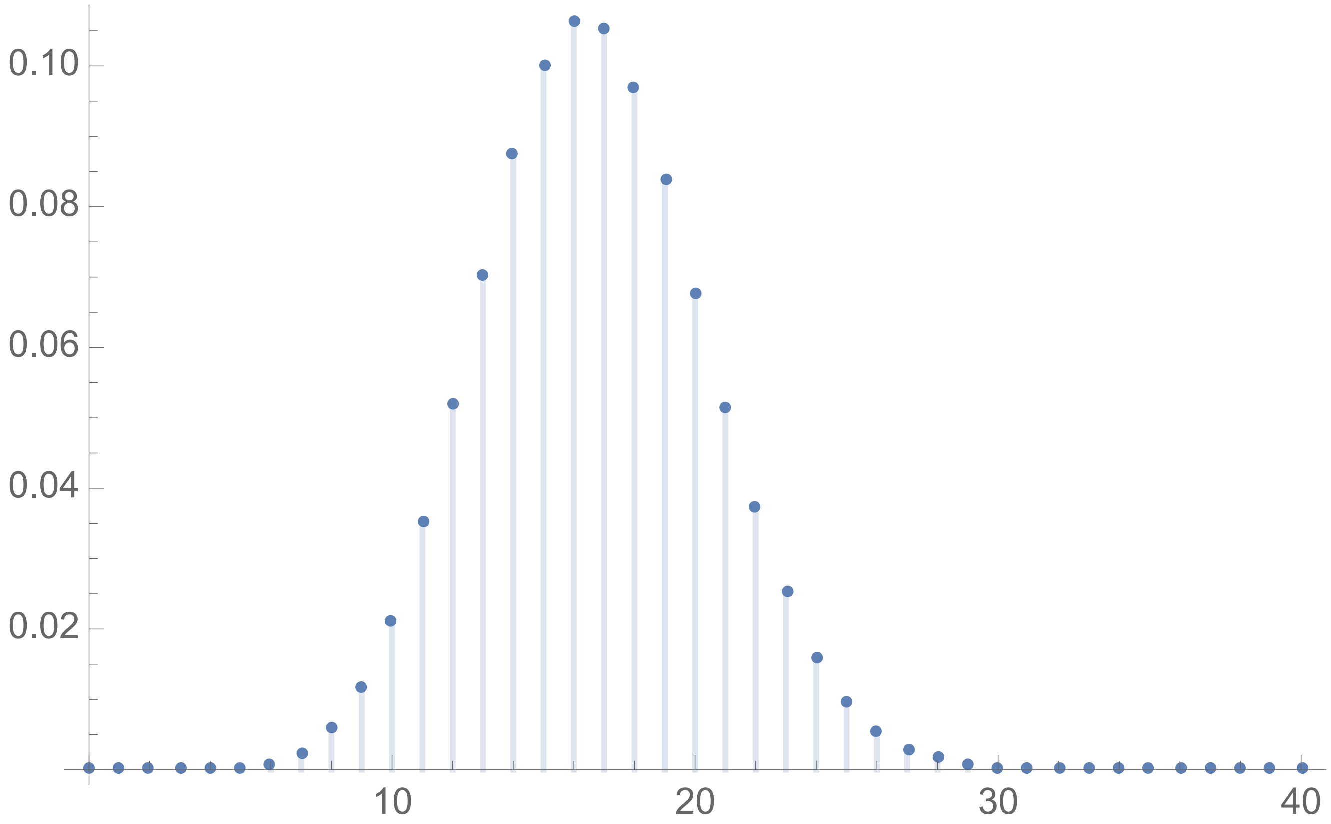
DATA→ THEORY

Prediction moves forwards →

← Inference moves backwards

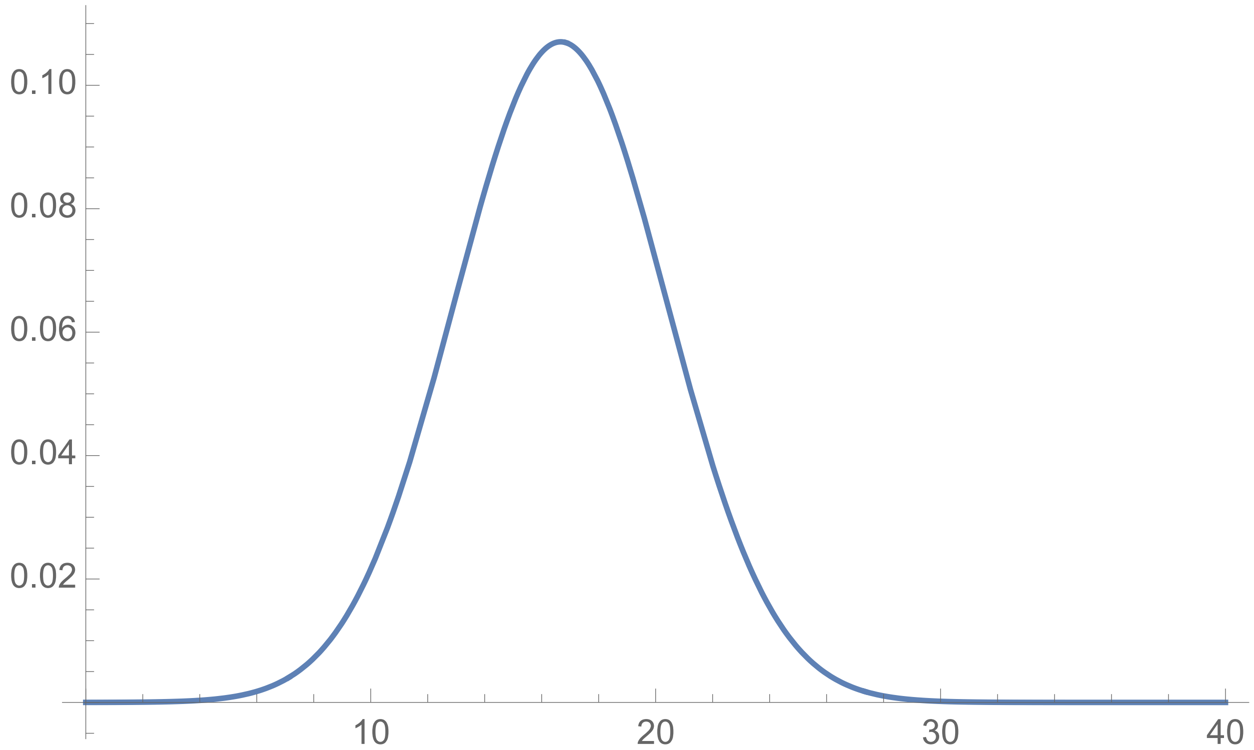
Final question: how to check you have learnt the maximum from your measurement and in a correct unbiassed way ? ➔ error evaluation

Solution: binomial distribution for p=1/6 and 100 trials.
at 20 the probability is about 6.5%



Done with Mathematica:
In[..] *DiscretePlot[PDF[BinomialDistribution[100, 0.16667], x], {x, 0., 40.}]*

Note that binomial is very similar to a gaussian with
μ=100*(1/6)=16.667 and σ^2=100*(1/6)*(5/6)=13.8889



In[..] *Plot[PDF[NormalDistribution[16.6667, 3.727], x], {x, 0., 40}]*

Consider

$$x = 22 \pm 5$$

What is the meaning of:

x ?

22 ?

5 ?

Theory and Real-Life are put together by introducing the quantity:

PDF : **Probability Density Function(s)**

$$P(A) = \int_A p.d.f.\ (\vec{x}) \bullet d\vec{x}$$

Where **P(A)** is the “probability” of the “event” **A** belonging to the space of the events Ω.
Then it should also be:

$$P(\Omega) = \int_\Omega p.d.f.\ (\vec{x}) \bullet d\vec{x} = 1$$

And one can assume (“axiomize”) **P(A)>0** for every event **A**, i.e. **p.d.f. (x)>0** for every **x**

And also the axiom that $P(A \cup B) = P(A) + P(B)$ if **A** and **B** are disjoints, i.e. $A \cap B = 0$

In reality ALL can be a PDF:
- ***just ask you the right question,***
- ***find the right random variable (i.e. the space Ω),***
- ***and find out the corresponding PDF***

A word on the concept of **RANDOM VARIABLE**:

The possible **outcome** of an experiment/measurement/question, which involve one or more random processes.
The outcome is called “**event**”.
They are usually “**elementary**” (opposite to “composite”) events
Very often they are “**independent**” events (in probabilistic sense)

Very relevant:

Any combination of random variables is itself a random variable.
Therefore, the PDF is itself a random variable, with its own PDF !
The combinations are usually called **composite** events.

The correct evaluation of the PDF allows you the to estimate the “sensitivity” of your measurement (i.e. estimate the error)

Before to elucidate further the concept of PDF, probability, inference etc.
we will concentrate on basic technicalities of the PDF.

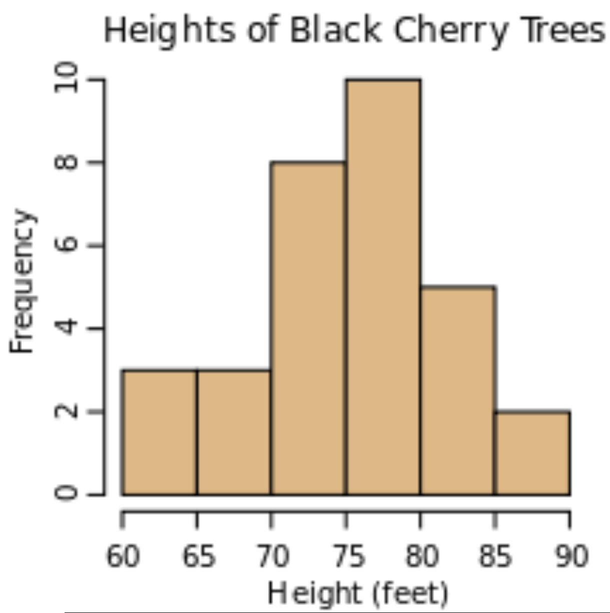
For a long time researchers have **just** tried out to evaluate the characteristics of the PDF.

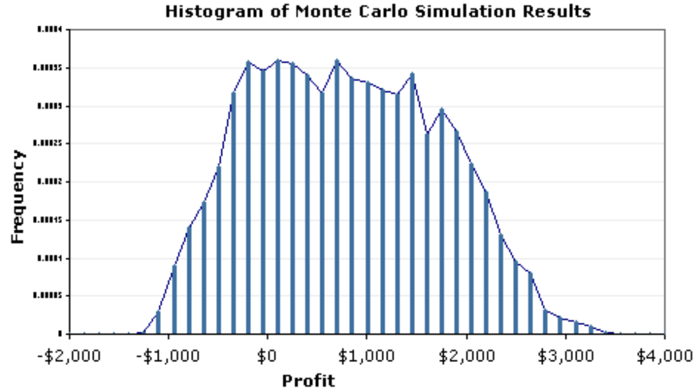
1)Make a (set of) measurement(s)
2)Construct a frequency plot
3)Extract the most "*probable*" value
4)Identify the latter with the "***result***" of the measurement
5)Extract an indication of the '"*dispersion"* of the frequency plot
6)Identify the latter with the "***error***"
7)Add the "*error*" to the "*result*"

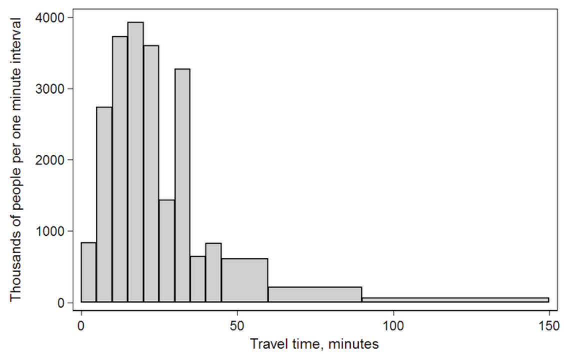
AAA: the PDF is the probability density of the **single** random variable.
Usually several measurements of the same quantity are made.
Each of them owns its own PDF.
Usually one assume that the measurements are INDEPENDENT and done in the same way, i.e. the single PDF is always the same.
Then, the *frequency plot* describes this unique PDF.
The final "error" has to be given by considering ALL the information collected.
E.G. "error"/$\sqrt{n}$
WHY THIS HAS TO BE CLEARLY UNDERSTOOD ? see later…

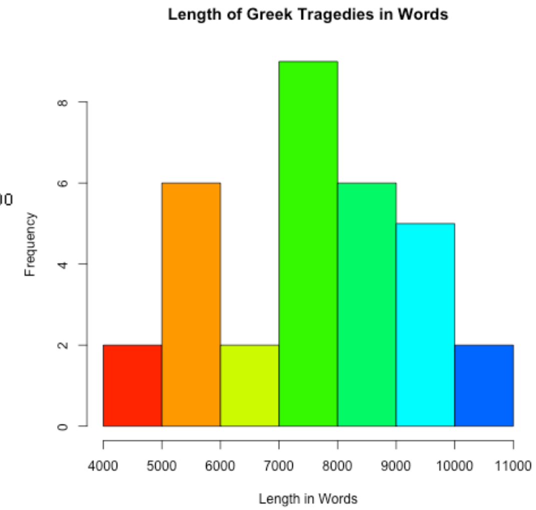
DATA DESCRIPTION

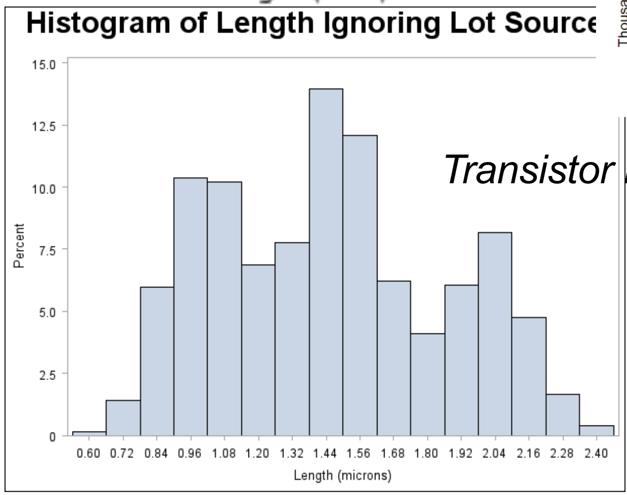
Usually, for the frequency plot, one find histograms like:



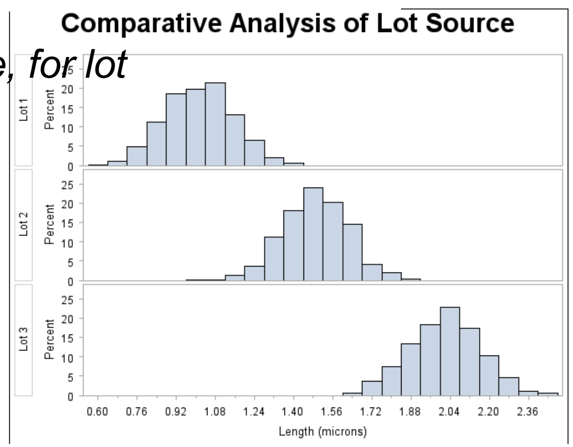








Transistor response, for lot



DATA DESCRIPTION ⇒"result"

(arithmetic) mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{OR} \quad \sum_{i=1}^{n} \frac{x_i}{n}$$

(sometime, wrongly indicated as x ⇒ μ)

Going to interval binning of measure:
(histogram in L bins):

$$\bar{x} = \sum_{j=1}^{L} x_j \frac{n_j}{n} = \sum_{i=1}^{L} x_j f_j(x_j)$$

and for the "continuous" case:

$$\bar{x} \Rightarrow E(x) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

Expectation value of the function f(*x*)

median $$x_{med} \perp \quad \int_{-\infty}^{x_{med}} f(x)dx = \int_{x_{med}}^{+\infty} f(x)dx$$

The median is NOT sensible to OUTLIERS, i.e. to the extreme values, not characteristic of the majority of data

mode ≡ maximum of the distribution

DATA DESCRIPTION ⇒"error"

The question is: how much "dispersed" is the frequency plot around the "result" value ?

First possibility: identify some percentile range, such that the integral in [a,b] is equal to the percentile $\int_{-a}^{+b} f(x)\,dx$

Example: identify the INTERQUARTILE range, IQR, by finding the median and the two sub-median $\int_{-a}^{+b} f(x)\,dx = 0.5$

$$\int_{-\infty}^{x_{med}} f(x)dx = \int_{x_{med}}^{+\infty} f(x)dx \quad \rightarrow x_{med}$$

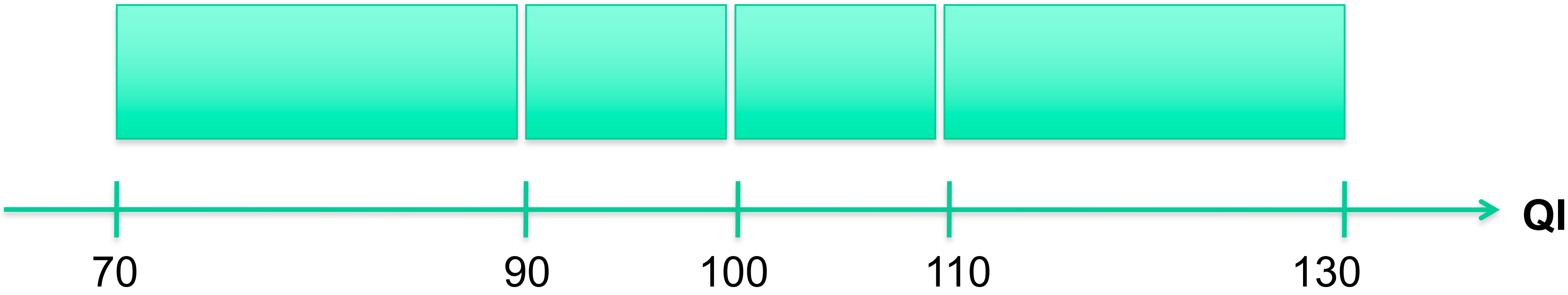
$$\int_{-\infty}^{a} f(x)dx = \int_{x_a}^{x_{med}} f(x)dx \quad \rightarrow x_a$$

$$\int_{x_{med}}^{b} f(x)dx = \int_{b}^{+\infty} f(x)dx \quad \rightarrow x_b$$

IQR=x_b-x_a

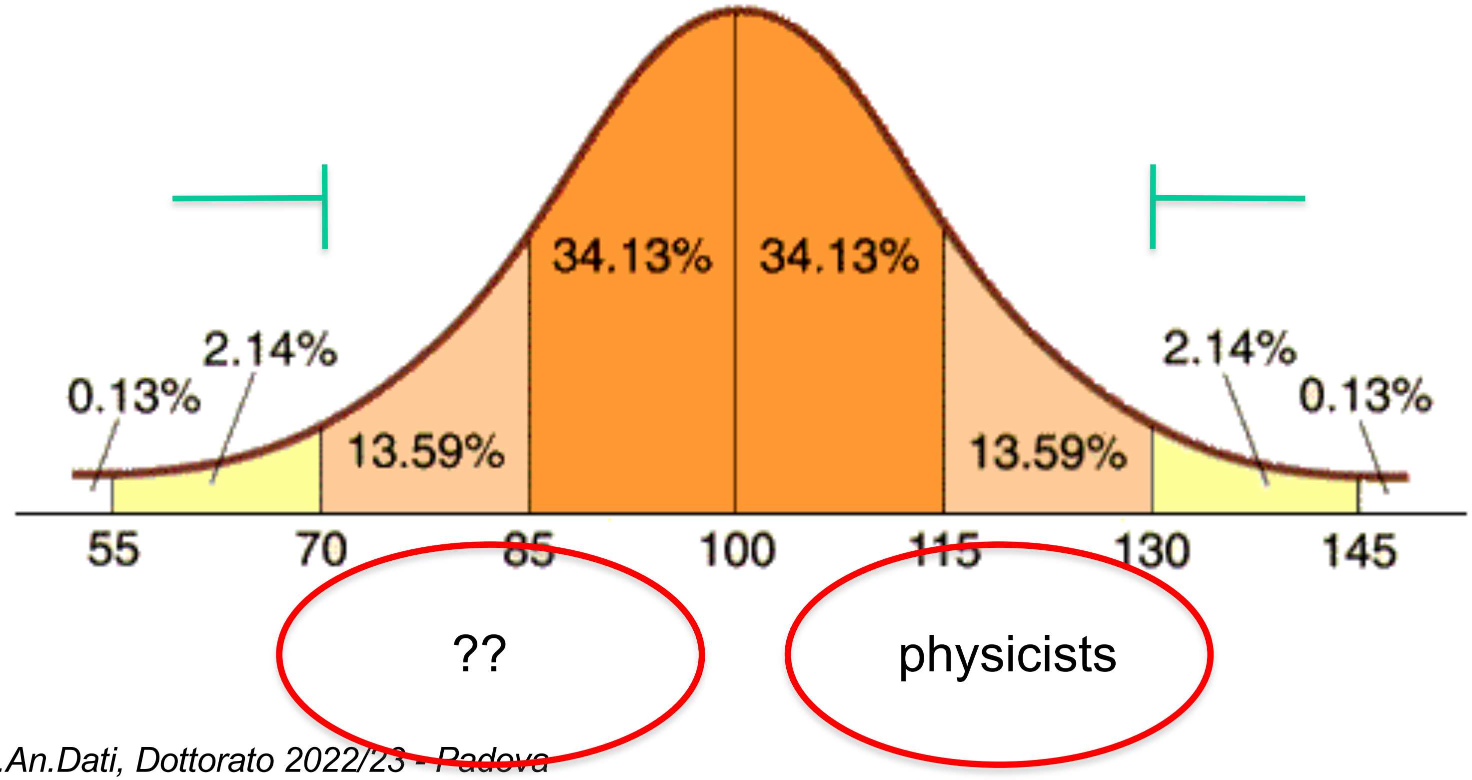
In practice, you have two find the 4 regions where the integrated PDF is equal to 0.25

QI distribution of human population :



Usually a relative QI is computed by considering the median at 100

IQR = 110 – 90 = 20



The nice plot of the previous slide introduces us to the GAUSSIAN distribution.

Usually one compute the **VARIANCE from the Mean Square Deviation**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

and the **STANDARD DEVIATION**

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

IF s → σ in the Gaussian PDF: $G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Passing to the continuous and going from "deviations" $z_i=\bar{x}-x_i$ to the probability density for the random variable x:

$$\sigma^2 \equiv E(x^2) - E^2(x) = \int_{-\infty}^{+\infty} z^2 \cdot f(z)\,dz$$

Expectation value of the 2° moment of the function f(x)

It is useful to compute the "distance" of a single measurement from the mean as the ***number of standard deviations from the mean***

$$z_i = \frac{x_i - \bar{x}}{s}$$

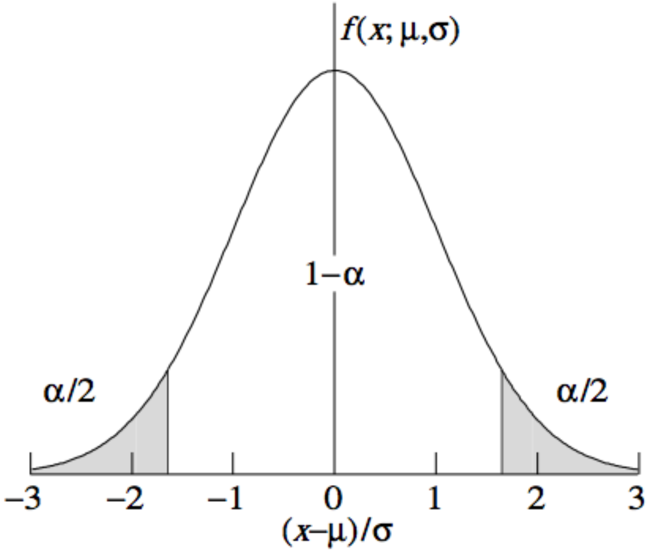


Figure 32.4: Illustration of a symmetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by α, are as shown.

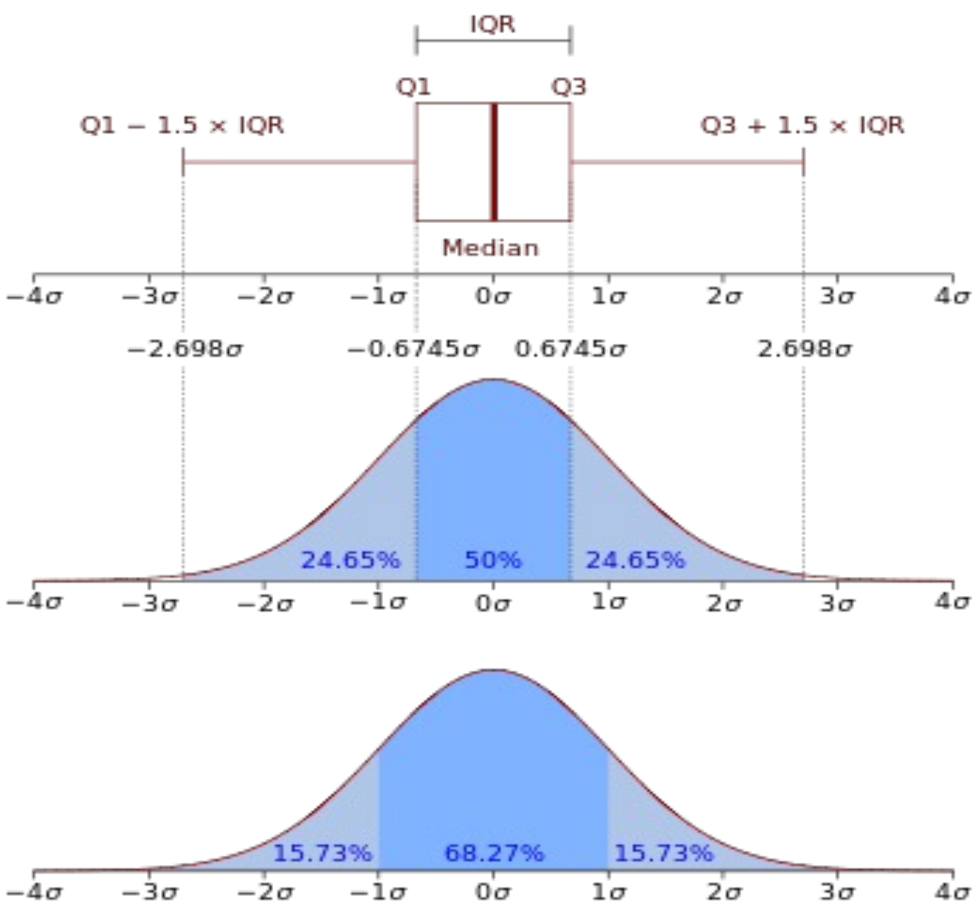
$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Table 32.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
4.55×10^{-2}	2σ	0.1	1.64σ
2.7×10^{-3}	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ

When reporting physical results one usually talk of "CONFIDENCE INTERVALS"
- at 1 sigma
- at 90% of CONFIDENCE LEVEL
- at 95% of C.L.

DATA DESCRIPTION ⇒"error"

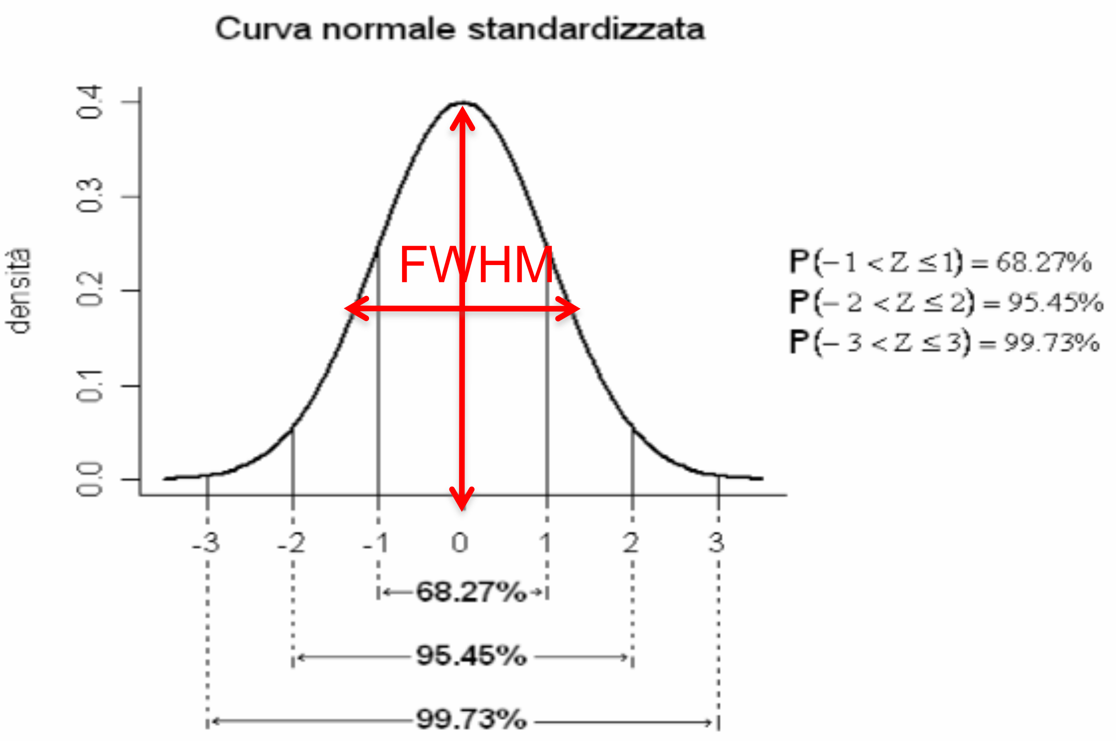


Modern approach:
•Define a *Confidence* ***Limit***, C.L. (how much probability you like to integrate)
•Find a "**centralized**" *Confidence* ***Interval***, C.I., such that P in [a,b] = C.L.
•Describe the result as, e.g. x_{max} in in [a,b], where x_{max} corresponds to P_{max}

Some useful characteristics of NORMAL DISTRIBUTION *(Loreti 8.2)*

FWHM: Full Width Half Maximum

Amplitude of the interval along the points x_1 e x_2 of abscissa $\mu \pm \sigma\sqrt{2\ln 2}$



It comes out:

$$2\sigma\sqrt{2\ln 2} \approx 2.35\sigma$$

Very useful for evaluations "de visu" !

WHY is the GAUSSIAN so relevant as PDF ?

➔ Theorem of the **CENTRAL LIMIT**
(De Moivre in 1733, dead and resurrected by Laplace in 1812,
dead and resurrected in the first years of XX century)

- Suppose you make several (***n***) measurements, each one described by a unique unknown PDF, f(x)
- Suppose that for the PDF mean μ and variance σ^2 exist
  (this it is not always true, e.g. the Breit-Wigner mean and variance do not exist)
- Compute the cumulative PDF of the ***n*** measurements, g(x)
  *(this correspond to the multiplicative convolution of **n** PDFs, see later)*
- Then, for ***n*** "sufficiently" large, g(x) **IS** the GAUSSIAN PDF
  with mean μ and variance σ^2/n !

Demonstration is tedious, but it is a matter of fact that observations fully support the result of the theorem: by accumulating more and more measurements ANY kind of cumulative PDF will behave more and more as a Gaussian.

The **C**entral **L**imit **T**heorem is the unofficial sovereign of probability theory.
However it created/creates a lot of confusion.

Anybody thinks that it can be applied everywhere anytime.
Even more relevant the fact that one usually compute the “error” as the standard deviation, i.e. its “estimator” from the Mean Square Deviation, whatever be the original PDF.

This is badly wrong !

Almost nobody pay attention to the following:

The **CLT** may induce researchers to assume as “error” the standard deviation of the Gaussian density, i.e. the 68% of C.L.

Let us repeat:
the error is usually given by the range that corresponds to the 68% of the PDF of the random variable.

However it is usually NOT true that for a PDF its Variance provide a range of 68% !
This (un)property is called (un)**coverage**.

Very tricky: the convolution of *n* PDF corresponds to a Gaussian, but if you interested to estimate the error of the single PDF, its σ does not corresponds to 68%

DATA DESCRIPTION ⇒"other semi-qualitative descriptions of the PDF"

In general, for almost every PDF the expectation values of order ***n*** can be computed. They are called, the *moments* α_n:

$$\alpha_n \equiv E\left[x^n\right] = \int_{-\infty}^{+\infty} x^n \cdot f(x)\,dx$$

and the *central moments*: $m_n \equiv E\left[(x-\mu)^n\right]$ where μ:mean

Then, two more (obsolete) quantities are defined:

SKEW: $\gamma_1 = \dfrac{m_3}{\sigma^3}$ *(possible asymmetry of the PDF)*

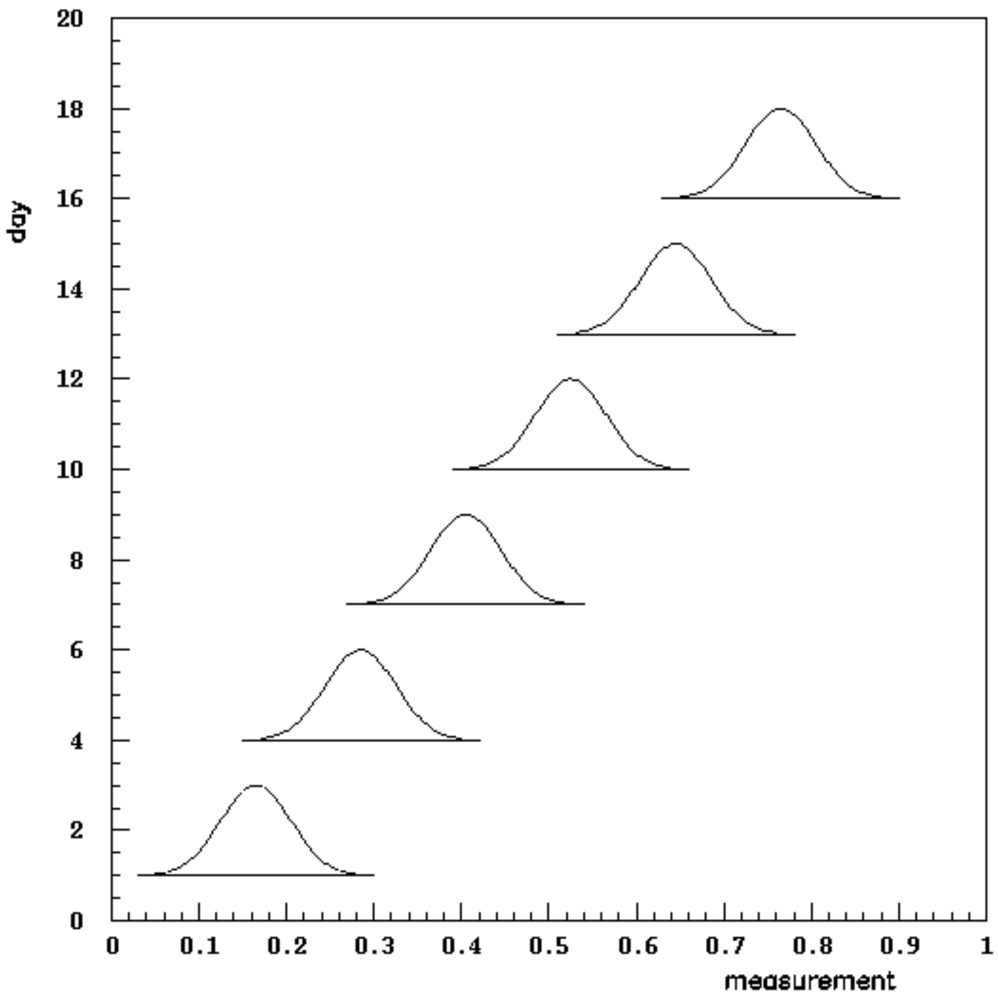
KURTOSIS: $\gamma_2 = \dfrac{m_4}{\sigma^4} - 3$ *(wideness of the tails with respect to the Gaussian that has γ_2=0 by construction)*

$\gamma_2 > 0$ *: leptokurtic distribution (wider tail than G, e.g. Cauchy/Breit-Wigner)*

$\gamma_2 < 0$ *: platykurtic distribution (more centralized that G, e.g. box PDF))*

DATA DESCRIPTION ⇒"parametrized" PDF

Often it happens that the set of measurements is taken as function of some parameters. Think e.g. to take a measurement every day and its results varies linearly with the day.



- The PDF of each-day measurement is assumed to be the same.
- However one is interested in evaluating the dependence law, f(day).
- We, logically, introduce some unknown parameter in the PDF.
- In the illustrated example: μ(day)
- We take a PRINCIPLE, e.g. the **Least Square Errors**, or the "*Maximum Probability*" (MLE):
- One introduce a new form of "Probability": the **LIKELIHOOD**
- Define an "ESTIMATOR", i.e. a function of the measurements, to extract f(x) and compute the "estimate"
- Usually, dispersion may not be unique

(Loreti 8)

The UNIFORM distribution

$$\begin{cases} f(x) = 0 & \text{per } x < a \text{ e per } x > b; \\ f(x) = \dfrac{1}{b-a} = \text{cost.} & \text{per } a \le x \le b. \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \begin{cases} 0 & \text{per } x < a; \\ \dfrac{x-a}{b-a} & \text{per } a \le x \le b; \\ 1 & \text{per } x > b. \end{cases}$$

Cumulative function

$$\begin{cases} E(x) = \dfrac{a+b}{2} \\ \mathrm{Var}(x) = \dfrac{(b-a)^2}{12} \end{cases}$$

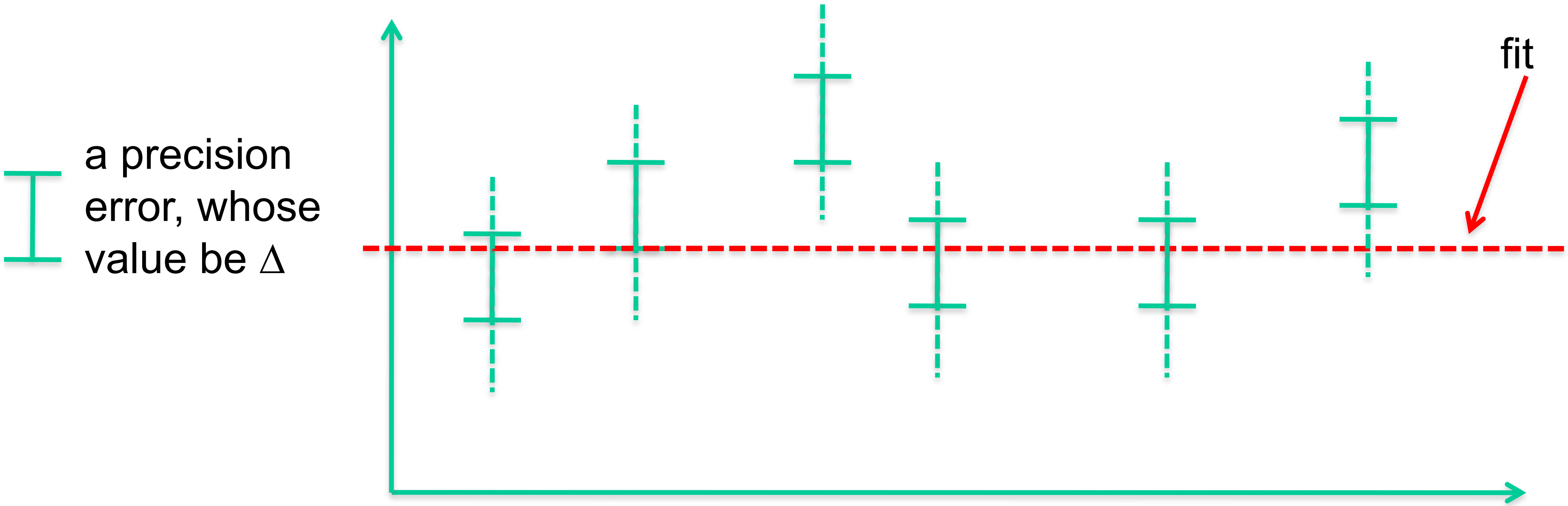
$$\sigma^2 = E(x^2) - E^2(x) = \int_{-\infty}^{+\infty} \frac{x^2}{b-a}dx - \left[\frac{a+b}{2}\right]^2 = \frac{1}{b-a} \bullet \left.\frac{x^3}{3}\right|_a^b - \frac{(a+b)^2}{4}$$

$$= \frac{1}{3}\frac{b^3-a^3}{b-a} - \frac{(a+b)^2}{4} = \frac{1}{12}\left[4\left(a^2+ab+b^2\right) - 3(a^2+2ab+b^2)\right]$$

For a physicist it is important to keep memory that:

$$\sigma = \Delta / \sqrt{12}$$ $\approx 0.3\ \Delta$

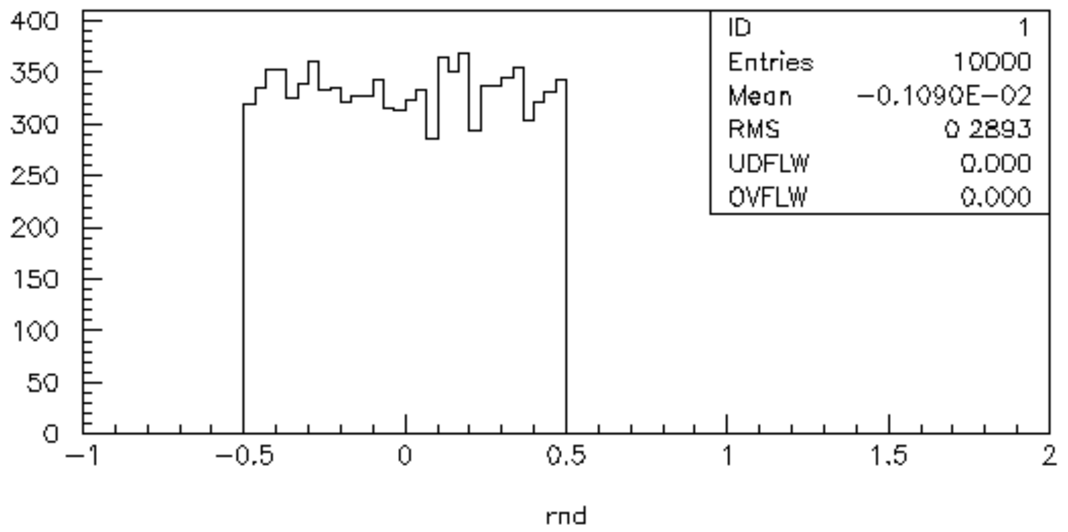
When we deal with n measurements, each with uniform distribution, We have to use the variance so defined. Example:

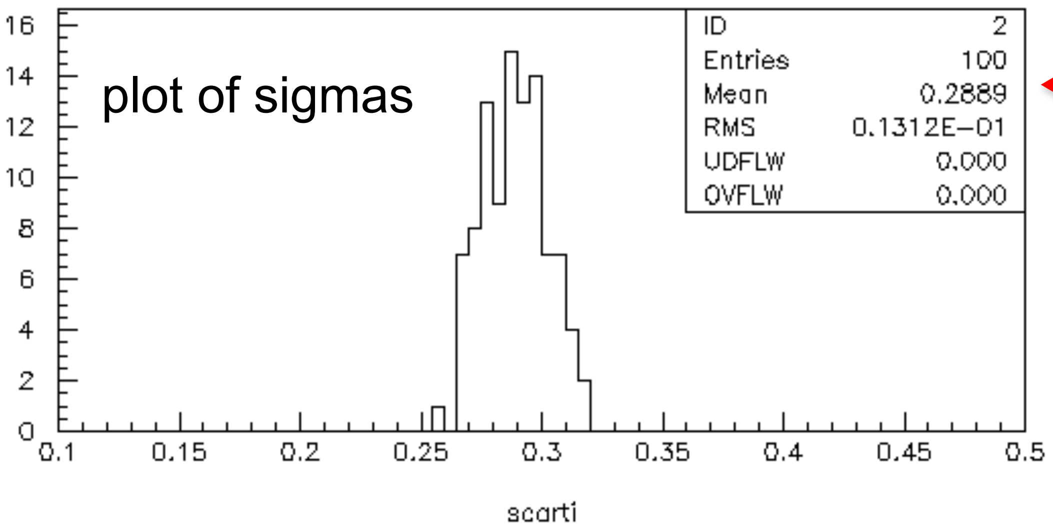


If the σ is mistakened the fit ***result*** will be wrong, i.e. its final error estimation

Demonstration:
- random uniform generation in the interval of width 1
- compute the distribution of residuals with respect to 0,
- consider 100 sets of samplings of “measurements”..

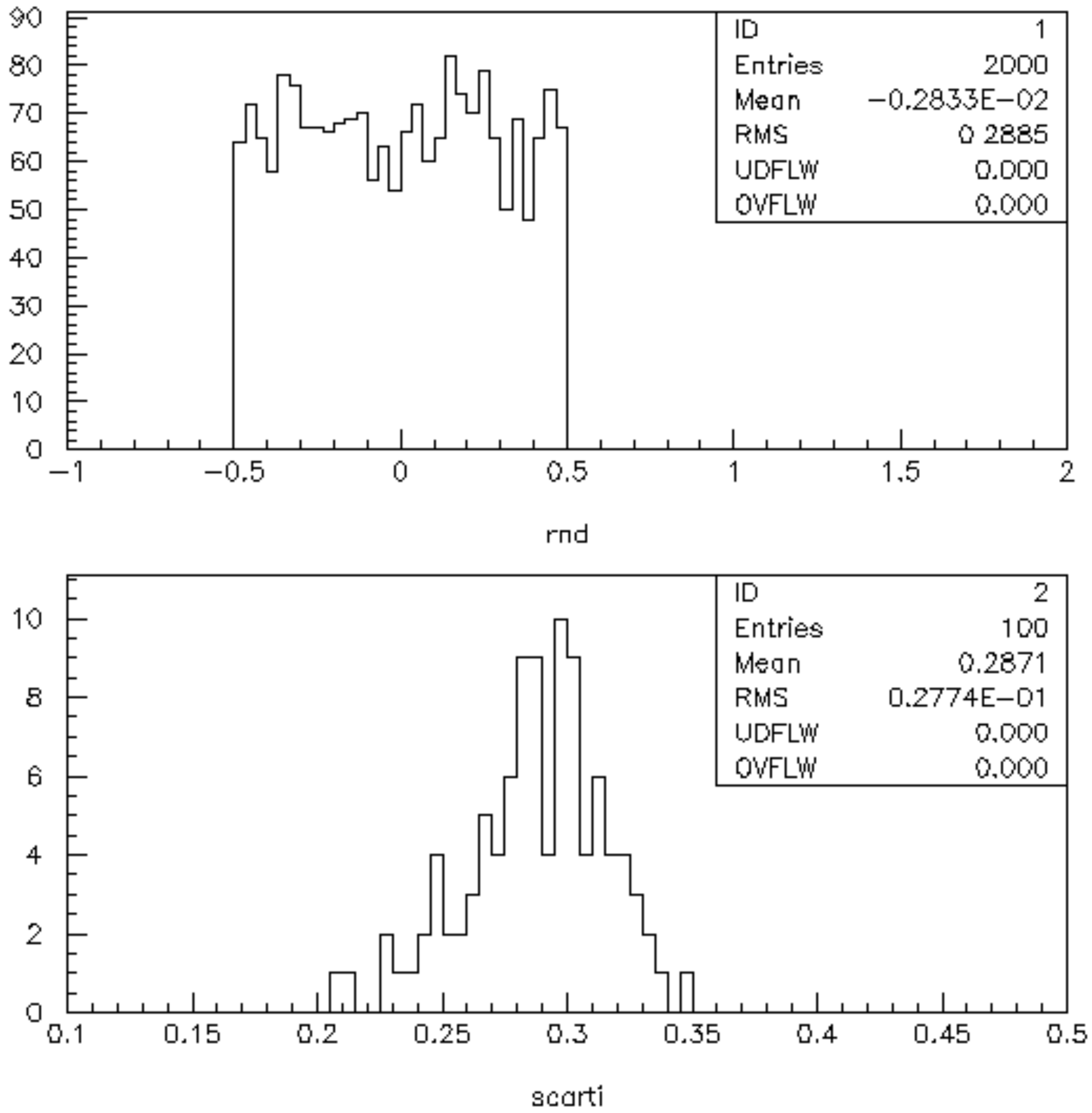
$$\sigma = \sqrt{\frac{(z - z_i)^2}{N - 1}}$$





!

Obviously the sampling of measurements is not large.
If e.g. we consider 20 measurements the distribution of the sigmas is :



Moreover, there is another critical issue. Compute: $\int_{-\sigma}^{+\sigma} f(x) \cdot dx$

i.e. the probability between $\pm \frac{\Delta}{\sqrt{12}}$

One obtain $\int_{-\Delta/\sqrt{12}}^{+\Delta/\sqrt{12}} \frac{1}{\Delta} \cdot dx = \frac{2}{\sqrt{12}} = 0.578 \neq 0.683$

The probability to measure the true value in ±σ is NOT equal to 68%, ie. what one usually assume !

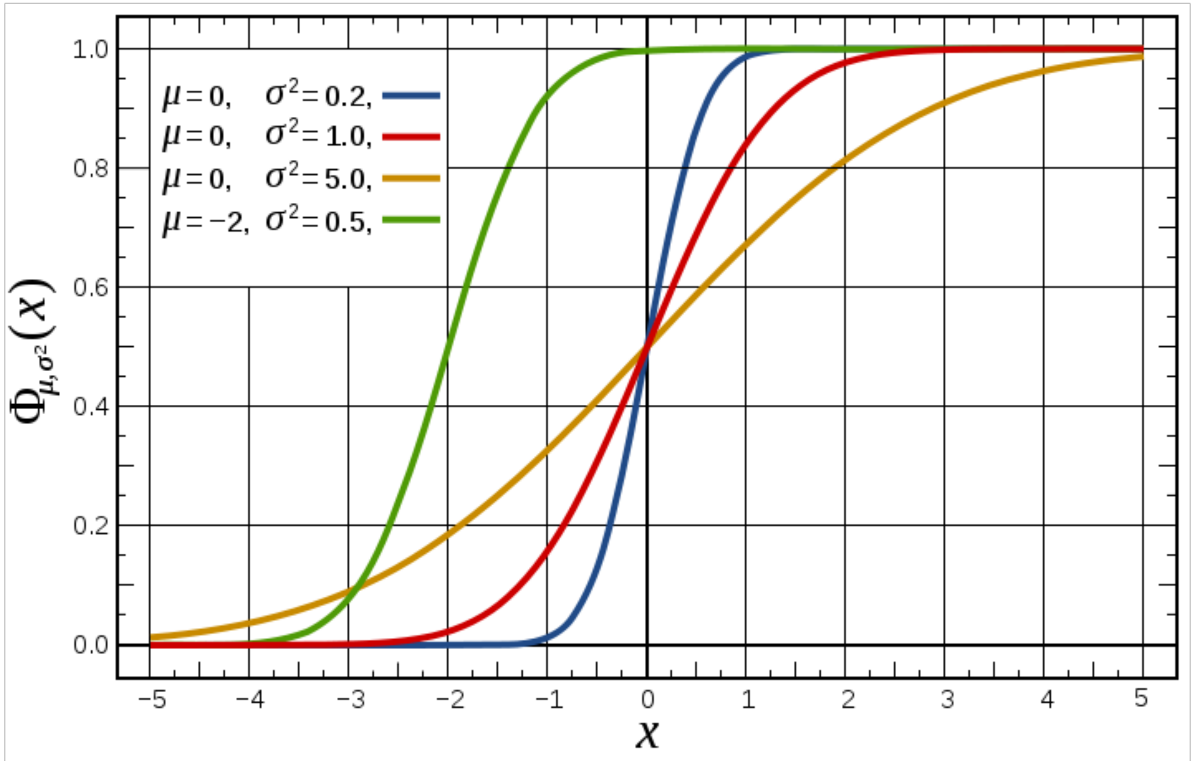
Very important is the concept of CUMULATIVE Distribution FUNCTION
(the probabiliy that x ≤ a)

$$F(a) = \int_{-\infty}^{a} f(x)\, dx \; .$$

For the NORMAL DISTRIBUTION the solution is:

if we define the function ERF: $$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

$$F(x; 0, 1) = \tfrac{1}{2} \left[1 + \mathrm{erf}(x/\sqrt{2}) \right]$$



Every CUMULATIVE function F(x) owns a UNIFORM distribution !

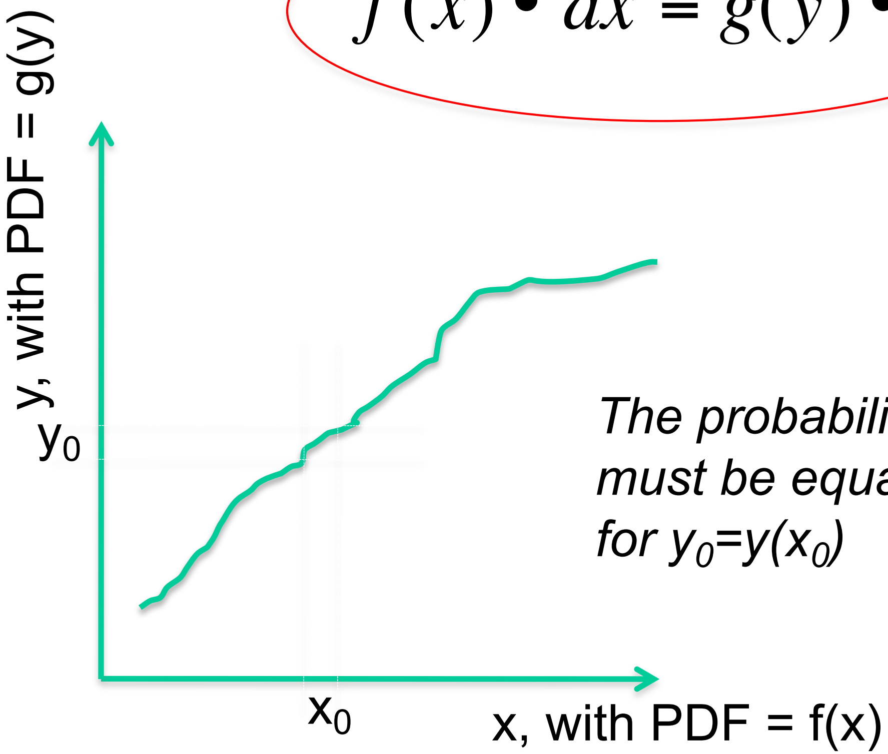
$$F(x) = \int_{-\infty}^{x} f(x') \bullet dx'$$

(Loreti 8.1.2)

Be y = F(x) the random variable, we want to compute the PDF g(y) of y.
For random variables the following rule holds for monotonic shapes:

$$f(x) \bullet dx = g(y) \bullet dy = g(y) \bullet |y'(x)| \bullet dx$$

Then g(y) = 1 since y'(x)=f(x).



The probability of events in dx, i.e. $f(x_0)dx$ must be equal to the probability in dy, i.e $g(y_0)dy$ for $y_0=y(x_0)$

Usefull for simulations" (Monte Carlo)

One can usually generate pseudo-random* numbers
With uniform distribution in the range [0,1]. And few more functions.

Suppose that a f(x) correspond to a physical phenomenum,
Then one can compute its sampling by
generating uniformily y in [0,1] and
by computing the inverse of the cumulative
Function F(x): $x = F^{-1}(y)$

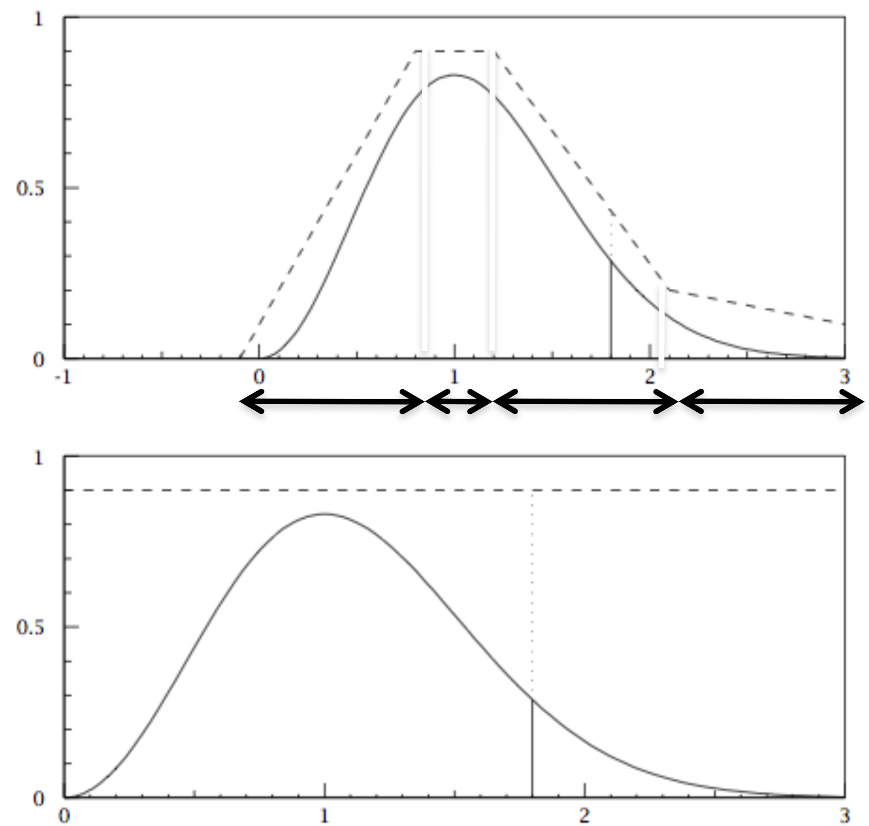
$$y = F(x) \rightarrow dy = f(x)dx$$
$$\rightarrow \int dy = \int f(x)dx$$

example: $\dfrac{d\sigma}{dx} \propto \dfrac{1}{x}$

$$\log x = \log x_0 + (\log x_1 - \log x_0)\cdot \xi$$

When inversion cannot be made:

FIGURA 8b - La scelta di un numero a caso con distribuzione prefissata mediante tecniche numeriche (la densità di probabilità è la stessa della figura 4d); la funzione maggiorante è una spezzata (superiormente) o la retta $y = 0.9$ (inferiormente).



*they own a finite period

The BINOMIAL distribution it answers the probabilistic question: is **that** true or not ?

Bernoulli trial,

PROVIDED IT HAS THESE CRITICAL PROPERTIES:

1) THE RESULT OF EACH TRIAL MAY BE EITHER A SUCCESS OR A FAILURE

2) THE PROBABILITY p OF SUCCESS IS THE SAME IN EVERY TRIAL.

3) THE TRIALS ARE *INDEPENDENT:* THE OUTCOME OF ONE TRIAL HAS NO INFLUENCE ON LATER OUTCOMES.

STARTING WITH A BERNOULLI TRIAL, WITH PROBABILITY OF SUCCESS p, LET'S BUILD A NEW RANDOM VARIABLE BY REPEATING THE BERNOULLI TRIAL.

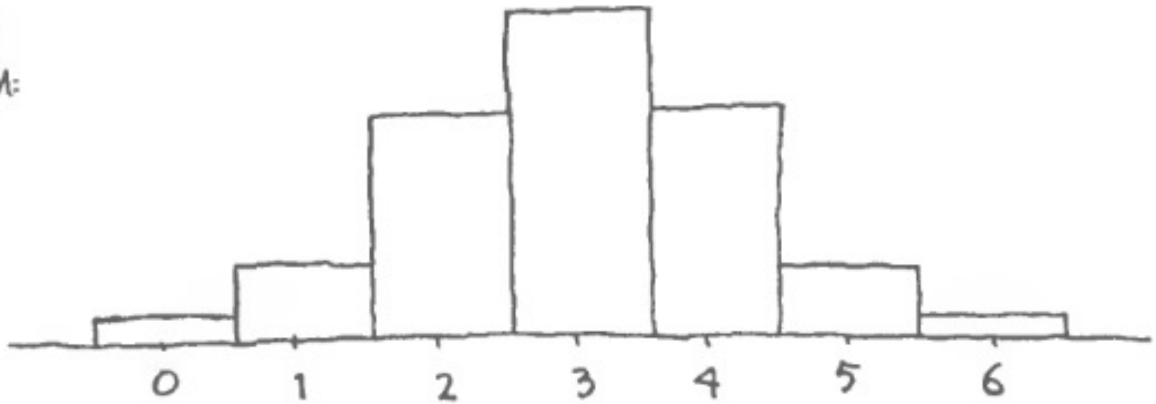
The binomial random variable

X IS THE *NUMBER OF SUCCESSES* IN n REPEATED BERNOULLI TRIALS WITH PROBABILITY p OF SUCCESS.

WHEN $p = .5$, THE BINOMIAL'S PROBABILITY DISTRIBUTION IS PERFECTLY SYMMETRICAL. FOR 6 COIN FLIPS, FOR INSTANCE, IT'S

k = #HEADS	0	1	2	3	4	5	6
$Pr(X=k)$	$\left(\frac{1}{2}\right)^6$	$\left(\frac{1}{2}\right)^6 6$	$\left(\frac{1}{2}\right)^6 \cdot 15$	$\left(\frac{1}{2}\right)^6 \cdot 20$	$\left(\frac{1}{2}\right)^6 \cdot 15$	$\left(\frac{1}{2}\right)^6 6$	$\left(\frac{1}{2}\right)^6$

WITH THIS HISTOGRAM:



THE *MEAN* AND *VARIANCE* OF THE BINOMIAL DISTRIBUTION ARE

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

NOTE THAT THE MEAN MAKES INTUITIVE SENSE: IN n BERNOULLI TRIALS, THE EXPECTED NUMBER OF SUCCESSES SHOULD BE np. THE VARIANCE FOLLOWS FROM THE FACT THAT THE BINOMIAL IS THE SUM OF n INDEPENDENT BERNOULLI TRIALS OF VARIANCE $p(1-p)$.

Expected number of successes = ΣnP_n = Np,

as is obvious

Variance of no. of successes = Np(1-p)

Variance ~ Np, for p~0

~ N(1-p) for p~1

NOT Np in general. NOT n $\pm\sqrt{n}$

e.g. 100 trials, 99 successes, NOT 99 ± 10

$$E(x^2) = \int_{0,1} x^2 \bullet \frac{d\Pr}{dx} \bullet dx = p$$

and then $\sigma^2(x) = E(x^2) - E^2(x) = pq$

Note: 1-p ≡ q

Interesting example in Physics: the **RADIACTIVE DECAYS**

(Loreti 8.4.1)

Be Λ_t the constant probability of an unstable nucleus to decay into a time interval ***t***, then, given ***N*** nuclei, the probability to get a certain nb of decays in the time ***t*** is given by the *binomial distribution* (the average nb of decays in ***N*** nuclei is $N\Lambda_t$)

Hypothesis: $\Lambda_t \propto t$ then $\Lambda_t = \lambda \cdot t$

Then, the nb of nuclei **N** changes as $dN = -N \cdot \lambda \cdot dt$

Therefore, the nb of not decayed nuclei is: $N(t) = N_{t=0} \cdot e^{-\lambda t} = N_0 e^{-t/\tau}$

The probabilistic question is: given ***t*** how many times do I get ***M(t)*** decay ?

➔ BINOMIAL

A more interesting, slightly different, probabilistic question is: given the time ***t*** how many ***M(t)*** nuclei decay? ➔ POISSONIAN

Repeat to yourself the two questions the needed number of times to rightly understand which are the two different random variables

Average number of decays at time t: $N_0 - N(t) = N_0\left(1 - e^{-t/\tau}\right)$

Binomial at time t: $P(x;t) = \binom{N_0}{x} p^x (1-p)^{N_0 - x}$ with $p = 1 - e^{-t/\tau}$

Example, ^{60}Co (amu=59.9338222), (half-life=1925.20±0.25 day), source of 1 gr.
For t=180 d, it holds

$$p = 1 - e^{-180 \cdot \ln 2 / 1925.2} = 0.06275169$$

p(t =1 d)=3.6x10^{-4}
M=3.6x10^{18}
σ=1.9x10^{9}

And $$N_0 = \frac{1}{59.9338222} 6.02214086 \cdot 10^{23} = 100.479840 \cdot 10^{20}$$

Maximum P is for

$$M = 0.06275169 * 100.479840 \cdot 10^{20} = 6.305281 * 10^{20}$$

With a dispersion of

$$\sigma = \sqrt{0.062752 \cdot (1 - 0.062752) \cdot 100.479 \cdot 10^{20}} = 2.4 * 10^{10} << \delta M$$

AAA use of significant digits

$$\delta amu : 1\,ppb$$
$$\delta N_A : 1\,ppb$$
$$\delta d : 4\,ppm\,(1\min\,over\,t)$$
$$\delta hl / hl : 1.3\times10^{-4}$$

$$\left(\frac{\delta N_0}{N_0}\right)^2 = \left(\frac{\delta N_A}{N_A}\right)^2 + \left(\frac{\delta amu}{amu}\right)^2 \approx (2\,ppb)^2$$

$$\left(\frac{\delta p}{1-p}\right)^2 = \left(\frac{\ln 2}{hl}\right)^2 \delta d^2 + \left(\frac{d\cdot\ln 2}{hl^2}\right)^2 \delta hl^2 = \left(\frac{d\cdot\ln 2}{hl}\right)^2 \left[\left(\frac{\delta d}{d}\right)^2 + \left(\frac{\delta hl}{hl}\right)^2\right]$$

$$\approx \left(6.5\% * 1.3\times10^{-4}\right)^2 \approx (8.5\,ppm)^2$$

$$\left(\frac{\delta M}{M}\right)^2 = \left(\frac{\delta p}{p}\right)^2 + \left(\frac{\delta N_0}{N_0}\right)^2 \approx (8.5\,ppm)^2$$

The Poisson distribution it answers the probabilistic question: how many ? *

Prob of n independent events occurring in time t when rate is r (constant)

e.g. events in the single bin of an histogram and NOT the radioactive decay for t ~ τ

$P_n = e^{-rt} (r\,t)^n / n! \;= e^{-\mu} \mu^n/n! \quad (\mu = r\,t)$

$\langle n \rangle = r\,t = \mu$ (No surprise!)

$\sigma^2_n = \mu$ "$n \pm \sqrt{n}$" **(note: 0 ± 0 has no meaning, 1 ± 1 is wrong)**

*** if the sample is limited the answer is provided by the *binomial***

Limit of Binomial ($N \rightarrow \infty$, $p \rightarrow 0$, $Np \rightarrow \mu$ constant, i.e. Poisson)

$\mu \rightarrow \infty$: Poisson→ Gaussian, with mean = μ

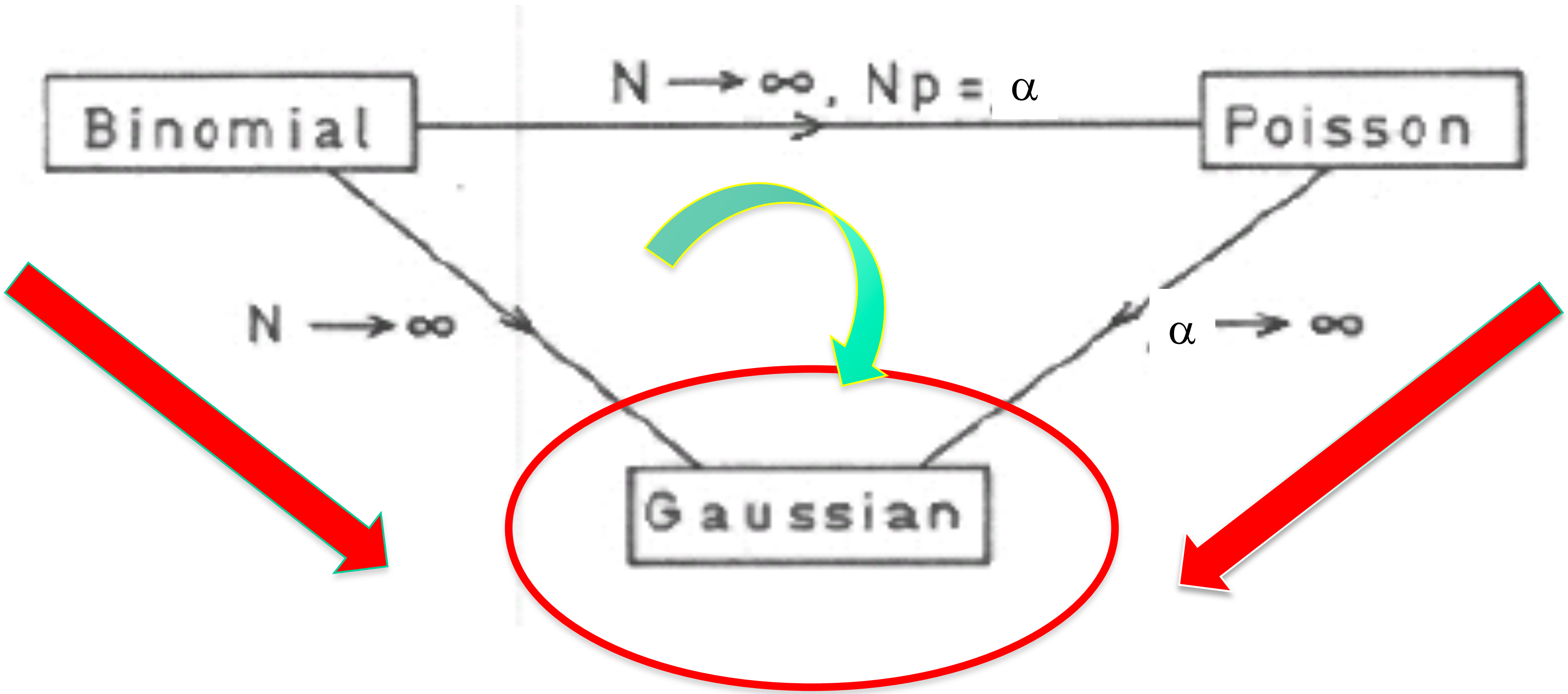
Important for χ^2 correct computation (i.e. the correctness of the error estimation)

Interlude

The discrete convolution formula:
-be x with its f(x) and y with g(y) two independent discrete random variables ≥0
-define the new z=x+y random variable
-what is the PDF of z, h(z)?

⟹ $$h(z) = \sum_{x=0}^{z} f(x) \cdot g(z-x)$$

As a corollary, if x_1 and x_2 are Poissonian random variables with μ_1 and μ_2, then $x_1 + x_2$ is Poissonian with $\mu=\mu_1 + \mu_2$



The actual logical transition goes from Binomial to Poissonian and then to Gaussian…..

Considering the example of ^{60}Co:

$$\alpha = N_0 \cdot p = 100.480 \cdot 10^{20} \cdot 0.062752 = 6.305 * 10^{20}$$

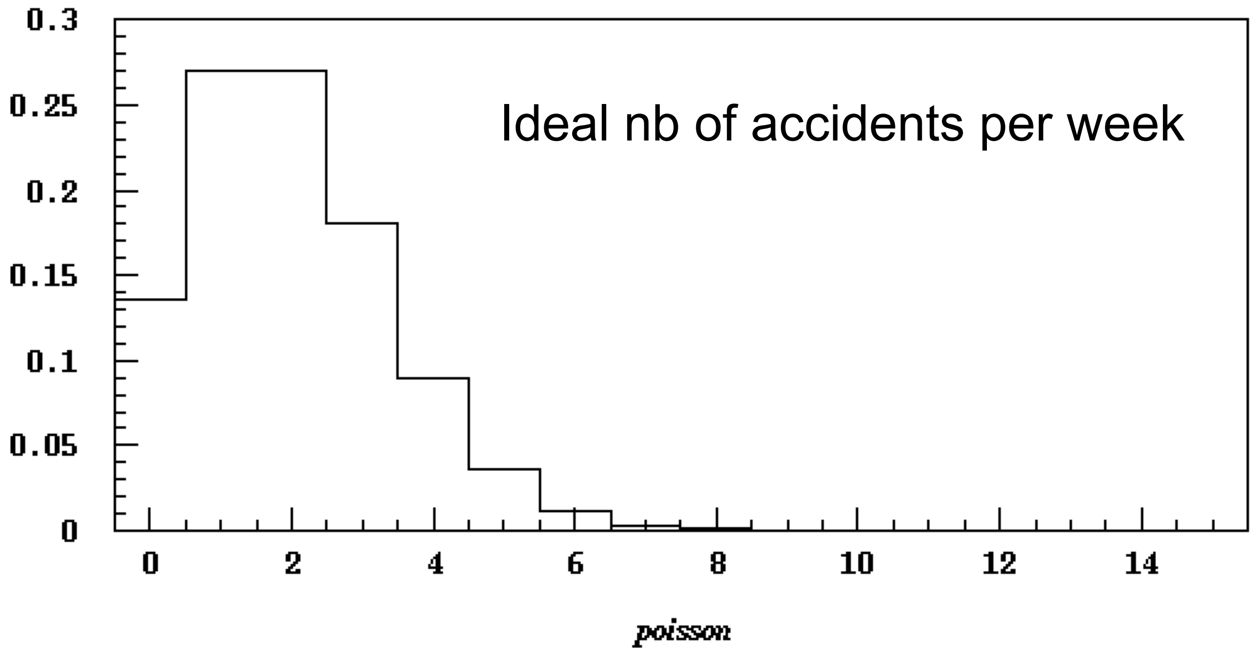
with dispersion:

$$\sigma = \sqrt{\alpha} = \sqrt{0.062752 \cdot 100.479 \cdot 10^{20}} = 2.511 * 10^{10}$$

Simulation of a Poisson process:
Test the number of car-crashed per week in a fixed town, for a total of 30 weeks.
The number is distributed as a Poisson function, since:
1)It depends of Δt (and Δt is "small compared to 30 weeks)
2)It does not depend on what happened before and what will happen after

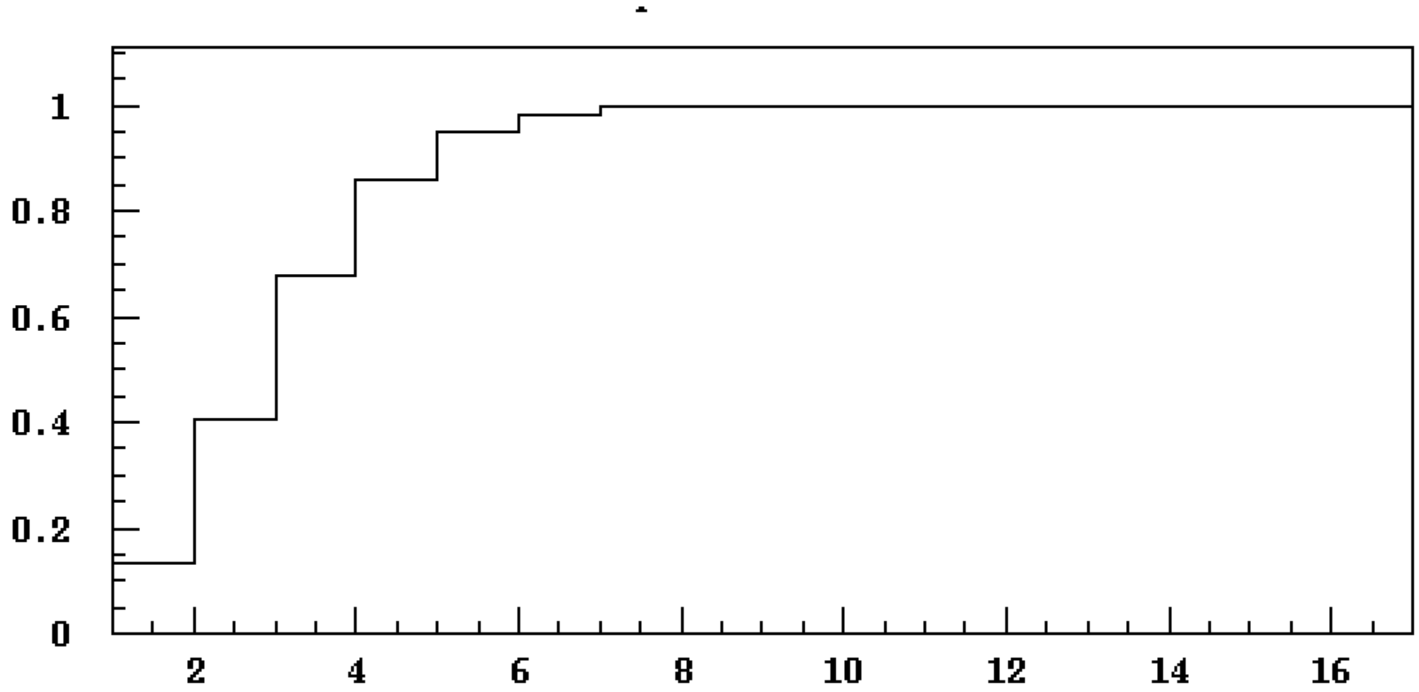
Suppose the average number of car-crashes per week is α=2. Then

$$P(x;t) = \frac{e^{-2} 2^x}{x!}$$



How the distribution actually looks for a set of 30 measures ?

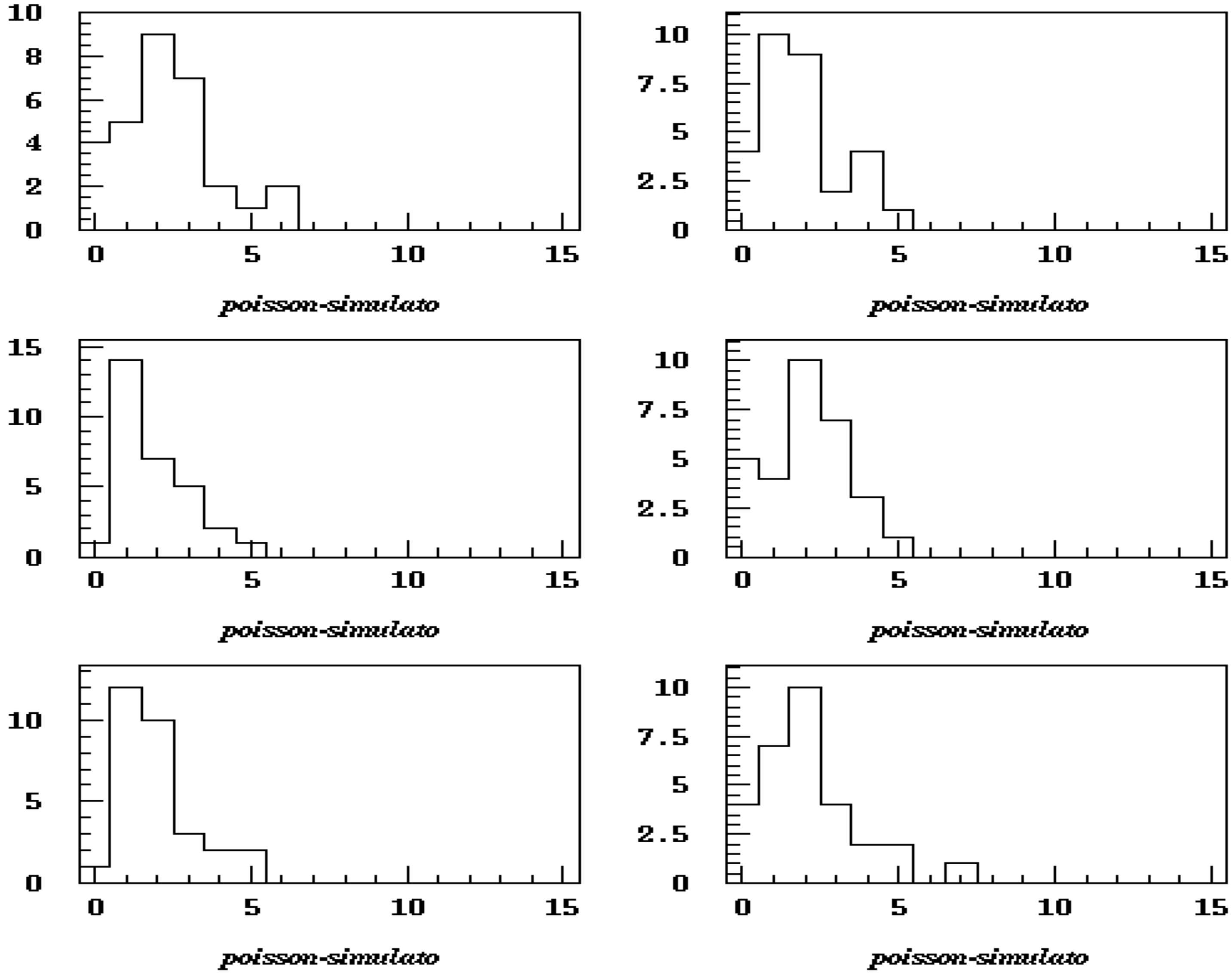
To simulate that, we first compute the cumulative function



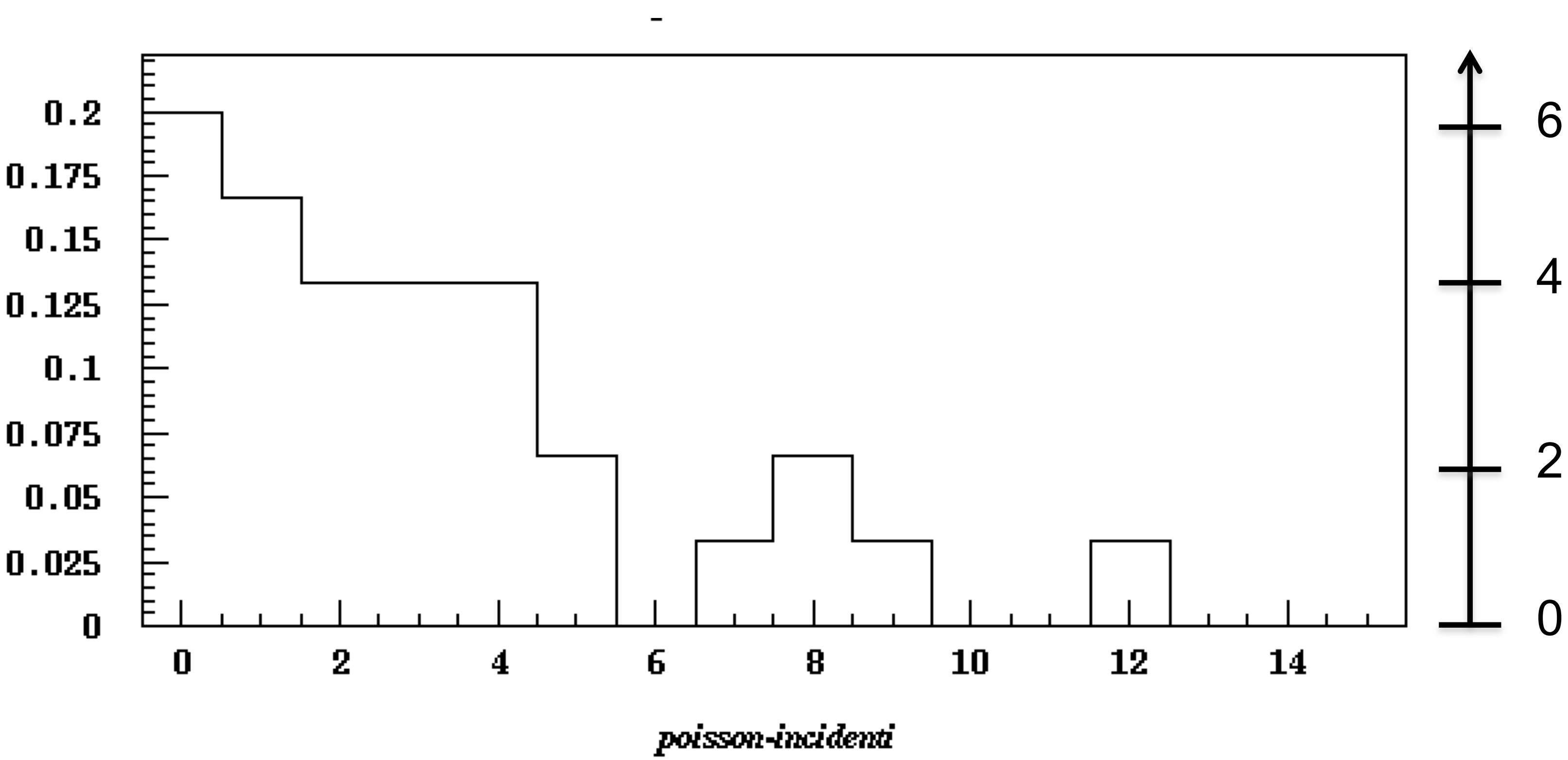
Then, we extract random numbers from the uniform distribution [0,1]
(the only way we know to generate pseudo-randoms)

To each extracted **x** into [0,1] we evaluate the corresponding **n**

Hera are examples of 30 observations:



QUESTION: does this simulation follow a Poisson distribution ?



➔ Test of Hypothesis

Using the χ^2:

$$P(o;t) = \frac{e^{-2}2^0}{0!} = e^{-2}$$

$$\chi^2 = \sum_{x=0}^{x=15} \frac{\left(n_i - p_i(x)\right)^2}{\sigma_i^2} = \frac{\left(6 - 30 \cdot 0.135\right)^2}{30 \cdot 0.135} + \ldots$$

Then compute the probability of the χ^2 for 15 degrees of freedom:
χ^2=29139.4 and P(χ^2,15)=0.

WHY IS SO LARGE ??
Prob at x=12 is $7 \cdot 10^{-6}$

Instead, the simulated sets give:
χ^2=11.38 and P(χ^2,15)=0.725
χ^2=4.67 and P(χ^2,15)=0.994
χ^2=9.53 and P(χ^2,15)=0.848
…

approximate because the PDF of each point is NOT a Gaussian
(the errors are estimated only with approximation)

Another possibility: make a fit leaving free μ and computing P(χ^2,14)

A very good application (very relevant for Physics) *(Loreti 8.5.1)*

We study the amount N_0 of protons in a time t. No decay is observed. What is the lower limit we can quote on the mean-life of the proton, τ, with a probability of 95% ?

Averaged number of decays expected in the time-range t from the binomial:

$$\alpha = N_0\left(1 - e^{-\frac{t}{\tau}}\right) \approx N_0 \frac{t}{\tau}$$

The probability to observe 0 events is given by the Poissonian:

$$P(0) = \frac{\alpha^0}{0!} e^{-\alpha} = e^{-\alpha}$$

What we have to compute, assuming that the proton be instable, is the minimum value the proton livetime owns such that the probability be at least of 95% not to observe anything. This happens when:

$$P(0) \geq 0.95$$

$$P(0) = e^{-\alpha} \approx e^{-N_0 \frac{t}{\tau}} \geq 0.95$$

$$-N_0 \frac{t}{\tau} \geq \ln 0.95$$

$$\tau \geq -\frac{N_0 t}{\ln 0.95}$$

→ τ ≥ 20 * Detector largeness* time range of data taking

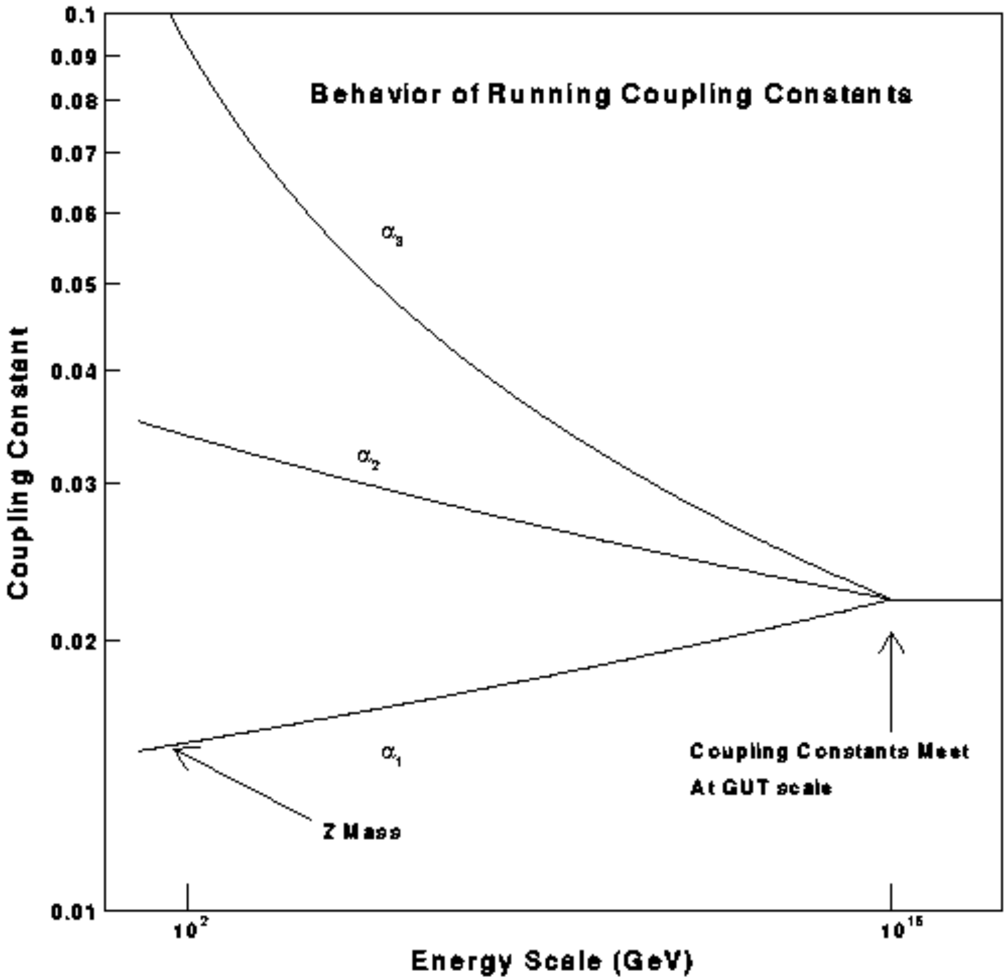
Best limit from SK (Super-Kamiokande): 50,000 m^3 di acqua

Table 1. Proton decay search detectors. Water Cherenkov detectors (Kamiokande, IMB-3, and Super-Kamiokande) and iron tracking detectors (Fréjus and Soudan 2) are listed. Partial lifetime limits have been set at 90% confidence level.

detectors	fiducial mass [kt]	exposure [kt·yr]	limit on $p \to e^+\pi^0$ [10^{31}yrs]	limit on $p \to \bar{\nu}K^+$ [10^{31}yrs]
Kamiokande	1.04	3.76	26 (2)	10 (2)
IMB-3	3.3	7.6	54.0 (3)*	15.1 (3)
Super-Kamiokande	22.5	52.2	330 (4)†	
		33		67 (5)
Fréjus	0.6	1.58	7.0 (6)	
		1.3		1.5 (7)
Soudan 2	0.77	3.56		4.3 (8)

SK recent result : 5.4*10^{33} (90% C.L.) with 0.1 megaton-year (Mt · year)

Why is so interesting?
Test of Grand Unified Theories (GUT)



α_1 is the $U(1)_Y$ coupling constant: $\alpha_1(M_Z) = \frac{5}{3}\frac{\alpha(M_Z)}{1-sin^2\theta_W(M_Z)}$

α_2 is the $SU(2)_L$ coupling constant: $\alpha_2(M_Z) = \frac{\alpha(M_Z)}{sin^2\theta_W(M_Z)}$

α_3 is the $SU(3)_c$ coupling constant: $\frac{\alpha_3(M_Z)}{1+\frac{\alpha_3(M_Z)}{4\pi}} = \alpha_s(M_Z)$

The present limit from SK excludes the most semplified "versions" of GUTs….

Two entanglements and a different point-of-view:

Existence of BACKGROUND events !
And, moreover, ***convolution*** with the estimated errors …

Computation of the POISSON PDF starting from the number of events effectively observed:

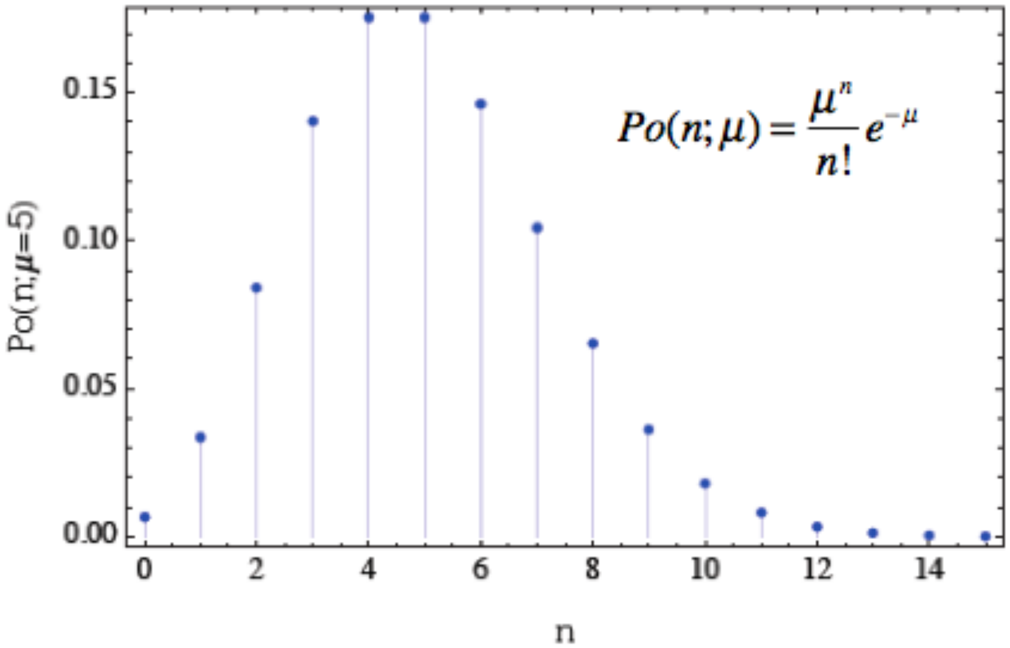
$$P(background) \geq P(n)$$

(see later the concept of p-value)

http://www.pit.physik.uni-tuebingen.de/grabmayr/workshop/talks/gomez-statistics.pdf

Poisson statistics

$$N \sim 5 \text{ for } T_{1/2}^{0\nu} \sim 10^{27}\ y$$



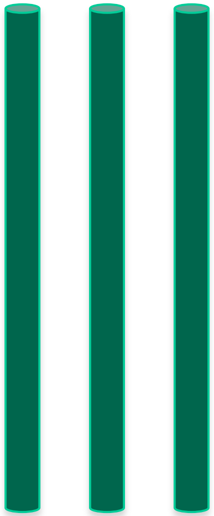
17.5% of the experiments will observe 5 events

17.5% of the experiments will observe 4 events

0.7% of the experiments will observe 0 events

Different identical experiments running for the same total exposure will observe different number of events

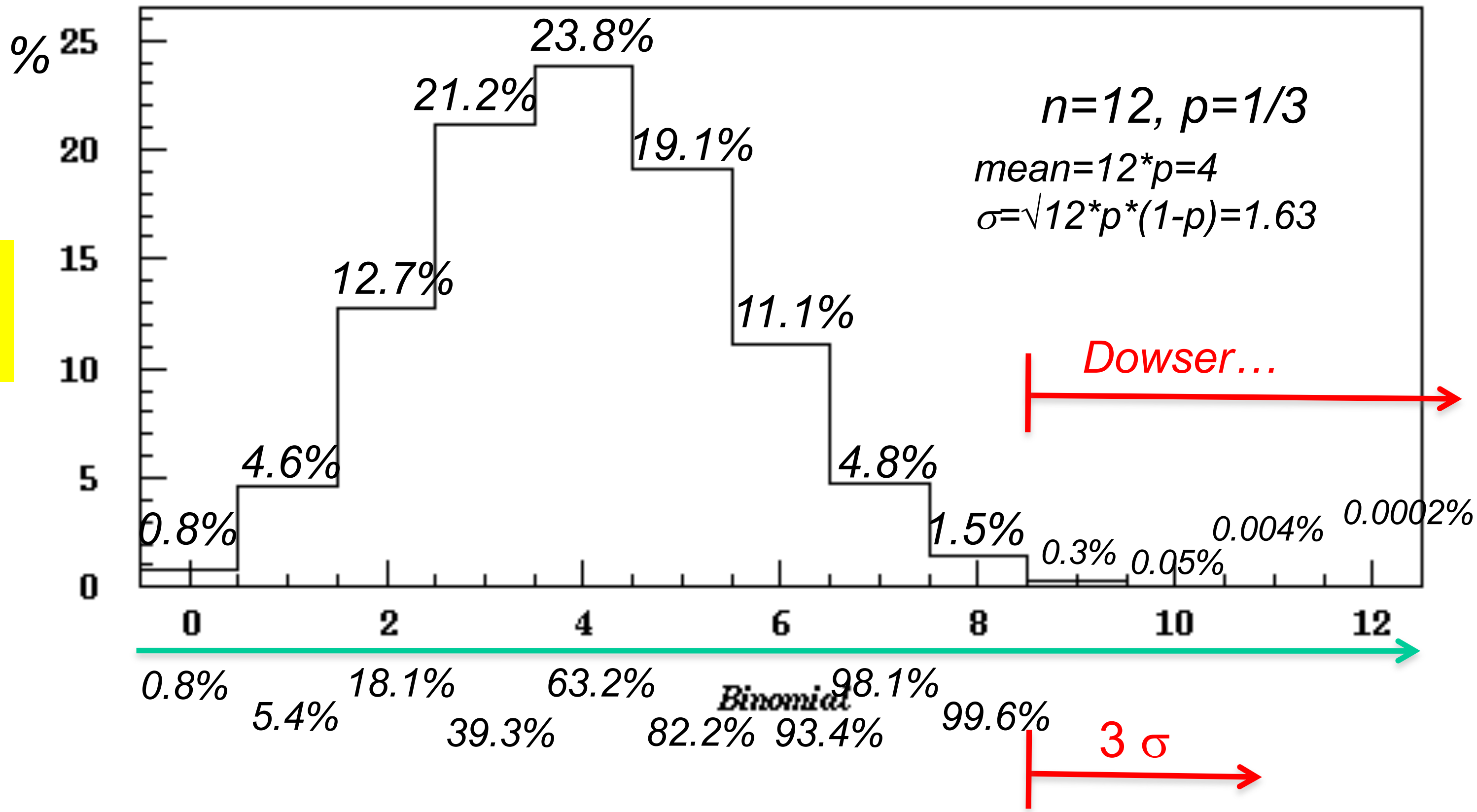
The dowser example



3 pipes conducting water randomly at once, a professed dowser has to predict the right pipe. 12 trials are foreseen. The statistician affirms that the candidate has to be right at least 9 times to be declared an effective dowser

$$P(x;12) = \binom{12}{x} (1/3)^x (2/3)^{12-x}$$

Otto Edler von Graeve in 1913



BUT on 1000 candidates 4 will randomly predict at least 9 times !

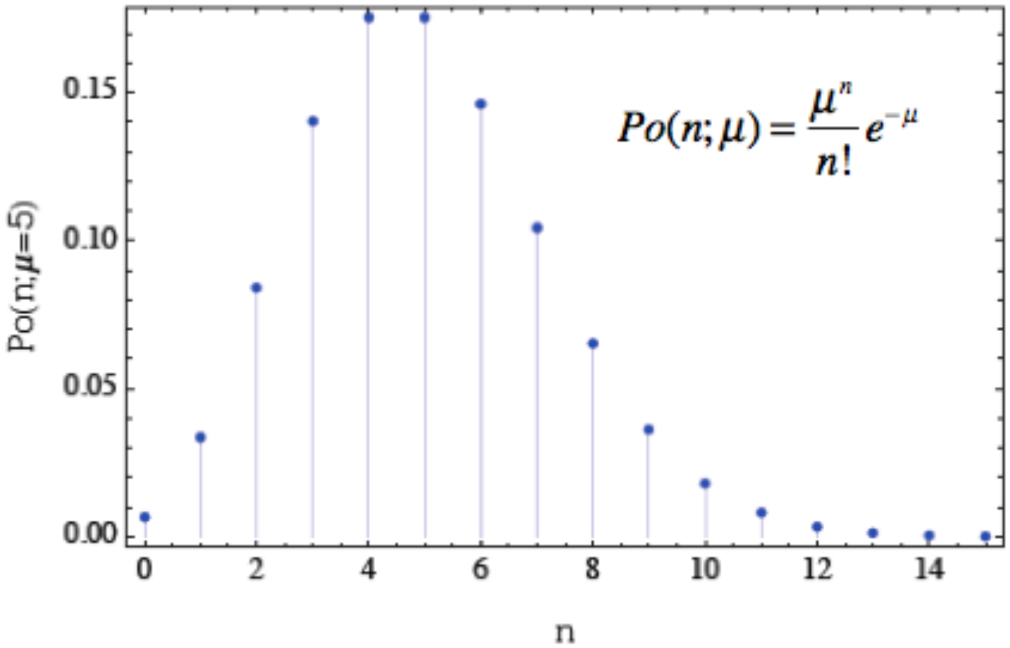
NOTE: Error / fluctuation on 4 !

Poisson

http://www.pit.physik.uni-tuebingen.de/grabmayr/workshop/talks/gomez-statistics.pdf

Poisson statistics

$$N \sim 5 \text{ for } T_{1/2}^{0\nu} \sim 10^{27}\ y$$



17.5% of the experiments will observe 5 events

17.5% of the experiments will observe 4 events

0.7% of the experiments will observe 0 events

Different identical experiments running for the same total exposure will observe different number of events

SIMULATION

We estimate that about 4 people over 1000 be able to give the right prediction at least 9 times.
The “significance” of the SINGLE PERSON is 3 σ

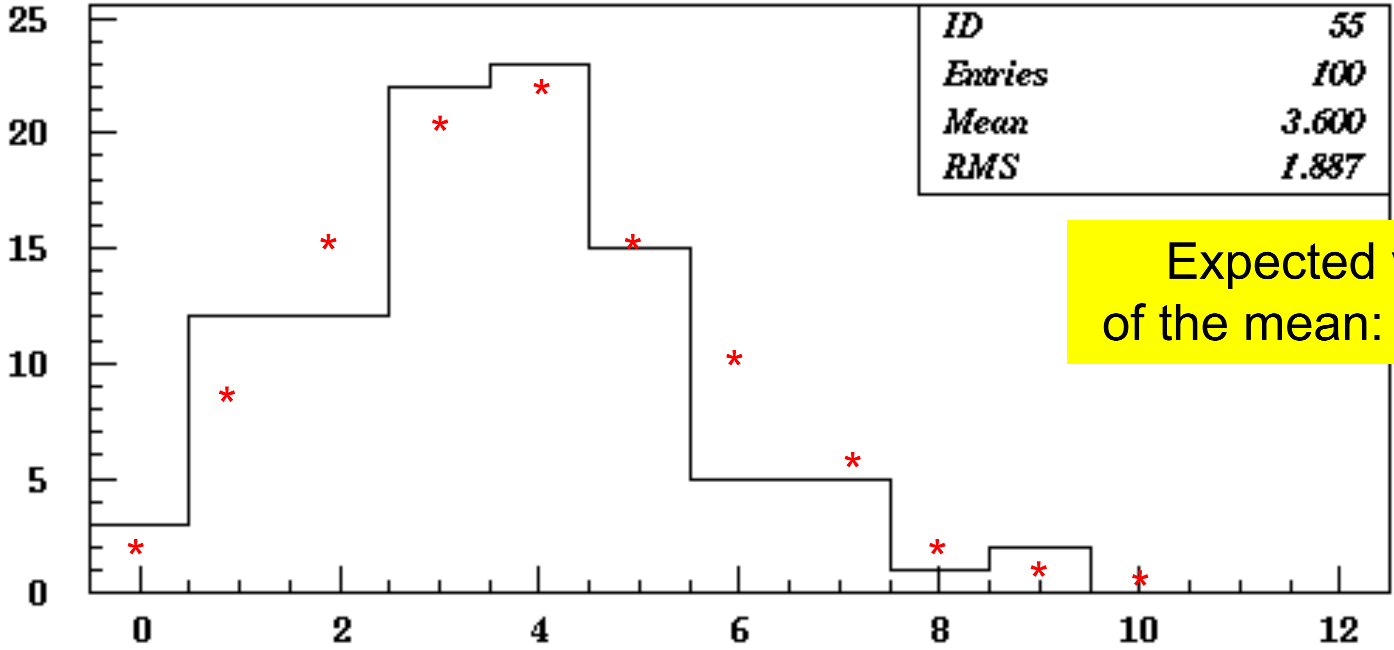
Actually the single bin of the Binomial distribution will follow Poissonian fluctuations.
(forgetting the constraint on the total nb. of “events”)
Moreover the distribution of the sum over the last 4 bins will also be a Poissonian.
(sum of Poissonians follows a Poissonian p.d.f.)

We simulate 1000 candidates making 1000 random extraction between 0 and 1:

Each extraction will simulate a trial corresponding to a specific result (1,2,3… 12)

Each set of 1000 candidates can be taken as ONE EXPERIMENT

Make 100 experiments: here is the distribution of the winning people (right prediction at least 9 times over 12)

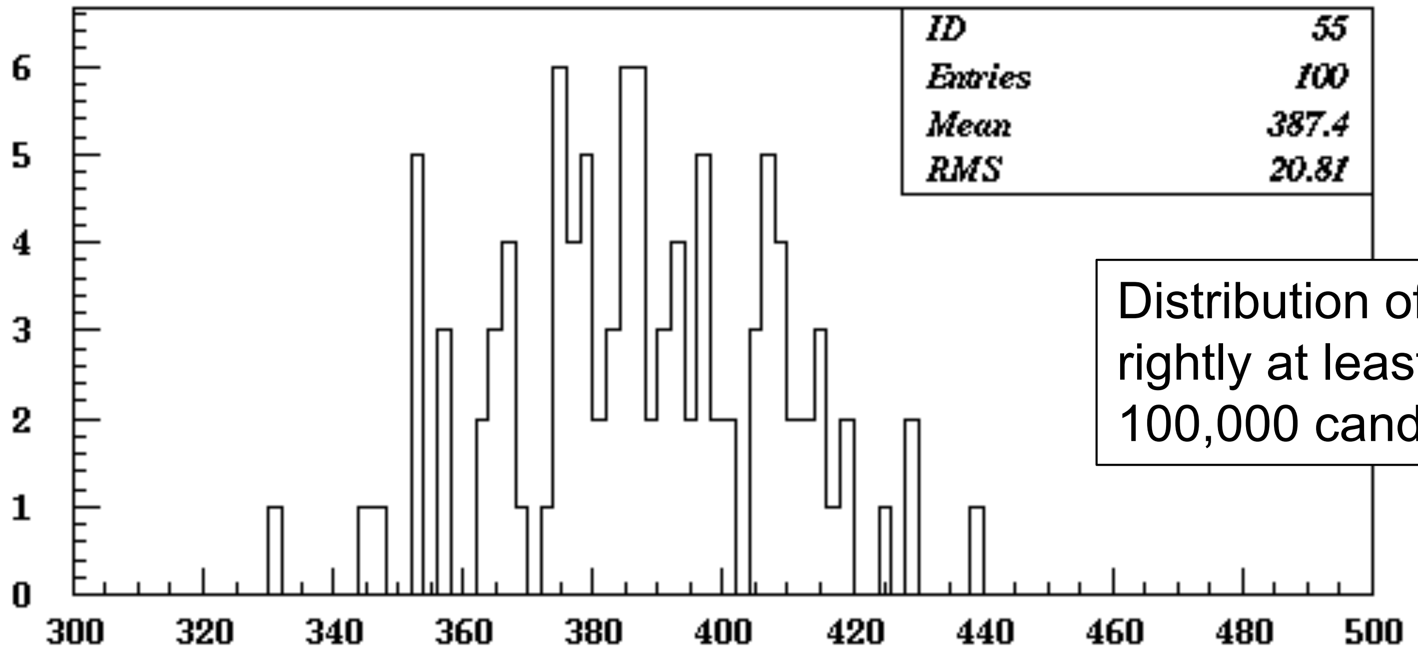


Expected value of the mean: 3.855…

Number of candidates over 1000 with at least 9 right answers

$$Po(n;\mu) = \frac{\mu^n}{n!} e^{-\mu}$$

Not so good description... Let us try 100 experiments, each with 100,000 candidates



Distribution of nb. of people guessing rightly at least 9 times in samples of 100,000 candidates each

The Poissonian becomes actually a Gaussian-like and therefore the true p.d.f. is … the t-Student

$$\sigma_{Poisson} = \sqrt{385.59} = 19.64 \approx \sigma_{Gauss}$$

$$\sigma_{t-Student} = \sigma_{Guass} \times \sqrt{\frac{100}{100-2}} = 19.84$$

The Cauchy distribution

(also known in Particle Physics as the Breit-Wigner)

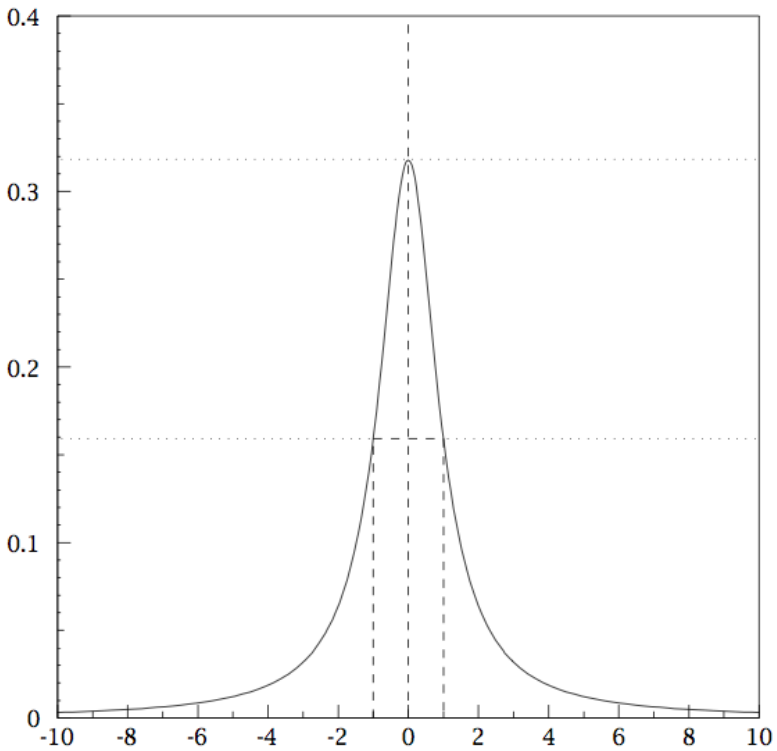
it answers the probabilistic question: what is the probability distribution of a resonant phenomenon ?

$$f(x;\theta,d) = \frac{1}{\pi d}\frac{1}{1+\left(\frac{x-\theta}{d}\right)^2}$$

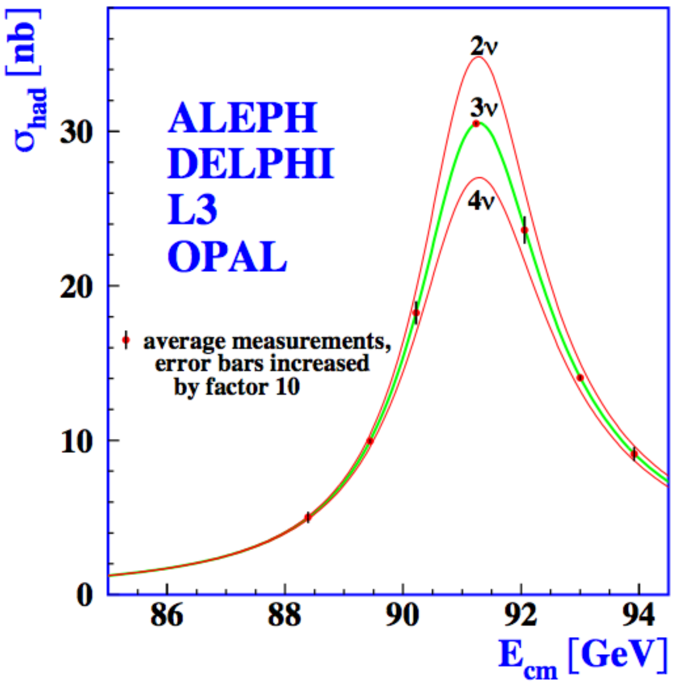
$$F(x;\theta,d) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{x-\theta}{d}\right)$$

$E(x^i)$ goes to ∞, for any $i \geq 1$

r $\theta = 0$ e $d = 1$.



http://lepewwg.web.cern.ch/LEPEWWG/1/physrep.pdf



Z^0 gauge boson LINESHAPE as measured at LEP-CERN ~ 1992-1994

The t-Student distribution

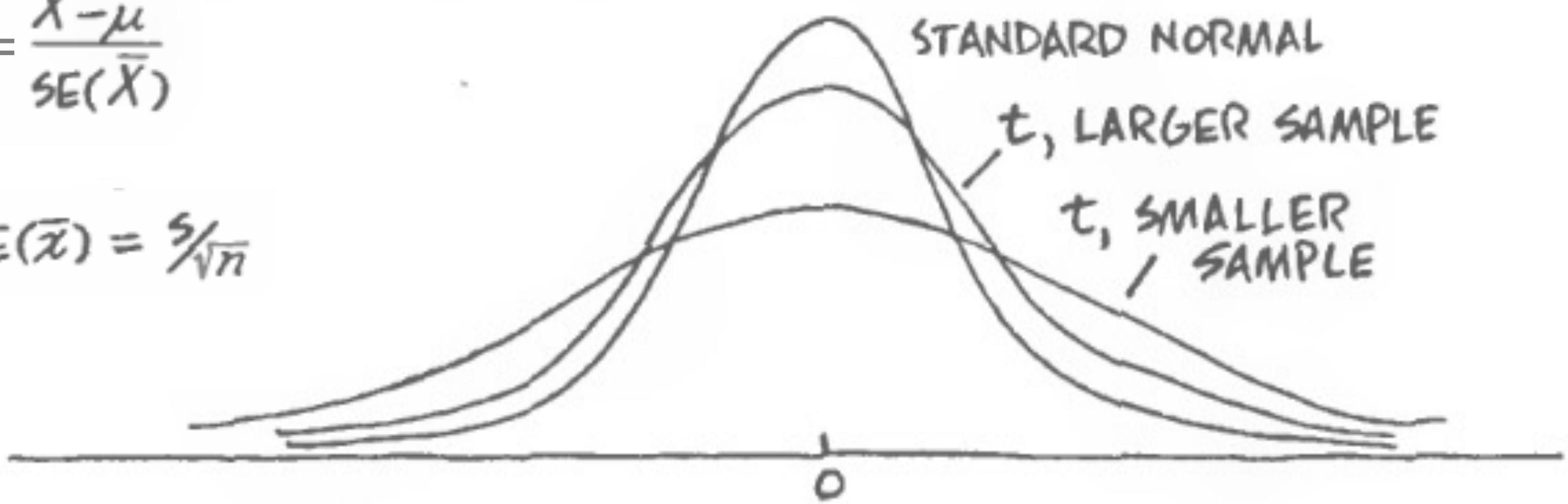
it answers the probabilistic question: what is the probability distribution of a finite set of measurements, each one following the Gaussian distribution ?

The practical problem, solved by William Gosset in 1908, is to correctly estimate the dispersion, i.e. the √Variance, from the histogram of a Gaussian-like distribution

From a set of ***n*** measurements compute its mean $\bar{x}$ and its standard deviation, ***s***. Then, define the random variable ***t*** (residuals):

$$t = \frac{\bar{X} - \mu}{SE(\bar{X})}$$

$$SE(\bar{x}) = s/\sqrt{n}$$



(PDG)

31.4.5. *Student's t distribution* :

Suppose that x and $x_1, \ldots, x_n$ are independent and Gaussian distributed with mean 0 and variance 1. We then define

(minor approximation on n: degrees of freedom)

$$z = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad t = \frac{x}{\sqrt{z/n}} \,. \qquad (31.29)$$

← reduced χ^2

The variable z thus follows a $\chi^2(n)$ distribution. Then t is distributed according to Student's t distribution with n degrees of freedom, $f(t;n)$.

The Student's t distribution resembles a Gaussian with wide tails. As $n \to \infty$, the distribution approaches a Gaussian. If $n = 1$, it is a *Cauchy* or *Breit–Wigner* distribution. The mean is finite only for $n > 1$ and the variance is finite only for $n > 2$, so the central limit theorem is not applicable to sums of random variables following the t distribution for $n = 1$ or 2.

t-Student distribution with N degrees of freedom

$$f(t;N) = \frac{T_N}{\left(1 + \frac{t^2}{N}\right)^{\frac{N+1}{2}}} \quad \text{with} \quad t = \frac{u}{\sqrt{\frac{z}{N}}} \qquad \mathrm{Var}(t) = \frac{N}{N-2}$$

(T_N normalization constant)

Other distributions

Log-Normal, to work out the field x>0

(and the product of large nb of random variables, e.g. electrons in a calorimetry)

$$f(x) = \frac{1}{x \cdot \sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}$$

Landau, to work out the energy loss

(with $x=(\Delta-\Delta_0)/\xi$) where Δ is the energy loss, and Δ_0, ξ depend on actual case and the material)

$$f(x) = \frac{1}{\pi}\int_0^\infty e^{-u\ln u - xu} \sin(\pi u)\, du$$

Negative Binomial

(probability of x successes before k failures)

$$f(x;k,p) = \frac{(k+x-1)!}{x!(k-1)!} p^x q^k$$

Chi Squared, $\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma_i^2}$

$$f(z;n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}\Gamma(n/2)} \quad ; \qquad z \geq 0$$

AAA

BACKUP SLIDES

Example: compatibility of the intial OPERA result (11 September 2011)
and the previous measurement by MINOS about the velocity of the neutrinos

OPERA $$\delta = 60.8 \pm 6.9\,(stat.) \pm 7.4\,(sys.)\ ns \qquad at\ 68\%\ C.L.$$

MINOS $$\delta = 126 \pm 32\,(stat.) \pm 64\,(sys.)\ ns \qquad at\ 68\%\ C.L.$$

Earlier arrival after 730 km with respect to the time-of-light

$$\delta t = (60.7 \pm 6.9\ (\text{stat.}) \pm 7.4\ (\text{sys.}))\ \text{ns.}$$

Expected value : 0 nsec *
Measured value (meeam): 60.7 nsec

Error (mean squared): $\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2} = 10.1\ n\sec$

"Distance" in term of "SIGMA'S": |0.0-60.7| / 10.1 = 6.0 σ

** ns: nanoseconds: 10^{-9} sec*

Delay of arrival time after 734 km with respect to the time-of-light

$$\delta = -126 \pm 32 \text{ (stat.)} \pm 64 \text{ (sys.) ns} \quad 68\% \text{ C.L.}$$

Error (mean squared): $\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2} = 72\, n\sec$

"Distance" in term of "SIGMA'S":
|0.0-126| / 72 = 1.8 σ

Conclusion of MINOS:
Result is compatible with $v_{neutrino} = v_c$

If we go from sigmas to the probabilities:
AAA one-side/two-side issue or choice

From “sigma’s” to probabiliy

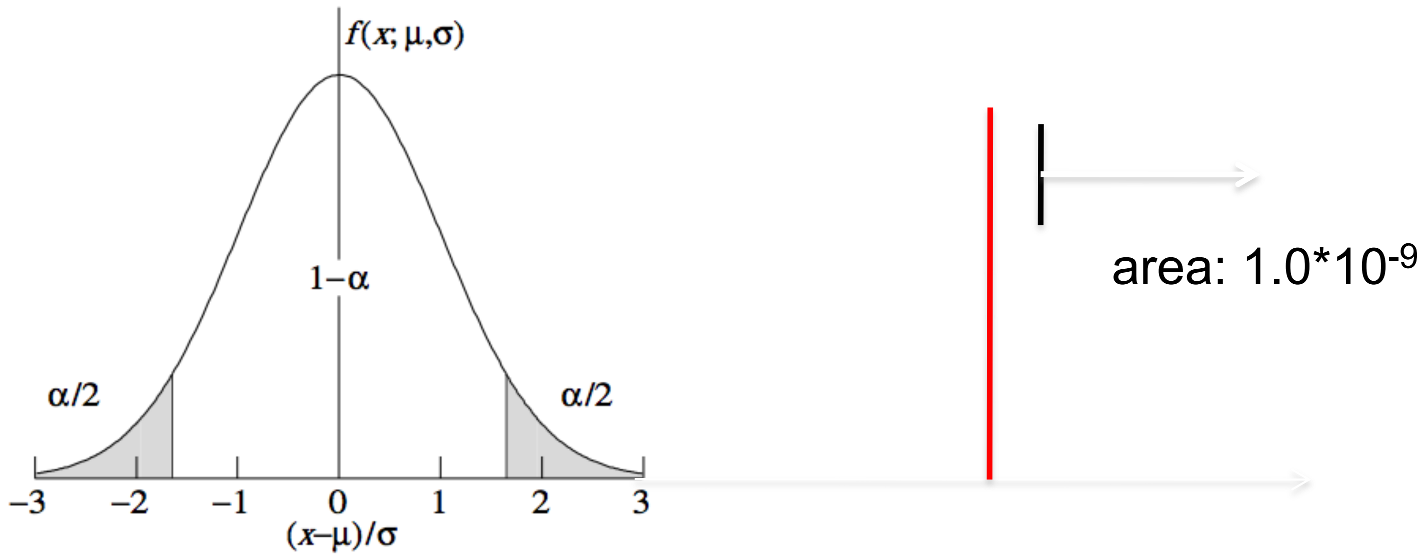


Figure 32.4: Illustration of a symmetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by α, are as shown.

One SIDE

Two SIDE

Two SIDE

24 February 1987: observation of supernova explosion type II SN1987A
3 hours before 23 neutrinos in 13 sec were observed
$v_{neutrino} = v_c$ while photons get out with some delay, when the shock wave reaches the star surface.

Whether OPERA result were true the RATIO would have been:

$$\delta c_\nu = (2.48 \pm 0.28_{\rm stat} \pm 0.30_{\rm syst}) \times 10^{-5} \qquad \text{(OPERA)}, \quad \text{where } \delta c_\nu \equiv (v_\nu - c)/c.$$

As SN1987A is far away 168,000 light-years neutrinos had to arrive

4.2±0.5±0.5 years before

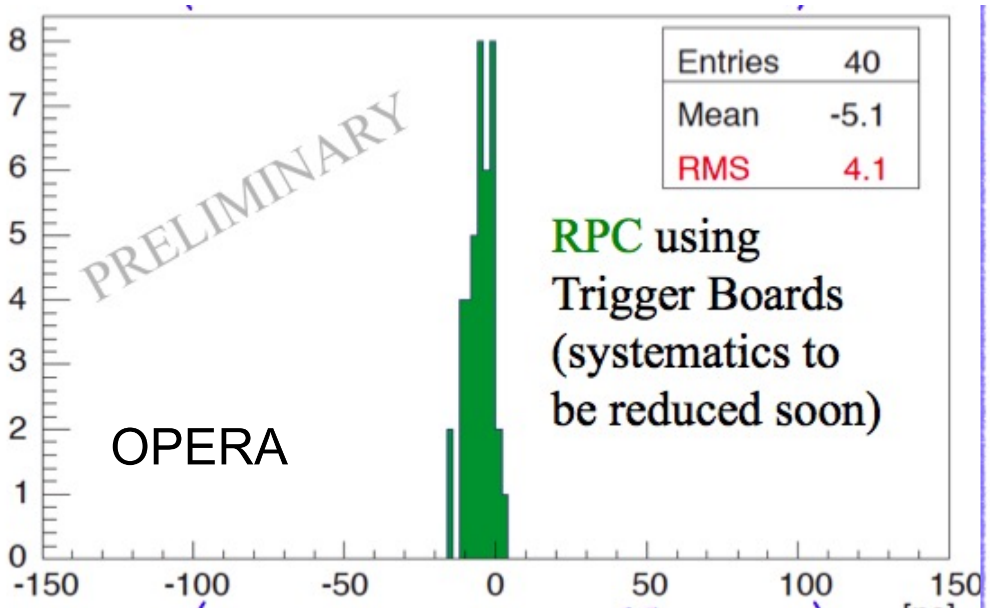
⇒ limit on the ratio between $v_{neutrino}$ e v_c : 10^{-9} (with an error around 1 order of magnitude)

Clearly incompatible, an energy dependence has to be introduced:

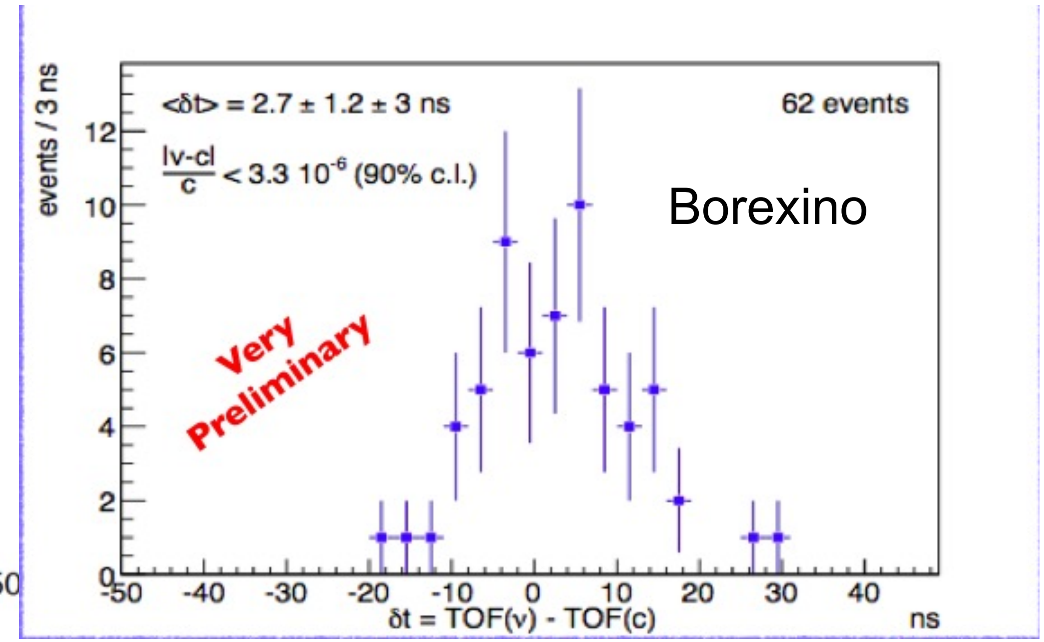
Mean energy of OPERA neutrinos: 17 GeV
Mean energy of supernova neutrinos: from 7.5 to 39 MeV

Or a flavor dependence…

- All experiments consistent with no measurable deviation from the speed of light for neutrinos:
  - **Borexino**: **δt = 2.7 ± 1.2 (stat) ± 3(sys) ns**
  - **ICARUS:** **δt = 5.1 ± 1.1(stat) ± 5.5(sys) ns**
  - **LVD:** **δt = 2.9 ± 0.6(stat) ± 3(sys) ns**
  - **OPERA:** **δt = 1.6 ± 1.1(stat) [+ 6.1, -3.7](sys) ns**



$$\delta t = \left(-5.1 \pm 0.6\ (stat.)^{+6.5}_{-5.1}\ (sys.)\right)\ \mathrm{ns}$$



THE VARIANCE OF THE SUM OF RANDOM VARIABLES HAS A SIMPLE FORM IN THE SPECIAL CASE WHEN THE VARIABLES X AND Y ARE *INDEPENDENT*. THE TECHNICAL DEFINITION OF INDEPENDENCE IS BASED ON THE PROBABILITY PROPERTY P(A AND B) = P(A)P(B)... BUT FOR US, INDEPENDENCE JUST MEANS THAT X AND Y ARE GENERATED BY *INDEPENDENT MECHANISMS*, SUCH AS FLIPS OF A COIN, ROLLS OF A DIE, ETC.

WHEN X AND Y ARE INDEPENDENT,
THEIR VARIANCES ADD:

$$\sigma^2(X+Y) = \sigma^2(X)+\sigma^2(Y)$$

(while the two PDFs multiply themselves !)

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

AND, WHEN THE X_i ARE ALL INDEPENDENT,

$$\sigma^2(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \sigma^2(X_i)$$

It is intuitive but we need a bit of Mathematics and some new definitions.

Functions GENERATRICE and CHARACTERISTIC *(Loreti, pag. 72)*

k-th Moment $\lambda_k = E(x^k) = \int_{-\infty}^{+\infty} x^k f(x)\,dx$

Function generatrice

(characteristic)

(with $e^{tx} \rightarrow e^{itx}$)

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx$$

Discrete variable

For a discrete variable it holds: $M_x(t) = \sum_i p_i \cdot e^{tx_i}$

and we use the McLaurin expansion of the exponential: $e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$

Then, in case all the Moments exist to any order with respect the origin (x=0) $M_x(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_k$

Finally, it follows: $\left. \frac{d^k M_x(t)}{dt^k} \right|_{t=0} = \lambda_k$

We can extend the result to more than two variables:

Theorem: linear combinations of random variables, all following a Normal distribution, and all statistically independent, follow also a Normal distribution

Be N normal variables, x_k (k=1, …, N), with their own μ_k and σ_k, if we consider the new random variable

$$y = \sum_{k=1}^{N} a_k x_k$$

By using the CHARACTERISTIC function it is easy to demonstrate the ***y*** owns Normal PDF with

$$\mu = \sum_{k=1}^{N} a_k \mu_k \qquad \sigma^2 = \sum_{k=1}^{N} a_k^2 \sigma_k^2 \qquad \text{i.e} \quad pdf(y) = \prod_i pdf(x_i)$$

widely used theorem in DATA ANALYSIS.

From the previous theorem it follows the well known method of AVERAGED MEAN

Have N normal measurements, x_k (k=1, …, N), all statistically independent, affected by random noise assumed gaussian distributed, the density probability of the N observations is given by

$$\prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x^* - x_i}{\sigma_i}\right)^2}$$

where ***x**** is the UNKNOWN true ***x*** and, a priori, each measurement owns its σ_i, computed e.g. via m.s.d.

The Likelihood functon is:

$$L(x_1, x_2, \ldots, x_N \mid x) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - x_i}{\sigma_i}\right)^2}$$

where ***x*** is now a parameter !

Let us compute the MAX of the Likelihood as function of variable x

$$\prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_i}{\sigma_i}\right)^2} \Rightarrow -2\ln\mathcal{L} = \sum_{i=1}^{N}\left(\frac{x-x_i}{\sigma_i}\right)^2 + 2\sum_{i=1}^{N}\ln\sigma_i + 2N\ln\sqrt{2\pi}$$

$$f(x) = \sum_{i=1}^{N}\left(\frac{x-x_i}{\sigma_i}\right)^2$$

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 2\sum_{i=1}^{N}\left(\frac{x-x_i}{\sigma_i}\right)\frac{1}{\sigma_i} = 2\left(x\sum_{i=1}^{N}\frac{1}{\sigma_i{}^2} - \sum_{i=1}^{N}\frac{x_i}{\sigma_i{}^2}\right)$$

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = 2\sum_{i=1}^{N}\frac{1}{\sigma_i{}^2} > 0\ .$$

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 2\left(Kx - \sum_{i=1}^{N}\frac{x_i}{\sigma_i{}^2}\right) = 0$$

con $K = \sum_{i=1}^{N}\frac{1}{\sigma_i{}^2}$

$$\Rightarrow \bar{x} = \frac{1}{K}\sum_{i=1}^{N}\frac{x_i}{\sigma_i{}^2}$$

and

$$\sigma_{\bar{x}}{}^2 = \sum_{i=1}^{N}\left(\frac{1}{K\sigma_i{}^2}\right)^2\sigma_i{}^2 = \frac{1}{K^2}\sum_{i=1}^{N}\frac{1}{\sigma_i{}^2} = \frac{1}{K} = \frac{1}{\sum_{i=1}^{N}\frac{1}{\sigma_i{}^2}}$$

Summary

For random ***independent*** processes it holds:

$$pdf(y) = \prod_i pdf(x_i)$$

It follows for **normal** PDF :

$$\mu = \sum_{k=1}^{N} a_k \mu_k \qquad \sigma^2 = \sum_{k=1}^{N} a_k{}^2 \sigma_k{}^2$$

and the method of the weighted average (with $a_k = 1/(\sigma_k^2 \Sigma_j 1/\sigma_j^2)$.

In case of a unique PDF, it holds the theorem of **Central Limit** for the cumulative PDF.

NOTE: take care of

- elementary and not-elementary events
- single and multiple PDFs