Statistics and Data Analysis

J.L. Tain

Jose.Luis.Tain@ific.uv.es
http://ific.uv.es/gamma/

Instituto de Física Corpuscular

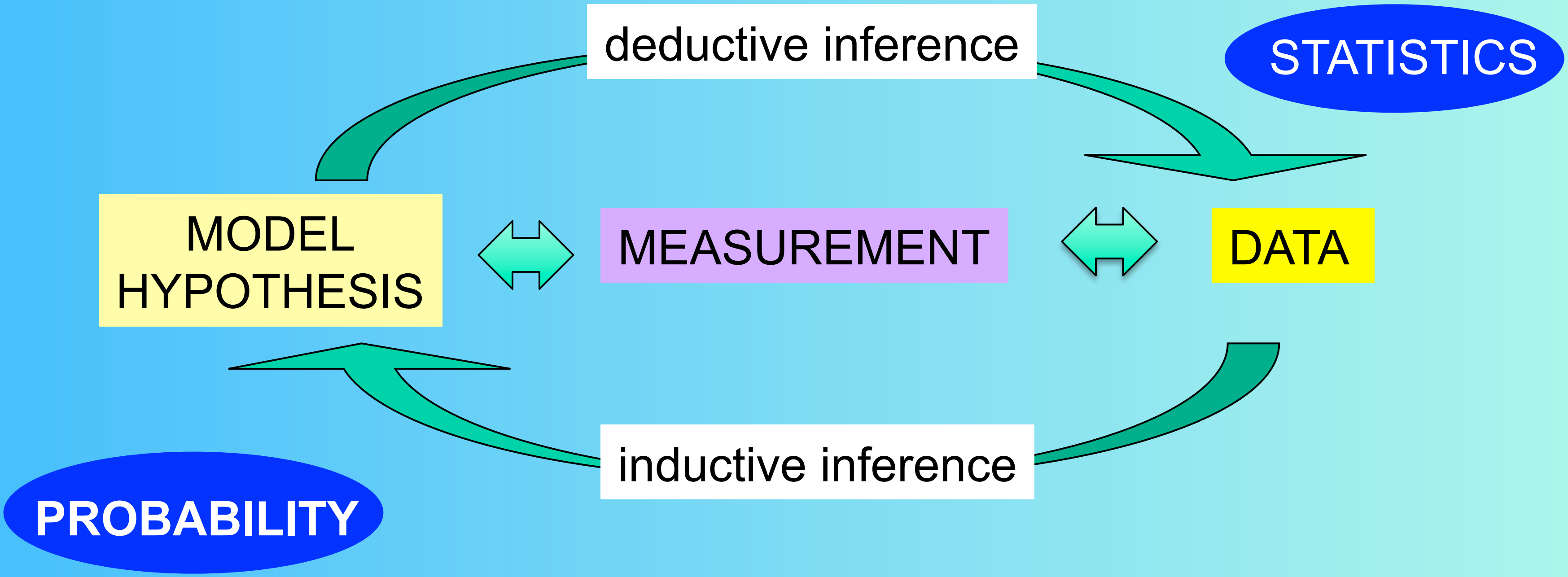
C.S.I.C - Univ. Valencia

Bibliography:

1) “Probability and Statistics in Particle Physics”, A.G. Frodesen, O. Skjeggestadt, H. Tofte, Universitetsforlaget, 1979
2) “Statistical Methods in Experimental Physics”, W.T. Eadie, D. Drijard, F.E. James, M. Roos, B. Sadoulet, North Holland, 1971
3) “Data Analysis: A Bayesian Tutorial”, D. S. Sivia, Clarendon Press, 1996
4) “The Data Analysis BriefBook”, http://rkb.home.cern.ch/rkb/titleA.html
5) “Probability and statistics Ebook”, http://wiki.stat.ucla.edu/socr/index.php/EBook

In experimental science we try to draw some conclusions on a theoretical construct (model, hypothesis, parameter,…) from the results of one or more experimental measurement:

- Experimental results are uncertain (i.e. the repetition of the same experiment gives different numerical results; different experiments give different results): how can we then characterize experimental data?
- How can we then infer something from data?


deductive inference

STATISTICS

MODEL HYPOTHESIS

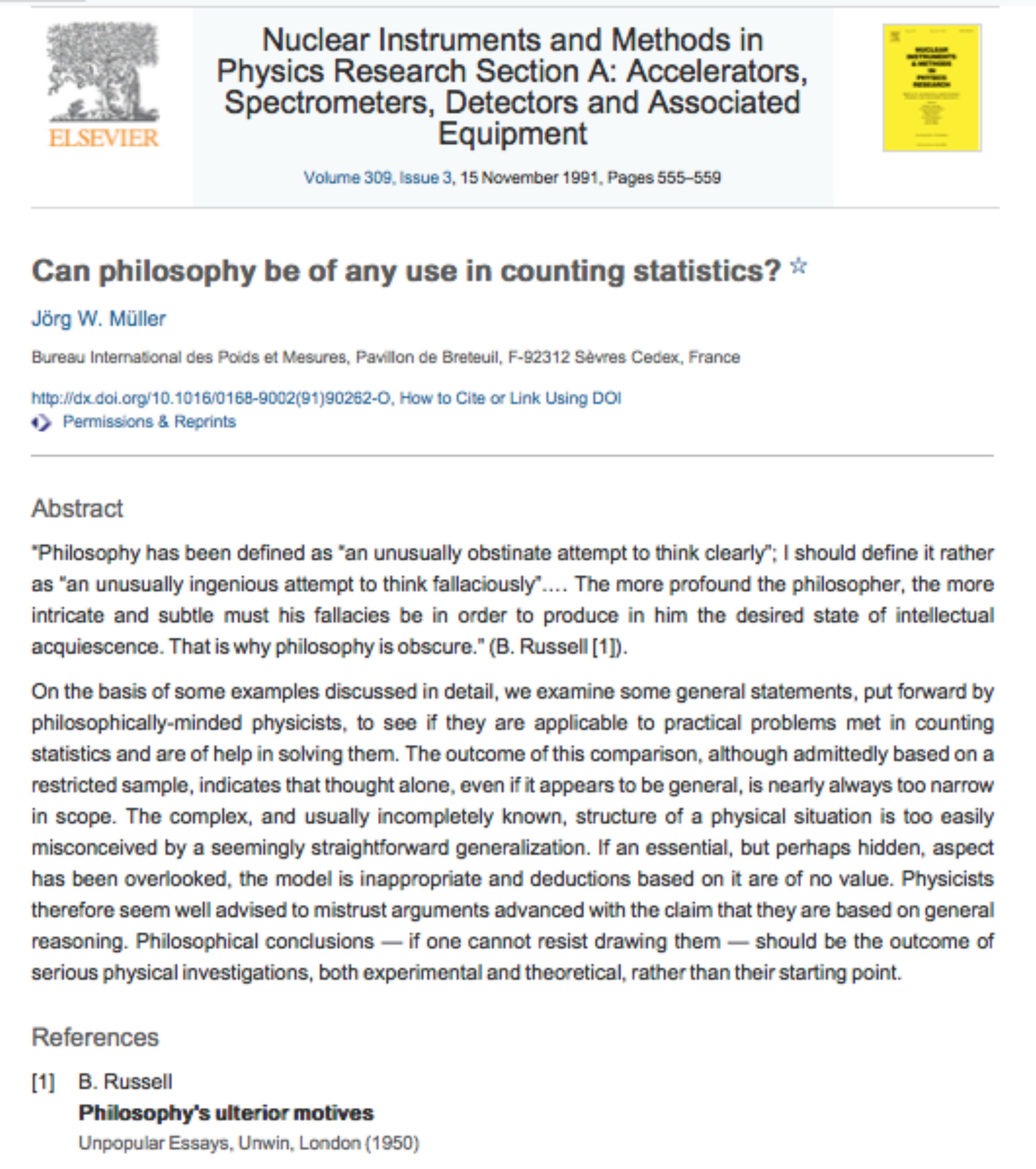
MEASUREMENT

DATA

inductive inference

PROBABILITY

This sounds a bit philosophical



ELSEVIER

Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment

Volume 309, Issue 3, 15 November 1991, Pages 555–559

Can philosophy be of any use in counting statistics? ☆

Jörg W. Müller

Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92312 Sèvres Cedex, France

http://dx.doi.org/10.1016/0168-9002(91)90262-O, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

"Philosophy has been defined as "an unusually obstinate attempt to think clearly"; I should define it rather as "an unusually ingenious attempt to think fallaciously".... The more profound the philosopher, the more intricate and subtle must his fallacies be in order to produce in him the desired state of intellectual acquiescence. That is why philosophy is obscure." (B. Russell [1]).

On the basis of some examples discussed in detail, we examine some general statements, put forward by philosophically-minded physicists, to see if they are applicable to practical problems met in counting statistics and are of help in solving them. The outcome of this comparison, although admittedly based on a restricted sample, indicates that thought alone, even if it appears to be general, is nearly always too narrow in scope. The complex, and usually incompletely known, structure of a physical situation is too easily misconceived by a seemingly straightforward generalization. If an essential, but perhaps hidden, aspect has been overlooked, the model is inappropriate and deductions based on it are of no value. Physicists therefore seem well advised to mistrust arguments advanced with the claim that they are based on general reasoning. Philosophical conclusions — if one cannot resist drawing them — should be the outcome of serious physical investigations, both experimental and theoretical, rather than their starting point.

References

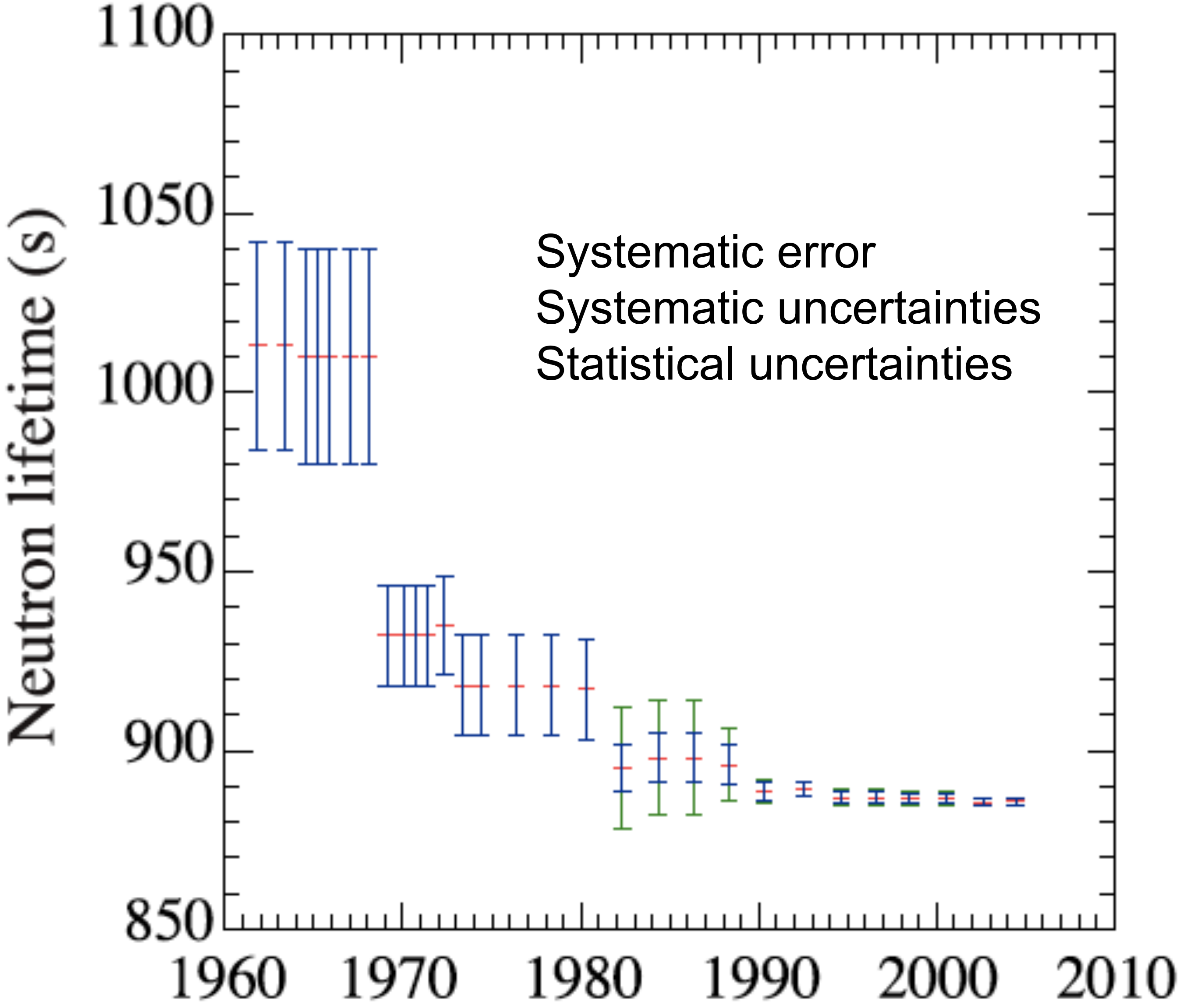
[1] B. Russell
Philosophy's ulterior motives
Unpopular Essays, Unwin, London (1950)

We want to answer these questions:

- How should we quantify the uncertainty on the measurement of certain parameter?
- How this uncertainty depends on other parameters used to obtain the result?
- If several parameters are obtained simultaneously from the same data how are their values correlated ?
- Do the results show a trend deviating from the expected ?
- How should we design a measurement in order to minimize uncertainties?

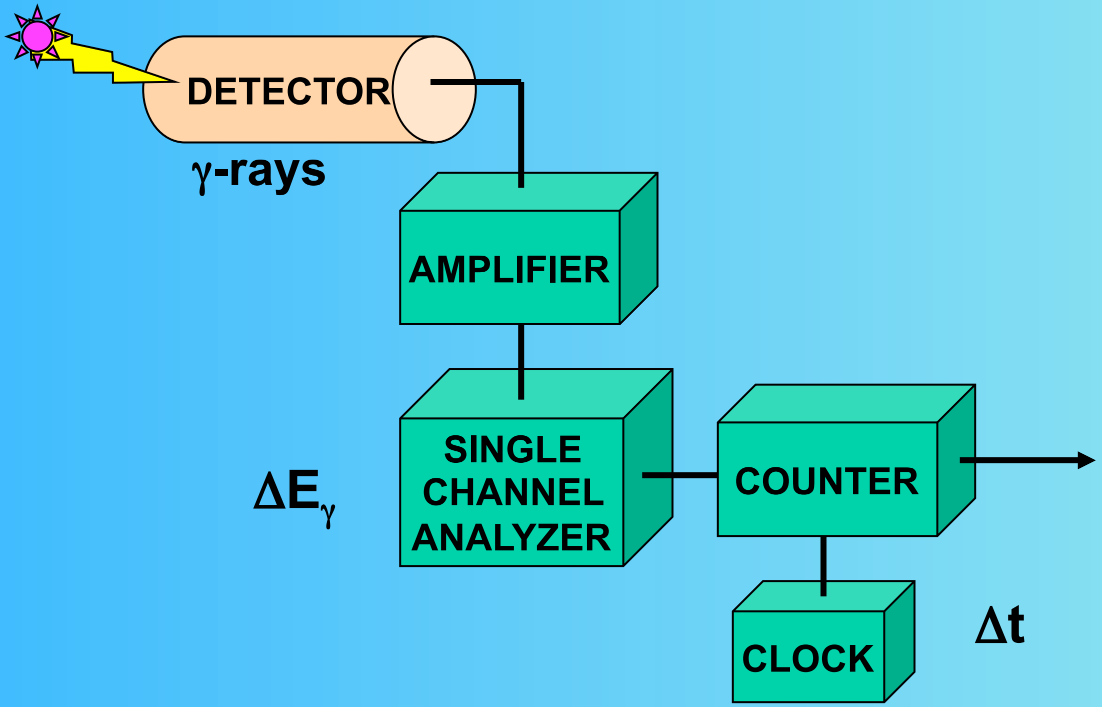
Vocabulary: uncertainty, error, precision, accuracy

- The repetition of a measurement under the *same* conditions usually leads to different outcomes: **uncertainty** *incertidumbre*
- If the conditions were really the same, the variations of the result can be related to the statistical nature of physical processes: **statistical uncertainty**
- If the conditions were actually varying between measurements (but this fact was unknown to us): **systematic uncertainty**
- If the measurement was faulty this could introduce a bias in the result: **systematic deviation**
- If the result of the measurement depends on not so well known parameters: **systematic error**
- We are assuming that the measurement has enough **precision** to allow distinguishing these variations *precision*
- The **accuracy** on the other hand measures the deviations of the measured value from the *true* value *"exactitud"*


Neutron lifetime (s)
1100
1050
1000
950
900
850
1960
1970
1980
1990
2000
2010
Systematic error
Systematic uncertainties
Statistical uncertainties

Examples of primary questions in NP and PP experiments :

- determine the amount of a radioactive isotope on a sample
- determine the half-life of a nuclear level or a particle
- determine the momentum distribution of certain reaction products


DETECTOR

γ-rays

AMPLIFIER

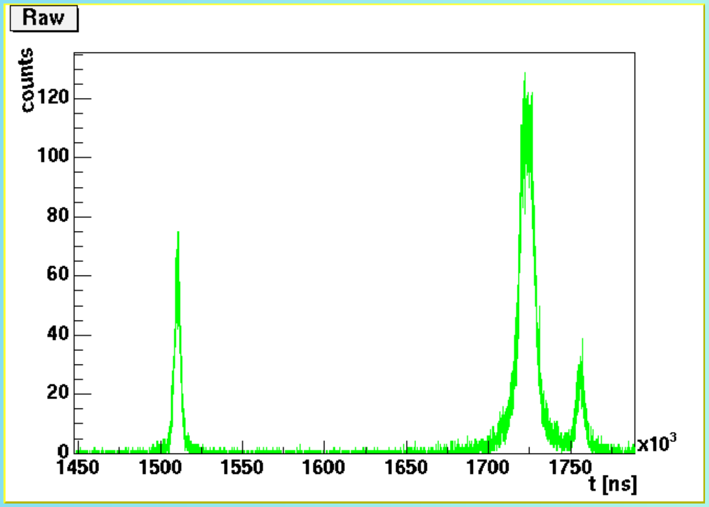
SINGLE CHANNEL ANALYZER

ΔE_γ

COUNTER

CLOCK

Δt



In nuclear and particle physics we are dealing with **counting experiments**: we register the number of counts in a given detector, produced by particles of a given type at a given time with a given momentum and under some other conditions. From this and other detector related information we obtain the requested data.

Some mathematical tools:

Random variables:

x, y, … represent variables (a certain magnitude)

{**x$_i$**}, {**y$_i$**}, … different values (the values it can take)

Probability density function (PDF):

P(x), P(y), P(x,y), … probability of obtaining x_i, or y_i, or x_i and y_i simultaneously

Discrete: x_1, x_2, … → Σ (e.g. number of events)

Continuous: x ∈ dx → ∫ (e.g. momentum of a particle)

Probability:

A function of the random variable which fulfill:

$$1)\; P(x) \geq 0\,, \quad 2) \int P(x)dx = 1, \quad 3)\; P(x_i)\,, P(x_j)\; indep.$$

Expected value of a function of the random variables:

$$E\left[f\right]=\int f(x,y,\ldots)P(x,y,\ldots)dxdy\ldots$$

Moments of the distribution:

algebraic: $E\left[x^k y^l \ldots\right]$

central: $E\left[\left(x-E[x]\right)^k\left(x-E[x]\right)^l\ldots\right]$

mean: $\bar{x}=E\left[x\right]=\int xP(x,y,\ldots)dxdy\ldots$ promedio

variance: $\sigma_x^2=E\left[\left(x-\bar{x}\right)^2\right]=\int\left(x-\bar{x}\right)^2 P(x,y,\ldots)dxdy\ldots$ varianza

skewness: $\gamma=\frac{1}{\sigma_x^3}\int\left(x-\bar{x}\right)^3 P(x,y,\ldots)dxdy\ldots$ sesgo

kurtosis: $\xi+3=\frac{1}{\sigma_x^4}\int\left(x-\bar{x}\right)^4 P(x,y,\ldots)dxdy\ldots$ kurtosis

median: the value that separates the probability distribution in two halves… *mediana*

covariance: $\sigma_{xy} = E\left[\left(x-\bar{x}\right)\left(y-\bar{y}\right)\right] = \int \left(x-\bar{x}\right)\left(y-\bar{y}\right)P(x,y,\ldots)dxdy\ldots$

correlation: $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

x and y are independent if $P(x,y)=P(x)P(y) \Rightarrow \sigma_{xy}=0$
X and y are uncorrelated if $\sigma_{xy}=0$ ≠ independent

confidence interval *[a,b]* and confidence level α

$$\alpha = \int_a^b dx \int P(x,y,\ldots)dydz\ldots$$

Binomial distribution

- Probability that out of N particles x disintegrate in a time interval Δt

$$x = \underbrace{\lambda \Delta t}_{p} N$$

- Probability that if there are $n_a n_b$ collisions there are x reactions

$$x = \underbrace{n_a n_b}_{N} \underbrace{\sigma / S}_{p}$$

x : success, N : trials, p : probability

$$P(x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \equiv \mathrm{B}(N, p)$$

$$\bar{x} = Np$$

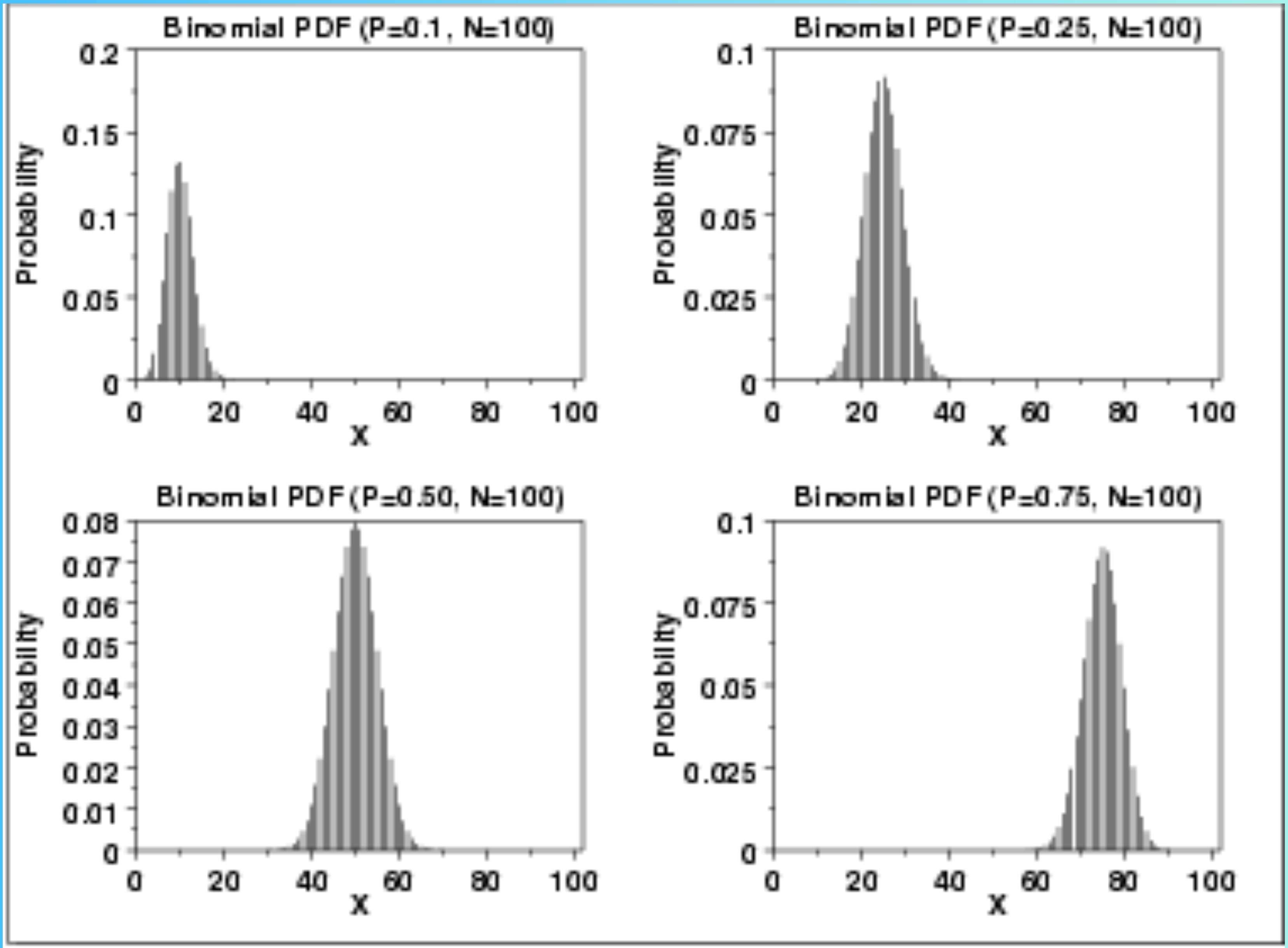
$$\sigma^2 = \bar{x}(1-p)$$

$$\gamma = \frac{1-2p}{\sigma}$$

$$\xi = \frac{1}{\sigma^2} - \frac{6}{N}$$

DISCRETE

The basic distribution of counting experiments


Binomial PDF (P=0.1, N=100)
Binomial PDF (P=0.25, N=100)
Binomial PDF (P=0.50, N=100)
Binomial PDF (P=0.75, N=100)
Probability
X
0
0.05
0.1
0.15
0.2
0.025
0.075
0.01
0.02
0.03
0.04
0.06
0.07
0.08
20
40
60
80
100

Poisson distribution

- Limiting case of binomial distribution when N >> and p<<

x : success, N : trials, p : probability

$\mu = Np$: mean

$$P(x) = \frac{\mu^x}{x!} e^{-\mu} \equiv \mathrm{P}(\mu)$$

$$\bar{x} = \mu$$

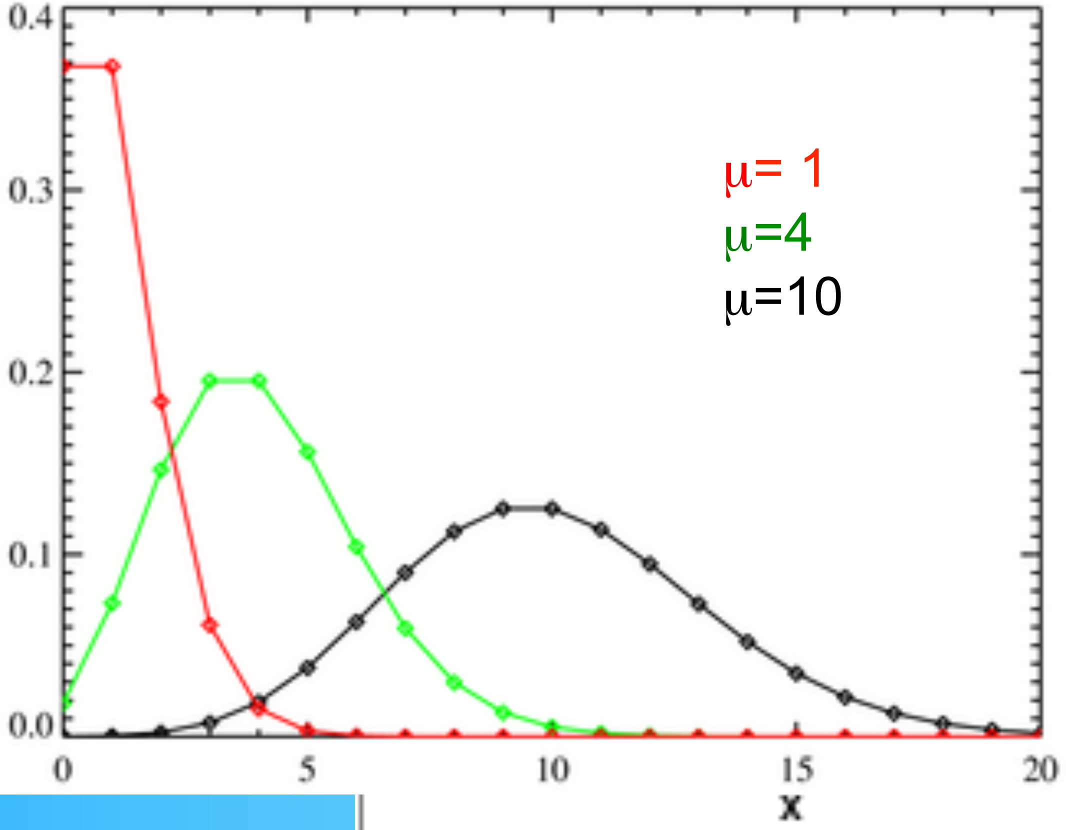
$$\sigma^2 = \mu$$

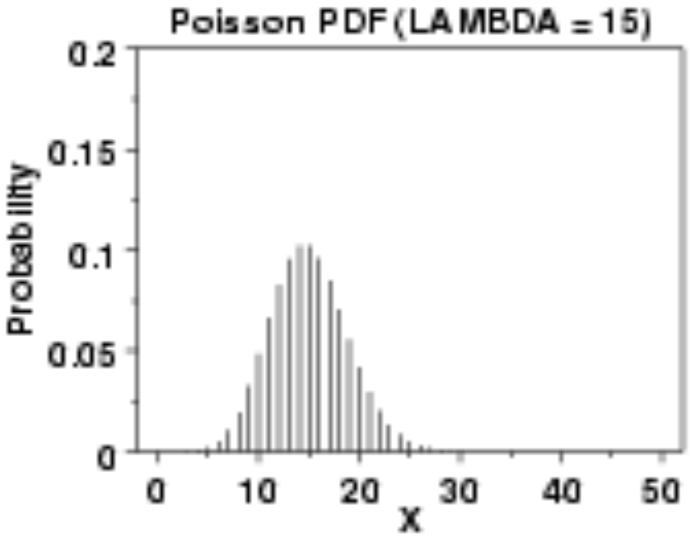
$$\gamma = \frac{1}{\sqrt{\mu}}$$

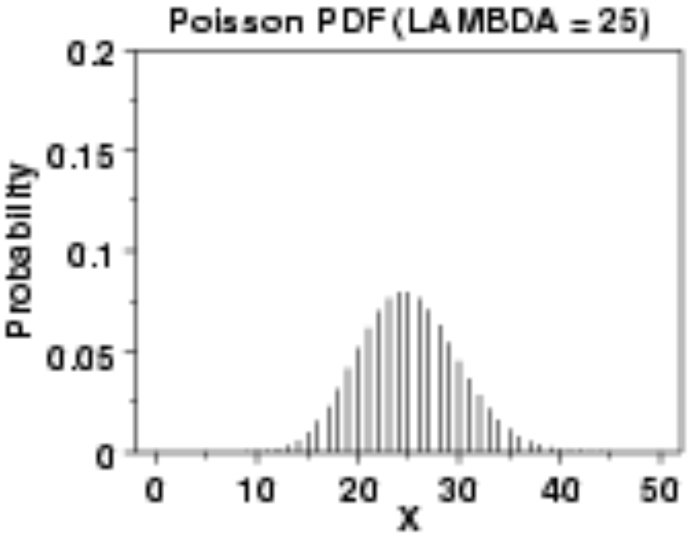
$$\xi = \frac{1}{\mu}$$

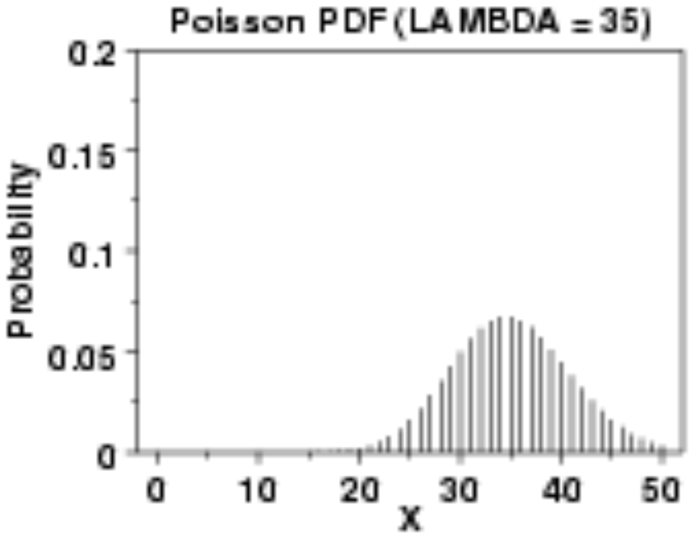
DISCRETE

The distribution used in NP and PP counting experiments


μ= 1
μ=4
μ=10
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
x


Poisson PDF (LAMBDA = 15)
Probability
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
X


Poisson PDF (LAMBDA = 25)
Probability
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
X


Poisson PDF (LAMBDA = 35)
Probability
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
X

Multinomial distribution

- Related to classification problems as histograms: probability that out of N events x_1 are of type 1, x_2 are of type 2, …

x_1, x_2, ... : events of type 1, 2, …
N : trials
p_1, p_2,... : probability of types 1, 2, …

$$P(x_1,x_2,\ldots) = \frac{N!}{x_1!x_2!\ldots} p_1^{\;x_1} p_2^{\;x_2} \ldots \equiv \mathrm{M}\left(N, p_1, p_2, \ldots\right)$$

$$\bar{x}_i = Np_i$$

$$\sigma_i^2 = p_i\left(1 - p_i\right)$$

$$\sigma_{ij} = -Np_i p_j$$

DISCRETE

Multinomial:
Poisson distribution for each channel with or without correlations

$$\mathrm{M}\left(N, p_1, p_2, \ldots\right) = \frac{\mathrm{P}\left(\mu_1\right)\mathrm{P}\left(\mu_2\right)\ldots}{\mathrm{P}\left(N\right)}$$

Normal or Gaussian distribution

- Appear as a consequence of the Law of Large Numbers. Good approximation of Binomial or Poisson distribution for large $\mu = Np$

μ: mean, σ: width

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \equiv \mathrm{N}(\mu, \sigma)$$

$$\overline{x} = \mu$$

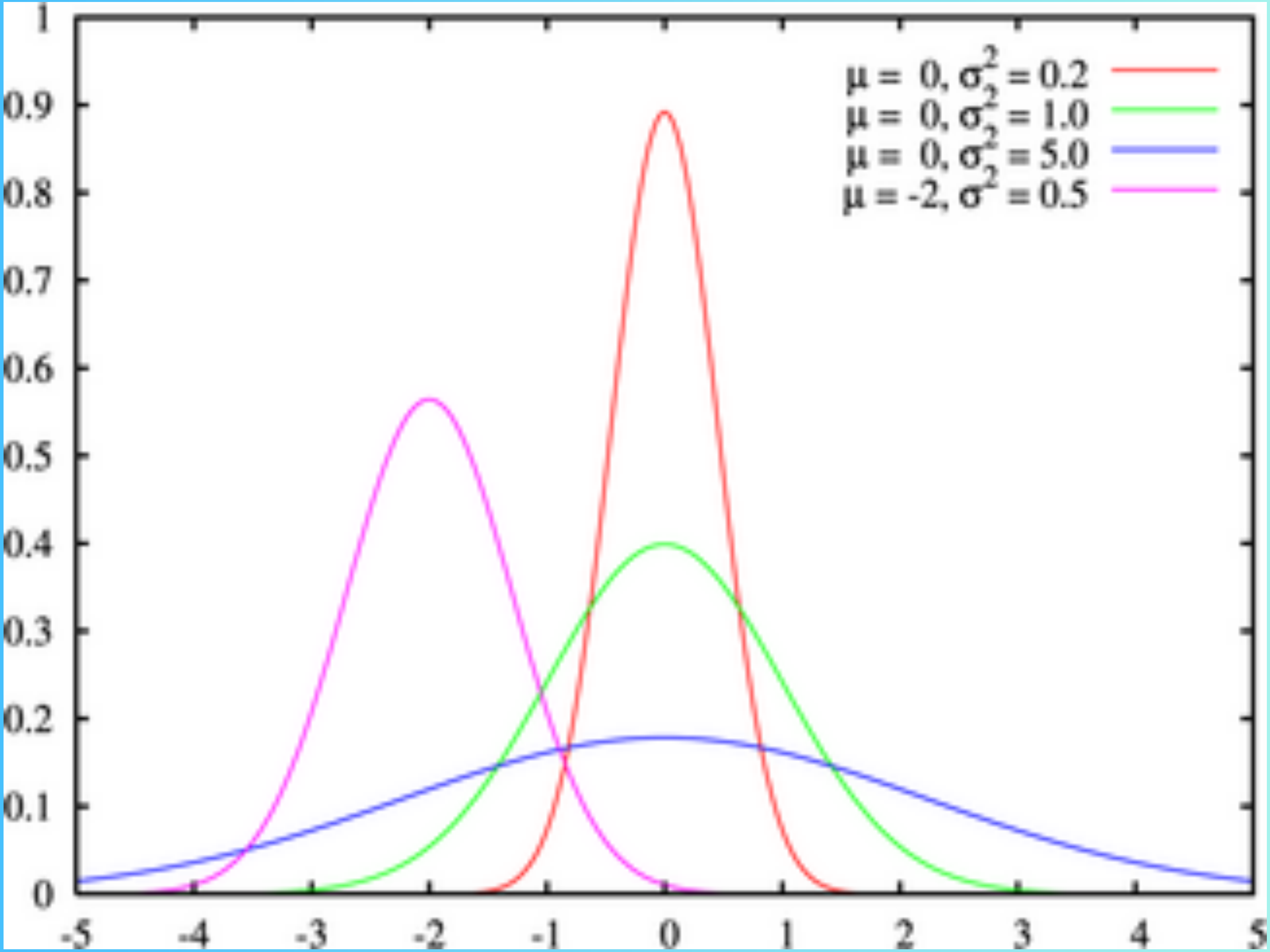
$$\sigma^2 = \sigma^2$$

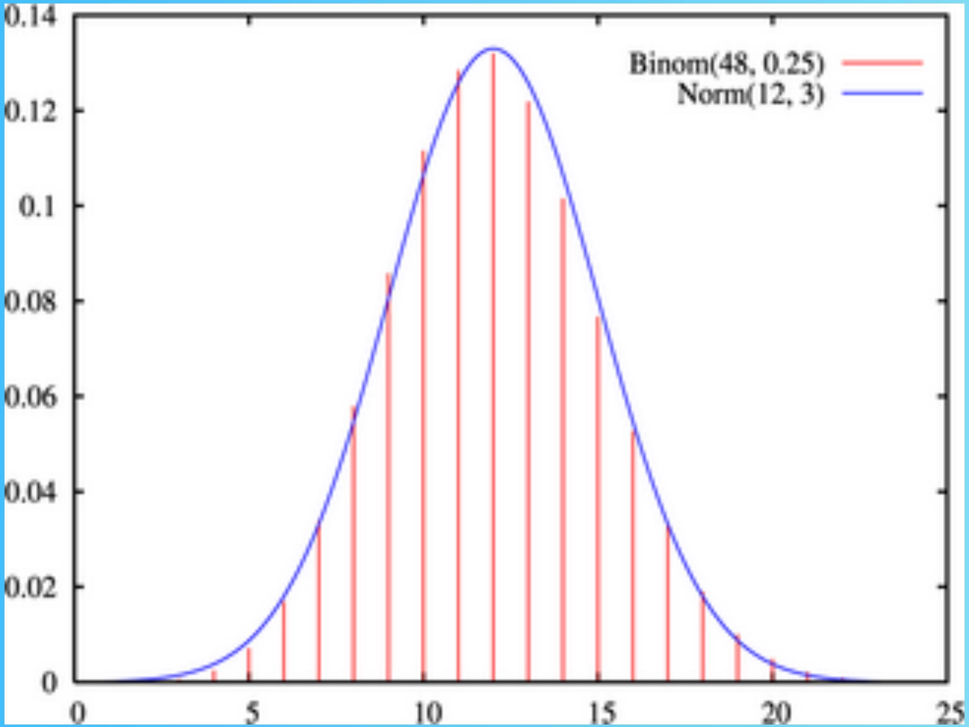
$$\gamma = 0$$

$$\xi = 0$$

CONTINUOUS

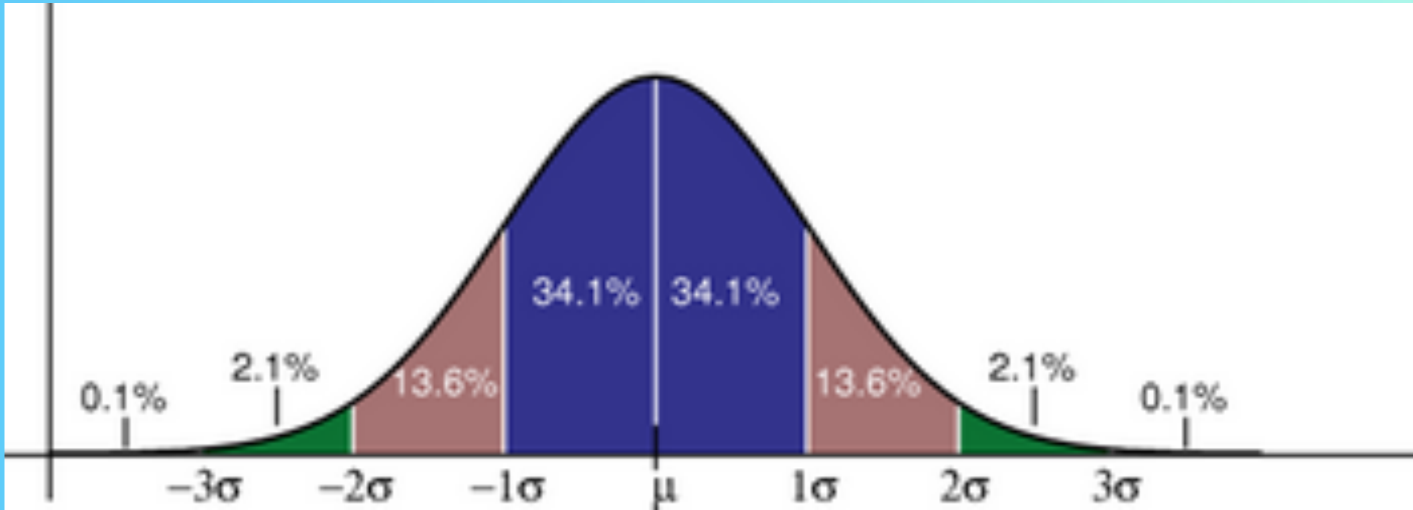
The most ubiquitous distribution in experimental science


$\mu = 0, \sigma^2 = 0.2$
$\mu = 0, \sigma^2 = 1.0$
$\mu = 0, \sigma^2 = 5.0$
$\mu = -2, \sigma^2 = 0.5$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5



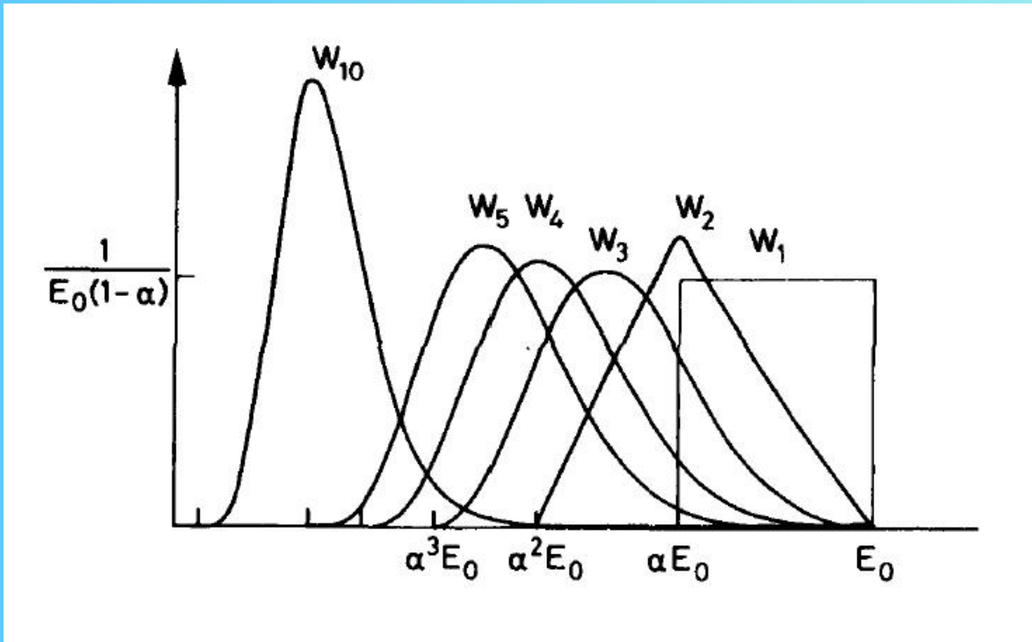
Good approximation of Binomial or Poisson when Np or μ ≥ ~10

Probability integrals



Central limit theorem

- The mean value calculated from a subset of a sufficiently large number of random samples will be approximately normally distributed
- The PDF of the sum of independent random variables is the convolution of the individuals PDF. The convolution of a large number of PDF tends to the normal distribution



χ^2 distribution

- Is the distribution followed by the sum of the square of ν independent random variables each with distribution *N(0,1)*

ν: degrees of freedom

$$P(x) = \frac{(x/2)^{\nu/2-1}}{2\Gamma(\nu/2)} e^{-\frac{x}{2}} \equiv \chi^2(\nu)$$

$$\bar{x} = \nu$$

$$\sigma^2 = 2\nu$$

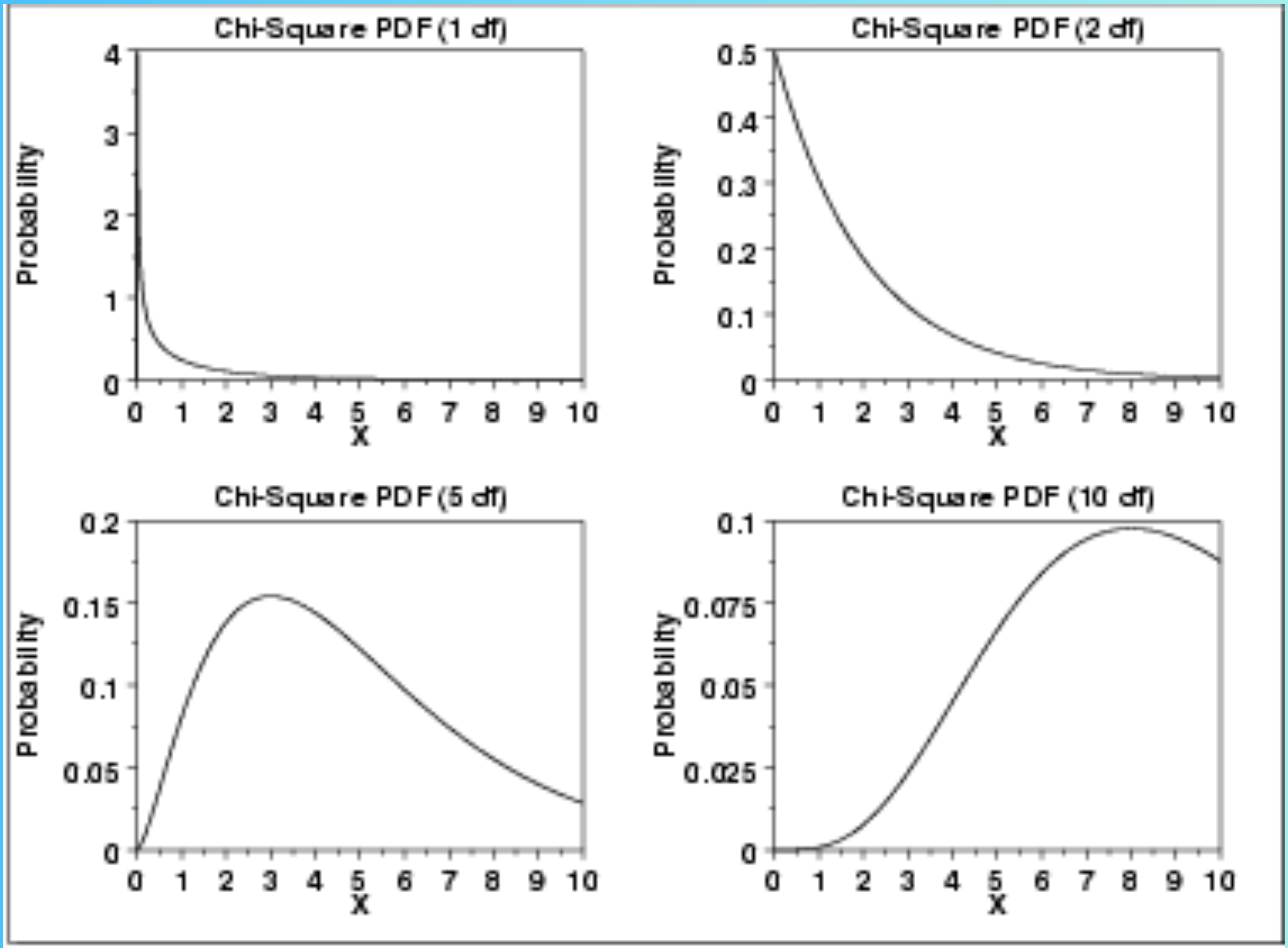
$$\gamma = 2\sqrt{2/\nu}$$

$$\xi = 12/\nu$$

CONTINUOUS

Useful for testing consistency of data points

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


Chi-Square PDF (1 df)
Chi-Square PDF (2 df)
Chi-Square PDF (5 df)
Chi-Square PDF (10 df)
Probability
X
0
1
2
3
4
0.5
0.4
0.3
0.2
0.1
0.15
0.05
0.075
0.025
5
6
7
8
9
10

Student t distribution

- Is the distribution followed by $\left(\hat{x}-\bar{x}\right)\Big/ s_x\sqrt{\left(\nu+1\right)}$ where $\hat{x},\ s_x^2$ are the mean and variance of a sample of size ν+1 whose parent distribution has mean value $\bar{x}$

ν: degrees of freedom

$$P(x)=\frac{\Gamma\left(\left(\nu+1\right)/2\right)}{\sqrt{\pi\nu}\,\Gamma\left(\nu/2\right)}\frac{1}{\left(1+x^2/\nu\right)^{(\nu+1)/2}}\equiv t\left(\nu\right)$$

$$\bar{x}=0$$

$$\sigma^2=\frac{\nu}{\nu-2}\quad,\quad\nu>2$$

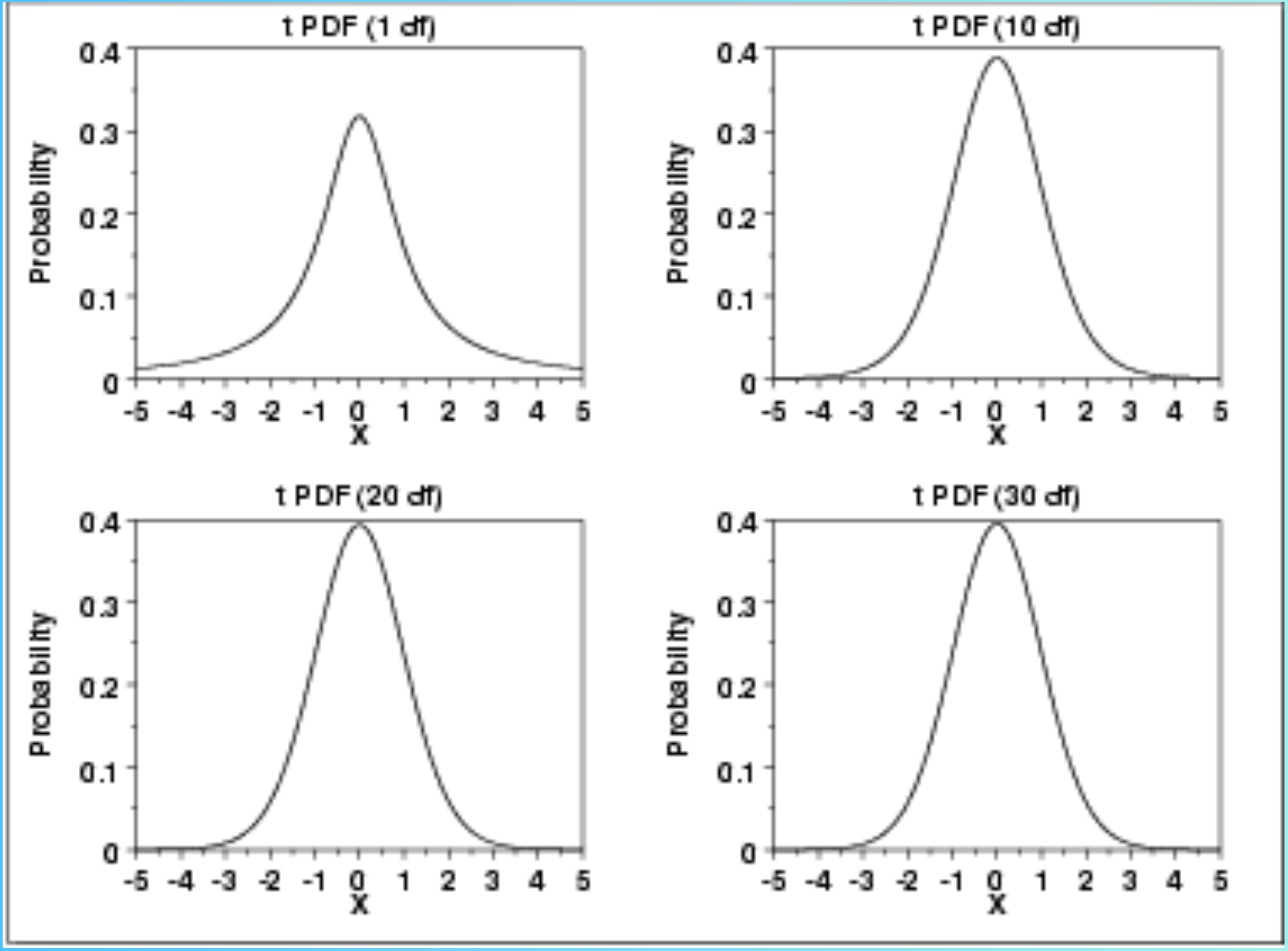
$$\gamma=0$$

$$\xi=\frac{6}{\nu-4}\quad,\quad\nu>4$$

Useful for testing the significance of the difference of two sample means

For ν=1 reduces to the Cauchy or Lorentzian distribution

For $\nu \to \infty$ approaches $N(0,1)$



What is probability?

The Frequentist point of view

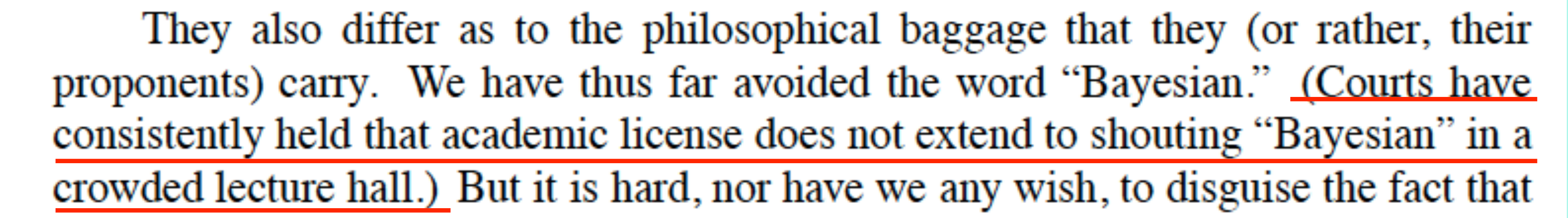
If we repeat and *infinite* number of times a measurement in *exactly the same conditions* we would obtain the PDF of the data

- **It is an “experimental” definition, not an abstraction**
- **Cannot be applied to parameters or hypothesis**

The Bayesian point of view

It measures the plausibility (objective) of, or the degree of belief (subjective) on anything

- **Based on Bayes theorem**
- **Can be applied to data, parameters or hypothesis**



They also differ as to the philosophical baggage that they (or rather, their proponents) carry. We have thus far avoided the word “Bayesian.” (Courts have consistently held that academic license does not extend to shouting “Bayesian” in a crowded lecture hall.) But it is hard, nor have we any wish, to disguise the fact that

NUMERICAL RECIPES: THE ART OF SCIENTIFIC COMPUTING: William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery

Bayes theorem:

$$P(C|E,I) = \frac{P(E|C,I)\,P(C|I)}{P(E|I)}$$

C: cause
E: effect
I: prior information

$P(C|I)$: prior probability

$P(E|C,I)$: likelihood function

$P(C|E,I)$: posterior probability

$P(E|I)$: normalization $\sum_{C} P(E|C,I)\,P(C|I)$

- ***C, E, I*** **: random variables (data or parameters) or propositions (hypothesis)**
- ***P*** **: degree of believe**
- **all probabilities are conditional in the subjective version (*I*!)**
- **allows to update the knowledge with new information**
- **intimately related to the objective of experimental science**

Parameter estimation

Estimator: Probability density function of the data sample and of the parameters which allows to estimate the latter.

Desired properties of estimator:

a) **Consistent**: if the sample increases the parameter value converges

b) **Unbiased**: in the limit of infinite sample size the parameter attains the “true” value

c) **Efficient**: the variance of the estimator is minimal (among the possible estimators)

d) **Robust**: the result (parameter value) is independent on the sample

Maximum likelihood estimator

Maximize the likelihood $L(\theta|x)$

The better one

$$\max L(\theta \mid x) = \max \prod_i P\left(x^i, \theta\right)$$

$P\left(x^i \middle| \theta\right)$: PDF of random variable x depending on parameter θ

For practical reasons often the log-likelihood is used:

$$\max \ln L\left(\theta \middle| x\right) = \max \sum_i \ln P\left(x^i \middle| \theta\right)$$

Application of ML estimator:

Fitting data: For Poisson distributed data two forms are usually employed:

$\ln L = \sum_i \ln f\left(x^i\right)$: event by event data (for low statistics)

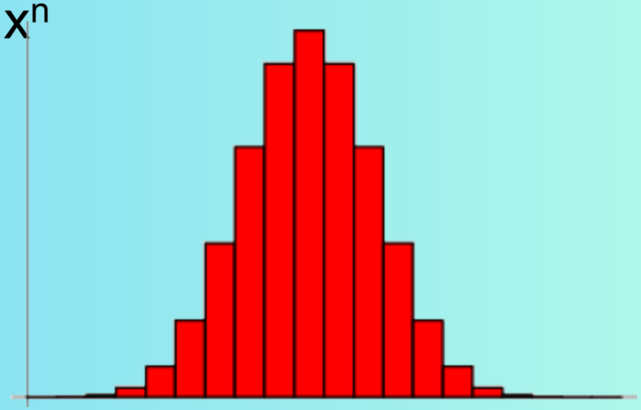
$\ln L = \sum_i n_i \ln f_i - f_i$: binned data (histograms)

Application of ML estimator:

Statistical sample characterization

How can we characterize the results of the repetition of the same experiment (sample)? ➔ <u>sample statistic</u>

N experiments to determine x, results: $x^1, x^2, \ldots, x^n$



Sample distribution: histogram

Sample mean:

$$\hat{x} = \frac{1}{N}\sum_i x^i \rightarrow \overline{x} = \hat{x}$$

Sample variance:

$$s_x^2 = \frac{1}{N}\sum_i \left(x^i - \hat{x}\right)^2 \rightarrow \sigma_x^2 = \frac{N}{N-1} s_x^2$$

Sample covariance:

$$s_{xy} = \frac{1}{N}\sum_i \left(x^i - \hat{x}\right)\left(y^i - \hat{y}\right) \rightarrow \sigma_{xy} = \frac{N}{N-1} s_{xy}$$

Application of ML estimator:

Combining measurements with different uncertainties

Weighted mean:

Set of measurements of the same quantity each one with a value and uncertainty:

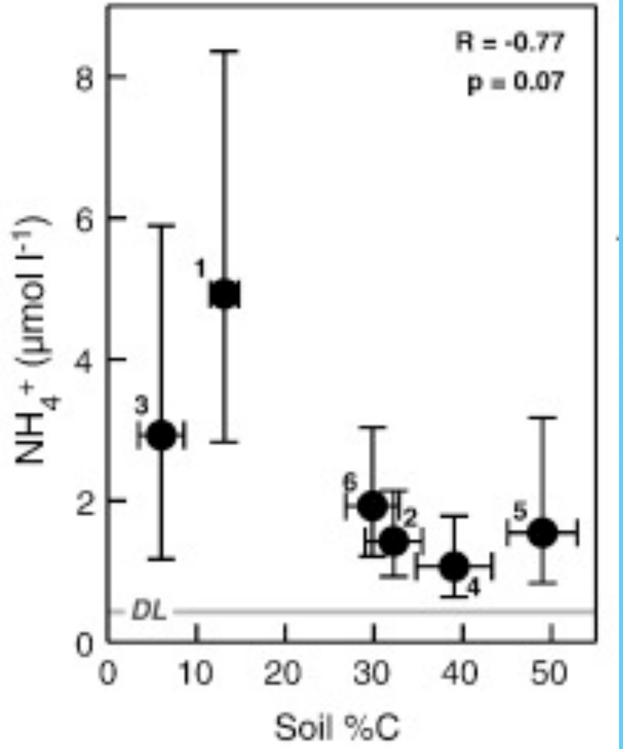
$$\mu_i \pm \sigma_i$$

Average value:

$$\mu = \frac{\sum_i \frac{\mu_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

Uncertainty of the average:

$$\sigma^2 = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}$$



Least squares estimator

The popular one

Minimize the squared deviations $Q^2(\theta|x)$

$$\min Q^2(\theta \mid x) = \sum_i \sum_j \left(x_i - \mathrm{x}_i(\theta)\right) V_{ij}^{-1} \left(x_j - \mathrm{x}_j(\theta)\right) \rightarrow \boxed{Q^2 = \sum \frac{\left(y_i - f(x_i)\right)^2}{\sigma_i^2}}$$

V_{ij} : covariance matrix

Can be deduced from ML for normally distributed data: $P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

To fit Poisson distributed data ($\sigma^2_i = y_i$) in histograms two forms are usually employed (counting experiments):

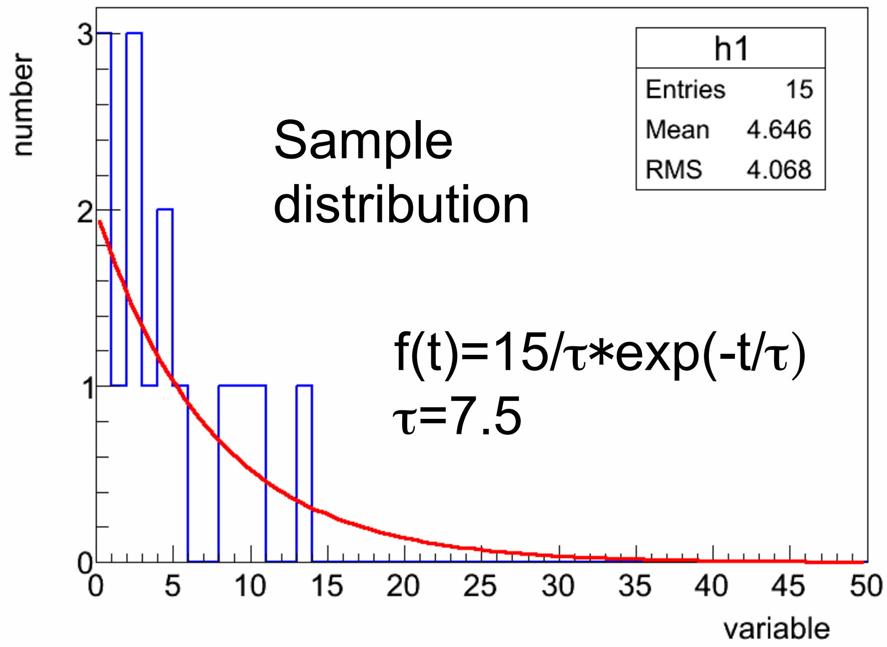
$$Q_a^2 = \sum \frac{\left(y_i - f(x_i)\right)^2}{y_i}$$

→ Cannot handle bins with cero counts: do not use for low statistics! (Popular solution: exclude those bins, biases the result!)

$$Q_b^2 = \sum \frac{\left(y_i - f(x_i)\right)^2}{f(x_i)}$$

→ Generally applicable (numerically more demanding)

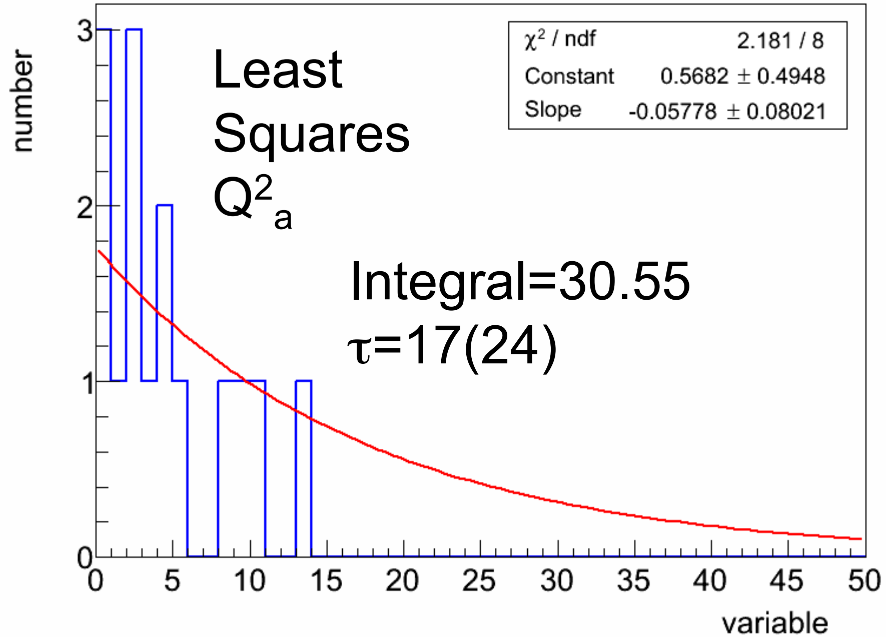
Examples of fits with low statistics:
Exponential distribution



h1	
Entries	15
Mean	4.646
RMS	4.068

Sample distribution

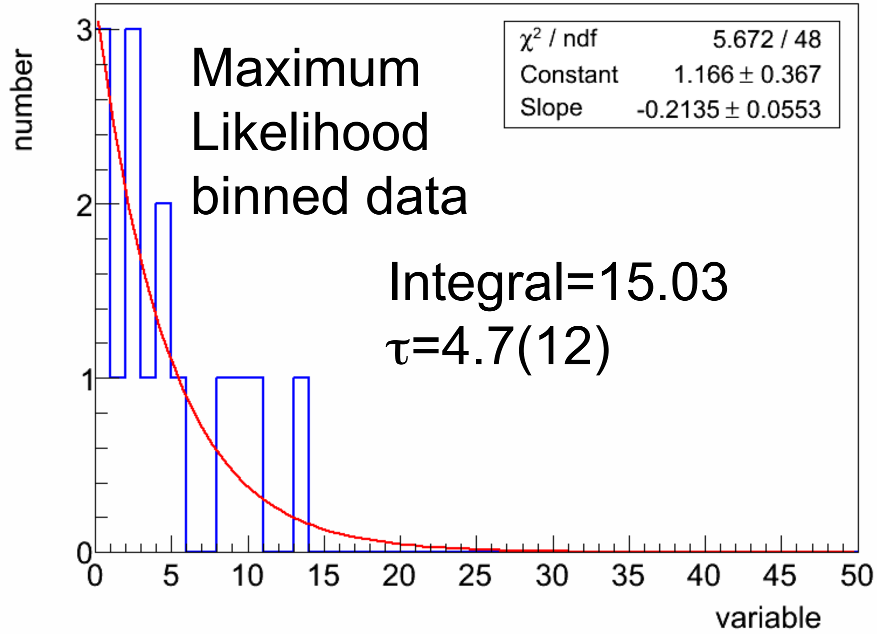
$f(t)=15/\tau*\exp(-t/\tau)$
$\tau=7.5$



χ^2 / ndf	2.181 / 8
Constant	0.5682 ± 0.4948
Slope	-0.05778 ± 0.08021

Least Squares Q^2_a

Integral=30.55
$\tau=17(24)$



χ^2 / ndf	5.672 / 48
Constant	1.166 ± 0.367
Slope	-0.2135 ± 0.0553

Maximum Likelihood binned data

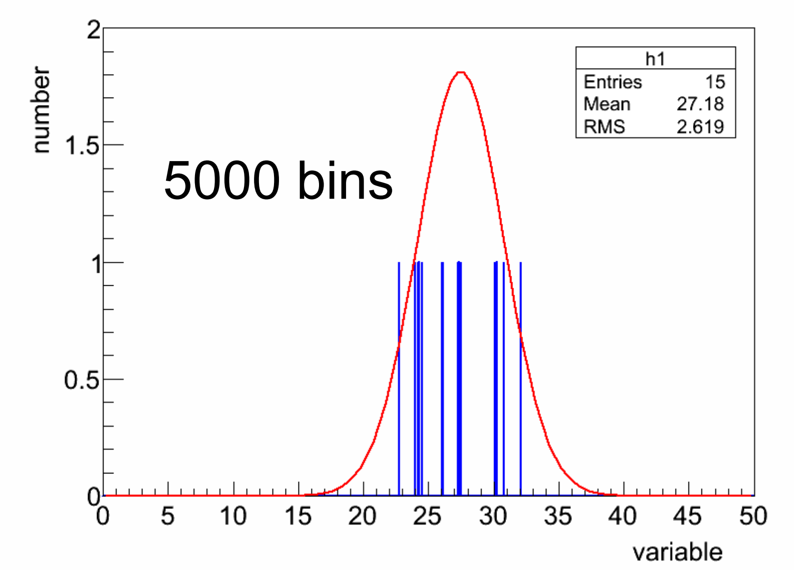
Integral=15.03
$\tau=4.7(12)$

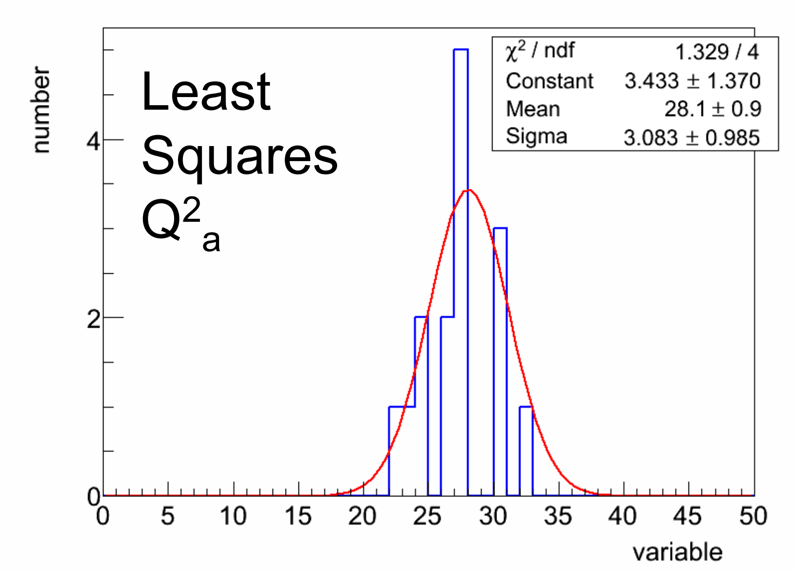
Example: Normal distribution

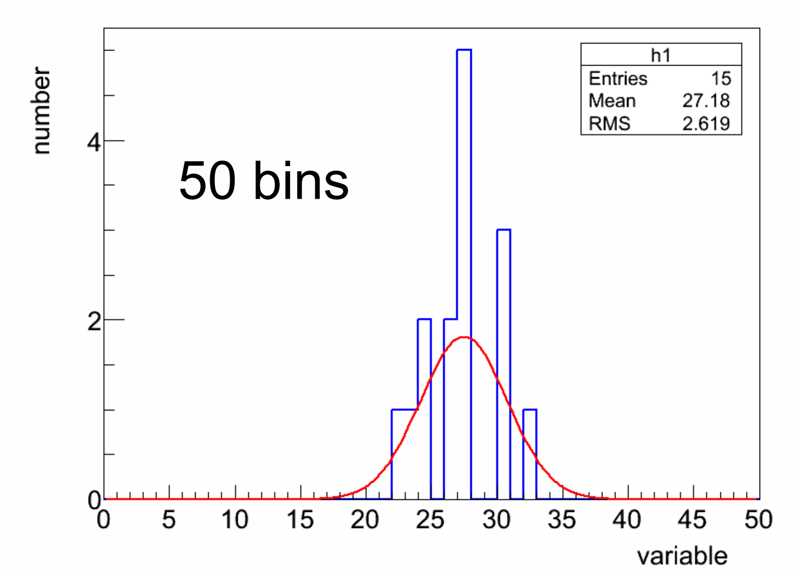
Drawn from N(27.5,3.3)

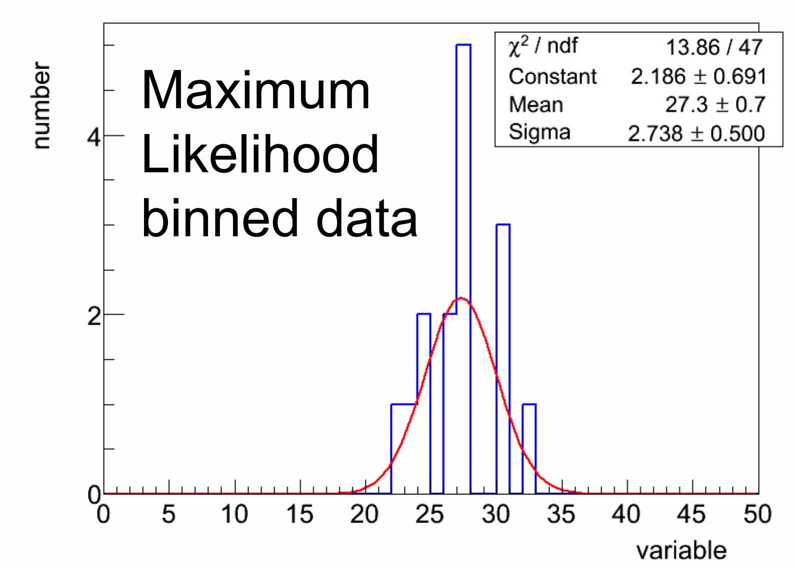
Sample[15]={30.7965, 26.0653, 30.0799, 27.4008, 30.2201, 27.3128, 24.5271, 27.2535, 27.5261, 26.1445, 32.0909, 24.2493, 27.3385, 22.737, 23.9998}

Sample mean= 27.183, RMS= 2.619









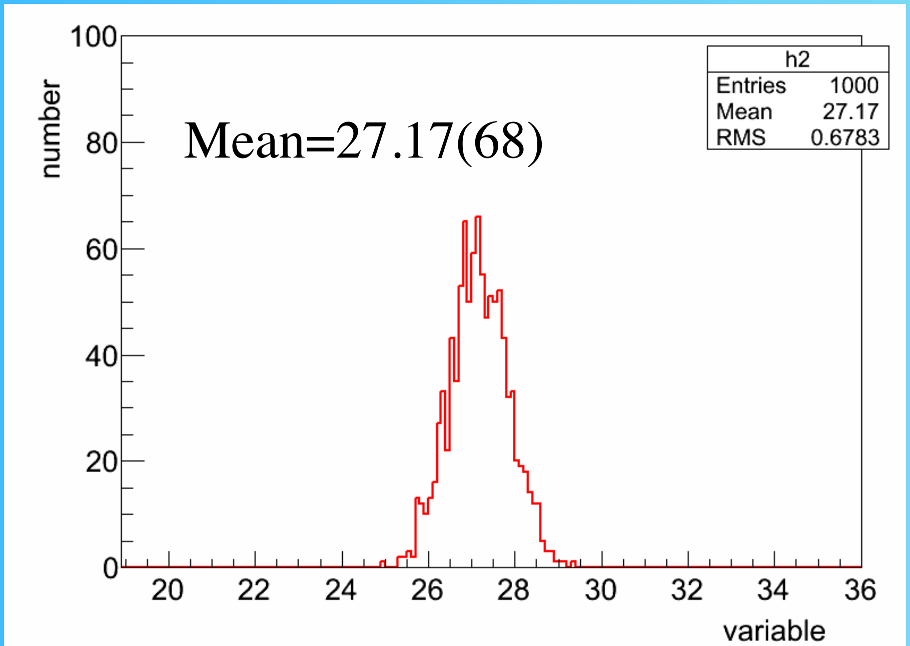
Bootstrap methods

- Monte Carlo method to evaluate sample statistic (mean, variance,…) and its uncertainty from samples
- Useful when the underlying distribution is unknown or the sample size is too small
- Consists of resampling with replacement the original sample many times, each time calculating the statistic, to finally compute the average of the distribution so obtained

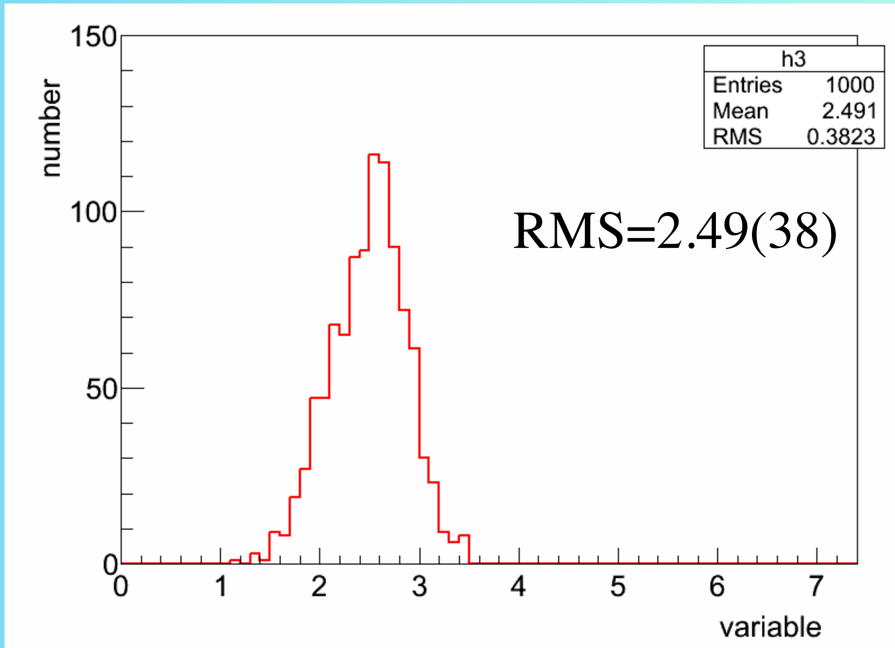
Example (previous sample):

Drawn from N(27.5,3.3)

N_B=1000



Mean=27.17(68)



RMS=2.49(38)

Derived magnitudes. Uncertainty propagation.

If the magnitude y is a function of other magnitudes with pdf $P(x_1, x_2, \ldots)$ what is the covariance on y coming from the covariance on $x_1, x_2, \ldots$?

Taylor expansion:

$$y(x_1, x_2, \ldots) = y(\bar{x}_1, \bar{x}_2, \ldots) + \sum_i \frac{\partial y}{\partial x_i}(x_i - \bar{x}_i) +$$

$$\sum_i \sum_j \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}(x_i - \bar{x}_i)(x_j - \bar{x}_j) + \ldots$$

Estimation of the covariance matrix:

$$\bar{y} = y(\bar{x}_1, \bar{x}_2, \ldots) + \mathrm{O}(2) + \ldots$$

$$\sigma_y^2 \cong \sum_i \sum_j \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\sigma_{x_i x_j}$$

$$\sigma_{y_k y_l} \cong \sum_i \sum_j \frac{\partial y_k}{\partial x_i}\frac{\partial y_l}{\partial x_j}\sigma_{x_i x_j}$$

If O(2)=0 use O(3)!

Approximation!:

- Non-zero 1st derivative
- Small 2nd derivative
- Small σ

Uncertainty propagation: some simple cases

$$z = ax + by \qquad \sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy}$$

$$z = axy \qquad \frac{\sigma_z^2}{\overline{z}^2} = \frac{\sigma_x^2}{\overline{x}^2} + \frac{\sigma_y^2}{\overline{y}^2} + 2\frac{\sigma_{xy}}{\overline{xy}}$$

$$z = a\frac{x}{y} \qquad \frac{\sigma_z^2}{\overline{z}^2} = \frac{\sigma_x^2}{\overline{x}^2} + \frac{\sigma_y^2}{\overline{y}^2} - 2\frac{\sigma_{xy}}{\overline{xy}}$$

$$z = ax^b \qquad \frac{\sigma_z^2}{\overline{z}^2} = b^2\frac{\sigma_x^2}{\overline{x}^2}$$

$$z = ae^{bx} \qquad \frac{\sigma_z^2}{\overline{z}^2} = b^2\sigma_x^2$$

$$z = a\ln(bx) \qquad \sigma_z^2 = a^2\frac{\sigma_x^2}{\overline{x}^2}$$

Beware of the correlations!

Inverse problems

Linear inverse problems:
$\mathbf{d} = \mathbf{R} \cdot \mathbf{f}$

Solution is not $\mathbf{f} = \mathbf{R}^{-1} \cdot \mathbf{d}$

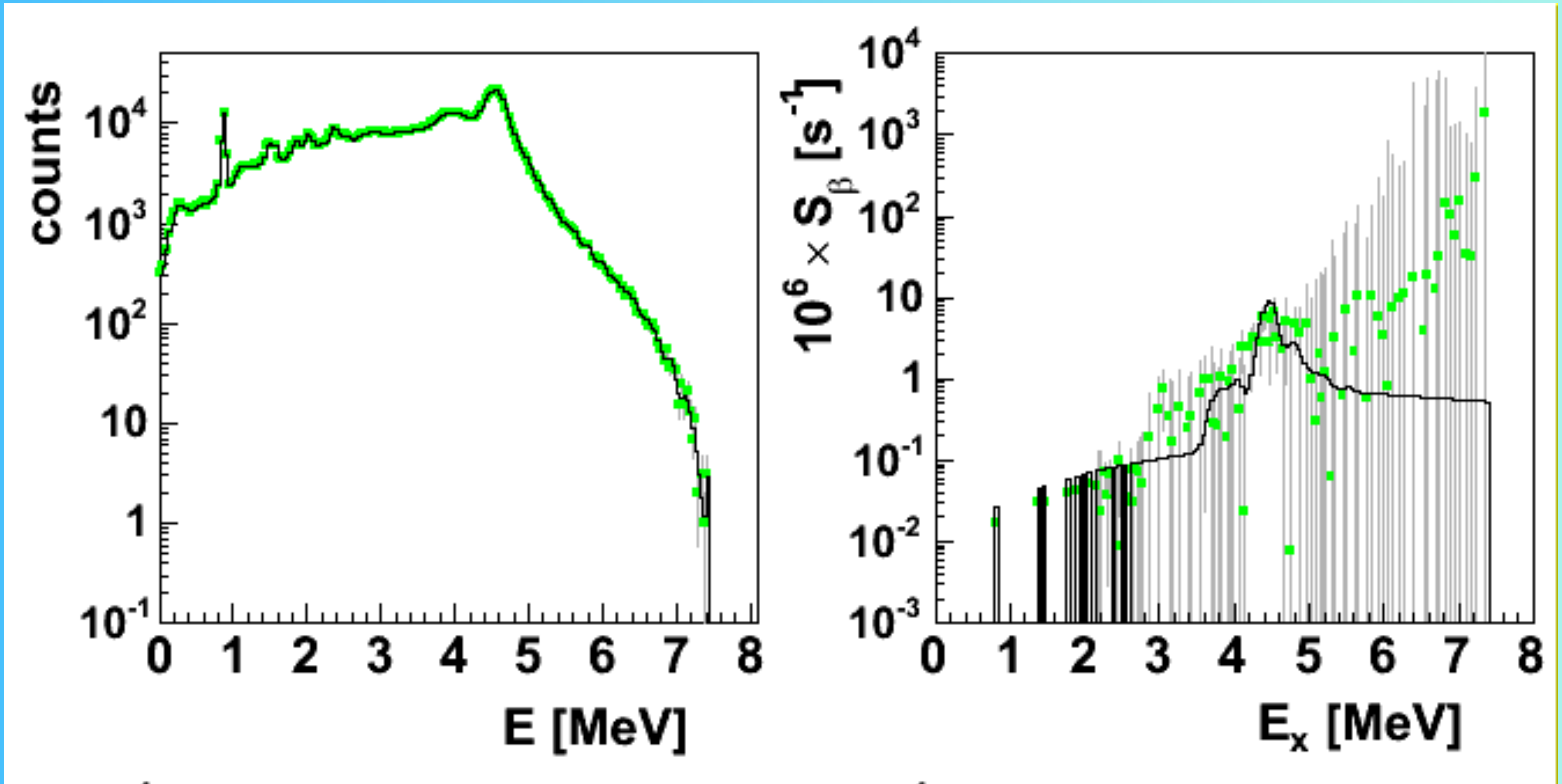
ill-posed or ill-conditioned problems

Problem:

- statistical nature of the problem
- numerical difficulties of the inversion

Solution:

- reproduce the data in χ^2 or maximum-likelihood sense
- use *a priori* information on the solution



Solution of linear inverse problems: $\mathbf{d} = \mathbf{R} \cdot \mathbf{f}$ (I)

Linear Regularization (LR) method:

- solution must be smooth: polynomial

$$\boxed{\min : \chi^2(\mathbf{f}) + \lambda \left|\mathbf{B} \cdot \mathbf{f}\right|^2}$$

λ: Lagrange multiplier

B: regularization matrix of order *o* (0,1,2,…)

Algorithm:

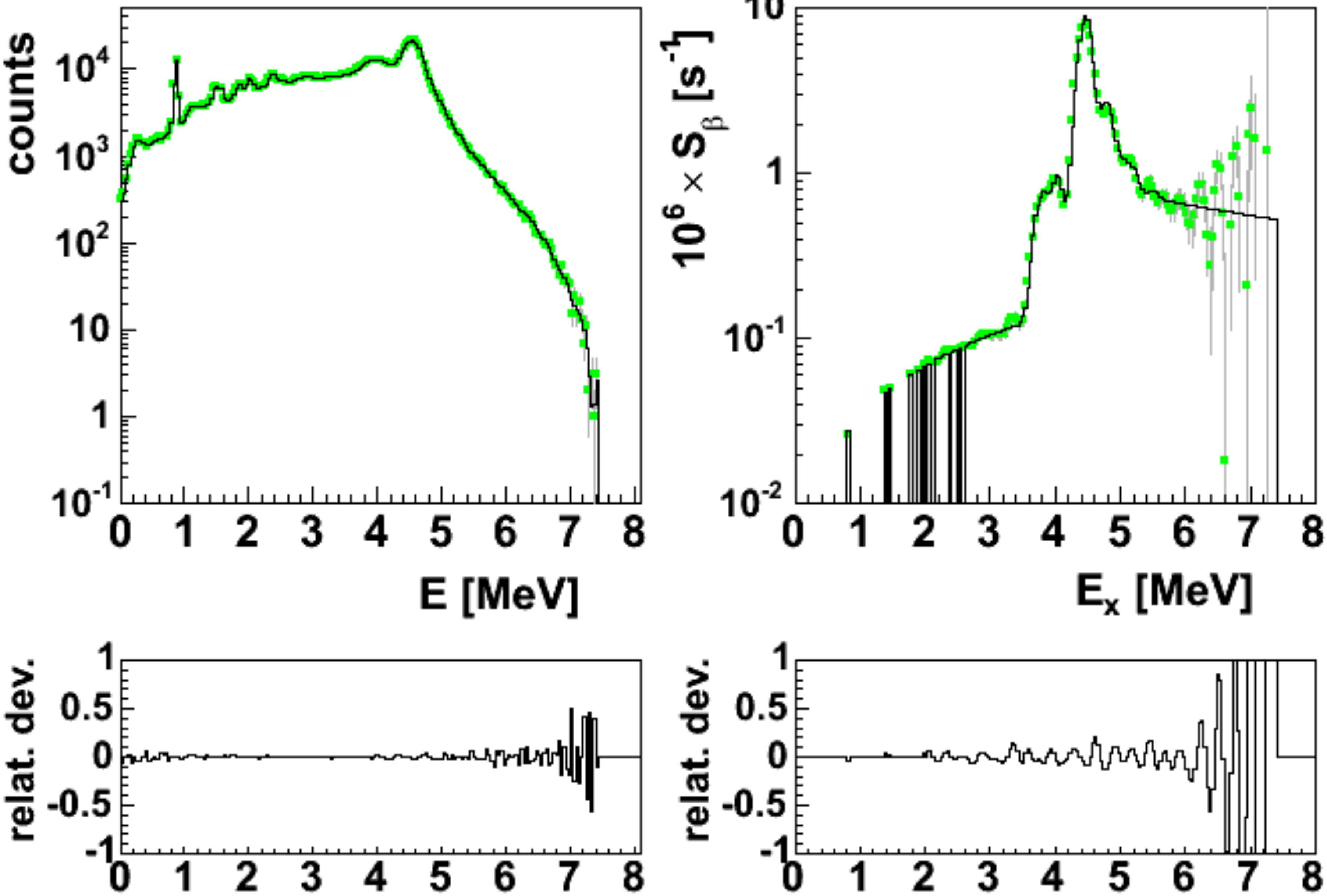
$$\mathbf{f} = \left(\mathbf{R}^{\mathrm{T}} \cdot \mathbf{V}_{\mathbf{d}}^{-1} \cdot \mathbf{R} + \lambda \mathbf{B}^{\mathrm{T}} \cdot \mathbf{B}\right)^{-1} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{V}_{\mathbf{d}}^{-1} \cdot \mathbf{d}$$

$\mathbf{V}_{\mathbf{d}}$: covariance matrix of data

Covariance of solution:

$$\mathbf{V}_{\mathbf{f}} = \left(\mathbf{R}^{\mathrm{T}} \cdot \mathbf{V}_{\mathbf{d}}^{-1} \cdot \mathbf{R} + \lambda \mathbf{B}^{\mathrm{T}} \cdot \mathbf{B}\right)^{-1} \cdot \mathbf{R}^{\mathrm{T}} \cdot \mathbf{V}_{\mathbf{d}}^{-1} \cdot \left(\mathbf{R}^{\mathrm{T}} \cdot \mathbf{V}_{\mathbf{d}}^{-1} \cdot \mathbf{R} + \lambda \mathbf{B}^{\mathrm{T}} \cdot \mathbf{B}\right)^{-1}$$

- set of non-singular linear equations
- solution and uncertainties depend on λ
- solution can be negative


counts
10^4
10^3
10^2
10
1
10^{-1}
0
1
2
3
4
5
6
7
8
E [MeV]
$10^6 \times S_\beta$ [s^{-1}]
10
1
10^{-1}
10^{-2}
E_x [MeV]
relat. dev.
1
0.5
0
-0.5
-1
relat. dev.

Solution of linear inverse problems: $\mathbf{d} = \mathbf{R} \cdot \mathbf{f}$ (II)

Maximum Entropy (ME) method:

- solution must maximize information entropy

$$\boxed{\max : S(\mathbf{f}) - \frac{1}{\lambda}\chi^2(\mathbf{f})}$$

λ: Lagrange multiplier

S(**f**): entropy, $S(\mathbf{f}) = -\sum_i \left(f_i \ln \frac{f_i}{h_i} - f_i + h_i \right)$

One possible algorithm:

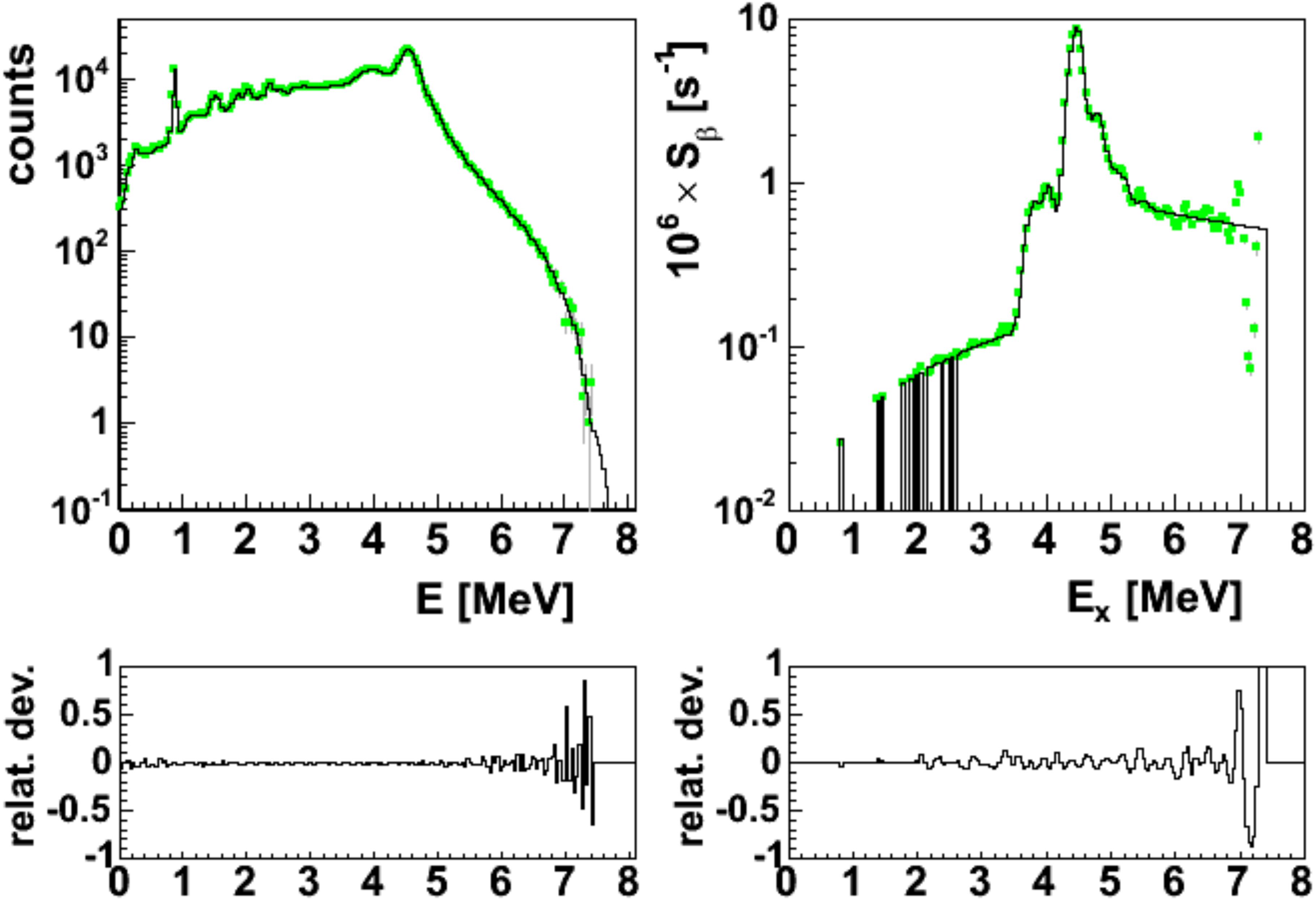
$$f_j^{(s+1)} = f_j^{(s)} \exp\left(\frac{2}{\lambda} \sum_i R_{ij} \left(d_i - \sum_k R_{ik} f_k^{(s)} \right) \Big/ \sigma_i^2 \right)$$

(uncorrelated data)

Covariance of the solution:

$$\sigma_{f_i f_j} \approx \frac{4}{\lambda} f_i f_j \sum_k R_{ki} R_{kj} \Big/ \sigma_{d_i}^2$$

- iterative solution: initial value & stopping criterion
- solution and uncertainties depend on λ
- solution is positive definite


counts
10^4
10^3
10^2
10
1
10^{-1}
0 1 2 3 4 5 6 7 8
E [MeV]
$10^6 \times S_\beta$ [s^{-1}]
10
1
10^{-1}
10^{-2}
0 1 2 3 4 5 6 7 8
E_x [MeV]
relat. dev.
1
0.5
0
-0.5
-1
0 1 2 3 4 5 6 7 8
relat. dev.
1
0.5
0
-0.5
-1
0 1 2 3 4 5 6 7 8

Solution of linear inverse problems: $\mathbf{d} = \mathbf{R} \cdot \mathbf{f}$ (III)

Expectation Maximization (EM) method:

- modify knowledge on causes from effects (Bayes Theorem)

$$\boxed{P(f_j | d_i) = \frac{P(d_i | f_j) P(f_j)}{\sum_j P(d_i | f_j) P(f_j)}}$$

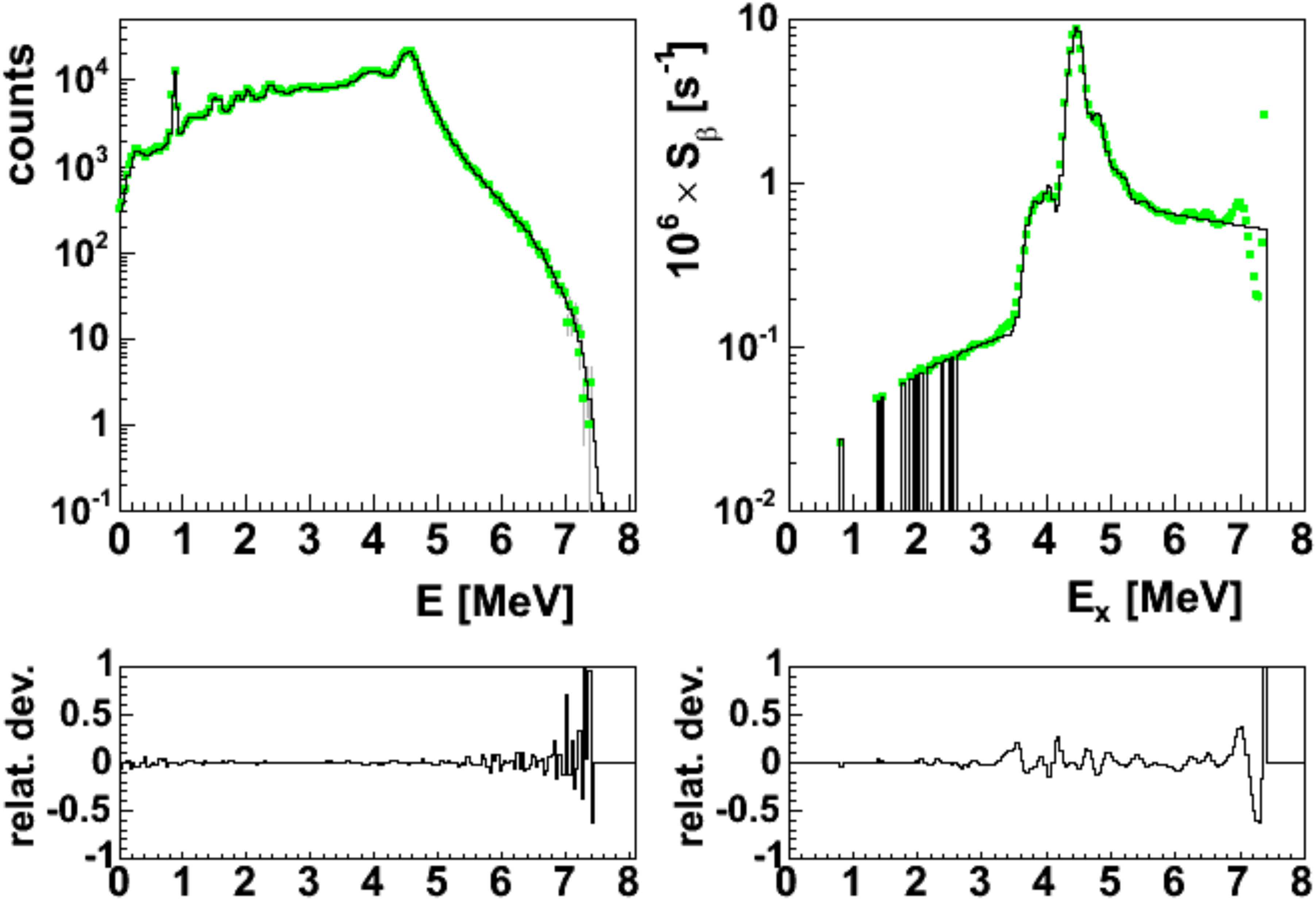
Algorithm:

$$f_j^{(s+1)} = \frac{1}{\sum_i R_{ij}} \sum_i \frac{R_{ij} f_j^{(s)} d_i}{\sum_k R_{ik} f_k^{(s)}}$$

Covariance of the solution:

$$\mathbf{V_f} = \mathbf{M} \cdot \mathbf{V_d} \cdot \mathbf{M}^{\mathrm{T}}$$
$$\left(\mathbf{f}^{(\mathrm{s}+1)} = \mathbf{M}^{(\mathrm{s})} \cdot \mathbf{d}\right)$$

- iterative solution: initial value & stopping criterion
- solution is positive definite

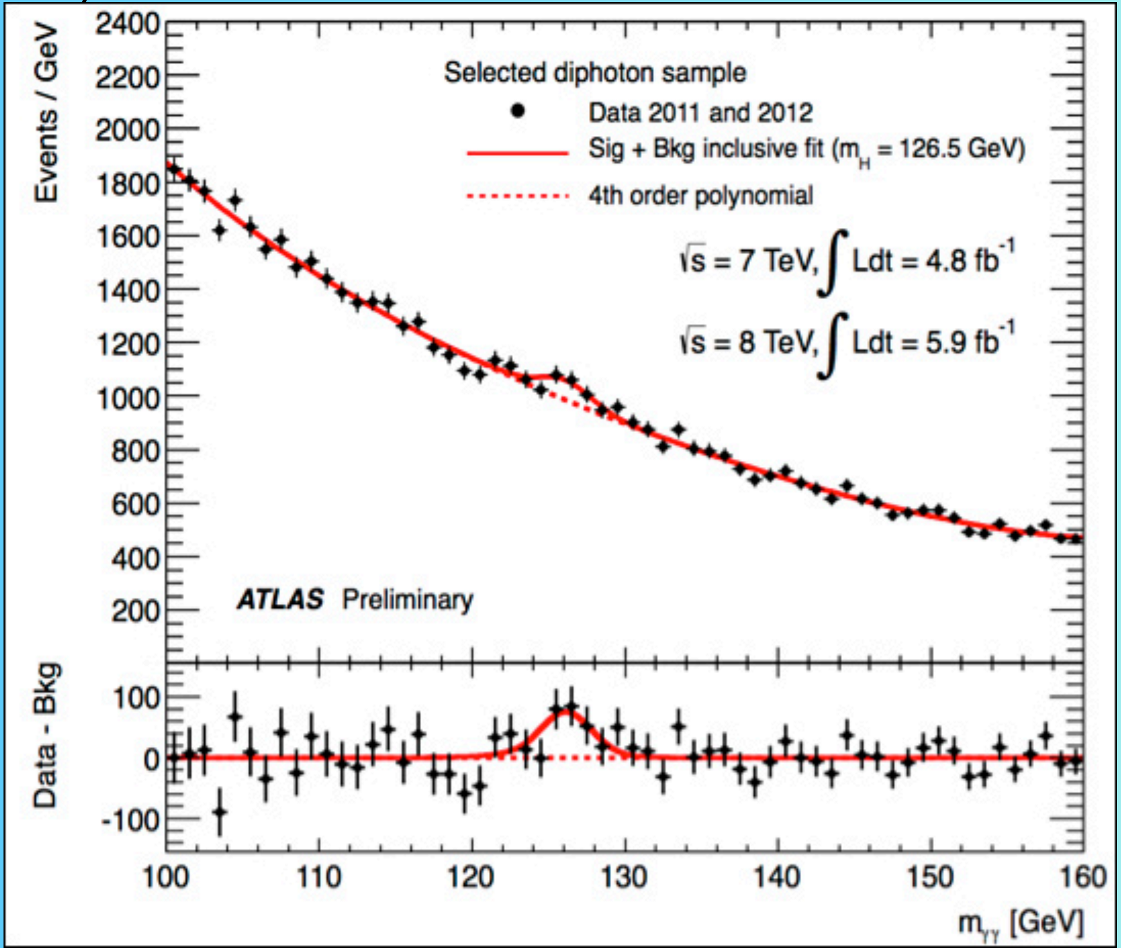


counts
10^4
10^3
10^2
10
1
10^-1
0 1 2 3 4 5 6 7 8
E [MeV]
10^6 × S_β [s^-1]
10
1
10^-1
10^-2
0 1 2 3 4 5 6 7 8
E_x [MeV]
relat. dev.
1
0.5
0
-0.5
-1
0 1 2 3 4 5 6 7 8
relat. dev.
1
0.5
0
-0.5
-1
0 1 2 3 4 5 6 7 8

Hypothesis testing and decision theory

Looking for a a small signal above background in a spectrum: S=T-B

- When can we say that a signal has been "seen" in a measurement? (a posteriori)
- What is the sensitivity of a measurement or the minimum detectable signal (a priori)



Nomenclature:

- **H_0** : null hypothesis (no signal: S=0)
- **H_1** : alternative hypothesis (there is signal: S>0 or signal: S>value)

a posteriori

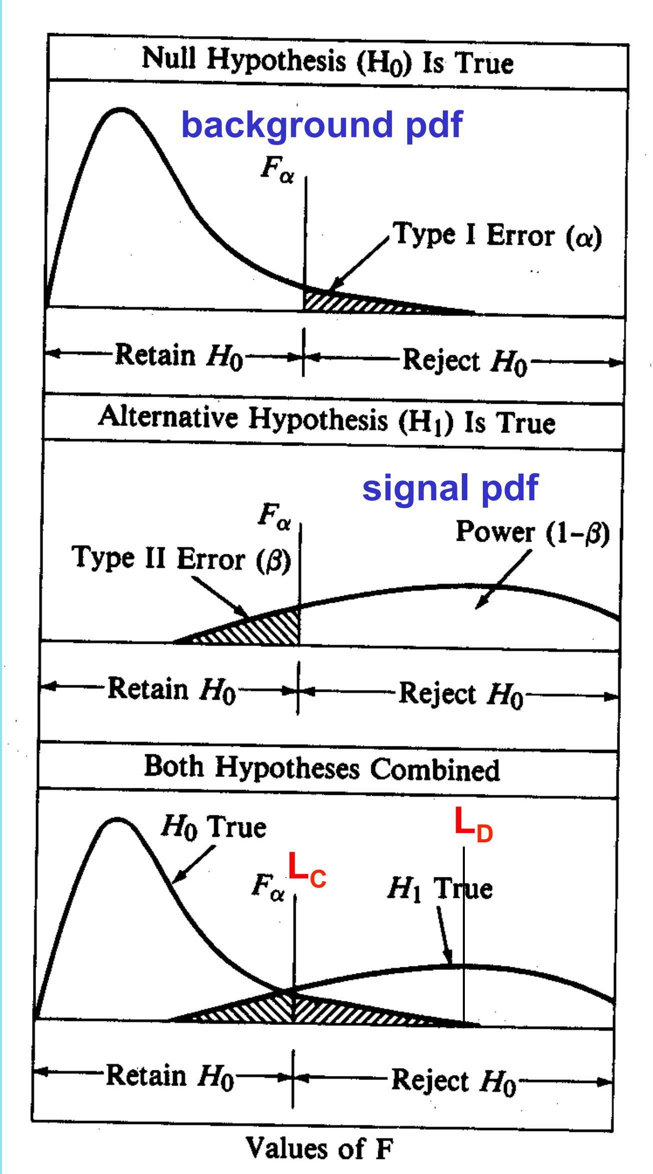
- **L_C** : critical level or decision limit
- α : integral of background PDF above L_c
- β : integral of signal PDF below L_c
- **Type I error (α)**: Probability of wrongly deciding that there was a signal
- **Type II error (β)**: probability of wrongly deciding that there was no signal

a priori

- **L_D** : detection limit or mean signal value with a probability 1-β of giving an actual value above **L_c**

Decision table	True H_0	False H_0
Reject H_0	Type I error α	Correct assessment
Fail to reject H_0	Correct assessment	Type II error β

BUT BEWARE OF DIFFERENT DEFFINITIONS



An example (simplistic approach):

Environmental lab trying to detect an isotope in a sample (for example looking for certain γ-ray peak)

B: background counts, Poisson distributed
T: total counts, Poisson distributed
S=T-B: signal counts

$$\sigma_B^2 = B$$

$$\sigma_T^2 = T = S + B$$

$$\sigma_S^2 = \sigma_T^2 + \sigma_B^2 = T + B = S + 2B$$

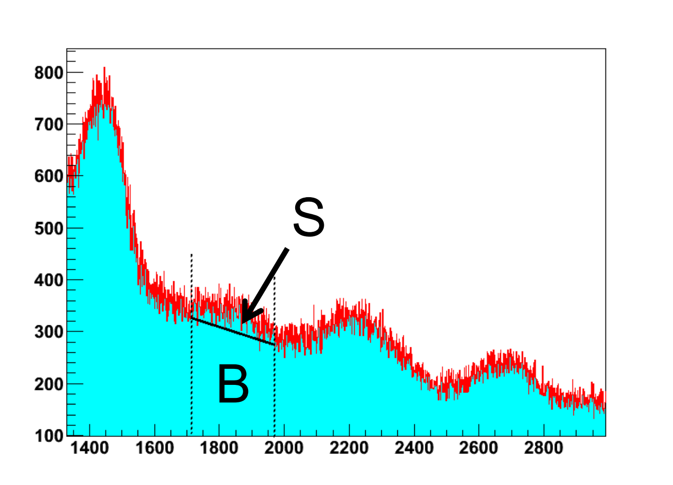
Confidence levels α, β: 5%

For mathematical simplicity: Poisson ➔Normal (wrong for low statistics!)

$$L_C = 1.645\sigma_B = 1.645\sqrt{B}$$ (net value)

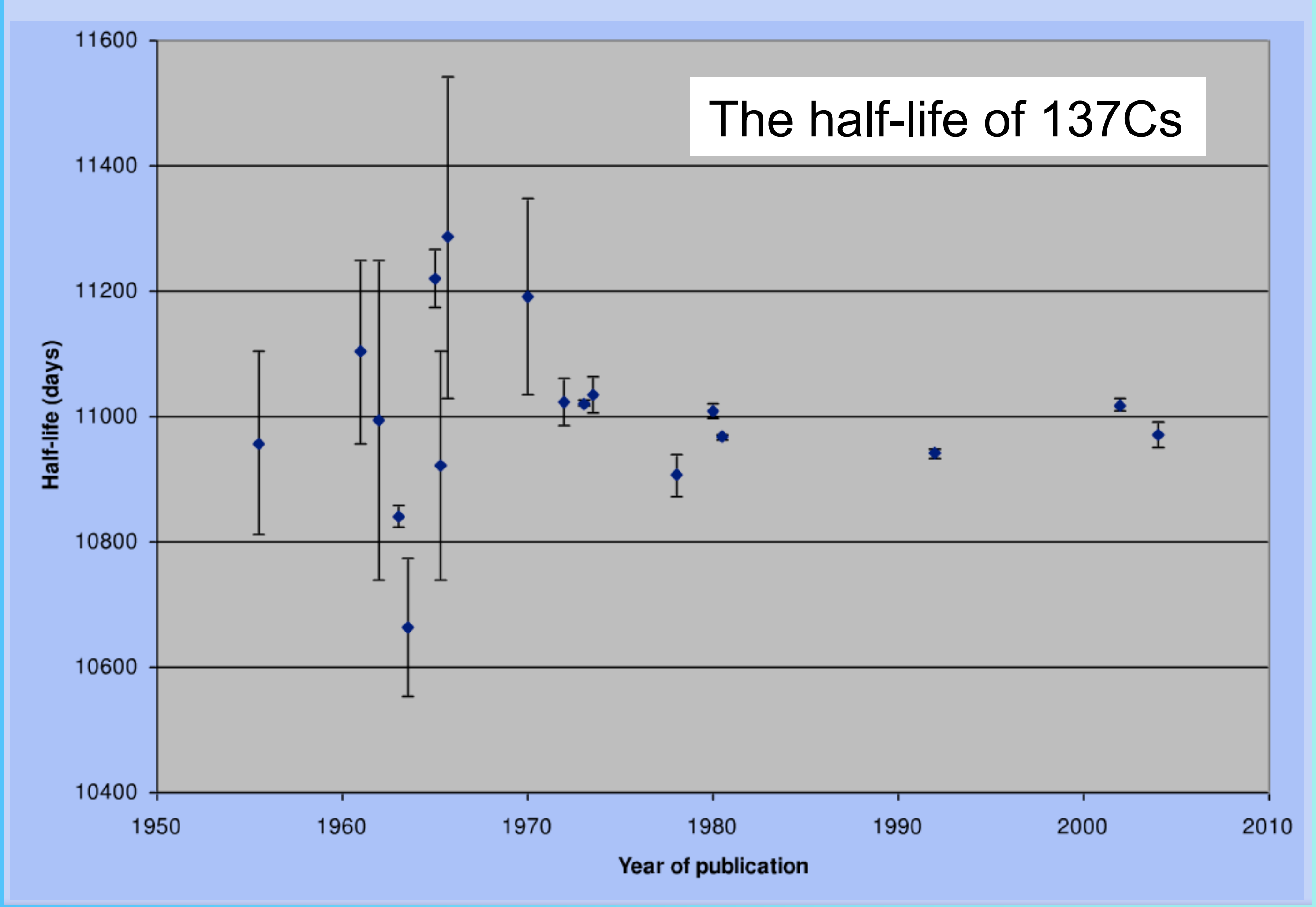
$$L_D = L_C + 1.645\sigma_S = 1.645\sqrt{B} + 1.645\sqrt{L_D + 2B}$$

If T>L$_c$: we then give a signal value S with uncertainty
If T<L$_c$: we give a maximum limit value S$_{max}$ for the signal



Discrepant data: Outliers *valores atipicos*

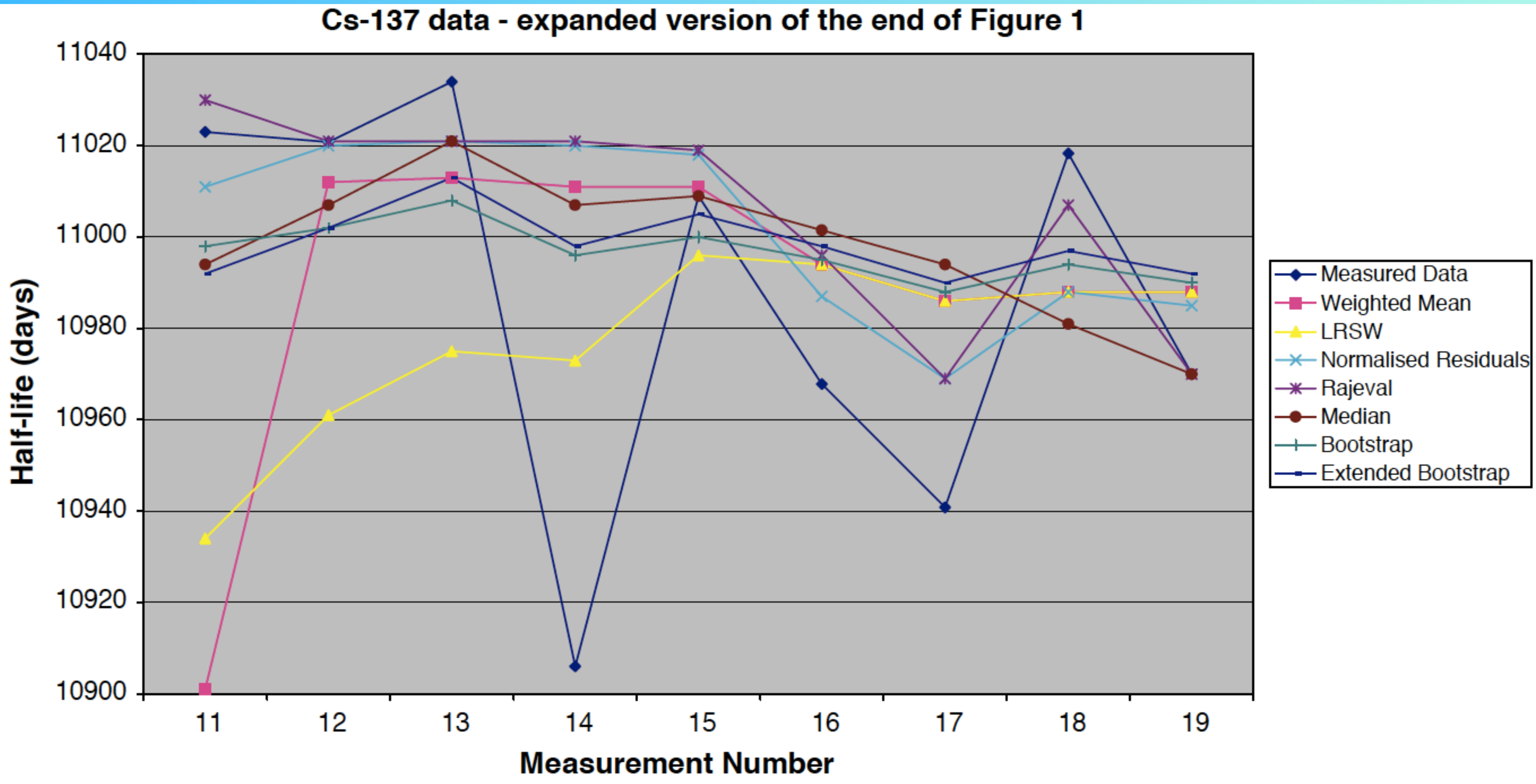
Data evaluation: How to combine the information from different measurements, when some of them deviate(*) from the rest?



(*): “the difference with the average value is much larger than the uncertainty”

- The problem is that either the quoted uncertainties are too small or there are unknown systematic
- Some times, after revision of measurement and analysis details, one is able to pinpoint the problem and eventually correct for it.
- More often this is not the case, then: how to calculate an average value an its uncertainty? There is no clear solution to the problem
- Several methods have been proposed: remove discrepant data, increase suspiciously low uncertainties (LRSW, Normalized Residuals, Rajeval), use median instead of mean, use Bayes theorem, use Bootstrap methods, …

MacMahon et al., App. Rad. Isot. 60 (2004) 275



MacMahon et al., App. Rad. Isot. 60 (2004) 275