STATISTICA PER FISICI

1. Calcolo delle probabilità
2a. Statistica frequentista
2b. Statistica bayesiana
3. Likelihood
4. Fondo e segnale
5. Metodi Bootstrap
6. *Unfolding*

Alberto Rotondi
Università di Pavia

What is Statistics ?

- ***a problem of probability calculus*: if $p = 1/2$ for having head in tossing a coin, what is the probability to have in 1000 coin tosses less than 450 heads?**
- ***the same problem in statistics*: if in 1000 coin tosses 450 heads have been obtained, what is the estimate of the true head probability?**

Statistical error: $s \approx \sigma$

$$\mu \pm \sigma = 500.0 \pm 15.8 \simeq 500 \pm 16 = [484, 516]$$

$$x \pm s = 450.0 \pm 15.7 \simeq 450 \pm 16 = [434, 466]$$

Physics and Statistics

- **Higgs mass (PDG 2000):**

$$m > 95.3\ GeV, CL = 95\%$$

- W **mass:**

$$m_W = 80.419 \pm 0.056\ GeV$$

These are
confidence intervals

The hystorical path

	FREQUENTISTS	BAYESIANS
1763		Thomas Bayes writes a fundamental paper. Bayesian age
1900	Karl Pearson proposes the χ^2 test	
1910	Robert Fisher invents Maximum Likelihood	
1937	The J. Neyman frequentist interval estimate	
1940	The Hypothesis testing of Pearson. Frequentist age The Popper scheme Frequentist teaching	
1990		rediscovering of the bayesian works of Jeffreys, De Finetti and Jaynes
now	the debate is open: see on Confidence Limits	the CERN Workshop (Geneva 2000) neo-Bayesian age?

Statistics 1.

Frequentist approach

Bayesian approach

What is Statistics ?

- ***a problem of probability calculus*: if $p = 1/2$ for having head in tossing a coin, what is the probability to have in 1000 coin tosses less than 450 heads?**
- ***the same problem in statistics*: if in 1000 coin tosses 450 heads have been obtained, what is the estimate of the true head probability?**

Statistical error: $s \approx \sigma$

$$\mu \pm \sigma = 500.0 \pm 15.8 \simeq 500 \pm 16 = [484, 516]$$

$$x \pm s = 450.0 \pm 15.7 \simeq 450 \pm 16 = [434, 466]$$

Statistics

We have 2 inferences

- parameter estimation: to estimate p from 1000 coin tosses
- hypothesis testing: in in two experiments of 1000 coin tosses 450 and 600 tosses have been obtained, how much is probable that the two experiments use two consistent coins?

Parametric Statistics: the probability depends on θ:

$$\mathcal{E}(\theta) \equiv (S, \mathcal{F}, P_\theta)$$

corresponding to a density

$$P\{X \in A\} = \int_A p(x;\theta)\,\mathrm{d}x$$

Physics and Statistics

- **Higgs mass (PDG 2000):**

$$m > 95.3\ GeV, CL = 95\%$$

- W **mass:**

$$m_W = 80.419 \pm 0.056\ GeV$$

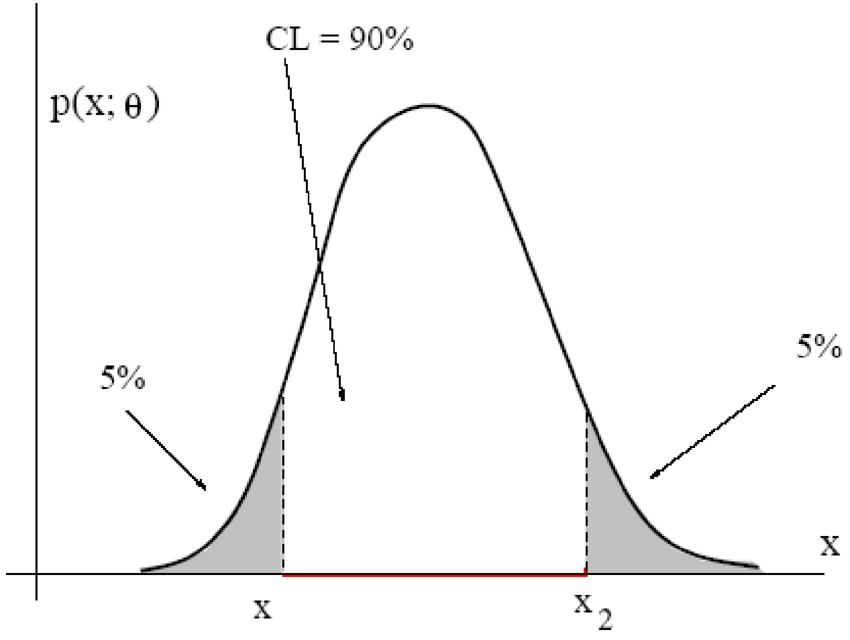
These are
confidence intervals

Frequentist Confidence Intervals

One (Neyman, 1937) starts from probability calculus

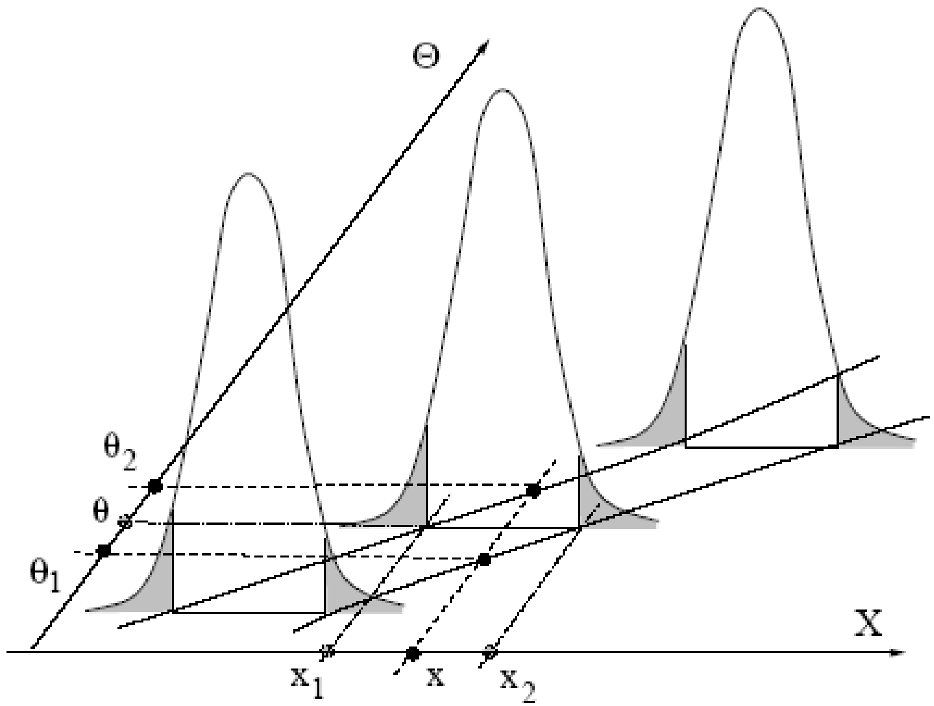
$$\int_{x_1}^{x_2} p(x;\theta)\,\mathrm{d}x = CL$$

and the procedure is repeated



for all the possible θ values

Frequentist confidence intervals



It is possible to show that

$$X \in [x_1, x_2] \text{ iff } \Theta \in [\theta_1, \theta_2]$$

Since

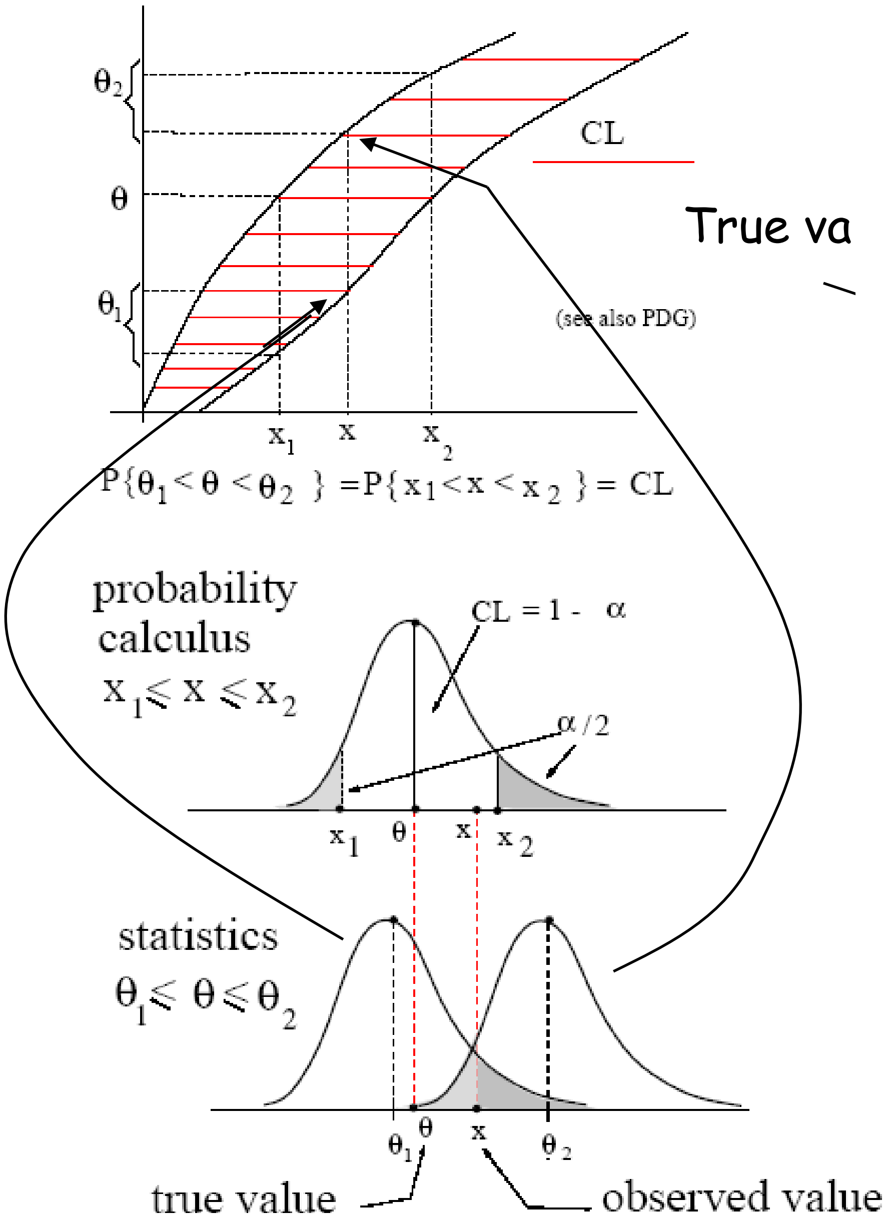
$$P\{X \in [x_1, x_2]\} = CL$$

then

$$P\{\Theta \in [\theta_1, \theta_2]\} = CL$$

Fundamental Neyman result (1937)

Cut and top views of the Neyman construction:



True va

interval

$x = x_1 \rightarrow \theta_1 < \theta < \theta$

$x = x_2 \rightarrow \theta < \theta < \theta_2$

$\theta_1 < \theta < \theta_2$ when $x_1 < x < x_2$

Frequentist Confidence Intervals
mathematical definitions

If T_1 and T_2 are two statistics, the interval

$$I = [T_1, T_2]$$

is a confidence interval for θ, with $0 < CL < 1$ confidence level, if, for all $\theta \in \Theta$, the probability that I contains θ (*coverage*) is CL:

$$P\{T_1 \leq \theta \leq T_2\} = CL \ .$$

If T_1 e T_2 are discrete variables, the confidence interval satisfies the *minimum overcoverage*

$$P\{T_1 \leq \theta \leq T_2\} \geq CL \ .$$

Note: $[T_1, T_2]$ are random variables, the θ parameter is fixed

Frequentist C.I.
right and wrong definitions

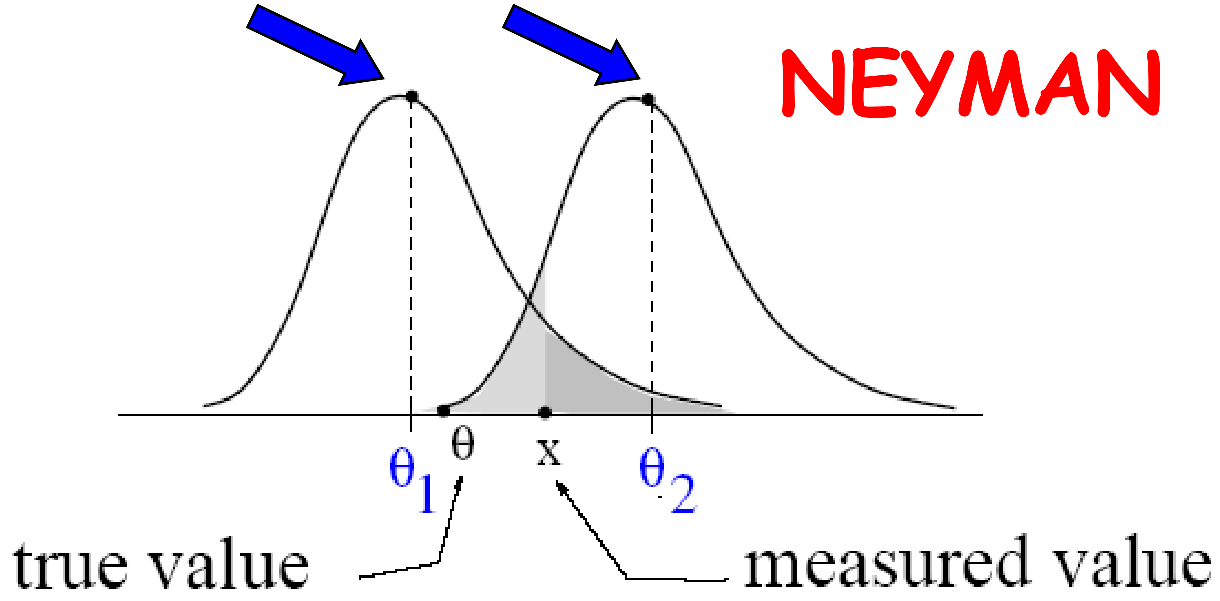
RIGHT quotations:

- CL is the probability that the random interval $[T_1, T_2]$ covers the true value θ;
- in an infinite set of repeated identical experiments, a fraction equal to CL will succeed in assigning $\theta \in [\theta_1, \theta_2]$;
- if $\theta \notin [\theta_1, \theta_2]$, one can obtain $\{I = [\theta_1, \theta_2]\}$ in a fraction of experiments $\leq 1 - CL$
- if $H_0 : \theta \notin [\theta_1, \theta_2]$ the probability to reject a true H_0 is $1 - CL$ (falsification).
  see upper and lower limits estimates.

WRONG quotations

- CL is the degree of belief that the true value is in $[\theta_1, \theta_2]$
- $P\{\theta \in [\theta_1, \theta_2]\} = CL$
  (θ is not a random variable!)

NEYMAN INTEGRALS



$$\int_x^{\infty} p(x;\theta_1)\,\mathrm{d}x = c_1 \qquad \int_{-\infty}^{x} p(x;\theta_2)\,\mathrm{d}x = c_2$$

where

$$\theta \in [\theta_1, \theta_2]\ , \quad 1-(c_1+c_2) = CL$$

MC techniques can be used: grid over θ to find the values θ_1 and θ_2 satisfying these integrals

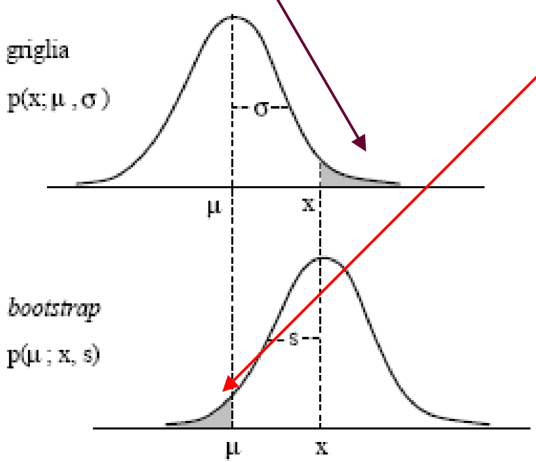
Important:

$$\int_{\theta_1}^{\theta_2} p(\theta;x)\,\mathrm{d}\theta = CL$$

WRONG!!!!

When the following property holds

$$1-\int_{-\infty}^{x} p(x;\theta)\,\mathrm{d}x = \int_{-\infty}^{\theta} p(\theta;x)\,\mathrm{d}\theta$$



$$1-F(x;\theta) = F(\theta;x)$$

Bootstrap property

$$\int_{x}^{\infty} p(x;\theta_1)\,\mathrm{d}x = c_1 \quad \int_{-\infty}^{x} p(x;\theta_2)\,\mathrm{d}x = c_2$$

$$c_1 = \int_{x}^{\infty} \ldots = 1-\int_{\theta_1}^{\infty} \ldots = \int_{-\infty}^{\theta_1} p(\theta;x)\,\mathrm{d}\theta$$

$$c_2 = \int_{-\infty}^{x} \ldots = 1-\int_{-\infty}^{\theta_2} \ldots = \int_{\theta_2}^{\infty} p(\theta;x)\,\mathrm{d}\theta$$

One can integrate along θ in a “bayesian” way

(see any elementary textbook)

$$CL = 1 - c_1 - c_2 = \int_{\theta_1}^{\theta_2} p(\theta;x)\,\mathrm{d}\theta$$

Pivot quantities

Avoid the calculation of the integrals

$$\int_A p(x;\theta)\,\mathrm{d}x = c_i$$

If $Q(x;\theta)$ is pivotal, $P\{Q \in A\}$ is independent of θ. Example:

$$Q = (X - \theta) \sim N(0, \sigma^2)$$

Method:

- **find $P\{q_1 \leq Q \leq q_2\} = CL$;**
- **invert the equation:**

$$Q(x;\theta) = q \rightarrow \theta = T(x;q)$$

- **Then:**

$$P\{q_1 \leq Q \leq q_2\} = P\{T_1 \leq \theta \leq T_2\} = CL$$

$$P\{\mu - \sigma \leq X \leq \mu + \sigma\} = P\{-\sigma \leq X - \mu \leq \sigma\}$$

$$= P\{X - \sigma \leq \mu \leq X + \sigma\}$$

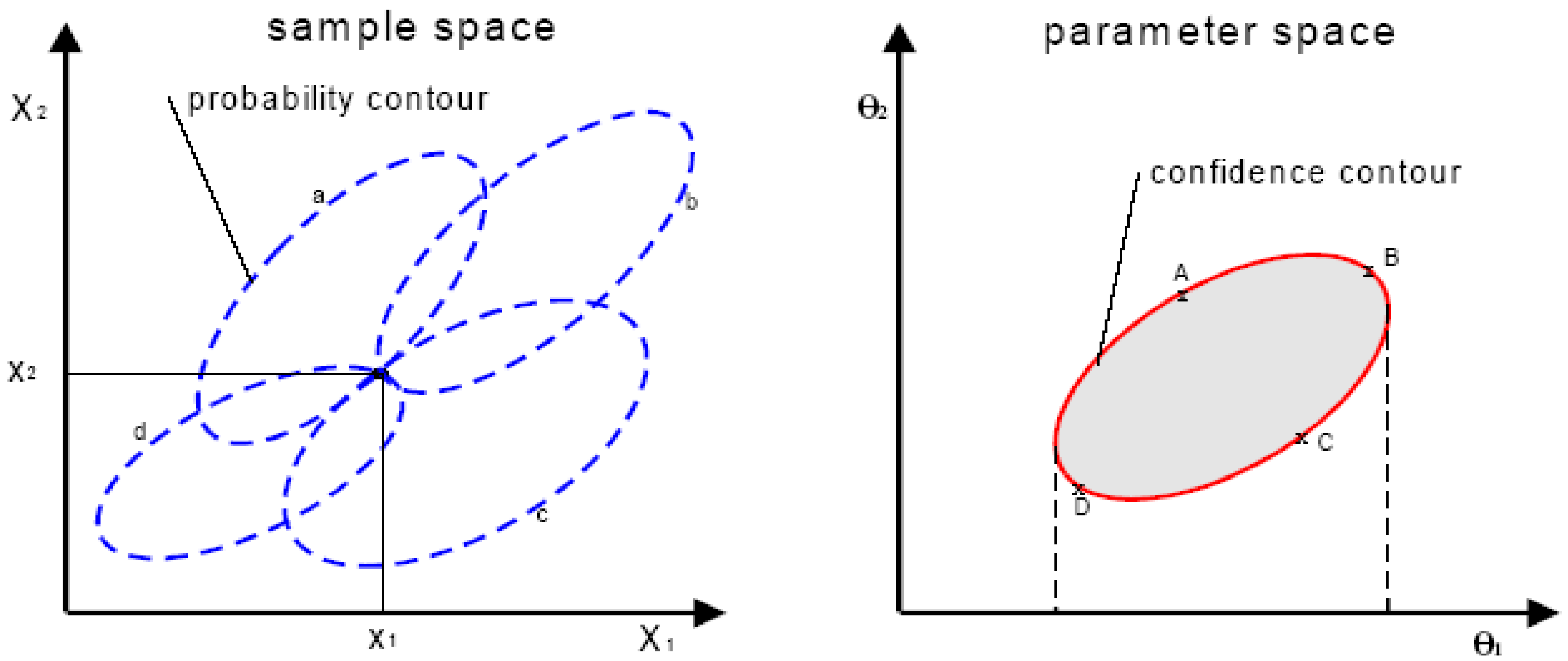


Fig. 2. Two parameter classical confidence limit for an observation x_1, x_2. The dashed contours labeled with small letters in the sample space correspond to probability contours of the parameter pairs labeled with capital letters in the parameter space

Probability estimate
big samples

$$\langle F\rangle = \frac{\langle X\rangle}{N} = \frac{Np}{N} = p$$

$$\mathrm{Var}[F] = \frac{\mathrm{Var}[X]}{N^2} = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N}$$

pivot quantity for $N \gg 1$:

$$T = \frac{F-p}{\sigma[F]} \sim N(0,1) \ ,$$

which can be inverted:

$$P\left\{\left|\frac{F-p}{\sigma[F]}\right| < 1\right\} = P\{F-\sigma F \le p \le F+\sigma[F]\}$$

$$\sigma[F] = \sqrt{\frac{p(1-p)}{N}} \approx \sqrt{\frac{f(1-f)}{N}}$$

If $N \gg 1$ then:

$$p = f \pm \sqrt{\frac{f(1-f)}{N}} \qquad \mathrm{CL} \approx 68\%$$

small samples

... first difficulties ....
there are no pivot quantities:

$$\sum_{k=x}^{n} \binom{n}{k} p_1^k (1-p_1)^{n-k} = c_1 \ ,$$

$$\sum_{k=0}^{x} \binom{n}{k} p_2^k (1-p_2)^{n-k} = c_2 \ .$$

Symmetric case: $c_1 = c_2 = (1-CL)/2 = \alpha/2$.
When $x = 0$, $x = n$, $c_1 = c_2 = 1 - CL$:

$$x = n \implies p_1^n = 1 - CL \ ,$$
$$x = 0 \implies (1-p_2)^n = 1 - CL \ .$$

all the attempts had success:

$$p_1 = \sqrt[n]{1-CL} \qquad p \in [p_1, 1]$$

no success:

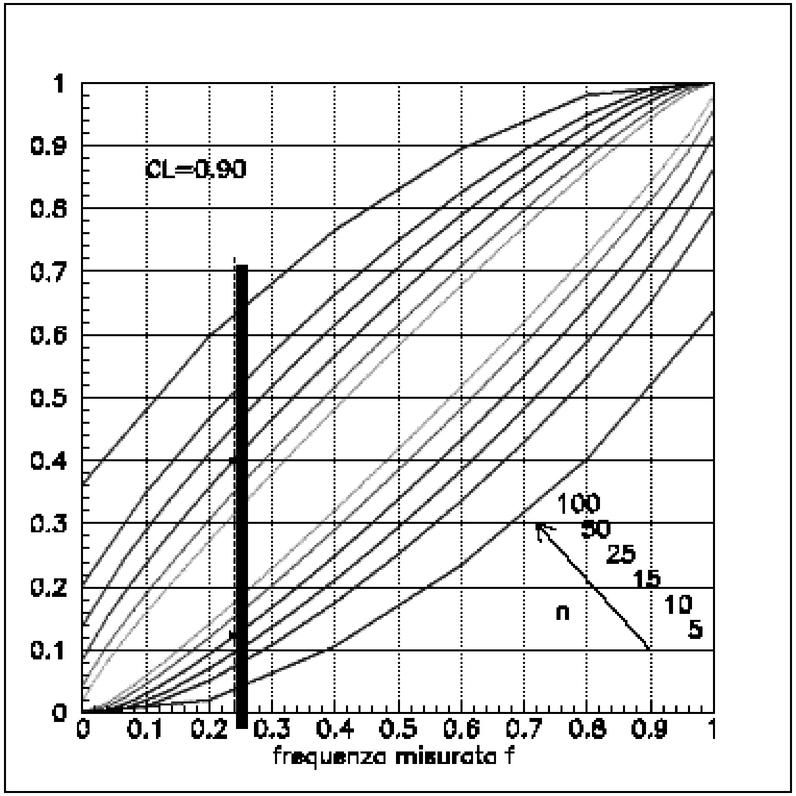
$$p_2 = 1 - \sqrt[n]{1-CL} \qquad p \in [0, p_2]$$

small samples

$$\sum_{k=x}^{n} \binom{n}{k} p_1^k (1-p_1)^{n-k} = c_1 \ ,$$

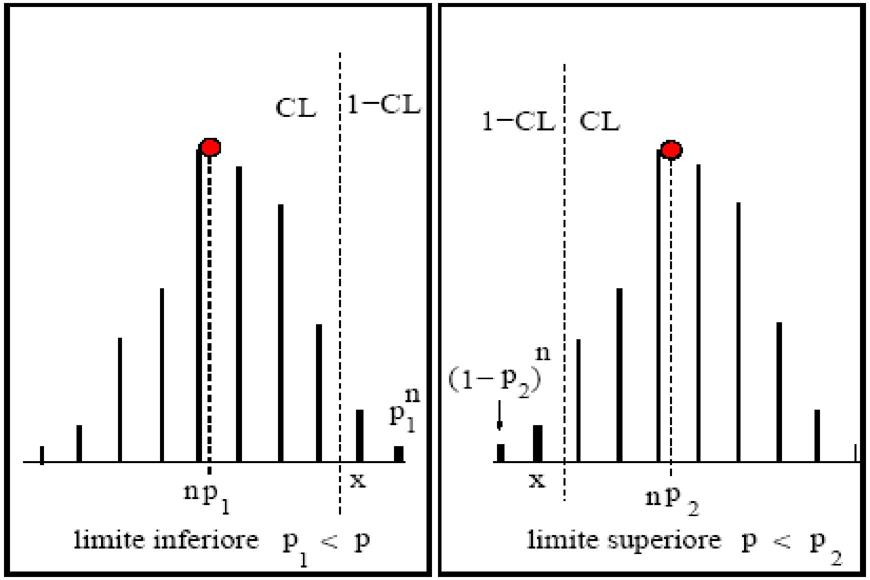
$$\sum_{k=0}^{x} \binom{n}{k} p_2^k (1-p_2)^{n-k} = c_2 \ .$$

Neyman curves $CL = 90\%$



Upper and lower limits

x **events are observed:**



- **lower limit** $p \in [p_1, 1]$: **if** $p < p_1$, **one can observe at least** x **events, but in a fraction of experiments** $< 1 - CL$. **If** $x = n$, $p_1^n = 1 - CL$;
- **upper limit** $p \in [0, p_2]$: **if** $p > p_2$, **one can observe up to** x **events, but in a fraction of experiments** $< 1 - CL$. **if** $x = 0$, $(1 - p_2)^n = 1 - CL$.

Poisson Limits

$$\sum_{k=x}^{\infty} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1 \ , \quad \sum_{k=0}^{x} \frac{\mu_2^k}{k!} \exp(-\mu_2) = c_2 \ ,$$

symmetric case: $c_1 = c_2 = (1 - CL)/2$.

Upper Limits to the mean number of events having obtained x events:

$$\sum_{k=0}^{x} \frac{\mu_2^k}{k!} \exp(-\mu_2) = 1 - CL \ .$$

For $x = 0, 1, 2$, where $\mu_2 \equiv \mu$

$$\begin{aligned} e^{-\mu} &= 1 - CL \ , \\ e^{-\mu} + \mu e^{-\mu} &= 1 - CL \ , \\ e^{-\mu} + \mu e^{-\mu} + \frac{\mu^2}{2} e^{-\mu} &= 1 - CL \end{aligned}$$

x	90%	95%	x	90%	95%
0	2.30	3.00	6	10.53	11.84
1	3.89	4.74	7	11.77	13.15
2	5.32	6.30	8	13.00	14.44
3	6.68	7.75	9	14.21	15.71
4	7.99	9.15	10	15.41	16.96
5	9.27	10.51	11	16.61	18.21

When $\mu > 2.3$, *one con observe no events* but in a number of experiments $< 10\%$.

Estimation of the sample mean

$$\mathrm{Var}[M] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] \ .$$

since $\mathrm{Var}[X_i] = \sigma^2 \quad \forall i,$

$$\mathrm{Var}[M] = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 = \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N} \ .$$

Due to the Central Limit theorem we have a pivot quantity when N>>1

$$\frac{\mu - M}{\sigma/\sqrt{N}} \sim N(0,1)$$

Hence:

$$P\left\{\left|\frac{\mu - M}{\sigma/\sqrt{N}}\right| \le 1\right\} = P\left\{M - \frac{\sigma}{\sqrt{N}} \le \mu \le M + \frac{\sigma}{\sqrt{N}}\right\}$$

$(N > 20 - 30)$:

$$\mu = m \pm \frac{\sigma}{\sqrt{N}} \simeq \mu = m \pm \frac{s}{\sqrt{N}} \qquad CL \simeq 68\%$$

The Bayes formula

$$P(B_k|A)P(A) = P(A|B_k)P(B_k)$$

if B_k are disjoint and cover the set S,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

then $P(B_k|A)$ can be written as:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \;, \quad P(A) > 0 \;.$$

Example trigger problem
A μ-π trigger has $\varepsilon(\pi) = 0.05$ e $\varepsilon(\mu) = 0.95$. If the beam is 90% π and 10% μ find efficiency and enrichment factor

$$P(\mu|T) = \frac{P(T|\mu)P(\mu)}{P(T|\mu)P(\mu) + P(T|\pi)P(\pi)}$$
$$= \frac{0.95 \cdot 0.10}{0.95 \cdot 0.10 + 0.05 \cdot 0.90} = \frac{0.095}{0.14} = 0.678$$

efficiency $= 14\%$, enrichment ≈ 6.8

Bayesian use of Bayes formula

An attempt to solve the trigger problem without knowing
the beam percentages ...!!

The Bayes formula is employed starting from
subjective probabilities

$$P(H_k|\text{data}) = \frac{P(\text{data}|H_k)P(H_k)}{\sum_{i=1}^{n} P(\text{data}|H_i)P(H_i)} \ .$$

an important step,

$$P(H_k|\text{data}) \rightarrow P_{n-1}(H_k)$$

iteration:

$$P_n(H_k|E) = \frac{P(E_n|H_k)P_{n-1}(H_k)}{\sum_{i=1}^{n} P(E_n|H_i)P_{n-1}(H_i)} \ ,$$

- **frequentist approach**:

subjective probabilities
for Hypotheses
never
are assumed.

$P(\mathrm{H}|\mathrm{data})$ NO!!!!

The gambler problem
Bayesian approach

$$P(\mathrm{Win}|C) = 1 \qquad P(\mathrm{Win}|H) = 0.5$$

Problem: to find the probability that the gambler is cheat, as a function of the number of consecutive wins $\{W_n\}$

$$P(H) \equiv P(H|W_0)\,, \quad P(C) \equiv P(C|W_0) \qquad P(H) = 1 - P(C)$$

Iteration:

$$P(C|W_n) = \frac{P(W_n|C)\,P(C|W_{n-1})}{P(W_n|C)\,P(C|W_{n-1}) + P(W_n|H)\,[1 - P(C|W_{n-1})]}$$

that is

$$P(C|W_n) = \frac{P(C|W_{n-1})}{P(C|W_{n-1}) + 0.5\,[1 - P(C|W_{n-1})]}$$

Bayes:

$P(C)/n$	**5**	**10**	**15**	**20**
1%	**24**	**91**	**99.7**	**99.99**
5%	**63**	**98**	**99.94**	**99.998**
50%	**97**	**99.9**	**99.997**	**99.999**

The gambler problem
Frequentist approach

Let us suppose 15 cosecutive wins

Hypothesis testing:
The null hypothesis H_0 (honest player) gives a significance level (p-value in this case)

$$0.5^{15} = 3.05\,10^{-5}$$

The probability to be wrong discarding the hypothesis is less then 0.003 %.
The player is cheat.

Cheat probability estimation:
with $n = 15$ and $CL = 90\%$ the probability is

$$p = (0.1)^{1/15} \approx 0.86 \; .$$

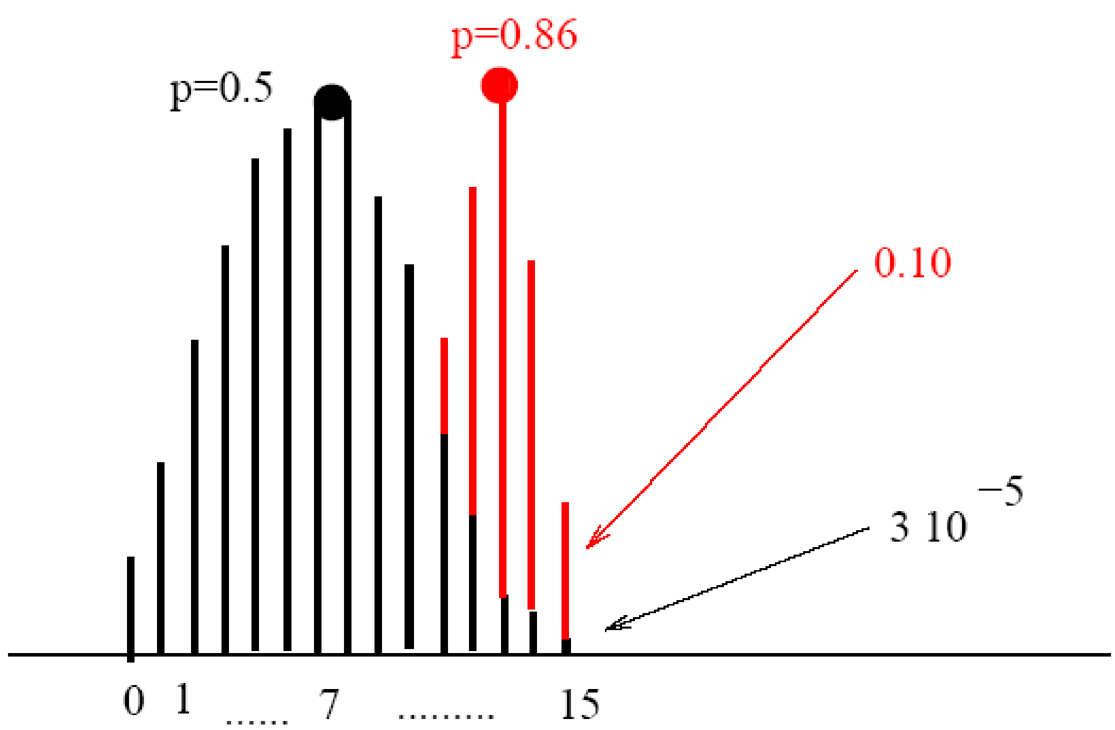
With a "cheat probability" $p < 0.86$ it is possible to win for 15/15 times, but in a percentage of plays $< 10\%$

$$0.86 < p < 1 \qquad CL = 90\%$$

The gambler problem
Frequentist approach

Black: hypothesis testing
Red: probability estimation

These conclusions are independent of any a priori hypothesis!



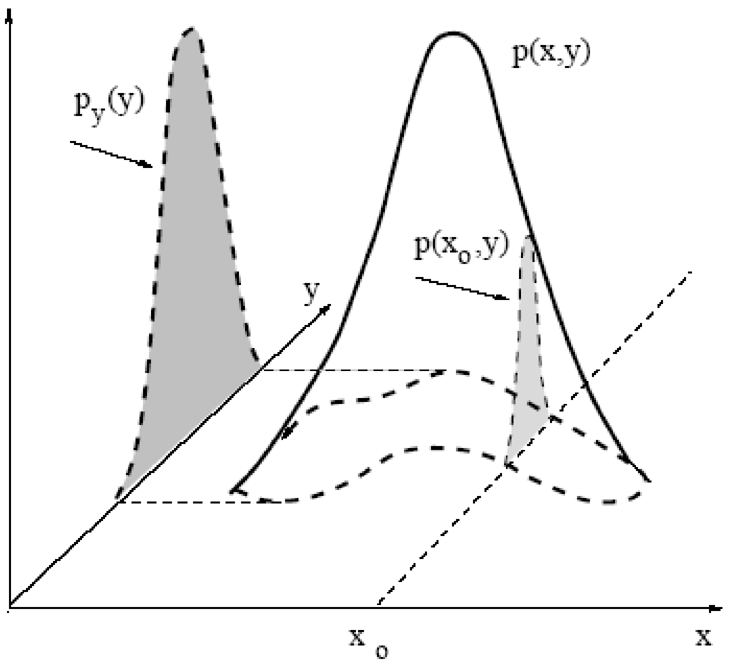
Marginal and conditional densities

probability product:

$$p(x, y) = p_Y(y)\, p(x|y) = p_X(x)\, p(y|x)$$

for independent variables:

$$p(x|y) = p_X(x) \ , \quad p(y|x) = p_Y(y) \ ,$$



Bayes for the continuum

$$p(x,y) = p_Y(y)\,p(x|y) = p_X(x)\,p(y|x)$$

hence

$$p(x|y) = \frac{p(y|x)\,p_X(x)}{p_Y(y)}$$

that is

$$p(x|y) = \frac{p(y|x)\,p_X(x)}{\int p(y|x)\,p_X(x)\,\mathrm{d}x}$$

Bayesian step:

$$p(\mu;x) = \frac{p(x;\mu)\,p_\mu(\mu)}{\int p(x;\mu)\,p_\mu(\mu)\,\mathrm{d}x}$$

The prior

$$p_\mu(\mu)$$

that is the subjective probability assigned to μ, is NEVER used by frequentists

that is

$$p(\mu;x) = \frac{\text{likelihood} \times \text{prior}}{\text{normalization}}$$

Bayesian Interval estimate

Degree of belief on μ for a measured x:

$$p(\mu; x) = \frac{L(x,\mu)\, p_\mu(\mu)}{\int L(x,\mu)\, p_\mu(\mu)\, \mathrm{d}p}$$

Estimate:

$$\mu \in [\mu_1, \mu_2]$$

with degree of belief

$$\int_{\mu_1}^{\mu_2} p(\mu; x)\, \mathrm{d}\mu = \text{degree of belief}$$

- **one integrates over μ considered as a random variable**
- **this coincides with the frequentist result if the prior $p_\mu(\mu)$ is uniform and the property**

  $$1 - F(\mu; x) = F(x; \mu)$$

  holds
- **but the interpretation is different!**

Bayesian coin tossing

$$p(p;n,x) = \frac{p^x(1-p)^{n-x}\, p_p(p)}{\int p^x(1-p)^{n-x}\, p_p(p)\,\mathrm{d}p}$$

With uniform prior,

$$p_p(p) = \mathrm{const} \quad 0 < p < 1$$

Recalling the β function:

$$\int_0^1 p^x(1-p)^{n-x}\,\mathrm{d}p = \frac{x!(n-x)!}{(n+1)!}$$

one obtains the degree of belief of p

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

$$\langle p \rangle = \frac{x+1}{n+1}$$

$$\mathrm{Var}[p] = \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2}$$

Bayesian Interval estimate

$$p \in [p_1, p_2]$$

with degree of belief

$$\int_{p_1}^{p_2} \frac{(n+1)!}{x!(n-x)!} p^x (1-p)^{n-x} \,\mathrm{d}p$$

$x = n$**:**

$$p(p; x=n) = (n+1)p^n \;, \; F(p) = \int_0^p (n+1)p^n \,\mathrm{d}p = p^{n+1}$$

The 90% bayesian lower bound

$$0.10 = p^{n+1} \rightarrow p = (0.10)^{1/(n+1)}$$

$x = 0$**:**

$$\begin{aligned} p(p; x=0, n) &= (n+1)(1-p)^n \;, \\ F(p) &= \int_0^p (n+1)(1-p)^n \,\mathrm{d}p = 1-(1-p)^{n+1} \end{aligned}$$

The 90% bayesian upper bound

$$0.10 = p^{n+1} \rightarrow p = 1-(0.10)^{1/(n+1)}$$

Frequentist → Bayesian

$$n \rightarrow n+1$$

but with a different meaning !!!

Elementary example I

In 20 drawings 5 successes have been obtained
Which is the estimation of the probability with CL=90%?

Frequentist result: x=5, n=20, CL=90%

$$\sum_{k=n}^{n} \binom{n}{k} p_1^k (1-p_1)^{n-k} = 0.05$$

$$\sum_{k=0}^{x} \binom{n}{k} p_2^k (1-p_2)^{n-k} = 0.05$$

p_1, p_2 → p=[0.117, 0.434]

Bayesian result:

What meaning??

$$\frac{\int_{p_1}^{p_2} p^x (1-p)^{n-x} \, \mathrm{d}p}{\int_0^1 p^x (1-p)^{n-x} \, \mathrm{d}p} = 0.90$$

p_1, p_2 → p=[0.133, 0.437]

Elementary example II

There is a large number of marbles, which are either white or black, and you wish information on the white fraction, μ.
You draw a single marble, and it is white. What is the fraction μ with 90% of confidence?

Classical:

$$p_1 = 1 - CL = 0.1 \quad \rightarrow \quad \mu \geq 0.1$$

Bayesian:

flat prior $\quad p^2{}_1 = 1 - CL = 0.1 \quad \rightarrow \quad \mu \geq 0.316$

$1/\mu$ prior $\quad p_1 = 1 - CL = 0.1 \quad \rightarrow \quad \mu \geq 0.100$

μ prior $\quad p^3{}_1 = 1 - CL = 0.1 \quad \rightarrow \quad \mu \geq 0.464$

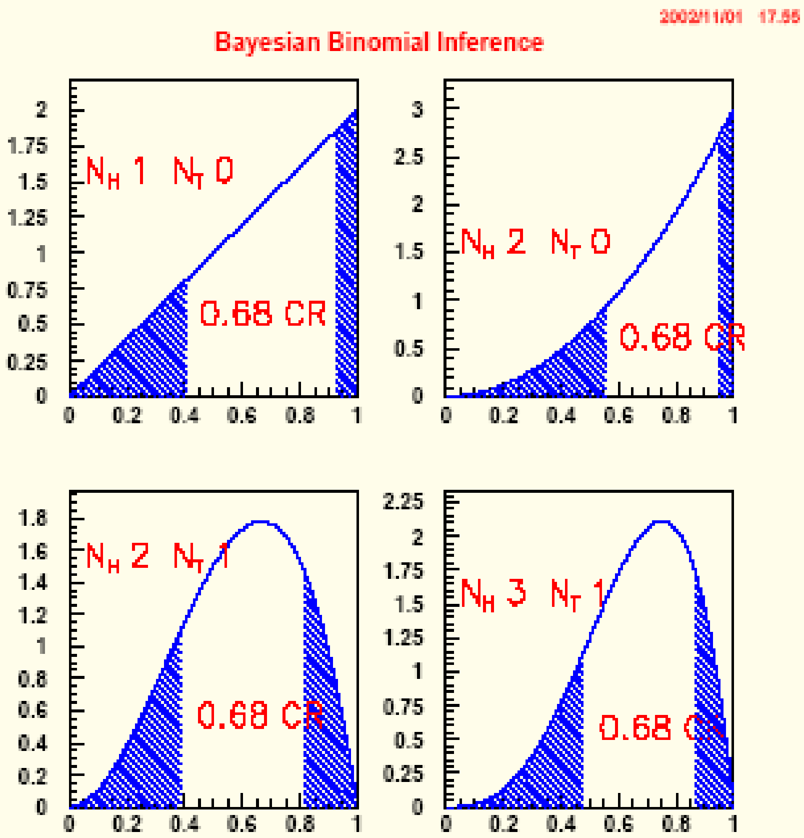
Bayesian coin tossing

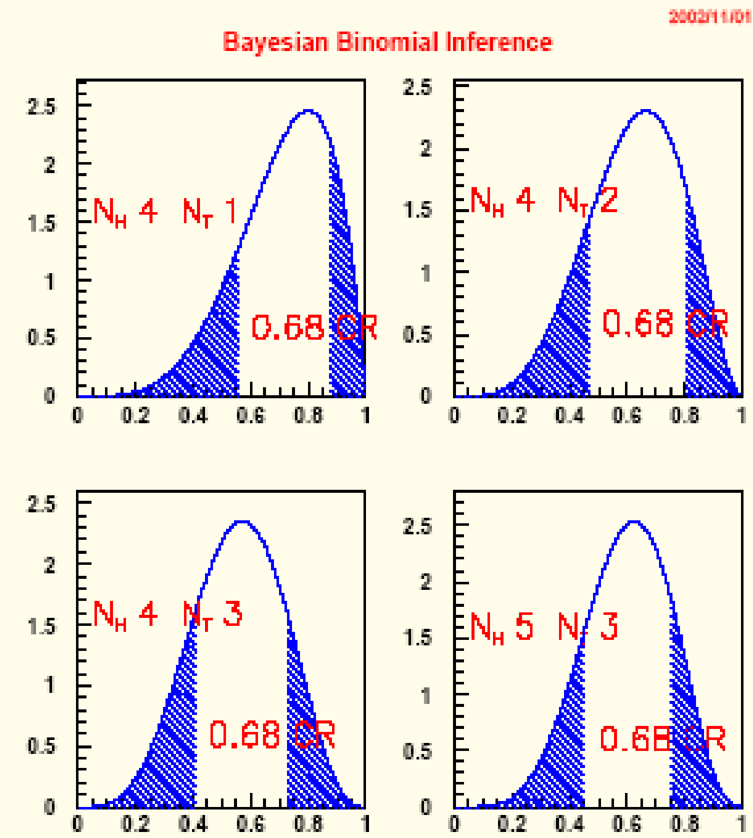
Bayes Theorem:

$$P(\theta \mid D, I) \propto P(D \mid \theta, I) \cdot P(\theta \mid I)$$

D is the *data*, I summarizes all the relevant information.
Assume a flat prior $P(\theta \mid I)$ in a Binomial experiment: cast a coin (p).
Bayes formula will update the information on p at each experiment:

$$\begin{aligned}
P(\theta \mid noData, I) &= P_0(\theta \mid I) = 1 \\
P(\theta \mid H, I) &\propto \theta \\
P(\theta \mid H, H, I) &\propto \theta^2 \\
P(\theta \mid H, H, , T, I) &\propto \theta^2(1-\theta) \\
&\cdots \\
P(\theta \mid nN, mT, I) &\propto \theta^n(1-\theta)^m
\end{aligned}$$





The plots are the posterior density after each measurement.
Also shown the 68% Credible Intervals.

The 3 event experiment I

A counting experiment registered 3 events

Find the estimate of μ:

$$p(x;\mu) = \frac{\mu^x}{x!}\,\mathrm{e}^{-\mu}$$

1. ~~"naif" physicist~~

$$\mu = 3 \pm \sqrt{3} = [1.27, 4.73]\ , \quad CL \simeq 68\%$$

wrong

2. standard (frequentist) physicist:

   solves the equations

$$\sum_{x=3}^{\infty} \frac{\mu_1^x}{x!}\,\mathrm{e}^{-\mu_1} = 0.16\ , \quad \sum_{x=0}^{3} \frac{\mu_2^x}{x!}\,\mathrm{e}^{-\mu_2} = 0.16$$

   Numerically one obtains the interval:

$$\mu = [1.37, 5.92]\ , \quad CL = 68\%$$

3. Bayesian physicist:

   solves the Bayes formula

$$\int_{\mu_1}^{\mu_2} \frac{\mu^3}{3!}\,\mathrm{e}^{-\mu}\,p_\mu(\mu)\,\mathrm{d}\mu = 0.68\ , \quad \text{with}$$

Last Informative Prior (LIP)

$$p_\mu(\mu) = 1\ , \quad \int_0^{\mu_1} = \int_{\mu_2}^{\infty} = 0.16$$

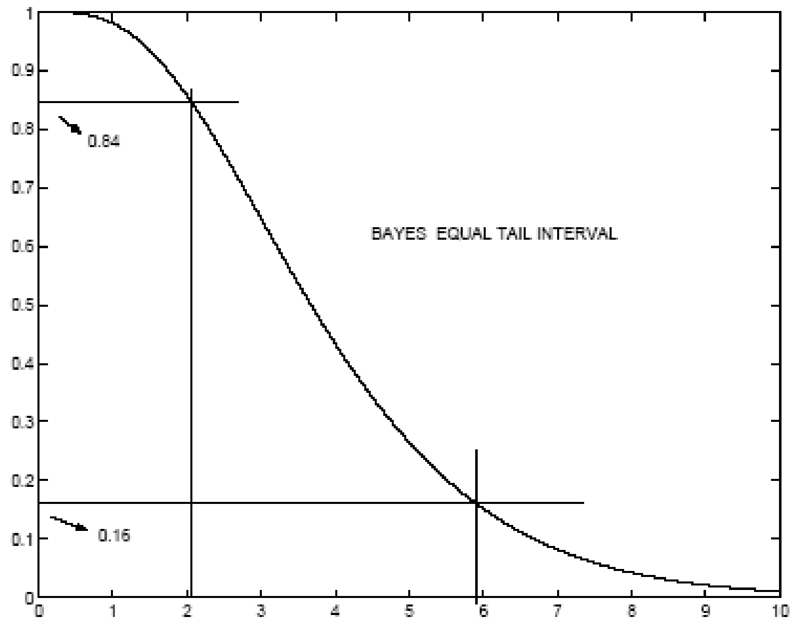
   The equal tail Bayesian interval:

$$\mu = [2.09, 5.92]$$

The important identity holds:

$$P(\mu) = \int_{\mu}^{\infty} \frac{\mu^x}{x!} \, \mathrm{e}^{-\mu} \, \mathrm{d}\mu = \mathrm{e}^{-\mu} \sum_{k=0}^{x} \frac{\mu^k}{k!}$$

$P(\mu)$ plot for $x = 3$:



$$\int_{0}^{\mu_1} = \int_{\mu_2}^{\infty} = 0.16$$

$$\mu = [2.09, 5.92]$$

Bayesian Interval estimate

3. Bayesian physicist:

solves the Bayes formula

$$\int_{\mu_1}^{\mu_2} \frac{\mu^3}{3!} \, e^{-\mu} \, p_\mu(\mu) \, d\mu = 0.68 \ , \quad \text{with}$$

$$p_\mu(\mu) = 1 \ , \quad \int_{\mu_1}^{\mu} = \int_{\mu}^{\mu_2}$$

The central Bayesian interval is:

$$\mu = [1.55, 5, 15]$$

4. Bayesian physicist:

Bayes formula with $p_\mu(\mu) = 1/\mu$, hence:

Uniform Jeffreys' Prior

$$d(1/\mu) = d\mu/\mu^2 \ , \quad \int_0^{\mu_1} = \int_{\mu_2}^{\infty} = 0.16$$

$$p(\mu; x) = \int L(\mu; x) p_\mu(1/\mu) \, d\mu$$

$$= \int_{\mu_1}^{\mu_2} \frac{\mu^3}{3!} \, e^{-\mu} \, \frac{1}{\mu^2} \, d\mu = 0.68$$

The central Bayesian interval is:

$$\mu = [1.37, 4.64]$$

The 3 event experiment
Summary

method	interval	coverage
naif	[1.27, 4.73]	NO
Neyman	[1.37, 5.92]	68%
Bayes, uniform $p_\mu(\mu)$	[2.09, 5.92]	? $< 68\%$
Bayes, uniform $p_\mu(1/\mu)$	[1.37, 4.64]	? $< 68\%$

- Bayes with uniform $p_\mu(\mu)$ gives the same frequentist upper limit (5.92);
- Bayes with uniform $p_\mu(1/\mu)$ gives the same frequentist lower limit (1.37);

The neutrino mass
...here Bayes helps!

An experiment with a Gaussian resolution of

$$\sigma = 3.3 \text{ eV}/c^2$$

measures the ν_e mass as:

$$m = -5.41 \text{ eV}/c^2$$

make the Bayesian estimate of m_ν.

Bayes formula

$$p(m_\nu; m, \sigma) = \frac{p(m; m_\nu, \sigma)\, p_\nu(m_\nu)}{\int p(m; m_\nu, \sigma)\, p_\nu(m_\nu)\, \mathrm{d}m_\nu}$$

Choosing the prior:

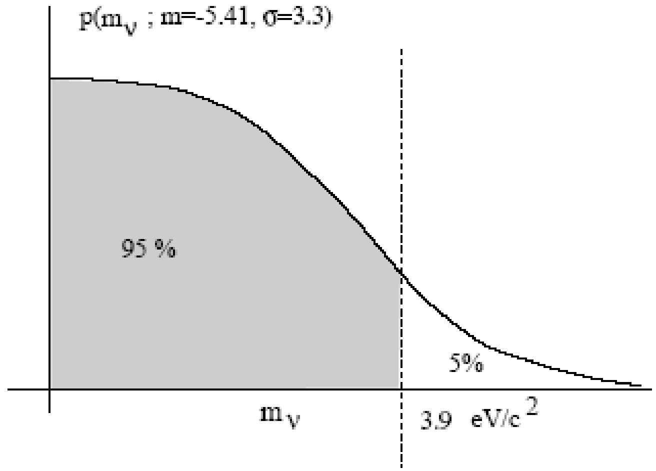
- **define** $0 \le m_\nu \le 20 - 30 \text{ eV}/c^2$ **;**
- **define** $\sigma_\nu = 10 \text{ eV}/c^2$
- **test three functional forms:**
  1. **uniform:** $p_\nu = p_u(m_\nu) = 1/30\ ,\ 0 \le m_\nu \le 30$
  2. **Gaussian:**
  $$p_\nu = p_g(m_\nu) = \frac{2}{2\pi\sigma_\nu} \exp[-m_\nu^2/(2\sigma_\nu^2)]$$
  3. **triangular:** $p_\nu = p_t(m_\nu) = \frac{1}{450}(30 - m_\nu)$**,**
  $0 \le m_\nu \le 30 \text{ eV}/c^2$

The neutrino mass II

For example, using the uniform $p_u(m_\nu)$ and $\sigma = 3.3$, $m = -5.41$ ev$/c^2$:

$$p(m_\nu; m, \sigma) = \frac{\exp\left[-\frac{(m-m_\nu)^2}{2\sigma^2}\right]\frac{1}{30}}{\int_0^{30}\exp\left[-\frac{(m-m_\nu)^2}{2\sigma^2}\right]\frac{1}{30}\,\mathrm{d}m_\nu}$$

one obtains, at 95% probability:



- uniform: $0 \le m_\nu \le 3.9$ eV$/c^2$;
- Gaussian: $0 \le m_\nu \le 3.7$ eV$/c^2$;
- triangular: $0 \le m_\nu \le 3.7$ eV$/c^2$.

result "independent" of the prior!
Here the prior represent the knowledge, not the ignorance!!!

The Gaussian Case

That is... Put in your analysis your KNOWLEDGE

Assume that we have made a measurement of the mean of a Gaussian variable and we have obtained: $x_1 \pm \sigma_1$. The posterior density, in the case of a flat prior is:

$$f(\mu \mid x_1, \sigma_1, I_b) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1-\mu)/2\sigma_1^2}$$

If now we make another measurement yielding: $x_2 \pm \sigma_2$ the posterior becomes:

$$f(\mu \mid x_1, \sigma_1, x_2, \sigma_2, I_b) = \frac{N(\mu; x_1, \sigma_1^2) \cdot N(\mu; x_2, \sigma_2^2)}{\int_{-\infty}^{\infty} N(\mu; x_1, \sigma_1^2) \cdot N(\mu; x_2, \sigma_2^2)\, d\mu}$$

The Gaussian Case

A boring calculation leads to the following result:

$$f(\mu \mid x_1, \sigma_1, x_2, \sigma_2, I_b) = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-(x_a-\mu)/2\sigma_a^2}$$

$$x_a = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

$$\sigma_a^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

The most probable value of μ is just the weighted mean.

Conclusions

- don't be dogmatic

- use Bayes to parametrize the a priori knowledge if any, not the ignorance

- in the case of poor a priori knowledge, use the frequentist methods

Quantum Mechanics: frequentist or bayesian? Born or Bhor?

$$\int |\psi|^2 dx$$

The standard interpretation is frequentist

Use of the likelihood principle in physics

Maximum Likelihood

Likelihood function:

$$L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}}) = p(x_{11}, x_{21}, .., x_{m1}; \boldsymbol{\theta})\, p(x_{12}, x_{22}, .., x_{m2}; \boldsymbol{\theta})\,. \\ \times\, p(x_{1n}, x_{2n}, .., x_{mn}; \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i; \boldsymbol{\theta}) \;,$$

the product covers
all the n values of the m variables $\boldsymbol{X}$.
Log-likelihood:

$$\mathcal{L} = -\ln\left(L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})\right) = -\sum_{i=1}^{n} \ln\left(p(\boldsymbol{x}_i; \boldsymbol{\theta})\right) \;,$$

Max L corresponds to Min $\mathcal{L}$.
For a given set of

$$\underline{\boldsymbol{x}} = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$$

observed values, from a

$$\underline{\boldsymbol{X}} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$$

sample with density $p(\boldsymbol{x}; \boldsymbol{\theta})$, the ML estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the maximum (if any) of the function

$$\max_{\Theta}\left[L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})\right] = \max_{\Theta}\left[\prod_{i=1}^{n} p(\boldsymbol{x}_i; \boldsymbol{\theta})\right] = L(\hat{\boldsymbol{\theta}};\, \underline{\boldsymbol{x}})$$

Maximum likelihood

$$\frac{\partial L}{\partial \theta_k} = \frac{\partial \left[\prod_{i=1}^{n} p(\boldsymbol{x}_i; \boldsymbol{\theta}) \right]}{\partial \theta_k} = 0$$

or

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_{i=1}^{n} \left[\frac{1}{p(\boldsymbol{x}_i; \boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}_i; \boldsymbol{\theta})}{\partial \theta_k} \right] = 0\,, \qquad (k = 1, 2, \ldots, p)\,.$$

- ***before the trial***, the likelihood function $L(\boldsymbol{\theta}; \underline{\boldsymbol{x}})$ is $\propto$ to the pdf of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \boldsymbol{X}_n)$;
- ***before the trial***, the likelihood function $L(\boldsymbol{\theta}; \underline{\boldsymbol{X}})$ is a random function of X;

- frequentist view: maximize the function

$$L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i; \boldsymbol{\theta}) \,, \quad \text{or} \quad \ln\left(L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})\right) = +\sum_{=1}^{n} \ln\left(p(\boldsymbol{x}_i; \boldsymbol{\theta})\right) \,,$$

or minimize

$$-2\ln\left(L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})\right) = -2\sum_{i=1}^{n} \ln\left(p(\boldsymbol{x}_i; \boldsymbol{\theta})\right)$$

w.r.t the parameters $\boldsymbol{\theta}$.

- Bayesian view:

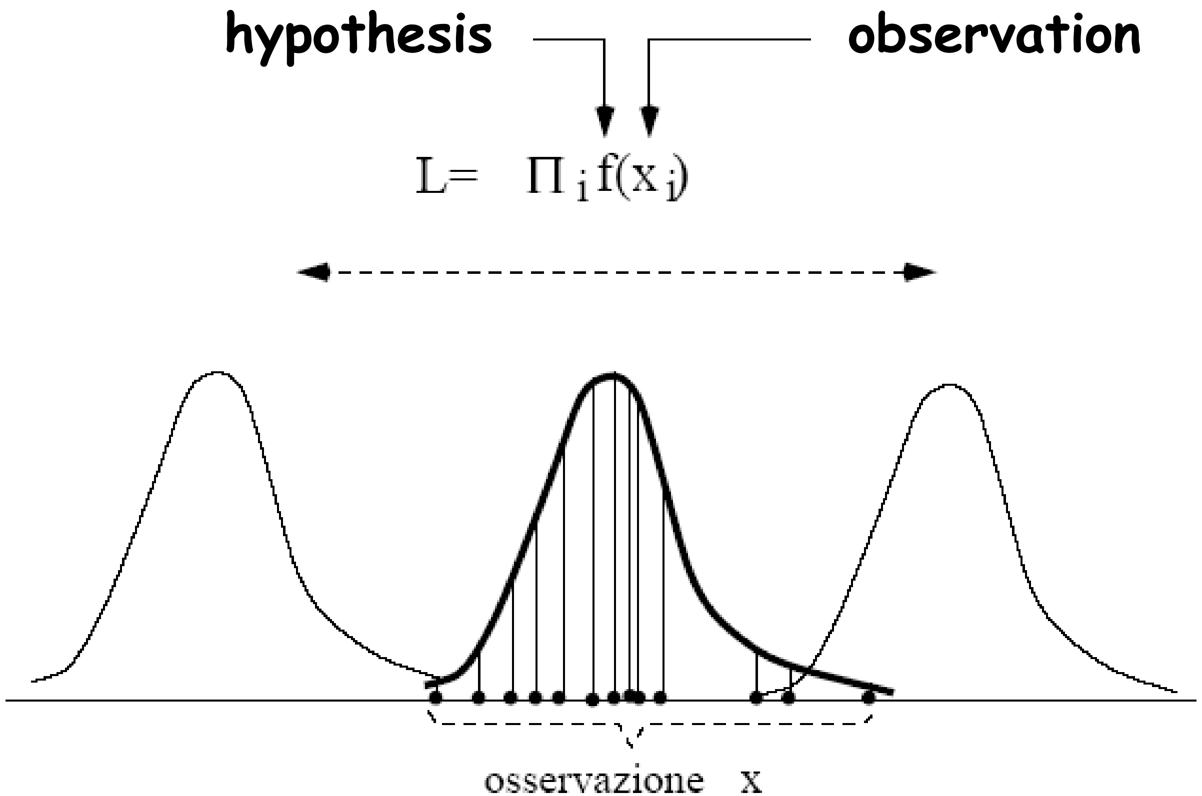
maximize the posterior probability

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\, p(\boldsymbol{\theta}')\, \mathrm{d}\boldsymbol{\theta}'} \propto L(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

- Bayes maximization updates the prior $p(\boldsymbol{\theta})$
- when the prior $p(\boldsymbol{\theta})$ is uniform (constant) technically the frequentist and the Bayesian approaches coincide because both maximize $L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})$ (but the meaning is different)
- Bayesian estimators are not independent of the transformation of the parameters, the frequentist ones are independent of them!

Bayesians vs Frequentists

Why ML does work?



The $p(x;\theta)$ form is fitted to data by maximizing the ordinates of the observed data

Example

An urn with three marbles

●●○ ○○●

$p = 1/3$ $p = 2/3$

An experiment with 4 drawings:

$$p(x; n = 4, p) = \frac{4!}{x!(4-x)!} p^x (1-p)^{4-x}$$

	x=0	x=1	x=2	x=3	x=4
$p(x; 4, p = 1/3)$	16/81	32/81	24/81	8/81	1/81
$p(x; 4, p = 2/3)$	1/81	8/81	24/81	32/81	16/81

The likelihood estimate:

$\hat{p} = 1/3$ **if** $0 \leq x \leq 1$

$\hat{p} = 2/3$ **if** $3 \leq x \leq 4$

no maximum if $x = 2$

Example

In n trial x successes have been obtained. Make the ML estimate of p.

Binomial density

$$\mathcal{L} = -x\ln(p) - (n-x)\ln(1-p) \ .$$

Minimum w.r.t. p:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} = -\frac{x}{p} + \frac{n-x}{1-p} = 0 \implies \hat{p} = \frac{x}{n} = f$$

Make the ML estimate of p when x_1 successes on n_1 trials and x_2 successes on n_2 trials have been obtained.

Two binomials with the same p:

$$L = p^{x_1}\, p^{x_2}\, (1-p)^{n_1-x_1}\, (1-p)^{n_2-x_2} \ .$$

With logarithms:

$$\mathcal{L} = -(x_1+x_2)\ln(p) - (n_1-x_1+n_2-x_2)\ln(1-p) \ ,$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} = -\frac{x_1+x_2}{p} + \frac{(n_1+n_2)-x_1-x_2}{1-p} = 0$$

$$\implies \hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

Estimators

- **Estimator of θ**

  If $\underline{\boldsymbol{X}}$ is a data sample with dimension n of a m-dimensional random variable $\boldsymbol{X}$ having $p(\boldsymbol{X};\theta)$ as a pdf, an estimator is a statistics

$$T_n(\underline{\boldsymbol{X}}) \equiv t_n(\underline{\boldsymbol{X}})$$

  for which $T : S \to \theta$.

- **Consistent estimator of θ**

$$\lim_{n\to\infty} P\{|T_n - \theta| < \epsilon\} = 1, \quad \forall\, \epsilon > 0\ .$$

- **Correct or unbiased estimator**

$$\langle T_n \rangle = \theta, \quad \forall\ n$$

- **The most efficient estimator**

  T_n is more efficient than Q_n if

$$\mathrm{Var}[T_n] < \mathrm{Var}[Q_n], \quad \forall\ \theta \in \Theta\ .$$

Theorems on $L(\theta; X)$

The mean value of the Score Function is zero:

$$\left\langle \frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right\rangle = 0 \ .$$

The variance of the Score Function is the Fisher information:

$$\begin{aligned} \mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta)\right] &= \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) - \left\langle \frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right\rangle \right)^2 \right\rangle \\ &= \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle \equiv I(\theta) \end{aligned}$$

These remarkable relations hold:

$$I(\theta) = \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle = -\left\langle \frac{\partial^2}{\partial\theta^2} \ln p(\boldsymbol{X};\theta) \right\rangle \ .$$

$$\left\langle \left(\frac{\partial}{\partial\theta} \ln L \right)^2 \right\rangle = \left\langle \left(\frac{\partial}{\partial\theta} \sum_i \ln p(\boldsymbol{X}_i;\theta) \right)^2 \right\rangle = n \left\langle \left(\frac{\partial}{\partial\theta} \ln p \right)^2 \right\rangle = nI(\theta) \ ,$$

The Cramér Rao theorem:

If T_n is an unbiased estimator

$$\mathrm{Var}[T_n] \geq \frac{1}{n \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle} = \frac{1}{nI(\theta)}$$

Binomial, Poisson, Gauss

$$\ln b(X;p) = \ln n! - \ln(n-X)! - \ln X! + X\ln p + (n-X)\ln(1-p)$$
$$\ln p(X;\mu) = X\ln\mu - \ln X! - \mu$$
$$\ln g(X;\mu,\sigma) = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2$$

These are random functions.

$$\begin{aligned}
\frac{\partial}{\partial p}\ln b(X;p) &= \frac{X}{p} - \frac{n-X}{1-p} = \frac{X-np}{p(1-p)} \\
\frac{\partial}{\partial \mu}\ln p(X;\mu) &= \frac{X}{\mu} - 1 = \frac{X-\mu}{\mu}\ , \\
\frac{\partial}{\partial \mu}\ln g(X;\mu,\sigma) &= -\frac{X-\mu}{\sigma}\left(-\frac{1}{\sigma}\right) = \frac{X-\mu}{\sigma^2}
\end{aligned}$$

according to $\left\langle \frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right\rangle = 0$

Information:

$$\begin{aligned}
I(p) &= \frac{1}{p^2(1-p)^2}\left\langle (X-np)^2\right\rangle = \frac{np(1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)}\ , \\
I(\mu) &= \frac{1}{\mu^2}\left\langle (X-\mu)^2\right\rangle = \frac{\sigma^2}{\mu^2} = \frac{1}{\mu} = \frac{1}{\sigma^2}\ , \\
I(\mu) &= \frac{1}{\sigma^4}\left\langle (X-\mu)^2\right\rangle = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}\ ,
\end{aligned}$$

Golden results

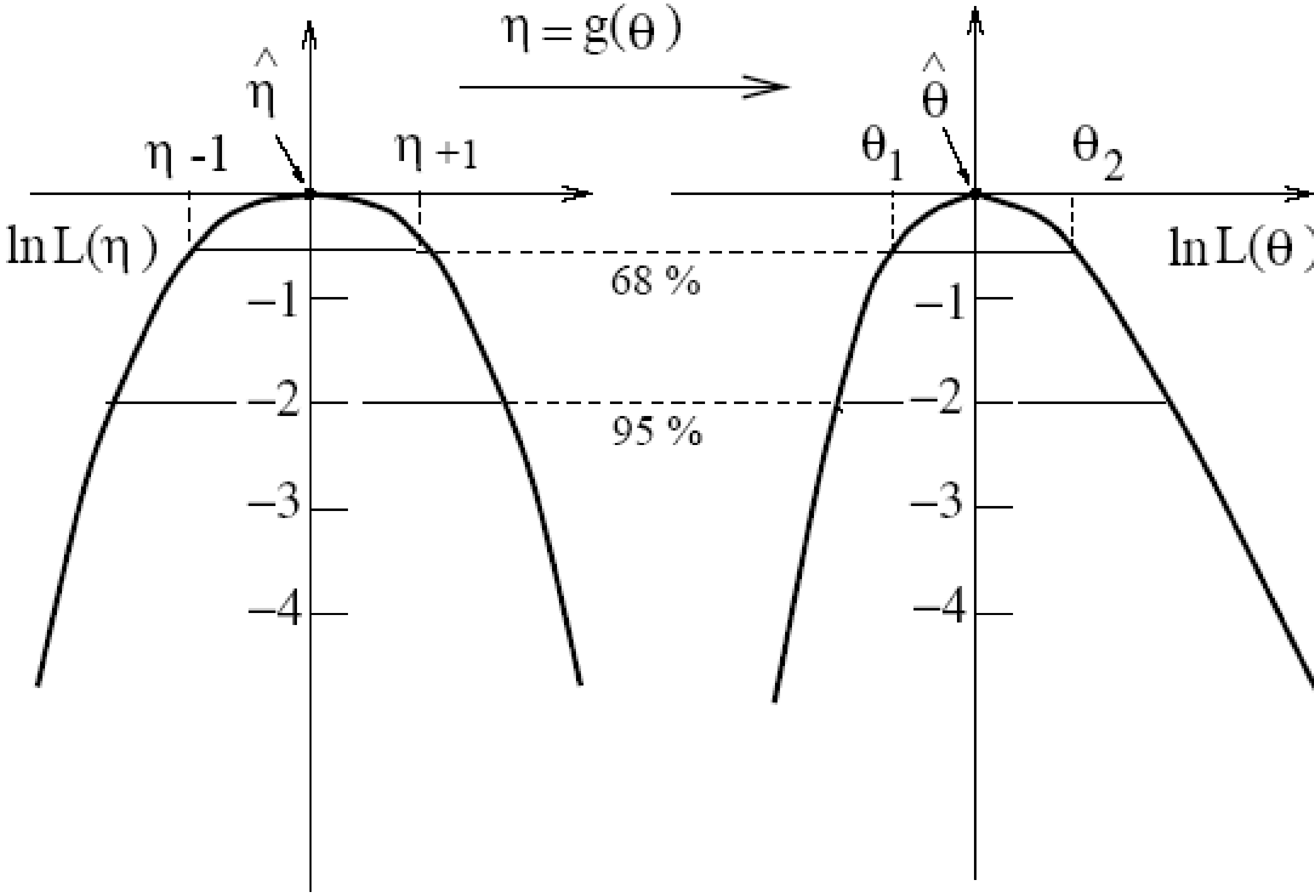
1. If T_n is the best estimator of $\tau(\theta)$, it coincides with the ML estimator (if any)
$$T_n = \tau(\hat{\theta}) \ .$$

2. the ML estimator is consistent

3. under broad conditions, the ML estimators are asymptotically normal. That is $(\theta - \hat{\boldsymbol{\theta}})$ is asymptotically normal with variance
$$\frac{1}{nI(\theta)}$$

4. the score function $\partial \ln L / \partial \theta$ has zero mean, $nI(\theta)$ variance and is asymptotically normal

5. the variable
$$2[\ln L(\hat{\boldsymbol{\theta}}) - \ln L(\boldsymbol{\theta})]$$
tends asymptotically to $\chi^2(p)$, where p is the dimension of θ



η = g(θ)
η̂
η -1
η +1
ln L(η)
θ̂
θ1
θ2
ln L(θ)
68 %
95 %
−1
−2
−3
−4
−1
−2
−3
−4

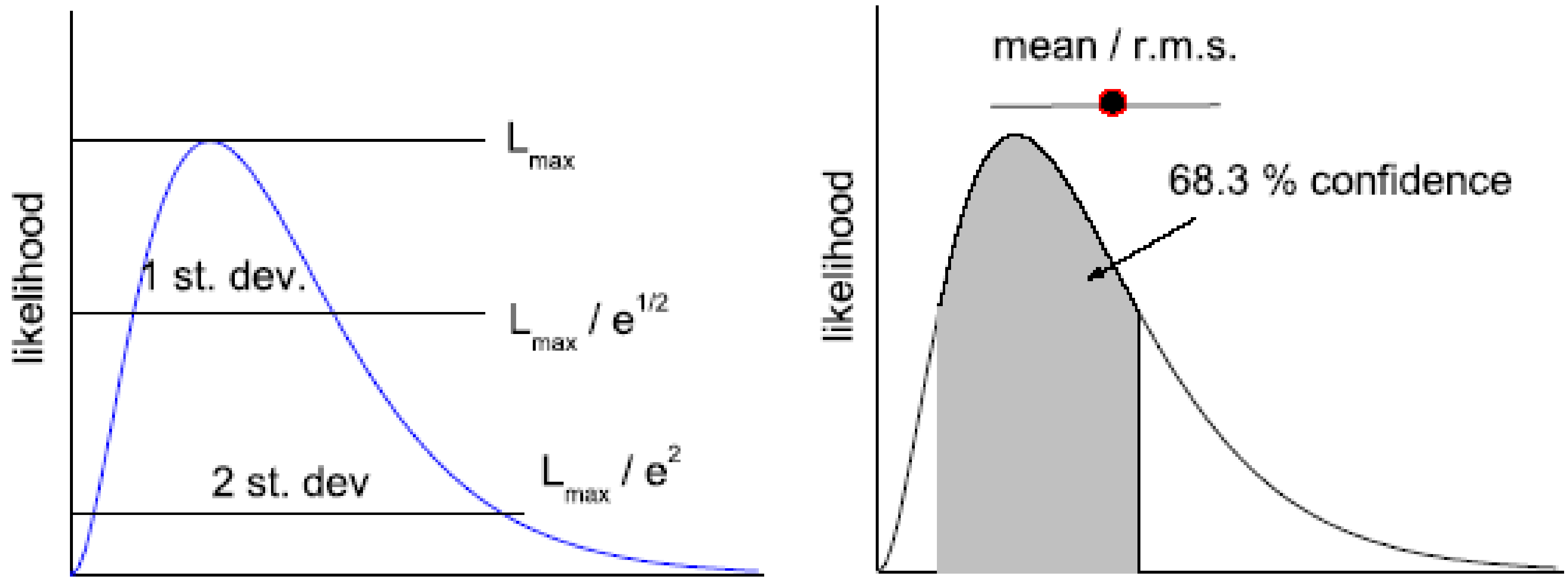
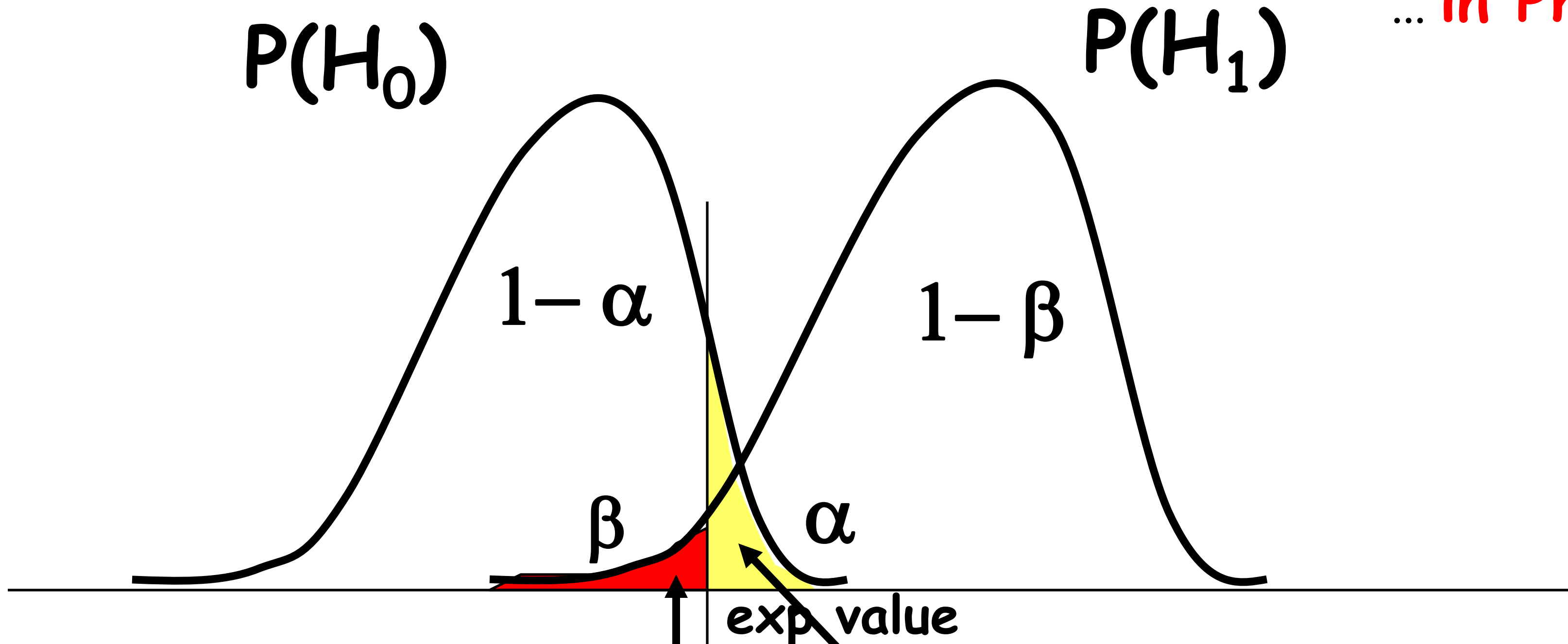


Fig. 18. Likelihood ratio limits (left) and Bayesian limits (right)

The other branch of Statistics: Hypothesis Testing

... **in Physics**



true hypothesis	Decision H_0	Decision H_1
H_0	correct decision	type I error
	$1-\alpha$	α
no effect	good rejection	contamination
H_1	type II error	correct decision
	β	$1-\beta$
effect	event loss	good acceptance

power ← $1-\beta$

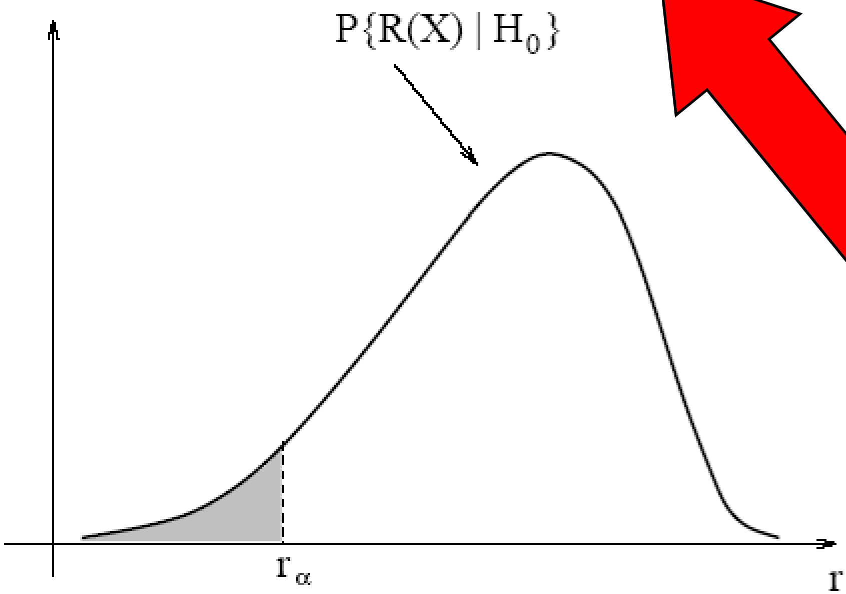
If H_1 is the discovery, the maximum power test maximizes the discovery probability, that is the good acceptance

When two **simple** hypotheses are given

$$H_0 : \theta = \theta_0 \ , \qquad H_1 : \theta = \theta_1 \ .$$

the most powerful test, for α given, is

$$\text{reject} \quad H_0 \quad \text{if} \quad \left\{ R(X) = \frac{L(\theta_0; X)}{L(\theta_1; X)} \leq r_\alpha \right\} \ ,$$



A Milestone:
the Neyman-Pearson theorem

Likelihood Ratio Test

That is:

the best ***test statistics*** is R

or ***any random variable*** T : $R = \psi(T)$.

A Milestone: the Neyman-Pearson theorem: limitations

- it holds for simple hypotheses
- for composite hypotheses like

$$
\begin{aligned}
H_0 &: \theta_1 = a \ , \quad \theta_2 = b \\
H_1 &: \theta_1 \neq a \ , \quad \theta_2 \neq b
\end{aligned}
$$

or

$$
\begin{aligned}
H_0 &: \theta = a \ , \\
H_1 &: \theta \geq a
\end{aligned}
$$

the NP ratio

$$R = \frac{L(\theta|H_0)}{\max_{[\theta \in \Theta_1]} L(\theta|H_1)}$$

is optimal, but only asymptotically

(theory is complicated!!)

- if H_1 has r free parameters more than H_0, then

$$-2 \ln R \sim \chi^2(r)$$

The powerful LR test is used usually on histograms with N_c channels:

$$Q = \frac{\Pi_{i=1}^{N_c} (s_i + b_i)^{n_i} \, e^{-(s_i+b_i)}/n_i!}{\Pi_{i=1}^{N_c} b_i^{n_i} \, e^{-b_i}/n_i!} \; , \qquad S_{\text{tot}} = \sum_{i=1}^{N_c} s_i \; .$$

where n_i is the number of observed events s_i and b_i are the expected signal and background events, b_i and s_i are obtained via MC

One obtains easily:

$$\ln Q = -S_{\text{tot}} + \sum_{i=1}^{N_c} n_i \ln \left(1 + \frac{s_i}{b_i} \right)$$

Usually one compare the quantity

$$-2 \ln Q \sim \chi^2 \qquad \text{(asymptotically)}$$

obtained experimentally (n_i = contents of the experimental bins) with the background ($n_i = b_i$) and the signal plus background ($n_i = s_i + b_i$) hypotheses. In this way, for an established signal to noise ratio, one performs the most powerful test, maximizing the signal discovery probability, *taking into account not only the global number of the events, but also the shape of the distributions* (*see LEP data*).

Likelihood Ratio

n_i from MC samples!

Steps of the likelihood ratio test

$$\ln Q = -S_{\text{tot}} + \sum_{i=1}^{N_c} n_i \ln\left(1 + \frac{s_i}{b_i}\right)$$

Determine the ratio s_i/b_i for each bin (model + MC simulation)

n_i

The Higgs at LEP in 2000

On 3 November 2000 in a seminar at CERN the LEP Higgs working group presented preliminary results of an analysis indicating a possible 2.9σ observation of a 115 GeV Higgs boson [1]. Based on this analysis the four LEP collaborations requested the continuation of LEP to collect more data at $\sqrt{s} = 208$ GeV. However, the arguments presented by the LEP collaborations did not convince the LEP management and in retrospect, it turned out that the LEP accelerator turn-off date of 2 November 2000 ended its eleven years of forefront research.

enough. However, the statistical arguments presented by the LEP Higgs working group were not based on these distributions, but rather on a sophisticated, though beautiful statistical analysis of the data. Two years after the event, when the last analysis of the LEP data indicated that the significance of a Higgs observation in the vicinity of 115 GeV went down to less than 2σ [2], it becomes apparent that the LEP Standard Model (SM) Higgs heritage will in fact be a lower bound on the mass of the Higgs boson. However, the LEP Higgs working group has taught us powerful and instructive lessons of statistical methods for deriving limits and confidence levels in the presence of mass dependent backgrounds from various channels and experiments. These lessons will remain with us long after the lower bound becomes outdated.

Available online at www.sciencedirect.com
SCIENCE DIRECT

PHYSICS LETTERS B

ELSEVIER

Physics Letters B 565 (2003) 61–75

www.elsevier.com/locate/npe

Search for the Standard Model Higgs boson at LEP

ALEPH Collaboration [1]
DELPHI Collaboration [2]
L3 Collaboration [3]
OPAL Collaboration [4]

The LEP Working Group for Higgs Boson Searches [5]

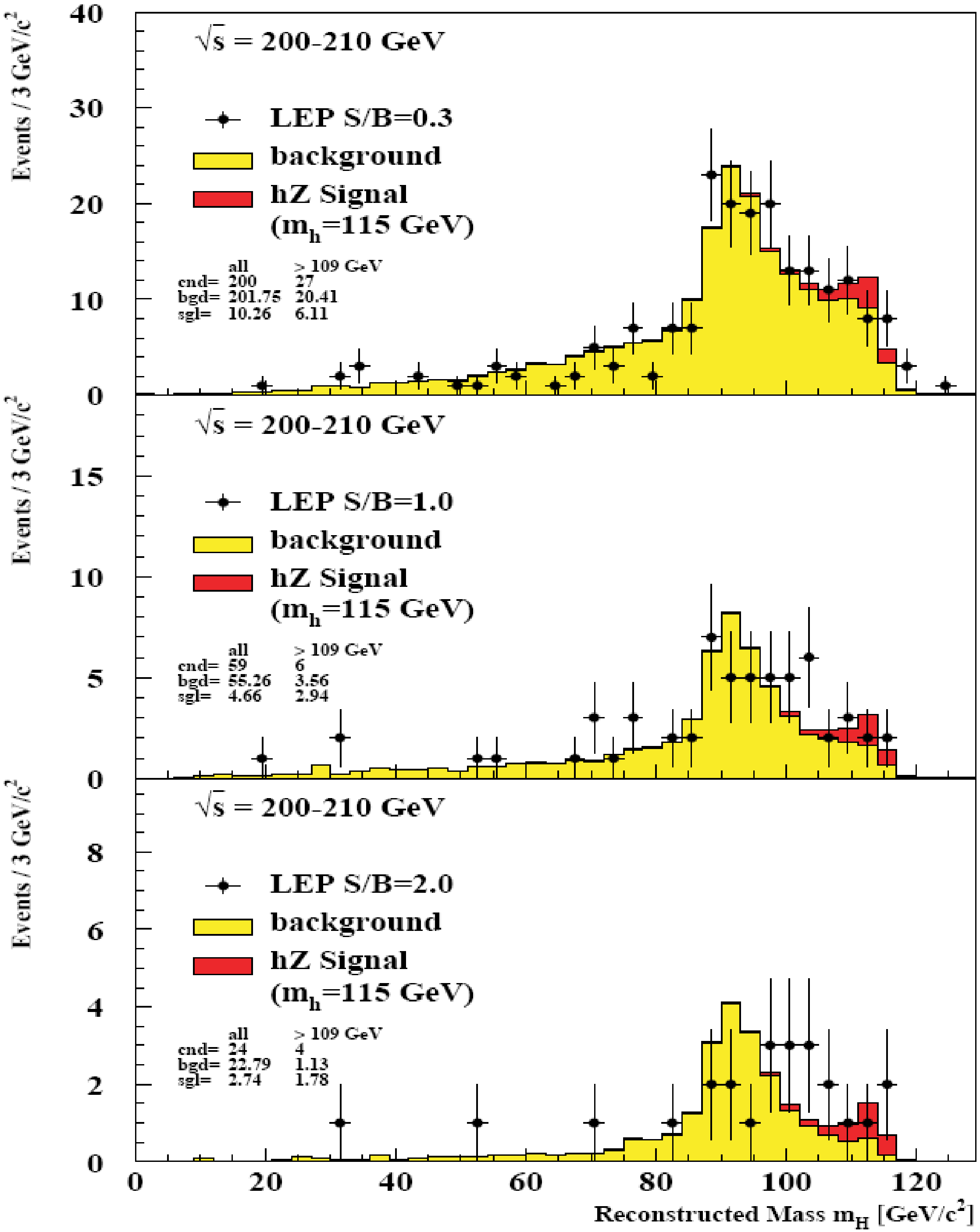
Received 7 March 2003; received in revised form 25 April 2003; accepted 28 April 2003

Editor: L. Rolandi

Abstract

The four LEP Collaborations, ALEPH, DELPHI, L3 and OPAL, have collected a total of 2461 pb^{-1} of e^+e^- collision data at centre-of-mass energies between 189 and 209 GeV. The data are used to search for the Standard Model Higgs boson. The search results of the four Collaborations are combined and examined in a likelihood test for their consistency with two hypotheses: the background hypothesis and the signal plus background hypothesis. The corresponding confidences have been computed as functions of the hypothetical Higgs boson mass. A lower bound of 114.4 GeV/c^2 is established, at the 95% confidence level, on the mass of the Standard Model Higgs boson. The LEP data are also used to set upper bounds on the HZZ coupling for various assumptions concerning the decay of the Higgs boson.

LEP real data



Three selections of the reconstructed Higgs mass of 115 GeV to obtain 0.5/1/2/ times as many expected signal as Background above 109 GeV

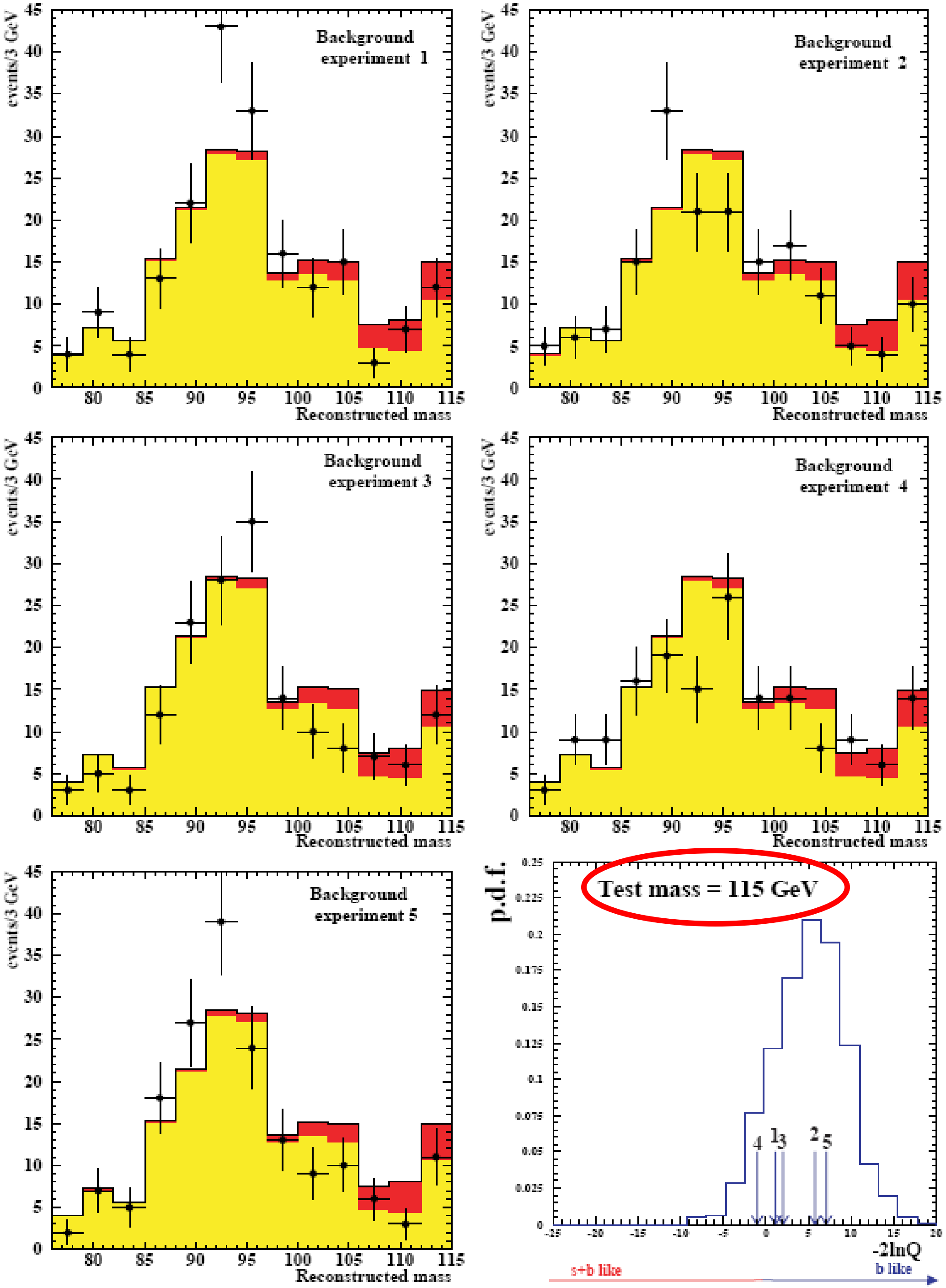
	Expt	E_{cm}	Decay channel	m_{rec} (GeV)	$\ln(1+s/b)$ at 115 GeV
1	ALEPH	206.6	4-jet	114.1	1.76
2	ALEPH	206.6	4-jet	114.4	1.44
3	ALEPH	206.4	4-jet	109.9	0.59
4	L3	206.4	E-miss	115.0	0.53
5	ALEPH	205.1	Lept	117.3	0.49
6	ALEPH	206.5	Taus	115.2	0.45
7	OPAL	206.4	4-jet	111.2	0.43
8	ALEPH	206.4	4-jet	114.4	0.41
9	L3	206.4	4-jet	108.3	0.30
10	DELPHI	206.6	4-jet	110.7	0.28
11	ALEPH	207.4	4-jet	102.8	0.27
12	DELPHI	206.6	4-jet	97.4	0.23
13	OPAL	201.5	E-miss	108.2	0.22
14	L3	206.4	E-miss	110.1	0.21
15	ALEPH	206.5	4-jet	114.2	0.19
16	DELPHI	206.6	4-jet	108.2	0.19
17	L3	206.6	4-jet	109.6	0.18

Table 1: Properties of the candidates with the highest weight at $m_H = 115$ GeV. Table is taken from [2].

Steps of the likelihood ratio test

$$\ln Q = -S_{\mathrm{tot}} + \sum_{i=1}^{N_c} n_i \ln\left(1 + \frac{s_i}{b_i}\right)$$

Determine the ratio s_i/b_i for each bin (model + MC simulation)



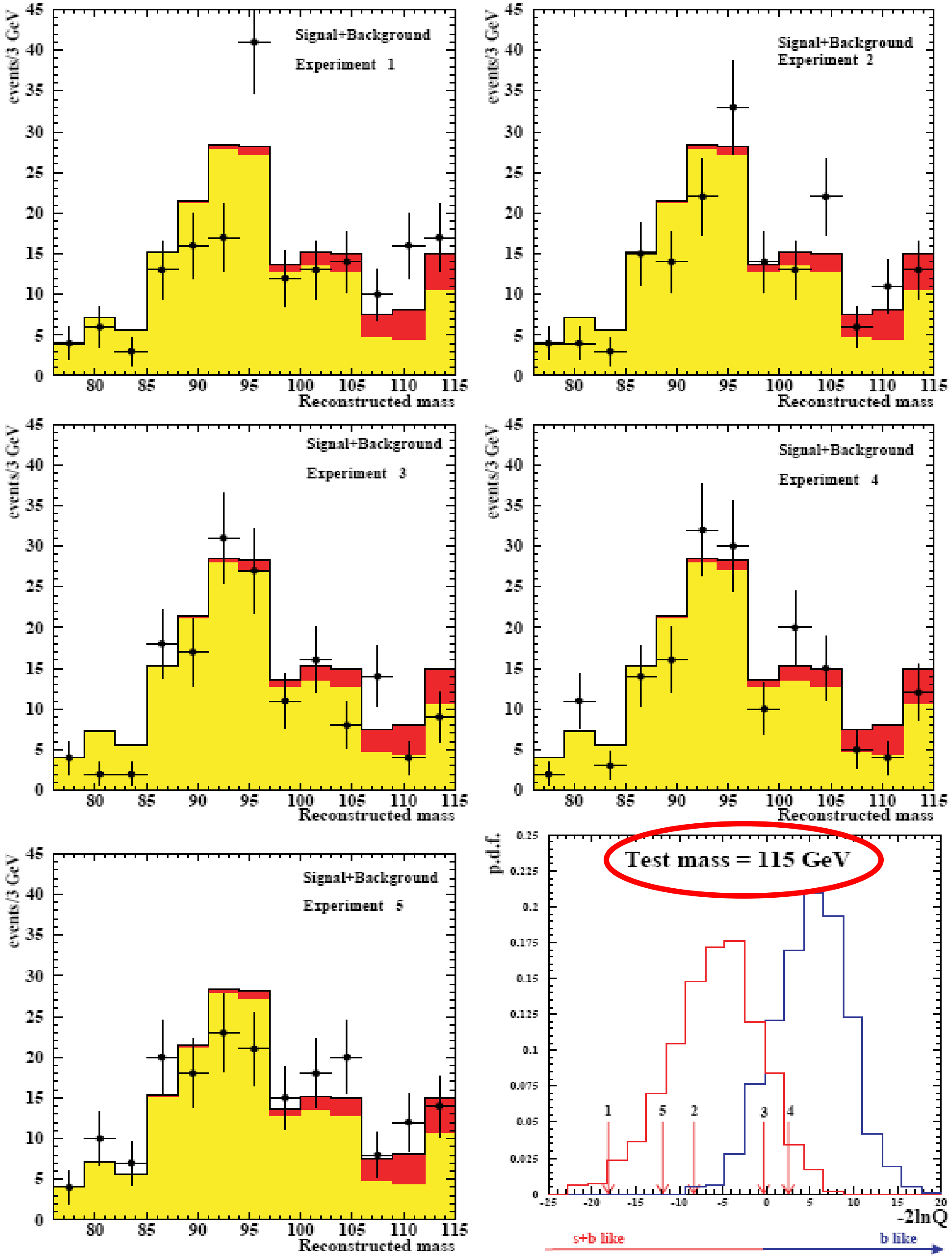
MC toy model

s_i red
b_i yellow

Crosses: MC data, Background only

ln(1+s/b) plot

1,2,3,4,5,....n



MC toy model

s_i red

b_i yellow

Crosses: MC data,

Background + Signal

m_H=115 GeV

ln(1+s/b) plot

1,2,3,4,5,....n

(in blue is the previous one with background only)

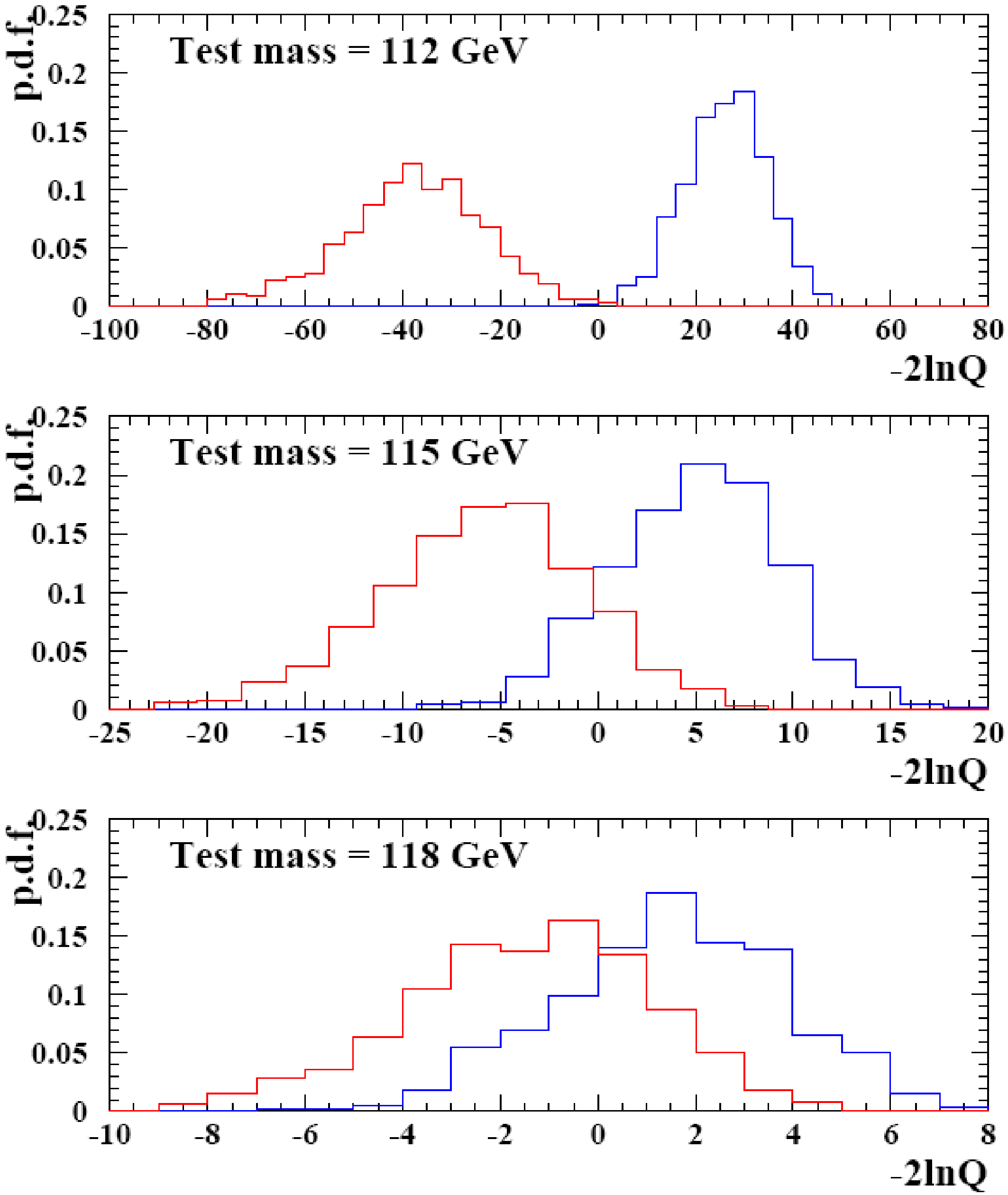


Figure 6: The separation between the Signal and the Background for various Higgs masses is shown by their likelihood p.d.f's.

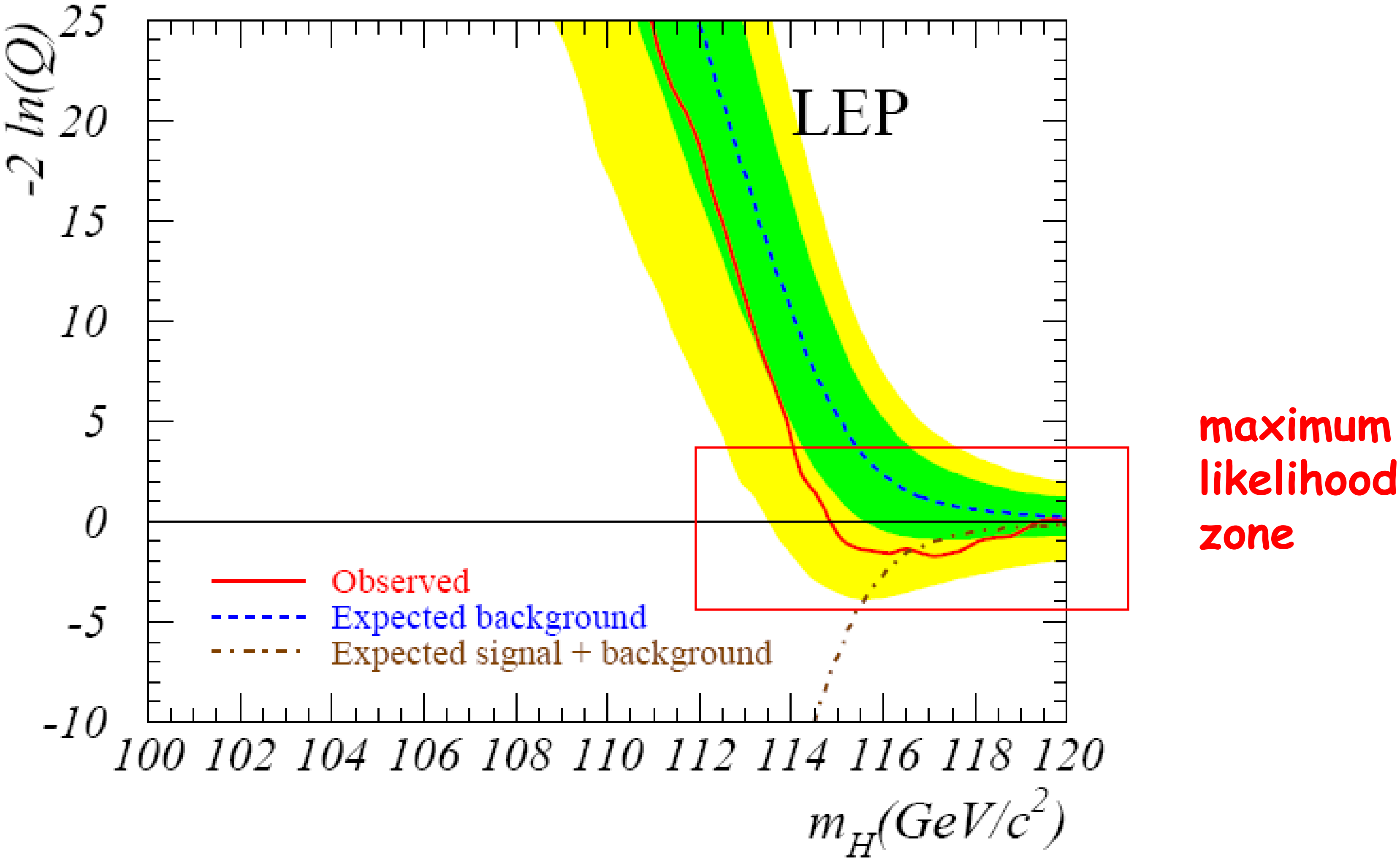


Figure 8: Observed and expected behavior of the likelihood $-2\ln Q$ as a function of the test-mass m_H for combined LEP experiments. The solid/red line represents the observation; the dashed/dash-dotted lines show the median background/signal+background expectations. The dark/green and light/yellow shaded bands represent the 1 and 2 σ probability bands about the median background expectation [2].

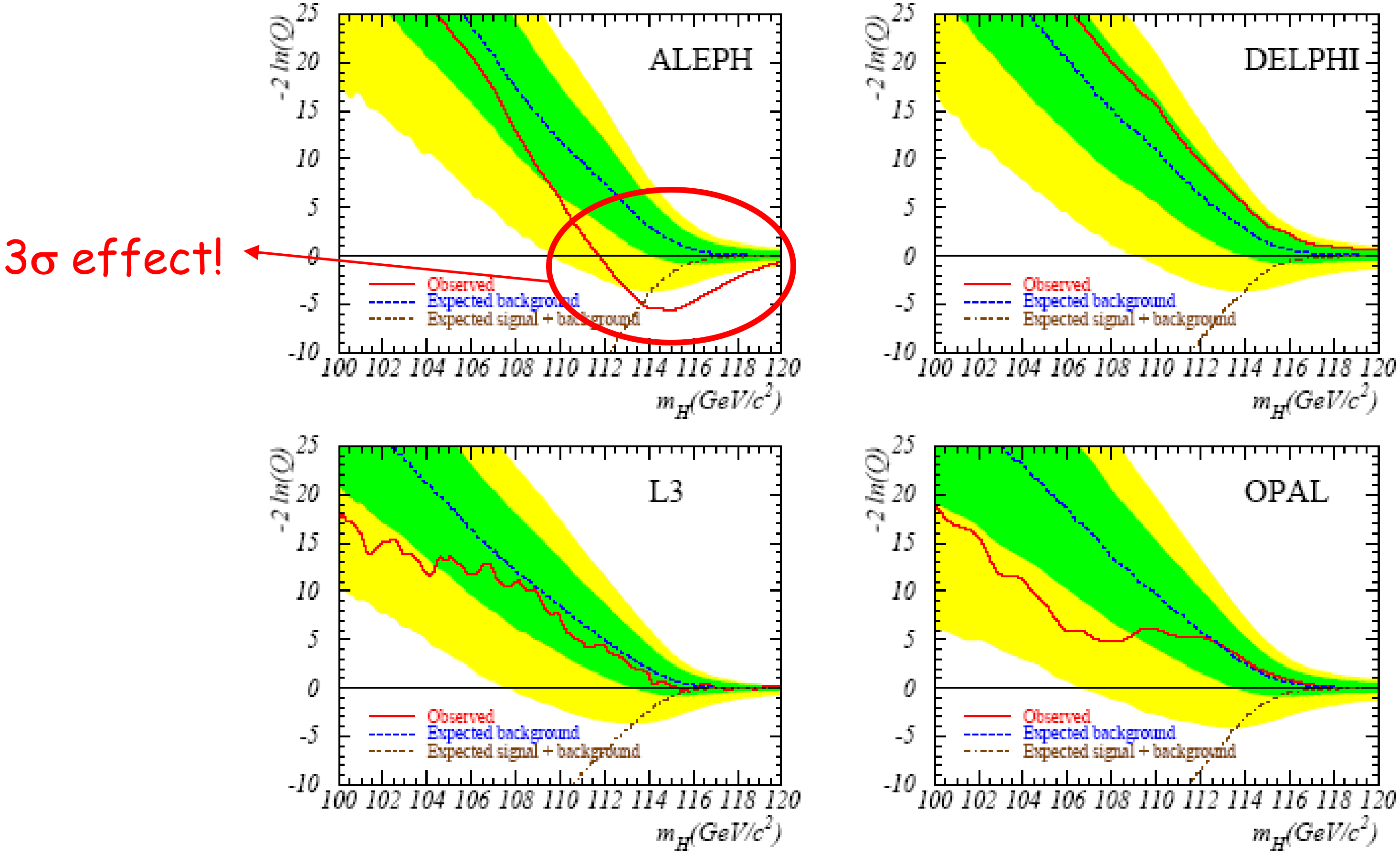
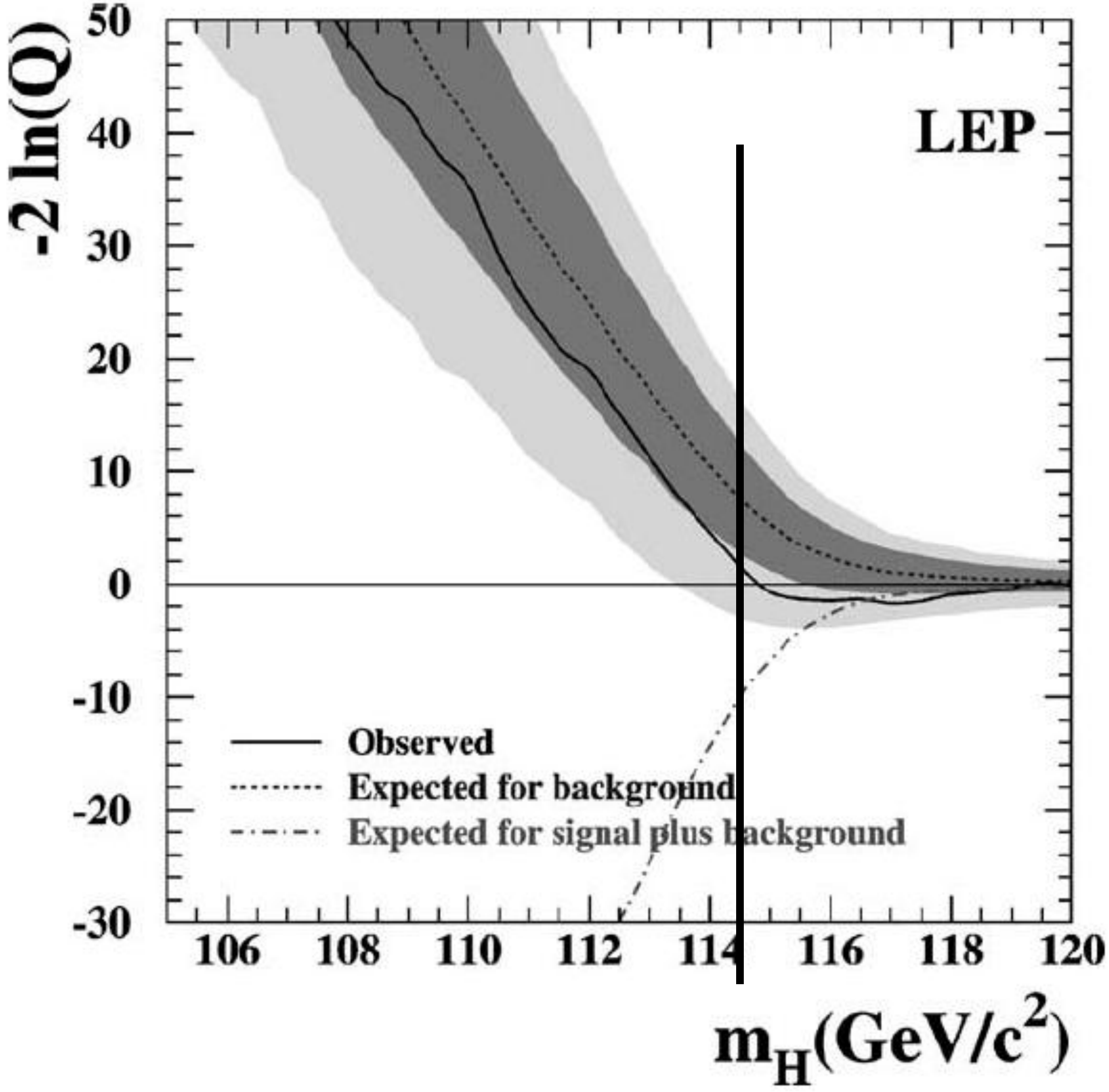


Figure 9: Observed and expected behavior of the likelihood $-2\ln Q$ as a function of the test-mass m_H for the various experiments. The solid/red line represents the observation; the dashed/dash-dotted lines show the median background/signal+background expectations. The dark/green and light/yellow shaded bands represent the 1 and 2 σ probability bands about the median background expectation [2].

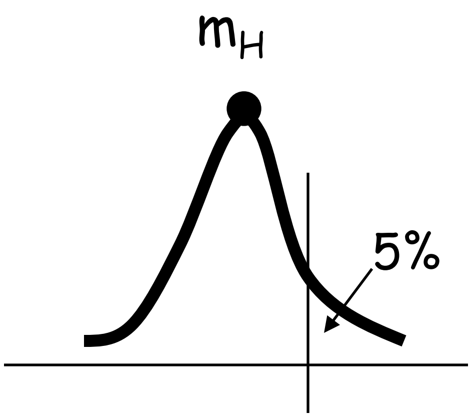
Conclusions

The broad minimum of the combined LEP likelihood from $m_H \sim 115 - 118$ GeV which crosses the expectation for $s+b$ around $m_H \sim 116$ GeV can be interpreted as a preference for a Standard Model Higgs boson at this mass range, however, at less than the 2σ level. When the LEP Higgs working group presented these results for the first time the significance was 2.9σ [1], and this relatively high significance generated a storm which unfortunately turned out to be in a tea cup...

The ALEPH observed likelihood has a 3σ signal-like behavior around $m_H \sim 114$ GeV, which led the collaboration to claim a possible observation of a SM Higgs boson [3]. This behavior originated mainly from the 4-jet channel and its significance is reduced when all experiments are combined. No other experiment or channel indicated a signal-like behavior.



ALEPH
DELPHI
L3
OPAL
2003



$m_H \geq 114.4\ GeV/c^2$ CL=95%

Bayesian Hypothesis test

1. the Bayesian refuses the concept of an ideal ensemble of repeated, identical experiments;

2. the probabilities of the errors of I and II kind are then replaced by the probabilities of the hypotheses

	test statistics	parameters
Bayesian	certain	random
frequentist	random	certain

A BIG problem:

$$P(H_0|\text{data}) = \frac{P(\text{data}|H_0)P(H_0)}{\Sigma_i \underbrace{P(\text{data}|H_i)}_{unknown!} P(H_i)}$$

A solution: the Relative belief updating ratio:

$$R = \frac{P(H_0|\text{data})}{P(H_1|\text{data})} = \frac{P(\text{data}|H_0)P(H_0)}{P(\text{data}|H_1)P(H_1)}$$

- the R values ***help*** the model choice, but the choice is subjective!!
- the $P(H_0)$, $P(H_1)$ priors are necessary
- α, β , $1-\beta$ are not calculated

Gravitational Bursts

(P.Astone, G.Pizzella,workshop (2000))

n_c counts are observed in a time T
r_b and r_s are the background and signal frequencies:

$$n_s = r_s T \text{ unknown} \; , \qquad n_b = r_b T \text{ measured}$$

Relative belief updating ratio
with $P(H_0) = P(H_1)$:

$$R(r_s; n_c, r_b, T) = \frac{e^{-(r_s+r_b)T}\,[(r_s+r_b)t]^{n_c}}{e^{-r_b T}\,[r_b T]^{n_c}} = e^{-r_s T}\left(1+\frac{r_s}{r_b}\right)^{n_c}$$

If $n_c = 0$

$$R = e^{-r_s T}$$

depends on the signal frequency only.
Arbitrary Standard Sensitivity Bound:

$$R = e^{-r_s T} = 0.05 \longrightarrow r_s = 2.99 \approx 3$$

Rule: this is the sensitivity of the experiment

Gravitational bursts

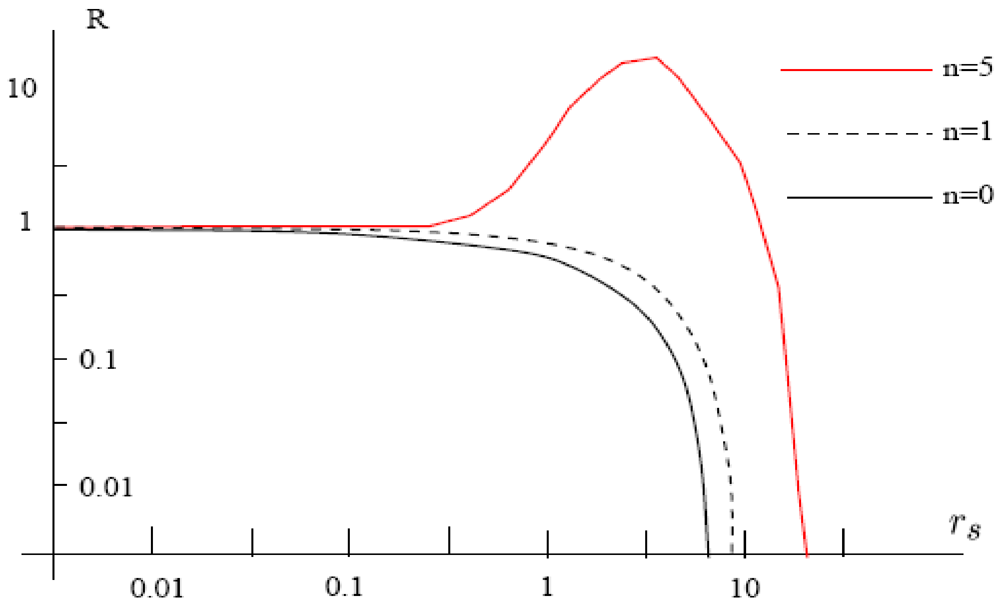


Figure 1: ratio R for the poisson intensity parameter r in units of events per month for an expected background rate $r_b = 1$ event/month and for $n = 0, 1, 5$ observed events

$$e^{-r_s T}\left(1+\frac{r_s}{r_b}\right)^{n_c}, \quad r_b = 1$$

Bayesian Conclusions:

- **If** $r_s < 0.1$ **the data are not relevant;**
- $r_s > 20$ **is excluded by the experiment;**
- **if** n=5 **the most probable hypothesis is** $r_s = 4$

Conclusions

- The maximum likelihood (ML) is the best estimator in the case of parametric statistics problems

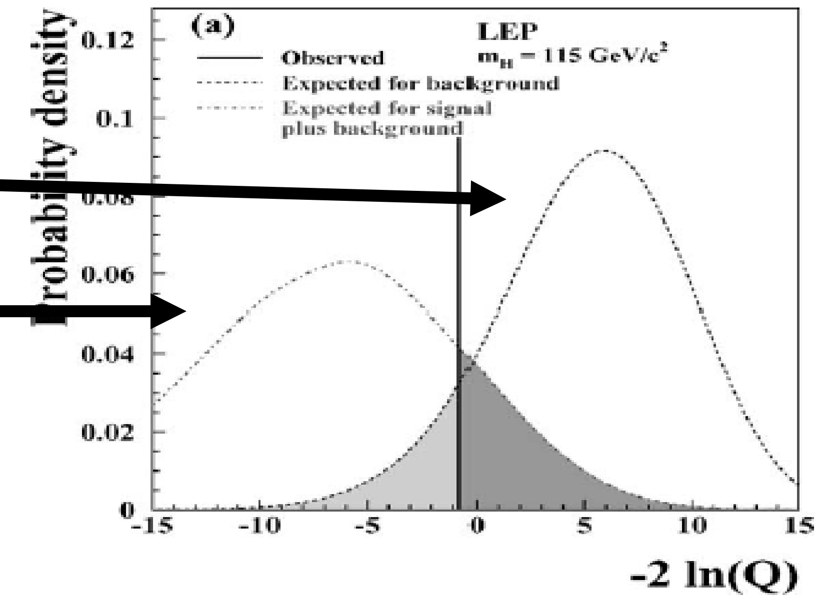
- The likelihood ratio is the maximum power test, that maximize the discovery potential

- The likelihood ratio permits to match toghether different experiments and to realize the Neyman frequentist scheme

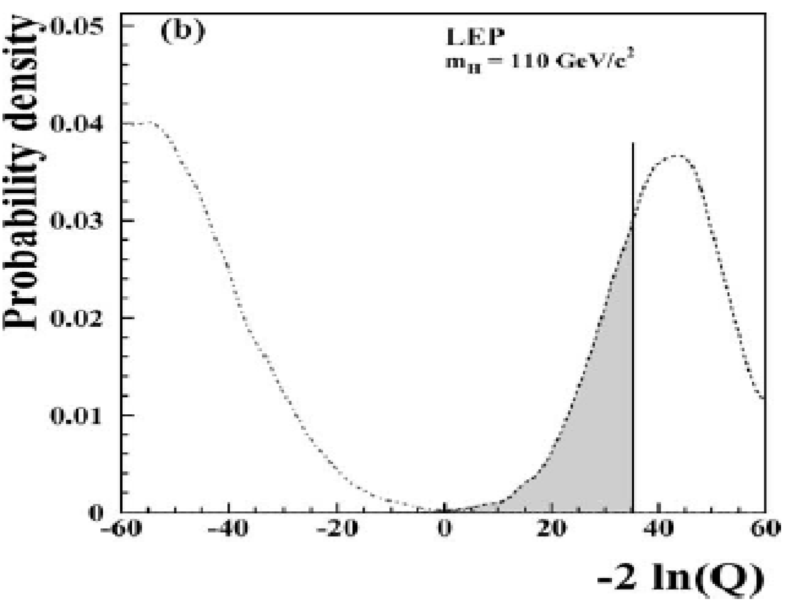
MC samples

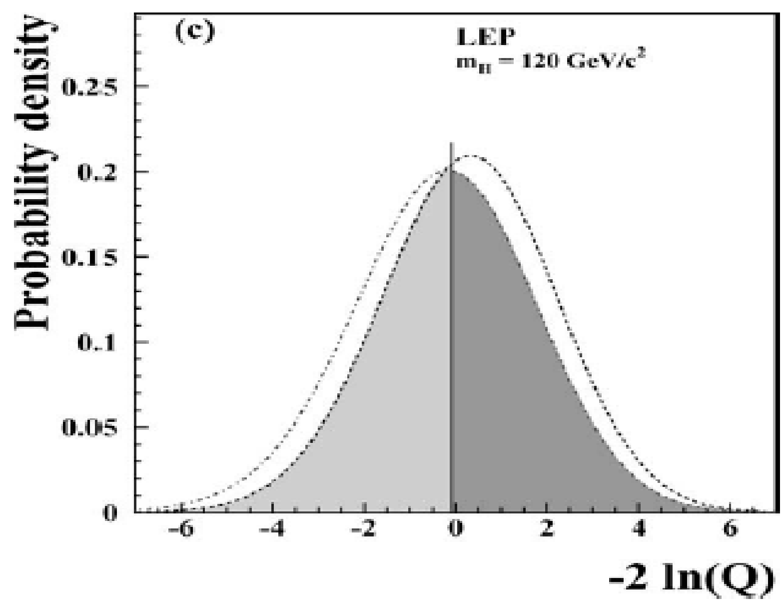
background

signal



With a mass of 116 GeV 10% of the background only experiments give the observed signal





With a Higgs mass of 120 GeV the data are not able to discriminate between the hypotheses

With a Higgs mass of 110 GeV the data are consistent with the background only hypothesis

From the n values x_i of a Gaussian variable, find the ML estimate of mean and variance

Likelihood function:

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2} .$$

The log-likelihood:

$$\mathcal{L}(\mu, \sigma) = +\frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 ,$$

Put the derivative =0:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\Longrightarrow \quad \hat{\mu} = \sum_{i=1}^{n}\frac{x_i}{n} \equiv m$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\Longrightarrow \quad \hat{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i - m)^2}{n}$$

LEP real data

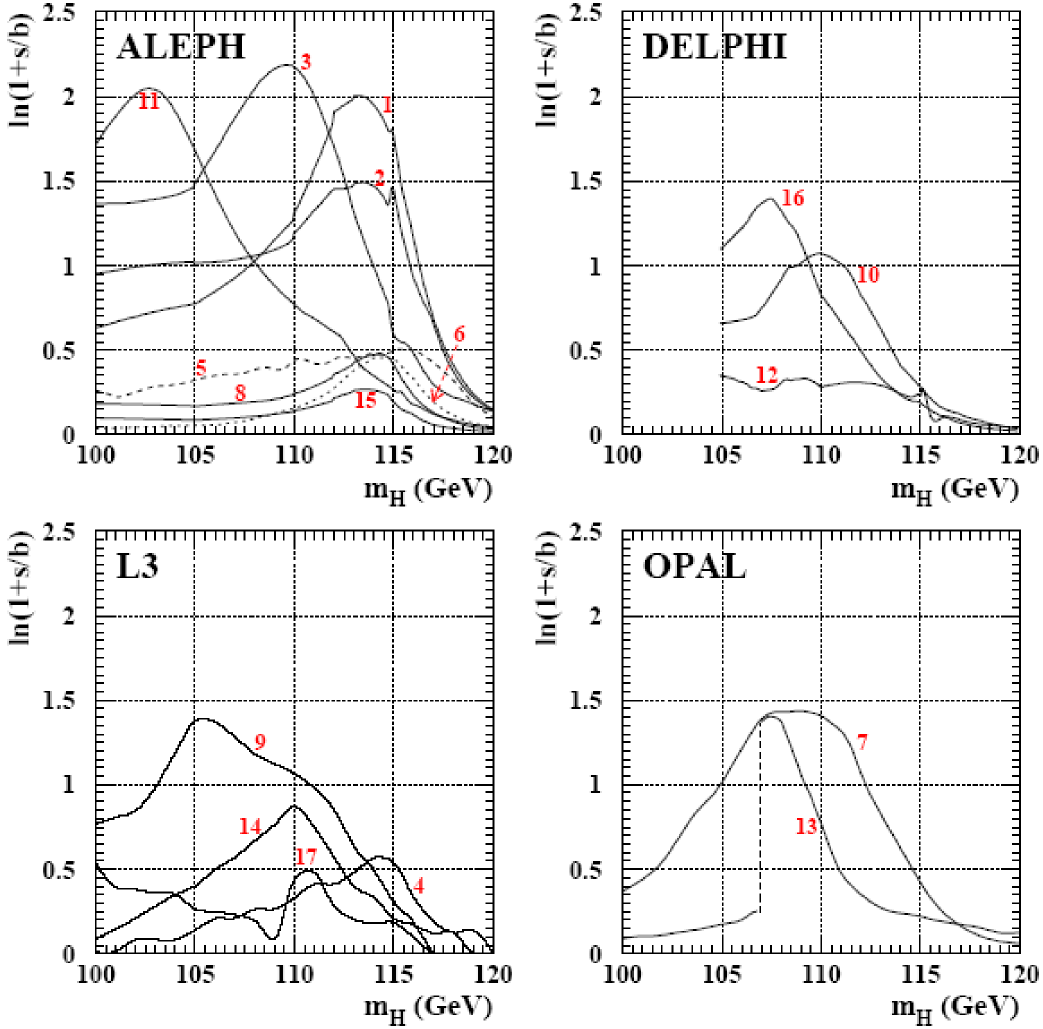


Figure 3: Evolution of the event weight $\ln(1 + s/b)$ with test-mass m_H for the events with the largest weight at $m_H = 115$ GeV. The labels correspond to the candidate numbers in the first column of Table 1. The sudden increase in the weight of the OPAL missing-energy candidate labeled "13" at $m_H = 107$ GeV is due to the switching from the low-mass to high-mass optimization of the search at that mass. A similar increase is observed in the case of the L3 four-jet candidate labeled "17" which is due to a test-mass dependent attribution of the jet-pairs to the Z and Higgs bosons.The Figure is taken from [2].

Signal over Background in Physics

How to count

Some case studies

The case of Pentaquark

The pentaquark is a baryon with five valence quarks.
The clearest signature is that of a

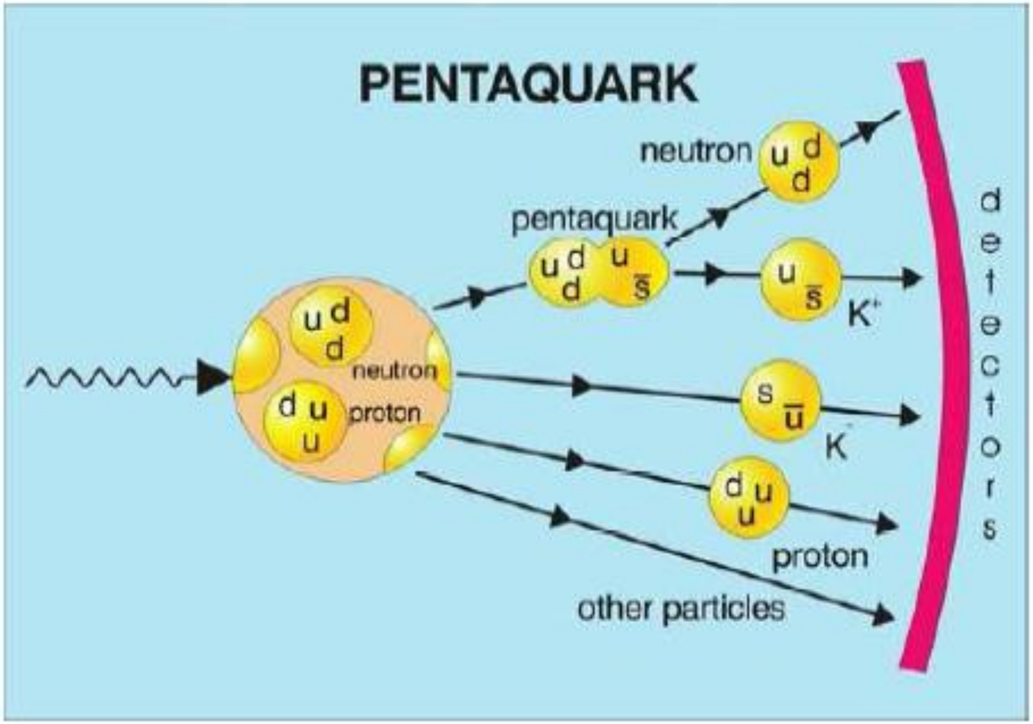
$$u\,u\,d\,d\,\bar{s}\;, \qquad S = +1$$

pentaquark, the unique baryon with positive strangeness.

The $\bar{s}$ antiquark cannot annihilate with the u or d quark by the strong interaction.
Some models predict a mass around 1.5 GeV and a very small width ($\simeq 0.015$ GeV)

The recent pentaquark saga began at 2002 PANIC conference when Nakano measured the following reaction on a Carbon nucleus

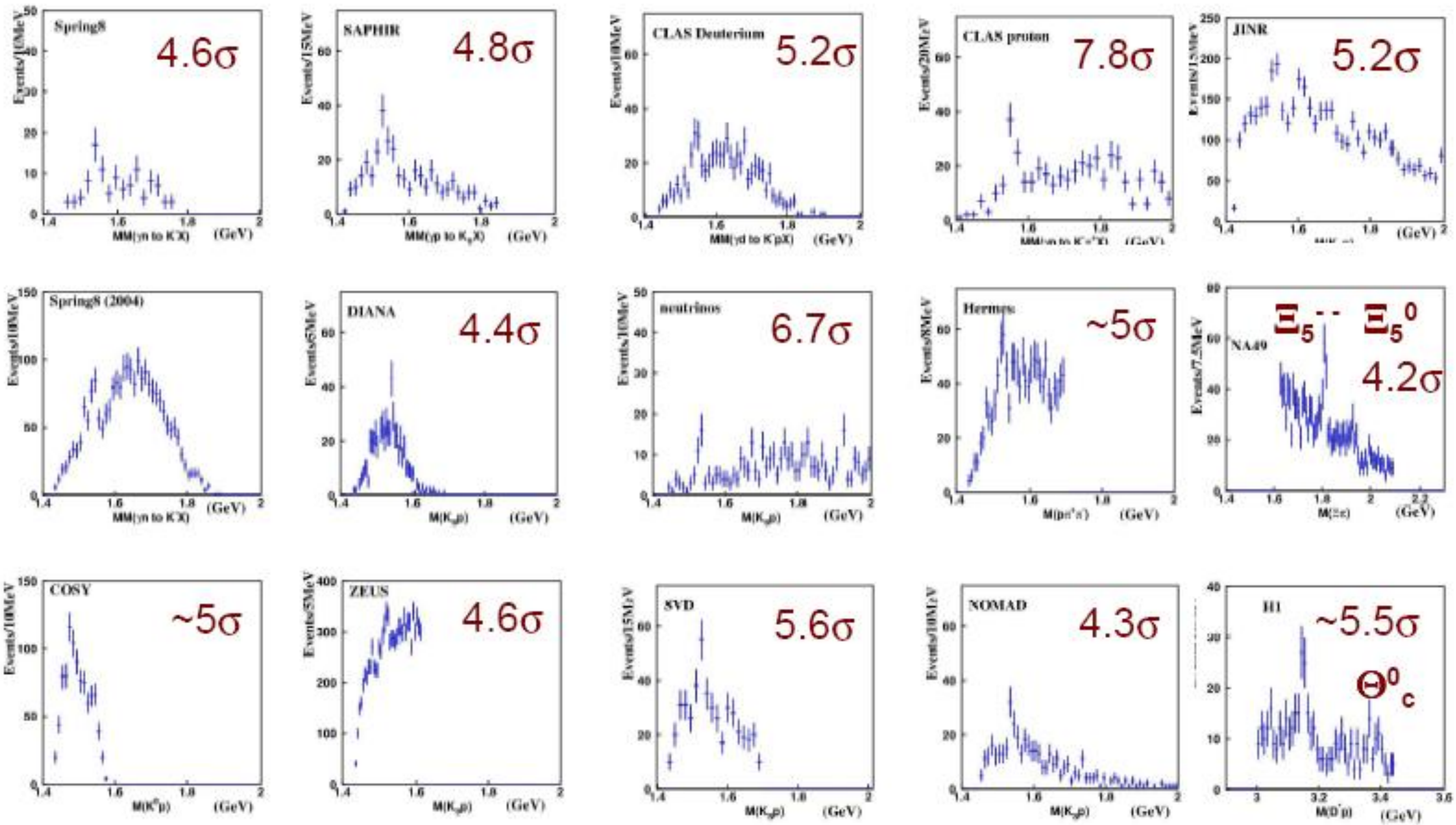
$$\gamma n \to \Theta^+ K^- \to K^+ K^- n$$



Before 2003 …. searches for flavor exotic baryons showed no evidence for such states.

1997: Diakonov, Petrov and Polykov use a chiral soliton model to predict a decuplet of pentaquark baryons. The lightest has S=+1 and a mass of 1530 MeV and expected to be narrow.
Zeit. Phys. A359, 305 (1997).

.. From the Curtis Meyer review (Miami 2004)



$\gamma p \rightarrow K^{+}\Lambda(1250) \rightarrow K^{+}K^{-}p$

$\gamma n \rightarrow K^{-}\Theta \rightarrow K^{-}K^{+}n$

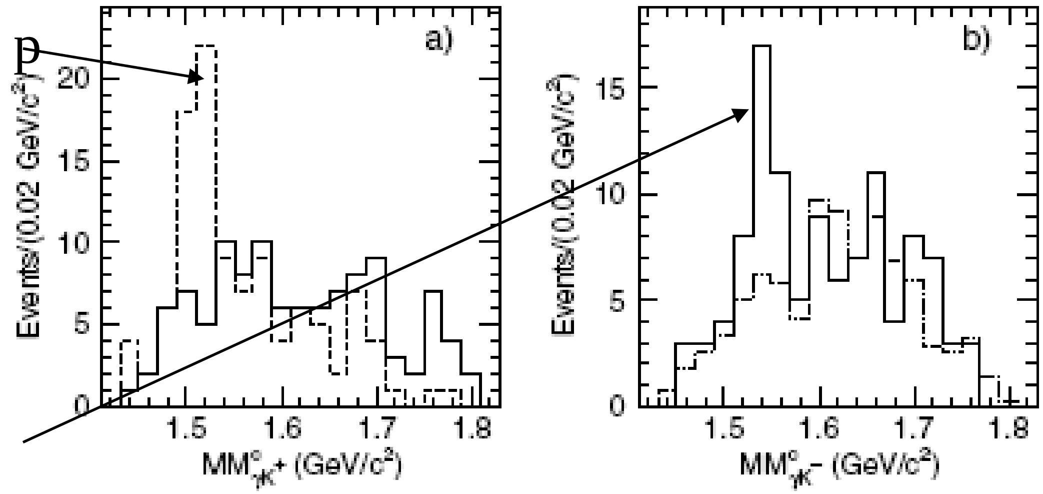


FIG. 3. (a) The $MM^{c}_{\gamma K^{+}}$ spectrum [Eq. (2)] for $K^{+}K^{-}$ productions for the signal sample (solid histogram) and for events from the SC with a proton hit in the SSD (dashed histogram). (b) The $MM^{c}_{\gamma K^{-}}$ spectrum for the signal sample (solid histogram) and for events from the LH_2 (dotted histogram) normalized by a fit in the region above 1.59 GeV/c^2.

The first result

PRL 91(2003)012002

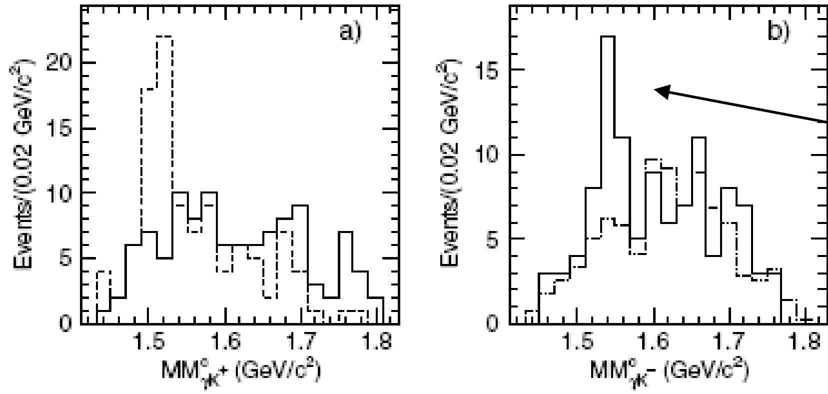
$$\left(\sum_{in} E_{in} - \sum_{fin} E_{fin}\right)^2 - \left(\sum_{in} \vec{p}_{in} - \sum_{fin} \vec{p}_{fin}\right)^2$$

The neutron presence was detected by the $MM_{\gamma K^{+}K^{-}}$ missing mass

The $\gamma p \rightarrow K^{+}K^{-}p$ reaction was eliminated by direct proton detection.

The neutron was reconstructed from the missing momentum and energy of K^{+} and K^{-}.

The background was measured from a LH_2 target.



4.6 sigma!

Is it convincing???

FIG. 3. (a) The $MM^c_{\gamma K^+}$ spectrum [Eq. (2)] for K^+K^- productions for the signal sample (solid histogram) and for events from the SC with a proton hit in the SSD (dashed histogram). (b) The $MM^c_{\gamma K^-}$ spectrum for the signal sample (solid histogram) and for events from the LH_2 (dotted histogram) normalized by a fit in the region above 1.59 GeV/c^2.

012002-3

The background level in the peak region is estimated to be $17.0 \pm 2.2 \pm 1.8$, where the first uncertainty is the error in the fitting in the region above 1.59 GeV/c^2 and the second is a statistical uncertainty in the peak region. The combined uncertainty of the background level is ± 2.8. The estimated number of the events above the background level is 19.0 ± 2.8, which corresponds to a Gaussian significance of $4.6^{+1.2}_{-1.0}\sigma$ $(19.0/\sqrt{17.0} = 4.6)$.

The signal over background

There are two way to count in Physics experiments

- **Poissonian counting**
  The samples are collected in runs of fixed time. The background is evaluated with MC methods, with ***blank*** runs, with ***sideband counting***, etc

- **Binomial counting** The runs collect a ***total number*** N_t of events and N_y of them pass the selection cuts (tagging) or the triggers.
  Signal and background have different probabilities to pass these cuts

To avoid mistakes the notation is very important

- N counts considered as a random variable
- n counts considered as the result of an experiment
- μ expected value of the counting distribution (Binomial or Poissonian).

Poissonian counting
Fundamental theorem

Let's count a Poisson variable N with mean λ with a detector of efficiency ε. The registered number of counts n follows the distribution

$$P(n|N)P(N) = \frac{e^{-\lambda}\lambda^N}{N!} \, \frac{N!}{n!(N-n)!} \varepsilon^n (1-\varepsilon)^{N-n}$$

By using the new variables

$$e^{-\lambda} = e^{-\lambda\varepsilon} e^{-\lambda(1-\varepsilon)}$$

$$m = N - n$$

$$\lambda^N = \lambda^{N-n}\lambda^n \equiv \lambda^m \lambda^n$$

one has

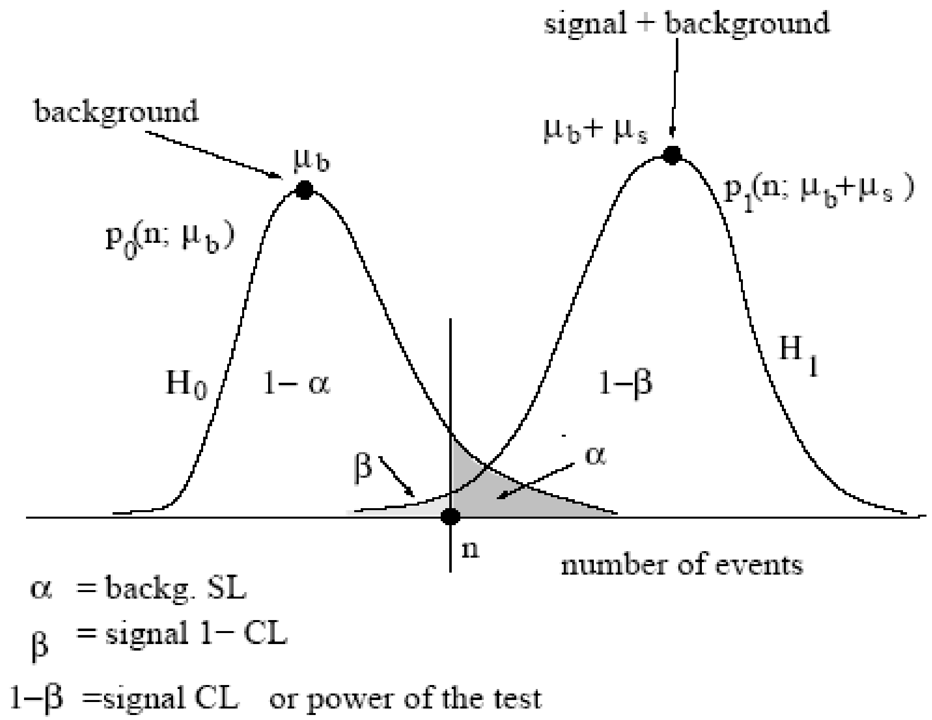
$$P(n|N)P(N) = \frac{e^{-\lambda\varepsilon}(\lambda\varepsilon)^n}{n!} \, \frac{e^{-\lambda(1-\varepsilon)}\lambda^m (1-\varepsilon)^m}{m!}$$

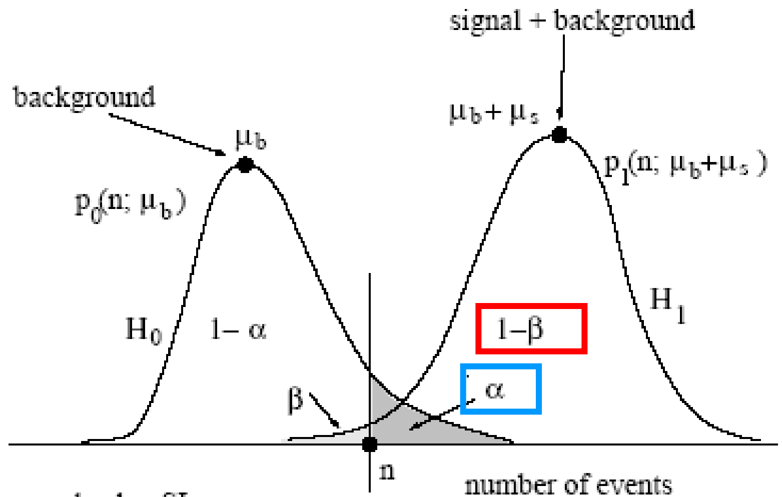
The number of counts n is still an independent Poisson variable with mean $\lambda\varepsilon$!
(also the lost counts m with mean $\lambda(1-\varepsilon)$)

$\{N = n\}$ **events are observed**, **that are supposed to come from a distribution with expected value** $\mu_b + \mu_s$, **where the expected amount of signal** μ_s **is unknown.**

$$p(n, \mu_b) = \frac{\mu_b^n e^{-\mu_b}}{n!} \quad (1)$$

$$p(n, \mu_b + \mu_s) = \frac{(\mu_b + \mu_s)^n}{n!} e^{-\mu_b + \mu_s} \quad (2)$$





α = backg. SL

β = signal $1-$ CL

$1-\beta$ = signal CL or power of the test

$$\alpha \leq 2.8 \cdot 10^{-7} \quad 5\sigma \text{ discovery}$$
$$\alpha \leq 1.3 \cdot 10^{-3} \quad 3\sigma \text{ strong evidence}$$
$$\alpha \leq 2.3 \cdot 10^{-2} \quad 2\sigma \text{ weak evidence}$$

true Hypothesis	Decision H_0	Decision H_1
H_0	correct decision	Type I error
	$1-\alpha$	α
background	**good rejection**	**false acceptance**
H_1	Type II error	Correct decision
signal +	β	$1-\beta$
background	**false exclusion**	**good acceptance**

Discovery Probability or Discovery Potential (DP): the power $1-\beta$ when the critical value n is decided *before* the measurement and when $p(n;\mu_b+\mu_s)$ is *true*.

Poissonian Signal detection

There are many formulas used for detecting a signal over the background (3σ, 5σ, 6σ, and so on)
$N = N_s + N_b$ are the registered counts

$$\frac{x-\mu}{\sigma} = \frac{N - N_b - 0}{\sqrt{N + N_b}}$$

$$S_0 = \frac{N - N_b}{\sqrt{N + N_b}} = \frac{N_b + N_s - N_b}{\sqrt{N + N_b}} = \frac{N_s}{\sqrt{N + N_b}}$$

$$S_b = \frac{N - \mu_b}{\sqrt{\mu_b}} = \frac{N_b + N_s - \mu_b}{\sqrt{\mu_b}} \simeq \frac{N_s}{\sqrt{\mu_b}}$$

This is the most common

$$S_s = \frac{N - \mu_b}{\sqrt{\mu_s}} = \frac{N_b + N_s - \mu_b}{\sqrt{\mu_s}} \simeq \frac{N_s}{\sqrt{\mu_s}}$$

Unclear (to me)

$$S_{sb} = \sqrt{N} - \sqrt{\mu_b} = \sqrt{N_s + N_b} - \sqrt{\mu_b}$$

Recently Proposed (hypothesis test)

Please take care of the notation: often μ is exchanged with N_b and so on, the formulae are obscure and used improperly!!

PRL 91(2003)012002

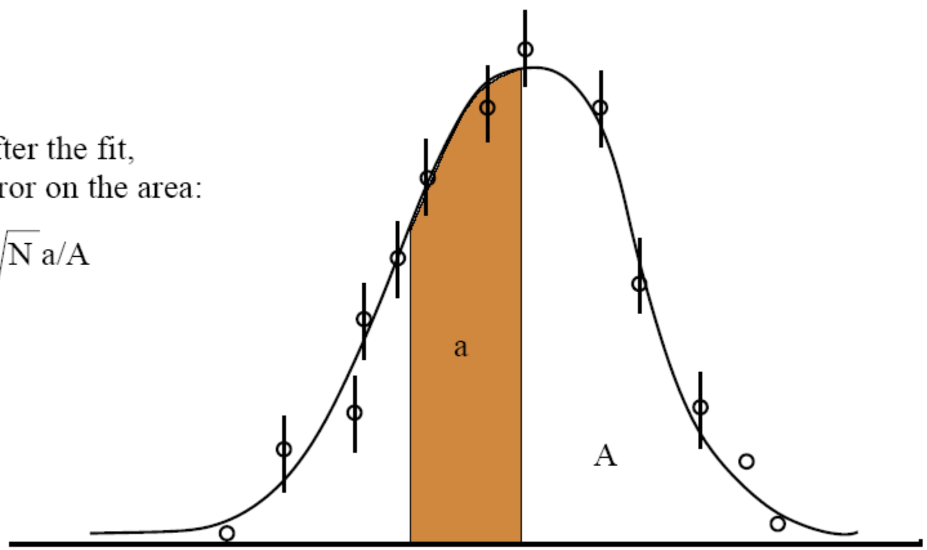
The background level in the peak region is estimated to be $17.0 \pm 2.2 \pm 1.8$, where the first uncertainty is the error in the fitting in the region above 1.59 GeV$/c^2$ and the second is a statistical uncertainty in the peak region. The combined uncertainty of the background level is ± 2.8. The estimated number of the events above the background level is 19.0 ± 2.8, which corresponds to a Gaussian significance of $4.6^{+1.2}_{-1.0}\sigma$ $(19.0/\sqrt{17.0} = 4.6)$.

$$\frac{19}{\sqrt{19+17+17}} = 2.6$$

$$\frac{19}{\sqrt{19+17+2.8^2}} = 2.9$$

after the fit,
error on the area:

$\sqrt{N}\, a/A$



Observation of an Exotic Baryon with $S = +1$ in Photoproduction from the Proton

V. Kubarovsky,[1,3] L. Guo,[2] D. P. Weygand,[3] P. Stoler,[1] M. Battaglieri,[18] R. DeVita,[18] G. Adams,[1] Ji Li,[1] M. Nozar,[3]
C. Salgado,[26] P. Ambrozewicz,[13] E. Anciant,[5] M. Anghinolfi,[18] B. Asavapibhop,[24] G. Audit,[5] T. Auger,[5] H. Avakian,[3]
H. Bagdasarvan,[28] J. P. Ball,[4] S. Barrow,[14] K. Beard,[21] M. Bektasoglu,[27] M. Bellis,[1] N. Benmouna,[15] B. L. Berman,[15]
C. S. Whisnant,[32] E. Wolin,[3] M. H. Wood,[32] A. Yegneswaran,[3] and J. Yun[28]

VOLUME 92, NUMBER 3 PHYSICAL REV

(CLAS Collaboration)

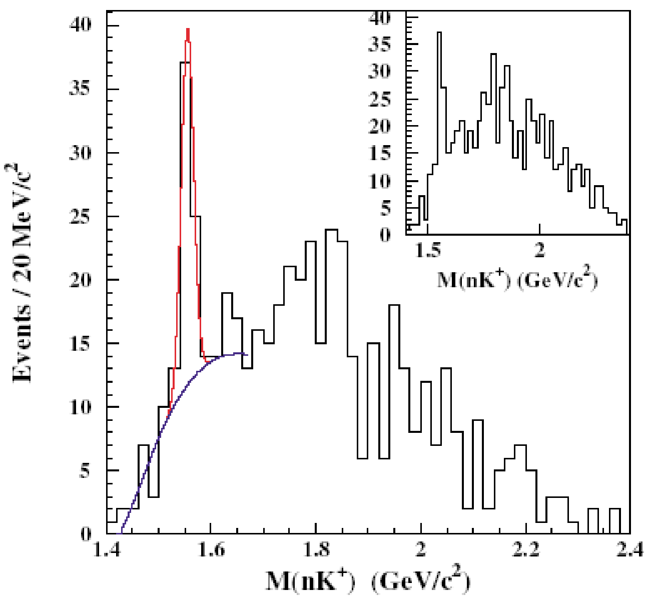
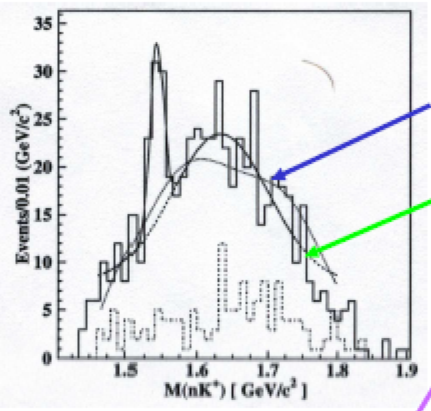


FIG. 4 (color online). The nK^+ invariant mass spectrum in the reaction $\gamma p \to \pi^+ K^- K^+ (n)$ with the cut $\cos\theta^*_{\pi^+} > 0.8$ and $\cos\theta^*_{K^+} < 0.6$. $\theta^*_{\pi^+}$ and $\theta^*_{K^+}$ are the angles between the π^+ and K^+ mesons and photon beam in the center-of-mass system. The background function we used in the fit was obtained from the simulation. The inset shows the nK^+ invariant mass spectrum with only the $\cos\theta^*_{\pi^+} > 0.8$ cut.

The final nK^+ effective mass distribution (Fig. 4) was fitted by the sum of a Gaussian function and a background function obtained from the simulation. The fit parameters are $N_{\Theta^+} = 41 \pm 10$, $M = 1555 \pm 1$ MeV$/c^2$, and $\Gamma = 26 \pm 7$ MeV$/c^2$ (FWHM), where the errors are statistical. The systematic mass scale uncertainty is estimated to be ± 10 MeV$/c^2$. This uncertainty is larger than our previously reported uncertainty [6] because of the different energy range and running conditions and is mainly due to the momentum calibration of the CLAS detector and the photon beam energy calibration. The statistical significance for the fit in Fig. 4 over a 40 MeV$/c^2$ mass window is calculated as $N_P/\sqrt{N_B}$, where N_B is the number of counts in the background fit under the peak and N_P is the number of counts in the peak. We estimate the significance to be $7.8 \pm 1.0\sigma$. The uncertainty of 1.0σ is due to

Statistical Fluctuation

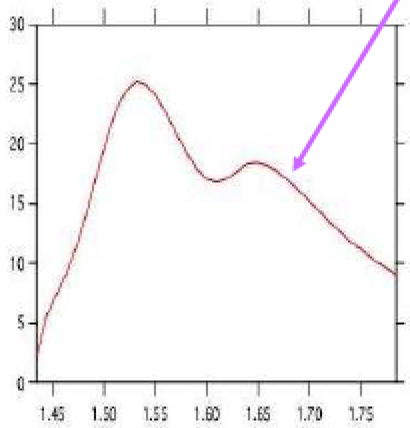
CLAS Published



Simple Physics Background

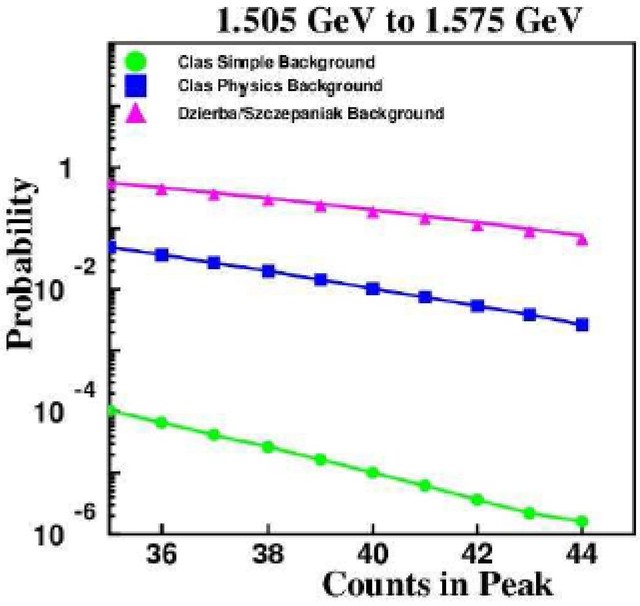
Naïve Background

Dzierba Background



You need to understand your background to claim a new discovery!

Chance of the Background Fluctuating into the observed signal



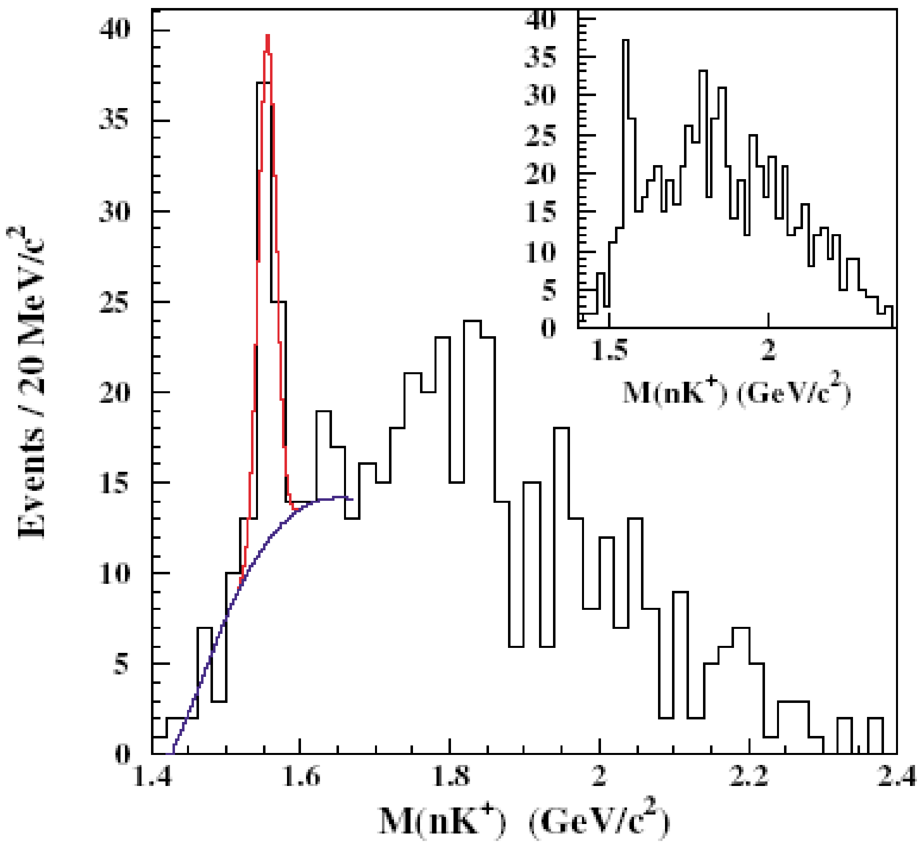


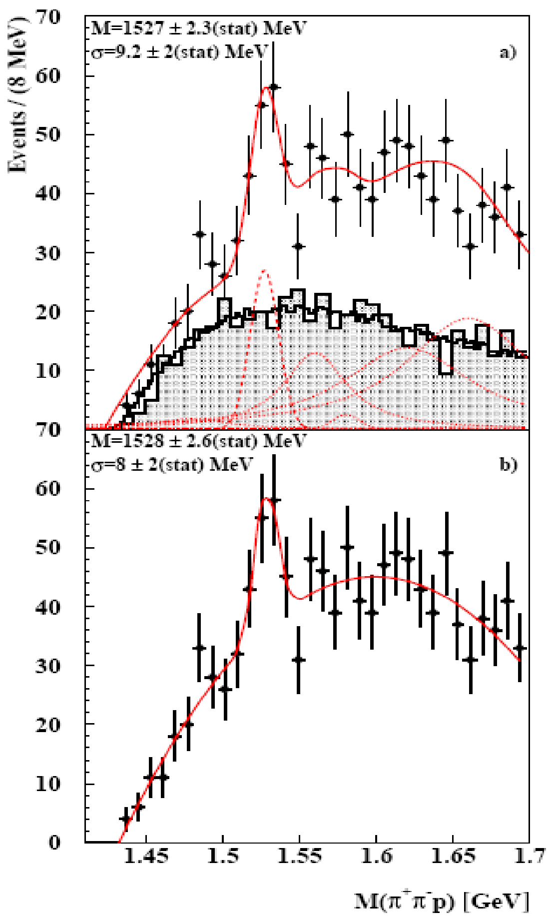
FIG. 4 (color online). The nK^+ invariant mass spectrum in the reaction $\gamma p \to \pi^+ K^- K^+ (n)$ with the cut $\cos\theta^*_{\pi^+} > 0.8$ and $\cos\theta^*_{K^+} < 0.6$. $\theta^*_{\pi^+}$ and $\theta^*_{K^+}$ are the angles between the π^+ and K^+ mesons and photon beam in the center-of-mass system. The background function we used in the fit was obtained from the simulation. The inset shows the nK^+ invariant mass spectrum with only the $\cos\theta^*_{\pi^+} > 0.8$ cut.

$$\frac{41}{\sqrt{27}} = 7.8 \quad ???$$

$$\frac{41}{\sqrt{41+27}} = 5$$

Evidence for a narrow |S| = 1 baryon state at a mass of 1528 MeV in quasi-real photoproduction

A. Airapetian,[31] N. Akopov,[31] Z. Akopov,[31] M. Amarian,[8,31] V.V. Ammosov,[23] A. Andrus,[16] E.C. Aschenauer,[8]
W. Augustyniak,[30] R. Avakian,[31] A. Avetissian,[31] E. Avetissian,[12] P. Bailey,[16] D. Balin,[22] V. Baturin,[22]
M. Beckmann,[7] S. Belostotski,[22] S. Bernreuther,[10] N. Bianchi,[12] H.P. Blok,[21,29] H. Böttcher,[8] A. Borissov,[18]
G. Ybeles Smit,[29] Y. Ye,[5] Z. Ye,[5] S. Yen,[27] W. Yu,[4] B. Zihlmann,[21] H. Zohrabian,[31] and P. Zupranski[30]
(The HERMES Collaboration)
[1] Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada



Photoproduction on a deuterium target

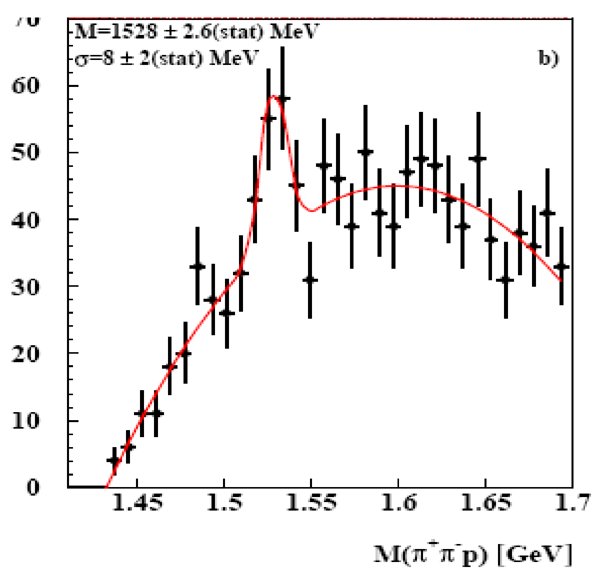
$$\Theta^+ \rightarrow pK_S^0 \rightarrow p\pi^+\pi^-$$

FIG. 2: Distribution in invariant mass of the $p\pi^+\pi^-$ system subject to various constraints described in the text. The experimental data are represented by the filled circles with statistical error bars, while the fitted smooth curves result in the indicated position and σ width of the peak of interest. In panel a), the PYTHIA6 Monte Carlo simulation is represented by the gray shaded histogram, the mixed-event model normalised to the PYTHIA6 simulation is represented by the fine-binned histogram, and the fitted curve is described in the text. In panel b), a fit to the data of a Gaussian plus a third-order polynomial is shown.

HERMES : 27.6 positron beam on deuterium

TABLE I: Mass values and experimental widths, with their statistical and systematic uncertainties, for the Θ^+ from the two fits, labelled by a) and b), shown in the corresponding panels of Fig. 2. Rows a') and b') are based on the same background models as rows a) and b) respectively, but a different mass reconstruction expression that is expected to result in better resolution. Also shown are the number of events in the peak N_s and the background N_b, both evaluated from the functions fitted to the mass distribution, and the results for the naïve significance $N_s^{2\sigma}/\sqrt{N_b^{2\sigma}}$ and realistic significance $N_s/\delta N_s$. The systematic uncertainties are common (correlated) between rows of the table.

	Θ^+ mass [MeV]	FWHM [MeV]	$N_s^{2\sigma}$ in $\pm 2\sigma$	$N_b^{2\sigma}$ in $\pm 2\sigma$	naïve signif.	Total $N_s \pm \delta N_s$	signif.
a)	$1527.0 \pm 2.3 \pm 2.1$	$22 \pm 5 \pm 2$	74	145	6.1σ	78 ± 18	4.3σ
a')	$1527.0 \pm 2.5 \pm 2.1$	$24 \pm 5 \pm 2$	79	158	6.3σ	83 ± 20	4.2σ
b)	$1528.0 \pm 2.6 \pm 2.1$	$19 \pm 5 \pm 2$	56	144	4.7σ	59 ± 16	3.7σ
b')	$1527.8 \pm 3.0 \pm 2.1$	$20 \pm 5 \pm 2$	52	155	4.2σ	54 ± 16	3.4σ



$$S_b = \frac{N - \mu_b}{\sqrt{\mu_b}} = \frac{N_b + N_s - \mu_b}{\sqrt{\mu_b}} \simeq \frac{N_s}{\sqrt{\mu_b}}$$

$$S_0 = \frac{N - N_b}{\sqrt{N + N_b}} = \frac{N_b + N_s - N_b}{\sqrt{N + N_b}} = \frac{N_s}{\sqrt{N + N_b}}$$

$$74/\sqrt{74+145+74} = 4.3$$

$$S_0 = \frac{N - N_b}{\sqrt{N + N_b}} = \frac{N_b + N_s - N_b}{\sqrt{N + N_b}} = \frac{N_s}{\sqrt{N + N_b}}$$

Several alternative expressions for the significance of the peak observed in Fig. 2 were considered. The first expression is the naïve estimator $N_s^{2\sigma}/\sqrt{N_b^{2\sigma}}$ used in Refs. [11, 12, 13, 14, 15, 16]. The corresponding result is listed in Table I. Because this statistic neglects the uncertainty in the background fit, it overestimates the significance of the peak [29]. A second estimator that was used in the analysis presented in Ref. [20], $N_s^{2\sigma}/\sqrt{N_s^{2\sigma} + N_b^{2\sigma}}$, gives a somewhat lower value, but may still underestimate the background uncertainty. A third estimate of the significance is given by $N_s/\delta N_s$, where N_s is now the full area of the peak from the fit and δN_s is its fully correlated uncertainty. This ratio measures how far the peak is away from zero in units of its own standard deviation. All correlated uncertainties from the fit, including those of the background parameters, are accounted for in δN_s. The results obtained with this expression are also given in Table I.

Statistics

$$\xi_1 = \frac{s}{\sqrt{b}}$$

$$\xi_2 = \frac{s}{\sqrt{s+b}}$$

$$\xi_3 = \frac{s}{\sqrt{s+2b}}$$

Experiment	Signal	Background	Significance			
			Publ.	ξ_1	ξ_2	ξ_3
Spring8	19	17	4.6σ	4.6	3.2	2.6
Spring8	56	162		4.4	3.8	2.9
SPAHIR	55	56	4.8σ	7.3	5.2	4.3
CLAS (d)	43	54	5.2σ	5.9	4.4	3.5
CLAS (p)	41	35	7.8σ	6.9	4.7	3.9
DIANA	29	44	4.4σ	4.4	3.4	2.7
ν	18	9	6.7σ	6.0	3.5	3.0
HERMES	51	150	4.3-6.2σ	4.2	3.6	2.7
COSY	57	95	4-6σ	5.9	4.7	3.7
ZEUS	230	1080	4.6σ	7.0	6.4	4.7
SVD	35	93	5.6σ	3.6	3.1	2.4
NOMAD	33	59	4.3σ	4.3	3.4	2.7
NA49	38	43	4.2σ	5.8	4.2	3.4
NA49	69	75	5.8σ	8.0	5.8	4.7
H1	50.6	51.7	5-6σ	7.0	5.0	4.1

No 5σ effect!!

Poissonian Signal detection

When the background is well known people use

$$S_b = \frac{N - \mu_b}{\sqrt{\mu_b}}$$

Recently Bityukov and Krasnikov (2000) proposed

$$S_{sb} = \sqrt{N} - \sqrt{\mu_b} = \sqrt{N_s + N_b} - \sqrt{\mu_b}$$

Proof: In gaussian approx ($\mu_b > 10$), the abscissa n satisfies the equation

$$t = \frac{n - N_b}{\sqrt{N_b}} = -\frac{n - N_s - N_b}{\sqrt{N_s + N_b}} ,$$

which implies

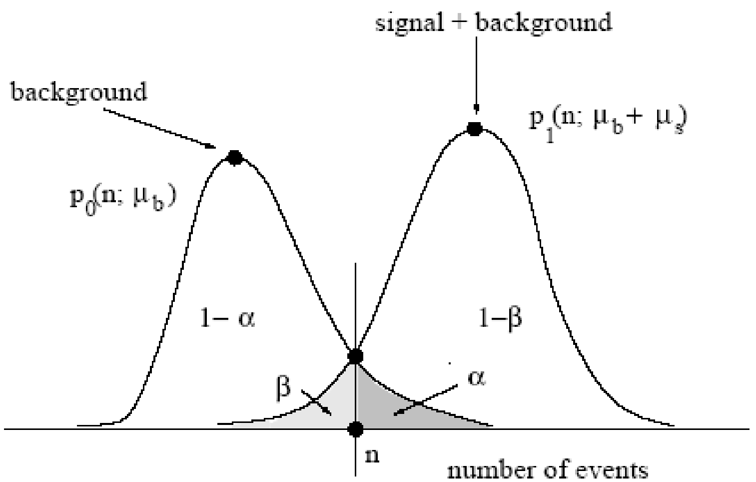
$$n = \sqrt{N_b(N_b + N_s)} \quad \text{and} \quad t = \sqrt{N_b + N_s} - \sqrt{N_b} .$$

Therefore, one can define the statistic

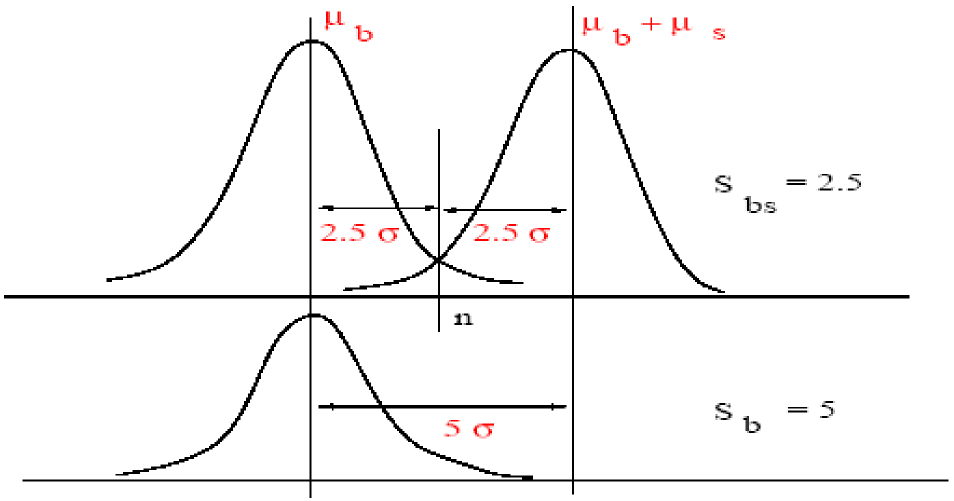
$$S_{bs} = \sqrt{N} - \sqrt{N_b} ,$$

with expectation value

$$\langle S_{bs} \rangle = \sqrt{\mu_b + \mu_s} - \sqrt{\mu_b} ,$$



Poissonian Signal detection



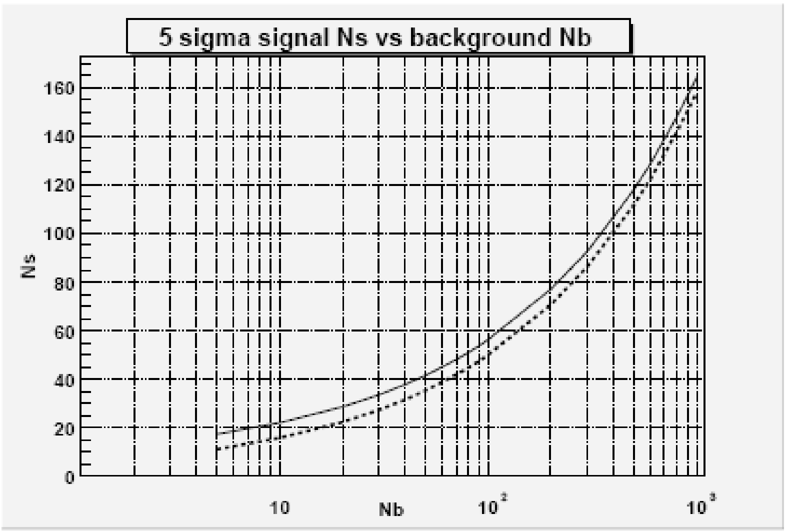
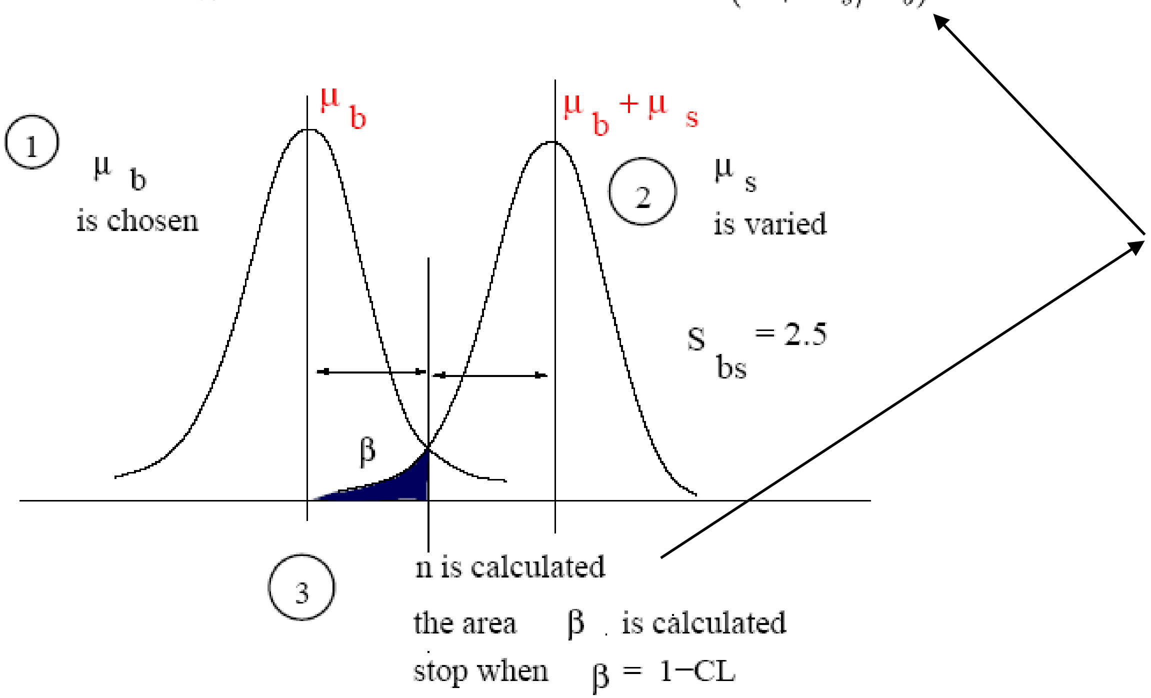


Figure 1: Number N_s of the signal events for $S_b = 5$ (dotted line) and $S_{bs} = 2.5$ (full line) versus the number N_b of background events.

Poissonian Signal Detection

$$\frac{N_b^n}{n!} e^{-N_b} = \frac{(N_b + N_s)^n}{n!} e^{-N_b + N_s} \quad \Longrightarrow \quad n = \frac{N_s}{\ln(1 + N_s/N_b)} .$$



μ_b	n	μ_s	α
10	19	22	0.004
20	32	26	0.005
30	45	33	0.005
60	80	43	0.006
200	234	73	0.008
500	554	111	0.008
x	$\mu_b + 2.33\sqrt{x}$	$2 \times 2.33\sqrt{x}$	0.010

Table 1: When the expected background value is μ_b, if n events are observed, μ_s is the upper limit of the signal intensity with $CL = 99\%$.

Binomial counting: candidate selection

A sample N_t can be considered as an ensemble of signal and background events:

$$N_t = N_s + N_b$$

The measurement is a linear operator $\boldsymbol{M}$ that acts on $N_s + N_b$ and divides this sample into events that pass the selection (the "yes" events N_y) and events that do not pass the selection (the "no" events N_n).

$$N_t = N_s + N_b = N_y + N_n$$

$$\begin{pmatrix} N_y \\ N_n \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} N_s \\ N_b \end{pmatrix}$$

$$\begin{pmatrix} N_y \\ N_n \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} N_s \\ N_b \end{pmatrix}$$

ε is the efficiency on the signal events and b that on the background:

$$N_{ys} = \varepsilon N_s \ , \quad N_{ns} = (1-\varepsilon) N_s$$

$$N_{yb} = b N_b \ , \quad N_{nb} = (1-b) N_b \ ,$$

Since

$$N_y = N_{ys} + N_{yb} = \varepsilon N_s + b N_b \ ,$$

$$N_n = N_{ns} + N_{nb} = (1-\varepsilon) N_s + (1-b) N_b \ ,$$

the M matrix becomes:

$$\boldsymbol{M} = \begin{pmatrix} \varepsilon & b \\ 1-\varepsilon & 1-b \end{pmatrix} \ .$$

The inverse of the measurement matrix is:

$$M^{-1} = \frac{1}{\varepsilon - b} \begin{pmatrix} 1-b & -b \\ \varepsilon - 1 & \varepsilon \end{pmatrix} ,$$

When the knowledge of the ε and b-efficiencies is achieved, one can solve the general Measurement Problem (MP):

having measured N_y and N_n from a sample $N_t = N_y + N_n$, what are N_s and N_b? :

$$N_s = \frac{(1-b)N_y - bN_n}{\varepsilon - b} = \frac{N_y - bN_t}{\varepsilon - b}$$

$$N_b = \frac{(\varepsilon - 1)N_y + \varepsilon N_n}{\varepsilon - b} = \frac{N_n - (1-\varepsilon)N_t}{\varepsilon - b} = \frac{\varepsilon N_t - N_y}{\varepsilon - b}$$

When $\varepsilon \gg b$ and $\varepsilon, b \ll 1$,

$$N_s = \frac{N_y}{\varepsilon} , \quad N_b = N_t - N_s .$$

The errors come from the binomial formula (N_t is not random):

$$\sigma[N_s] = \sigma[N_b] = \frac{1}{\varepsilon - b}\sigma[N_y] = \frac{1}{\varepsilon - b}\sqrt{N_y(1 - N_y/N_t)}$$

When there are more background sources

$$b \to b_{\rm tot} = \sum_i b_i w_i , \quad w_i = \frac{N_{b_i}}{\sum_i N_{b_i}} .$$

Problem: when $\varepsilon \simeq b$ the system is ill-conditioned!

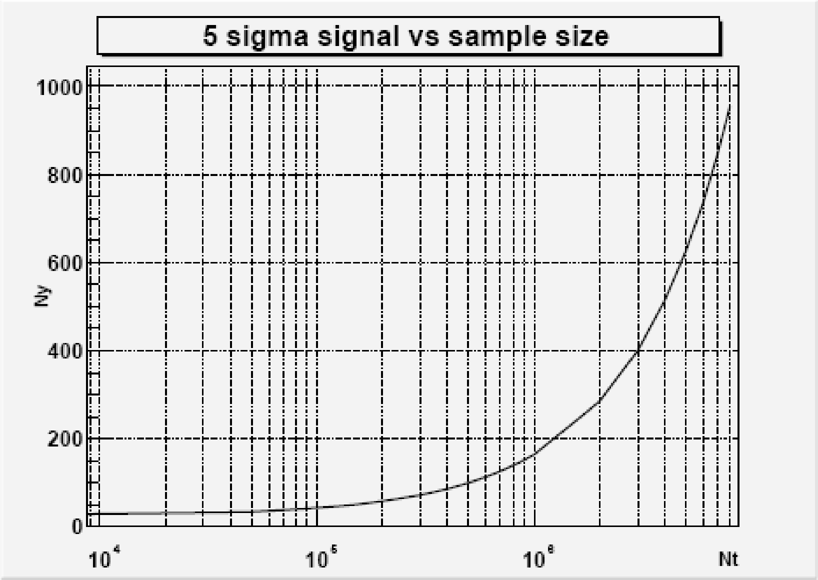


Figure 2: Number N_y of the events that pass the selection at a 5-σ level versus the sample size N_t with a b-efficiency $1 \cdot 10^{-4}$.

$$t = \frac{N_s}{\sigma[N_b]} = \frac{(\varepsilon - b)N_s}{\sqrt{N_y(1 - N_y/N_t)}}$$

$$= \frac{N_y - bN_t}{\sqrt{N_y(1 - N_y/N_t)}} \quad \left(\begin{array}{c} \geq 5 \text{ for the } 5\sigma \text{ discovery} \\ \text{binomial case} \end{array} \right) , (3)$$

If the b-efficiency is deduced from a background sample, it must be considered as a random variable:

$$b = \frac{N_{yb}}{N_t} ,$$

if the signal and background runs have the same N_t:

$$t = \frac{N_s}{\sigma[N_b]} = \frac{N_y - N_{yb}}{\sqrt{N_y(1 - N_y/N_t) + N_{yb}(1 - N_{yb}/N_t)}} \simeq \frac{N_y - N_{yb}}{\sqrt{N_y + N_{yb}}} ,$$

where the last term is the well known S_0 result.

Bayes formula

Having found N_s and N_b, the percentage of signal in the accepted events N_y can be found with the Bayes formula (used in a frequentist way, because $P(S)$ is not subjective)

$$P(S|T) = \frac{P(T|S)\,P(S)}{P(T|S)\,P(S) + P(T|B)\,P(B)}$$

where:

$$\begin{aligned}
P(S) &= N_s/N_t = \text{ percentage of events in the triggered sample} \\
P(B) &= N_b/N_t = \text{ percentage of background in the triggered sample} \\
P(T|S) &= \varepsilon = \text{ probability that a signal event passes the selection} \\
P(T|B) &= b = \text{ probability that a background event passes the selection} \\
P(S|T) &= \text{ probability that a selected event is the signal}
\end{aligned}$$

$$\begin{aligned}
P(S|T) &= \varepsilon\,\frac{N_y - bN_t}{(\varepsilon - b)N_y} = \frac{\varepsilon}{\varepsilon - b}\left(1 - b\,\frac{N_t}{N_y}\right) = \frac{\varepsilon N_s}{N_y} \\
P(B|T) &= 1 - P(S|T)\ ,
\end{aligned}$$

$$\sigma[P(\bar{H}|S)] = \frac{\varepsilon\,b}{\varepsilon - b}\,\frac{N_t}{N_y^2}\,\sqrt{N_y(1 - N_y/N_t)} \simeq \frac{\varepsilon\,b}{\varepsilon - b}\,N_t\,N_y^{-3/2}\ .$$

In summary,

$$P(S|T) \simeq \frac{\varepsilon}{\varepsilon - b}\left(1 - b\,\frac{N_t}{N_y}\right) \pm \frac{\varepsilon\,b}{\varepsilon - b}\,N_t\,N_y^{-3/2}$$

The top quark discovery of CDF

The CDF experiment claimed the op quark discovery (Phys. Rev. Lett.74(1995)2626) with two different selection methods of discriminating the signal

$$t\bar{t} \rightarrow W b W \bar{b}$$

from background:

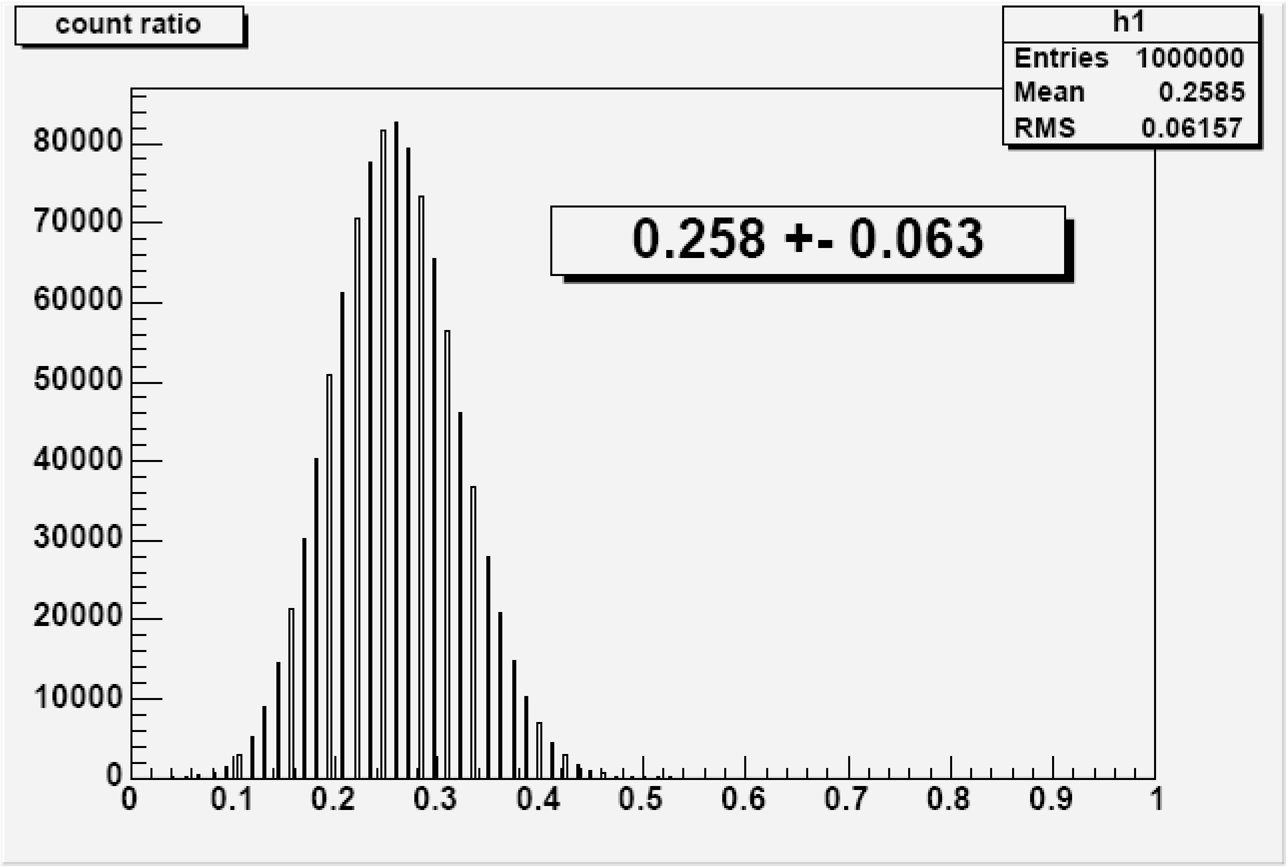
- **SVX tagging: b jets identification by searching for secondary vertices in the Silicon Vertex detector;**
- **SLT tagging: to search for an additional soft lepton from semileptonic b decay**

tag	N_t	N_y	ε %	b %	N_s/N_t	$N_t P(S\|T)$
SVX	203	27	42 ± 5	3.3 ± 0.1	$0.25^{+0.08}_{-0.07}$	$22.5^{+2.3}_{-2.9}$
SLT	203	23	20 ± 2	7.6 ± 0.1	$0.24^{+0.22}_{-0.18}$	$13.2^{+4.6}_{-6.0}$

The error on N_s/N_t from the standard formula is ± 0.06 for SVX and ± 0.18 for SLT, slightly underestimated.
To take into account the uncertainties on the efficiencies (nuisance parameters) a grid MC is necessary

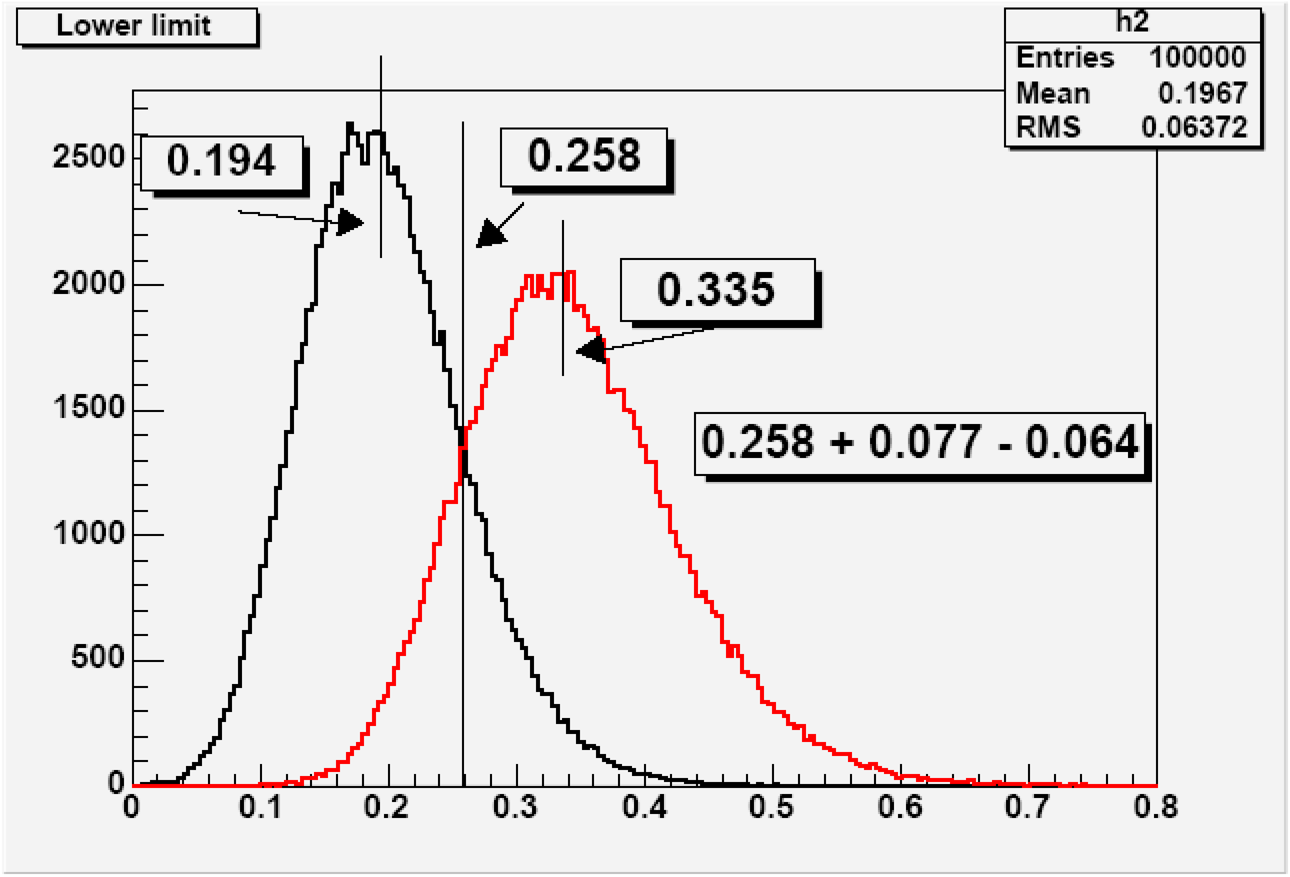
The bootstrap method for confidence levels

With fixed efficiencies we have the binomial/gaussian distribution



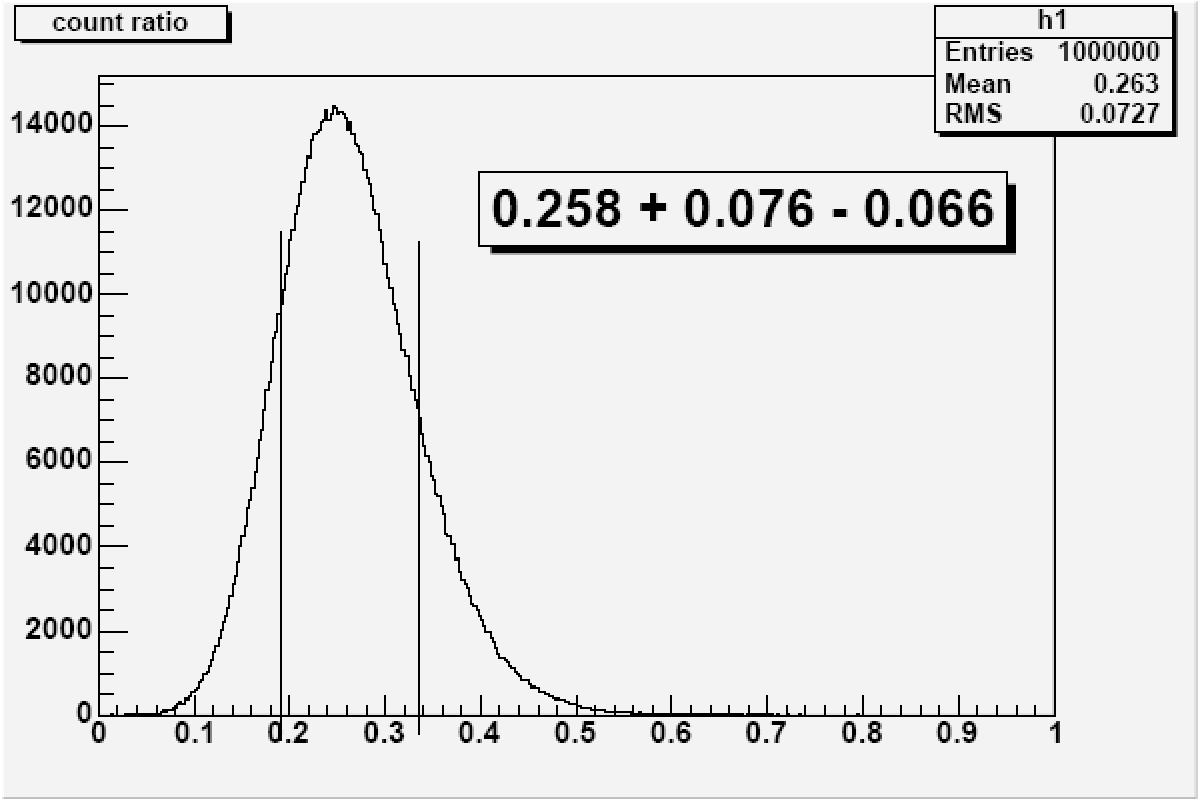
The grid method for confidence levels

For each value of $p = N_s/N_t$ a sample of $100\,000$ events is generated sampling randomly the ε and b efficiencies.



The bootstrap method for confidence levels

In this case also the approximate bootstrap method gives the same result.
This method is called Parametric Bootstrap



The bootstrap method for BR

- A=15 Reaction A events in a N=200 event sample
- B=30 Reaction B events in a N=200 event sample

Standard error propagation for 95% CL

$$\sigma(A/B) = 1.96 \times \frac{A}{B}\sqrt{\frac{\sigma^2(A)}{A^2} + \frac{\sigma^2(B)}{B^2}} = 0.15 \ \rightarrow \ [0.21, 079] \, , \ \ \mathrm{CL} = 95\%$$

$$\sigma(A) = \sqrt{A\left(1 - \frac{A}{N}\right)}$$

Bootstrap methods:

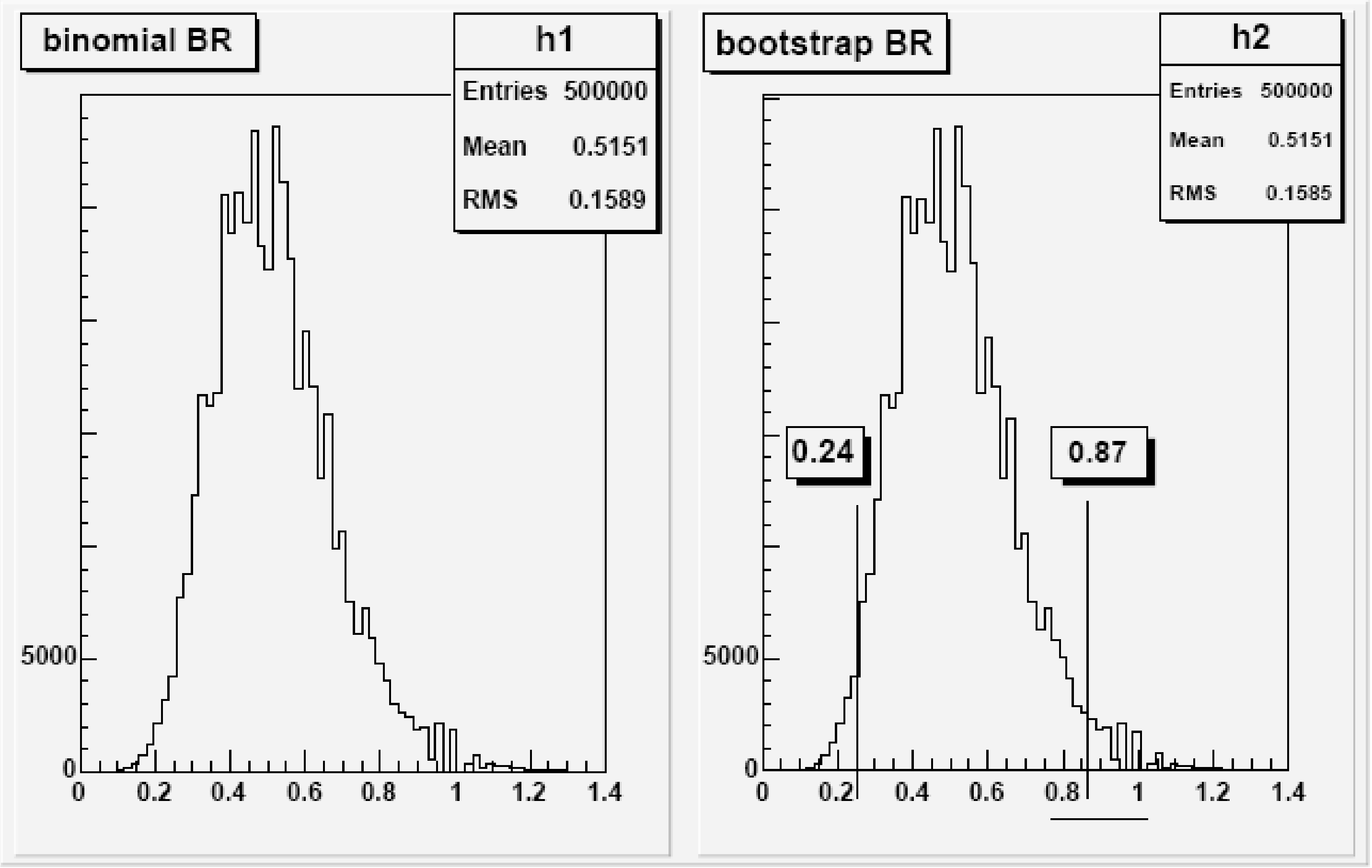
- parametric bootstrap: the events A and B are MC-sampled from two binomial distributions with $N = 200$ and $p_1 = A/N$ and $p_2 = B/N$;
- non parametric bootstrap: the events A and B are sampled with replacement from two experimental samples with $N = 200$ and A or B events $= 1$, the others $= 0$

Obviously, in this case the two methods give the same result:

$$[A/B] \in [0.24, 0, 86] \ , \quad CL = 95\%$$

Are the published BR really all RELIABLE??

The bootstrap method for BR



Consider a sample X containing N objects. We need an estimate of θ as $\hat{\theta}(X)$.
No model of the X distribution is known or considered
Statisticians have elaborated the following (non parametric) methods:

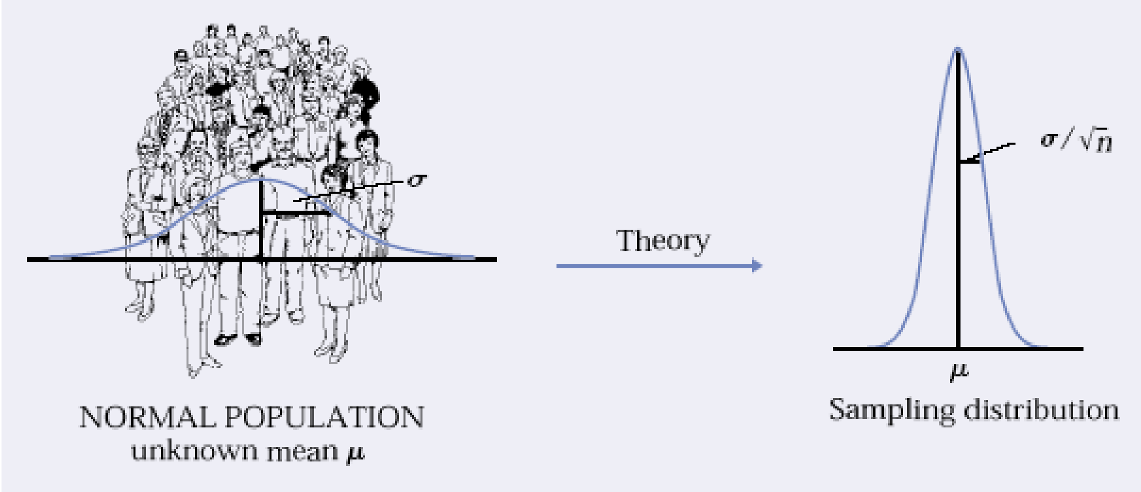
The non parametric Sampling methods

- **Jackknife** (Quenouille, 1949):
  N samples are generated leaving out one element at a time;
- **Subsampling**:
  S resamples of dimension N_B are created by repeatedly sampling **without replacement** from the ex... imental sample. Obviously one has N...
- **Bootstrap** (Efron 1979):
  S resamples of dimension N_B are created by repeatedly sampling **with replacement** from the experimental sample. Usually $N_B = N$ is set.
- **Permutation**:
  used in the test between two hypotheses, by resampling in a way that moves observations between the two groups, under the assumption that the null hypothesis is true

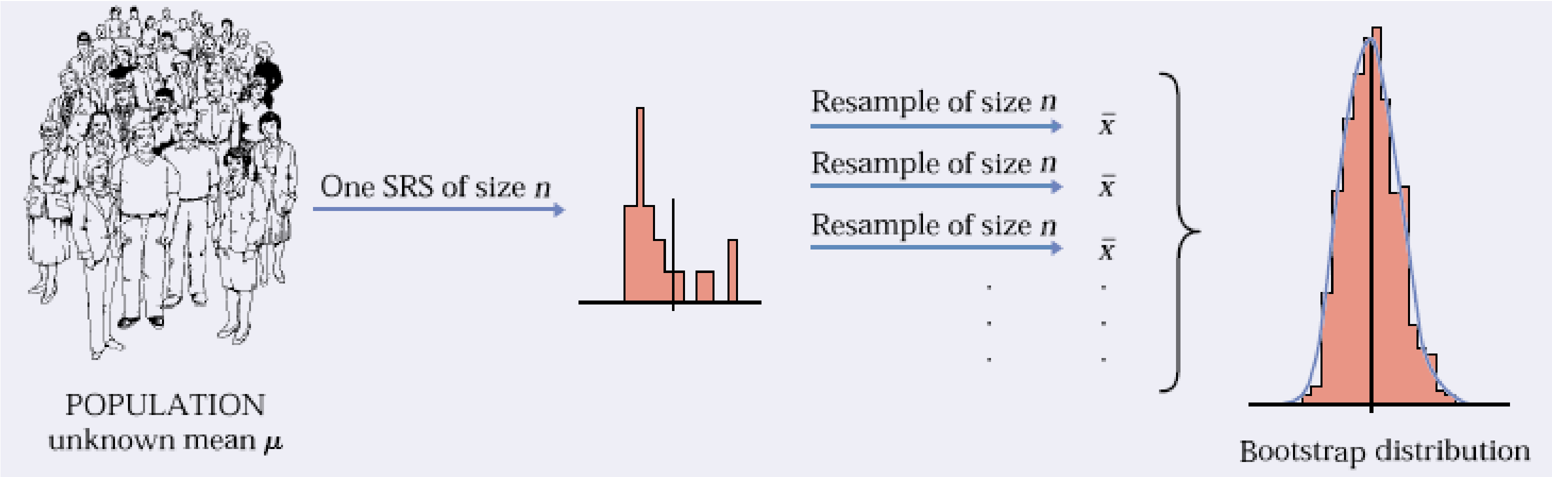
The best one !!!

These methods, familiar among statisticians, are practically not (yet) used by physicists (only 3 papers with non parametric Bootstrap!)

Non parametric Bootstrap



(b)



(c)

The non parametric BOOTSTRAP

Consider a sample X containing N objects. We need an estimate of θ as

$$\hat{\theta}(X)$$

Using the Bootstrap sample, we obtain the estimator

$$\hat{\theta}^* = \hat{\theta}(X^*)$$

The Bootstrap samples have expectation values $\hat{\theta}^*$ that differ from the true one θ (bias), but ...
the Bootstrap approximates the distribution of

$$\hat{\theta} - \theta$$

with the distribution of

$$\hat{\theta}^* - \hat{\theta}$$

obtained by resampling.

Limits of non parametric BOOTSTRAP

Drawback: the Bootstrap samples are correlated.
Some important results on this:

- the sharing of the same elements in different samples reduces the variance s_{res} of the (re)samples:

$$s^2_{\text{res}} \to (1-\rho)\sigma^2$$

  where $\rho = N_B/N$ in subsampling without replacement;

- the sampling with replacement in bootstrap increases the variance of the (re)samples:

$$s^2_{\text{res}} \to (1-\rho)\rho_1\sigma^2$$

- in many cases in the bootstrap the positive bias due to the within sample correlation and the negative bias due to the between sample correlation cancel exactly

$$\sqrt{1-\rho}\sqrt{\rho_1} \simeq 1$$

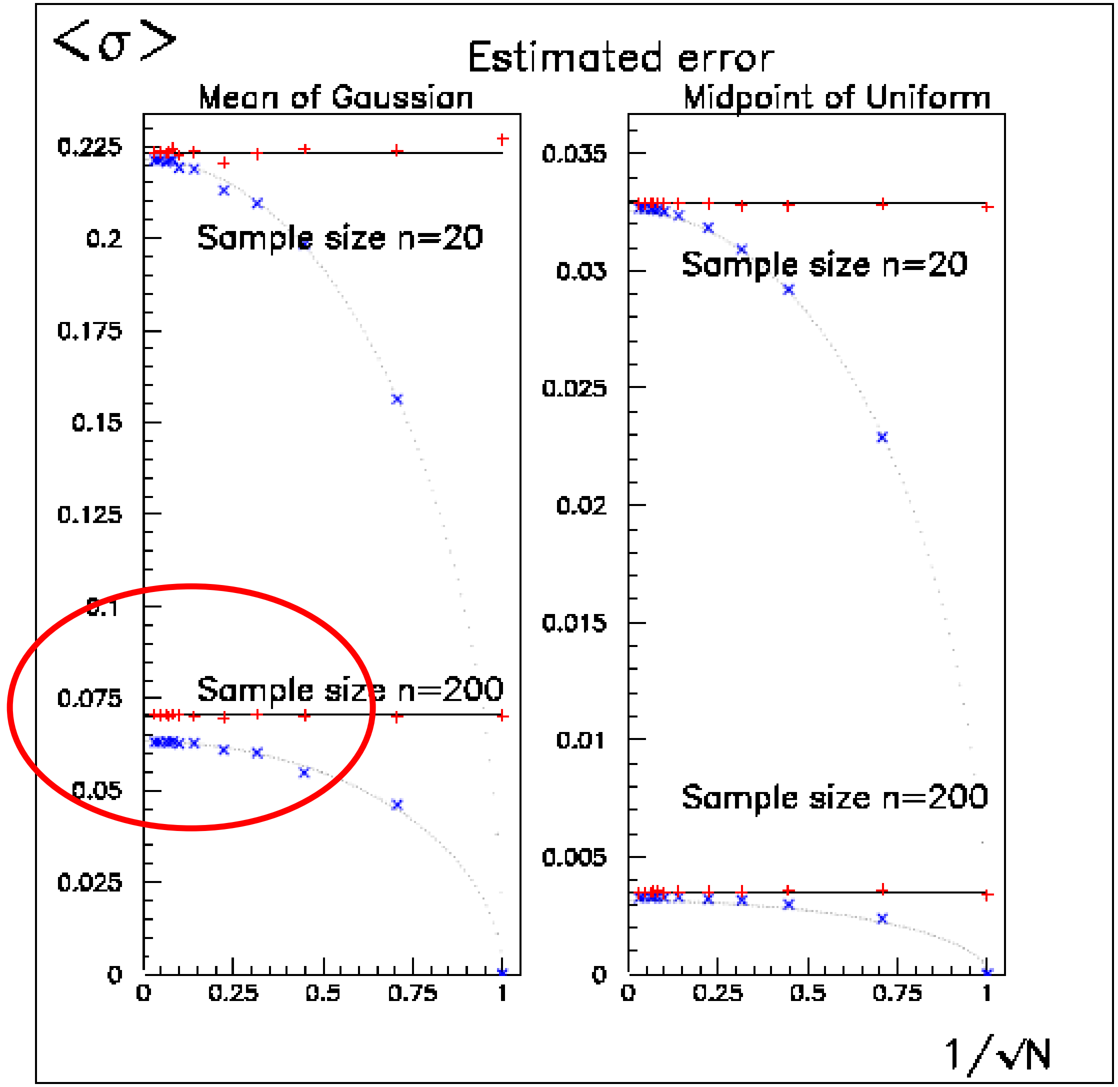


Figure 4: The mean estimated error as a function of the number of samples. Sampling was done 'without replacement'

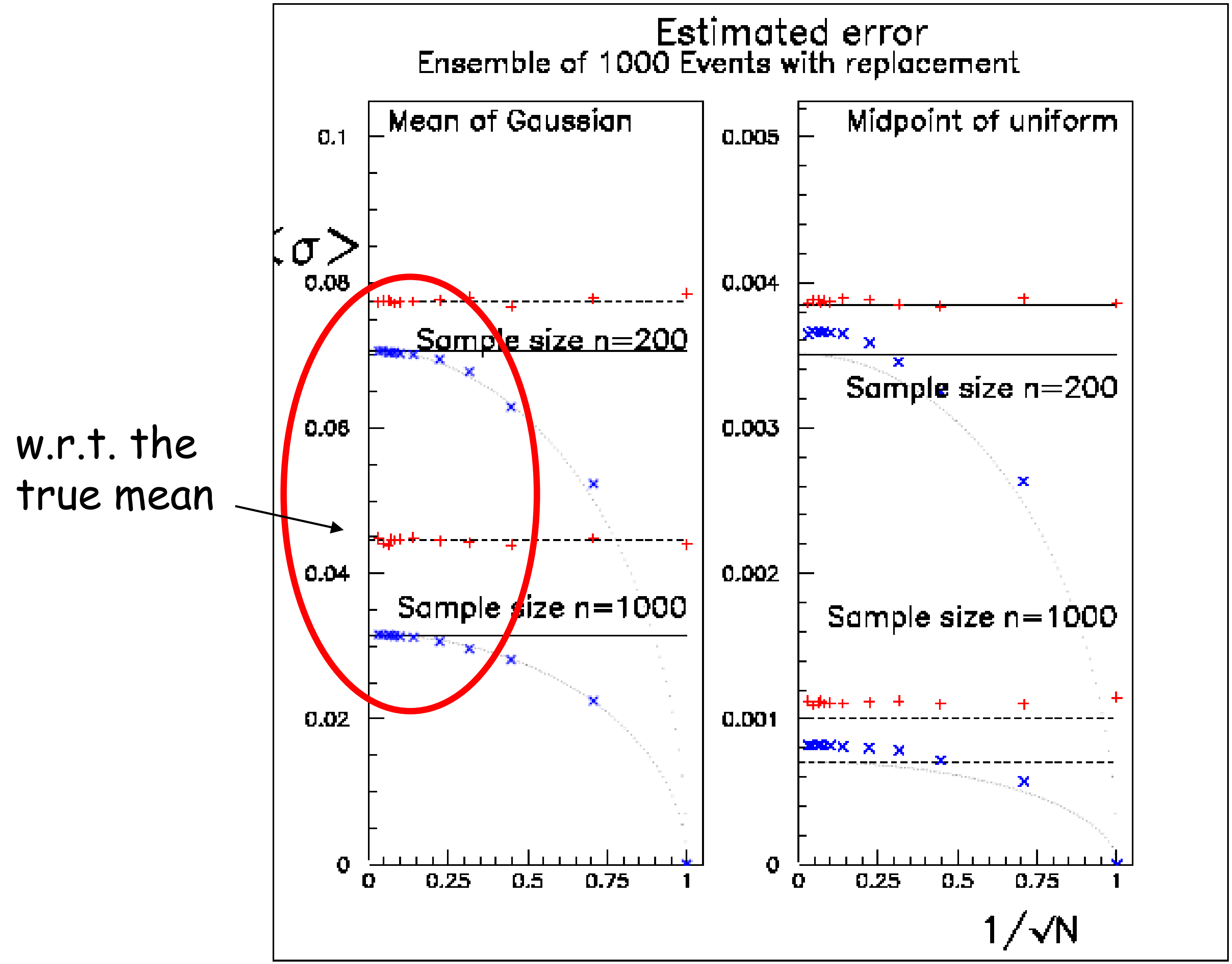


Figure 5: Mean estimated error using 'with replacement' samples

The non parametric BOOTSTRAP

When does the Bootstrap work?
For the consistency of the method, the reliability must be Bootstrap-checked, through the Bootstrap samples themselves!
The important checks are:

- **check the symmetry of the Bootstrap distribution, that assures the bootstrap property. Find if necessary a transformation h such as**

$$h(\hat{\theta}) - h(\theta) \quad \text{and} \quad h(\hat{\theta}^*) - h(\hat{\theta})$$

  are pivotal, that is follow the same distribution. Then make the estimate of the h intervals before anti-transforming with h^{-1}

- **make different estimates with different bootstrap samples (with replacement) $N_B \leq N$ and verify that the variances scales as $1/N_B$. This verify the condition**

$$\sqrt{1-\rho}\sqrt{\rho_1} \simeq 1$$

There exists a wide statistical literature on the subject....

PHYSICAL REVIEW D VOLUME 39, NUMBER 1 1 JANUARY 1989

Application of the bootstrap statistical method to the tau-decay-mode problem

Kenneth G. Hayes and Martin L. Perl
Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309

Bradley Efron
Department of Statistics, Stanford University, Stanford, California 94309
(Received 18 July 1988)

The bootstrap statistical method is applied to the discrepancy in the one-charged-particle decay modes of the tau lepton. This eliminates questions about the correctness of the errors ascribed to the branching-fraction measurements and the use of Gaussian error distributions for systematic errors. The discrepancy is still seen when the results of the bootstrap analysis are combined with other measurements and with deductions from theory. But the bootstrap method assigns less statistical significance to the discrepancy compared to a method using Gaussian error distributions.

At present there is a problem[1,2] in fully understanding the decay modes of the tau lepton to one-charged particle. The average directly measured value[1] of the inclusive, one-charged-particle, branching fraction B_1 is (86.6±0.3)%. The same number should be obtained by adding up the branching fractions of the individual one-charged-particle modes. Examples of these individual branching fractions are

$$B_e, \quad \tau^- \to \nu_\tau + e^- + \bar{\nu}_e \ ,$$

$$B_\mu, \quad \tau^- \to \nu_\tau + \mu^- + \bar{\nu}_\mu \ ,$$

$$B_\pi, \quad \tau^- \to \nu_\tau + \pi^- \ ,$$

$$B_\rho, \quad \tau^- \to \nu_\tau + \rho^- \to \nu_\tau + \pi^- + \pi^0 \ ,$$

$$B_{\pi 2\pi^0}, \quad \tau^- \to \nu_\tau + \pi^- + 2\pi^0 \ ,$$

$$B_{\pi 3\pi^0}, \quad \tau^- \to \nu_\tau + \pi^- + 3\pi^0 \ .$$

As shown in Table I from Ref. 2, this sum is less than (80.6±1.5)%, 6% less than the directly measured value of B_1. This is the τ-decay-mode problem.

Bootstrap of B_1 and B_i data

TABLE II. B_1 topological branching fractions in percent. The statistical error is given first, the systematic error second.

B_1 Measurement	Combined error	Energy (GeV)	Experimental group	Reference
84.0	±2.0	32.0–36.8	CELLO	H. J. Behrend *et al.*, Phys. Lett. **114B**, 282 (1982)
85.2±2.6±1.3	±2.9	14.0	CELLO	H. J. Behrend *et al.*, Z. Phys. C **23**, 103 (1984)
85.1±2.8±1.3	±3.1	22.0	CELLO	H. J. Behrend *et al.*, Z. Phys. C **23**, 103 (1984)
87.8±1.3±3.9	±4.1	34.6 average	PLUTO	Ch. Berger *et al.*, Z. Phys. C **28**, 1 (1985)
$84.7\pm1.1^{+1.6}_{-1.3}$	$^{+1.9}_{-1.7}$	13.9–43.1	TASSO	M. Althoff *et al.*, Z. Phys. C **26**, 521 (1985)
86.7±0.3±0.6	±0.7	29.0	MAC	E. Fernandez *et al.*, Phys. Rev. Lett. **54**, 1624 (1985)
86.9±0.2±0.3	±0.4	29.0	HRS	C. Akerlof *et al.*, Phys. Rev. Lett. **55**, 570 (1985)
86.1±0.5±0.9	±1.0	30.0–46.8	JADE	W. Bartel *et al.*, Phys. Lett. **161B**, 188 (1985)
87.9±0.5±1.2	±1.3	29.0	DELCO	W. Ruckstuhl *et al.*, Phys. Rev. Lett. **56**, 2132 (1986)
87.2±0.5±0.8	±0.9	29.0	Mark II	W. B. Schmidke *et al.*, Phys. Rev. Lett. **57**, 527 (1986)
84.7±0.8±0.6	±1.0	29.0	TPC	H. Aihara *et al.*, Phys. Rev. D **35**, 1553 (1987)

Trimmed mean 50%
Correlation between measurements
Weighted resampling Int($1/\sigma^2$) times
The error on measurements is not considered

Scope of the analysis: to test wether errors only or the data itself are unreliable

TABLE V. (a) Independent measurement of the $\tau^- \rightarrow e^- \bar{\nu}_e \nu_T$ and $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ branching fractions B_e and B_μ in percent. The statistical error is given first, the systematic error second. (b) Constrained or correlated measurements of the $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ branching fractions B_e and B_μ in percent. The statistical error is given first, the systematic error second.

B_e Measurement	B_e Combined error	B_μ Measurement	B_μ Combined error	Energy (GeV)	Experimental group	Reference
					(a)	
		17.5±2.7±3.0	±4.0	3.8–7.8	Mark I	M. L. Perl *et al.*, Phys. Lett. **70B**, 487 (977)
		22	$^{+10}_{-7}$	4.8		M. Cavalli-Storza *et al.*, Lett. Nuovo Cimento **20**, 337 (1977)
		15	±3.0	3.6–5.0	PLUTO	J. Burmester *et al.*, Phys. Lett. **68B**, 297 (1977)
16.0	±1.3			3.1–7.4	DELCO	W. Bacino *et al.*, Phys. Rev. Lett. **41**, 13 (1978)
		22	$^{+7}_{-8}$	6.4–7.4	Iron Ball	J. G. Smith *et al.*, Phys. Rev. D **18**, 1 (1978)
		21±5±3	±6	3.6–7.4	DELCO	W. Bacino *et al.*, Phys. Rev. Lett. **42**, 6 (1979)
19	±9.0	35	±14	12–31.6	TASSO	R. Brandelik *et al.*, Phys. Lett. **92B**, 199 (1980)
		17.8±2.0±1.8	±2.7	9.4–31.6	PLUTO	Ch. Berger *et al.*, Phys. Lett. **99B**, 489 (1981)
18.3±2.4±1.9	±3.1	17.6±2.6±2.1	±3.3	34.0	CELLO	H. J. Behrend *et al.*, Phys. Lett. **127B**, 270 (1983)
$20.4\pm3.0^{+1.4}_{-0.9}$	$^{+3.3}_{-3.1}$	$12.9\pm1.7^{+0.7}_{-0.5}$	±1.8	13.9–43.1	TASSO	M. Althoff *et al.*, Z. Phys. C **26**, 521 (1985)
13.0±1.9±2.9	±3.5	19.4±1.6±1.7	±2.3	34.6 average	PLUTO	Ch. Berger *et al.*, Z. Phys. C **28**, 1 (1985)
		17.4±0.6±0.8	±1.0	14.0–46.8	Mark J	B. Adeva *et al.*, Phys. Lett. B **179**, 177 (1986)
17.0±0.7±0.9	±1.1	18.8±0.8±0.7	±1.1	34.6 average	JADE	W. Bartel *et al.*, Phys. Lett. B **182**, 216 (1986)
19.1±0.8±1.1	±1.4	18.3±0.9±0.8	±1.2	29.0	Mark II	P. R. Burchat *et al.*, Phys. Rev. D **35**, 27 (1987)
					(b)	
18.9±1.0±2.8	±3.0	18.3±1.0±2.8	±3.0	3.8–7.8	Mark I	M. L. Perl *et al.*, Phys. Lett. **70B**, 487 (1977)
22.7	±5.5	22.1	±5.5	4.1–7.4	Lead-Glass Wall	A. Barbaro-Galtiero *et al.*, Phys. Rev. Lett. **39**, 1058 (1977)
18.5±2.8±1.4	±3.1	18.0±2.8±1.4	±3.1	3.9–5.2	DASP	R. Brandelik *et al.*, Phys. Lett. **73B**, 109 (1978)
17.6±0.6±1.0	±1.3	17.1±0.6±1.0	±1.3	3.5–6.7	Mark II	C. A. Blocker *et al.*, Phys. Lett. **109B**, 119 (1982)
18.2±0.7±0.5	±0.9	18.0±1.0±0.6	±1.2	3.8	Mark III	R. M. Baltrusaitis *et al.*, Phys. Rev. Lett. **55**, 1842 (1985)
17.4±0.8±0.5	±0.9	17.7±0.8±0.5	±0.9	29.0	MAC	W. W. Ash *et al.*, Phys. Rev. Lett. **55**, 2118 (1985)
18.4±1.2±1.0	±1.6	17.7±1.2±0.7	±1.4	29.0	TPC	H. Aihara *et al.*, Phys. Rev. D **35**, 1553 (1987)

Results

TABLE VI. Means and standard deviations (SD) for B_1, B_ρ, B_π, B_e, and B_μ. Both quantities are in percent.

	Analysis method					
	Bootstrap (method A)		Bootstrap with weighted measurements (method C)		Normal-error method from Ref. 1	
Branching fraction	Mean, 25% trimmed	SD	Mean, 25% trimmed	SD	Mean, not trimmed	Formal error
B_1	85.8	0.63	86.9	0.36	86.6	0.28
B_ρ	22.5	0.35	22.5	0.19	22.5	0.85
B_π	10.2	0.55	10.8	0.45	10.8	0.60
B_e	18.3	0.38	17.8	0.30	17.6	0.44
B_μ	18.2	0.56	17.8	0.21	17.7	0.41

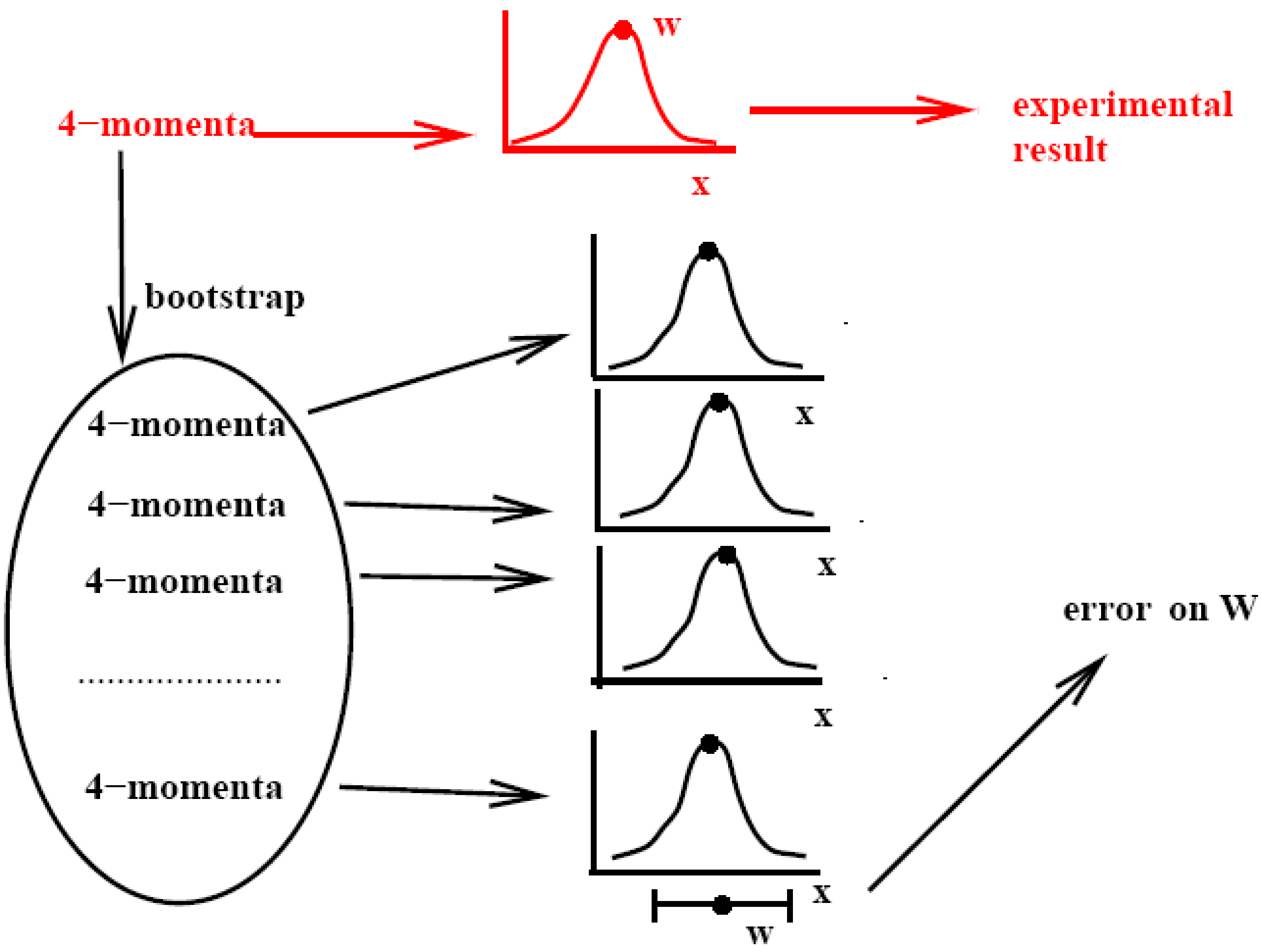
Bootstrap: $$\frac{85.8-81.3}{\sqrt{1.3^2+0.6^2}}=3.1$$

Standard analysis: $$\frac{86.6-80.6}{\sqrt{1.5^2+0.3^2}}=3.9$$

Some data are unreliable

The non parametric BOOTSTRAP

A possible use of the Bootstrap in Nuclear physics



BOOTSTRAP FOR COMPARING TWO POPULATIONS

Given independent SRSs of sizes n and m from two populations:

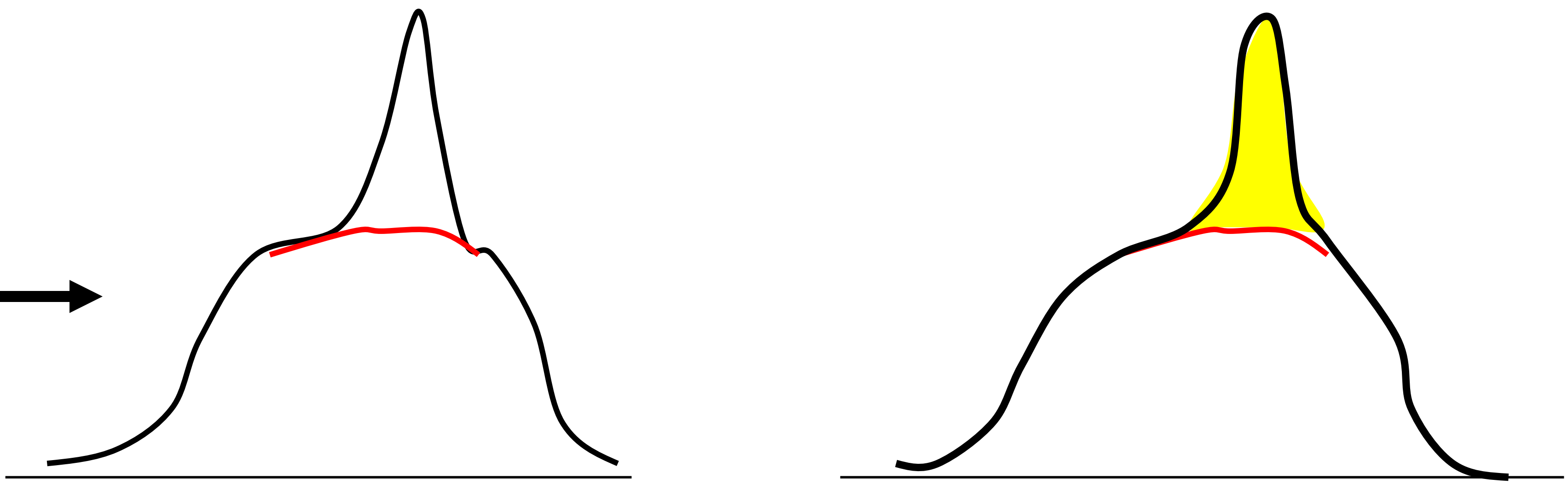
1. Draw a resample of size n with replacement from the first sample and a separate resample of size m from the second sample. Compute a statistic that compares the two groups, such as the difference between the two sample means.

2. Repeat this resampling process hundreds of times.

3. Construct the bootstrap distribution of the statistic. Inspect its shape, bias, and bootstrap standard error in the usual way.

Useful when the two samples are signal and background....

The dual Bootstrap



Fix the background on one sample and calculated the peak signal with another sample to avoid biases !!

Repeat on bootstrap samples (dual bootstrap)

Standard analysis in nuclear physics experiments

- the 4-momenta are reconstructed and the analysis is performed
- errors are calculated following the standard (gaussian) theory
- a MC toy model is invented and the analysis procedure is checked on this model
- at this point the procedure could be further checked on bootstrapped data!

Conclusions

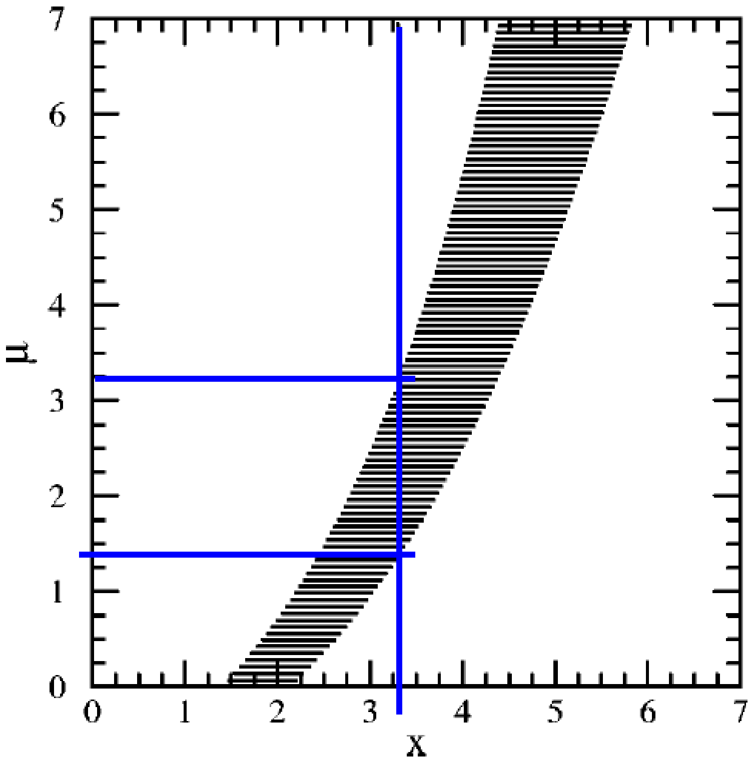
- Poissonian Counting: most of the tests do not consider the error on background and overestimate the signal. Often true (mean) values and measured values are improperly confused.

- Binomial counting: a general theory there exists and should be applied.

- The errors should be calculated by MC methods and the procedure checked with MC toy models

- Nonparametric Bootstrap methods should be used also by physicists

Some problems with frequentism and their cure



We restart from the
Neyman construction

True value

Observed value

Neyman's prescription:

Before doing an experiment, for each possible
Value of theory parameters determine a region of data that occurs
C.L. of the time, say 90%.

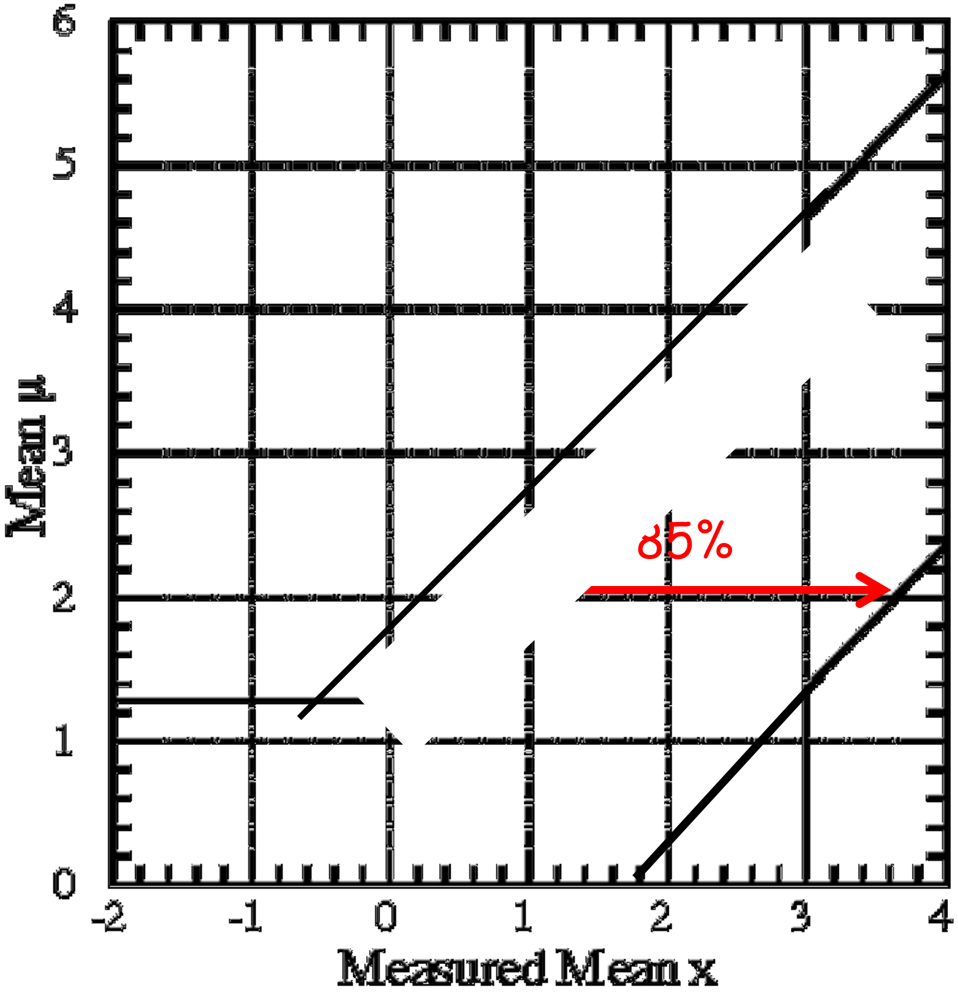
Some frequentist problems - I

A Gaussian upper limit estimate with $CL = 90\%$ when physics says that the true value is $\mu \geq 0$.

If a negative value $x < -1.28\,\sigma$ is obtained, we find an unphysical upper limit $\mu < 0$!

A solution, the FLIP-FLOP technique:
when $x < 0$ one assumes the upper limit for $x = 0$, that is $\mu =$ 1.28 s

WRONG!: one finds an 85% upper limit



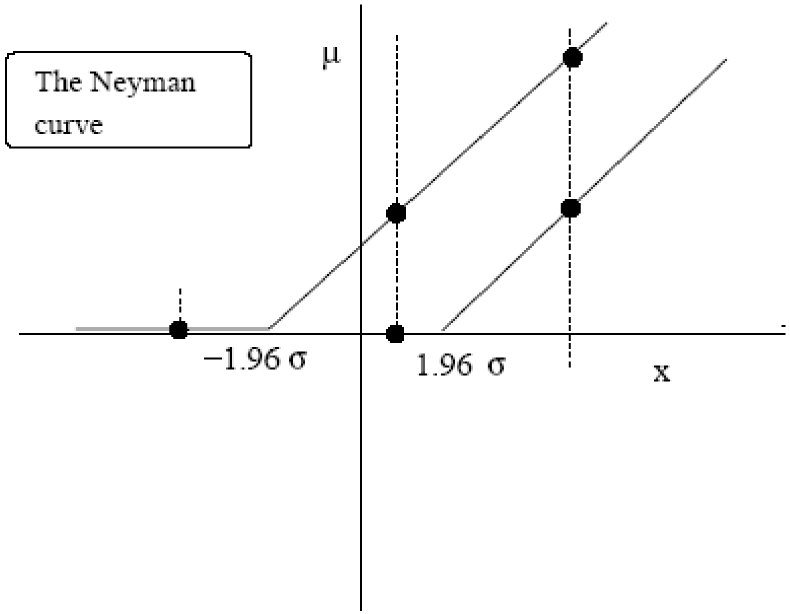
When μ =2 there is only 85% coverage!
Due to flip-flopping (deciding whether to use an upper limit or a central confidence region based on the data) these are not valid confidence intervals.

Some frequentist problems - I

If $\mu > 0$ an interval with correct coverage (i.e. $CL = 95\%$) is

$$\mu \in \begin{cases} (x - 1.96\,\sigma, x + 1.96\,\sigma) & \text{WHEN} \quad x > 1.96\,\sigma \\ (0, x + 1.96\,\sigma) & \text{WHEN} \quad -1.96\,\sigma < x < 1.96\,\sigma \\ 0 & \text{WHEN} \quad x < -1.96\,\sigma \end{cases}$$

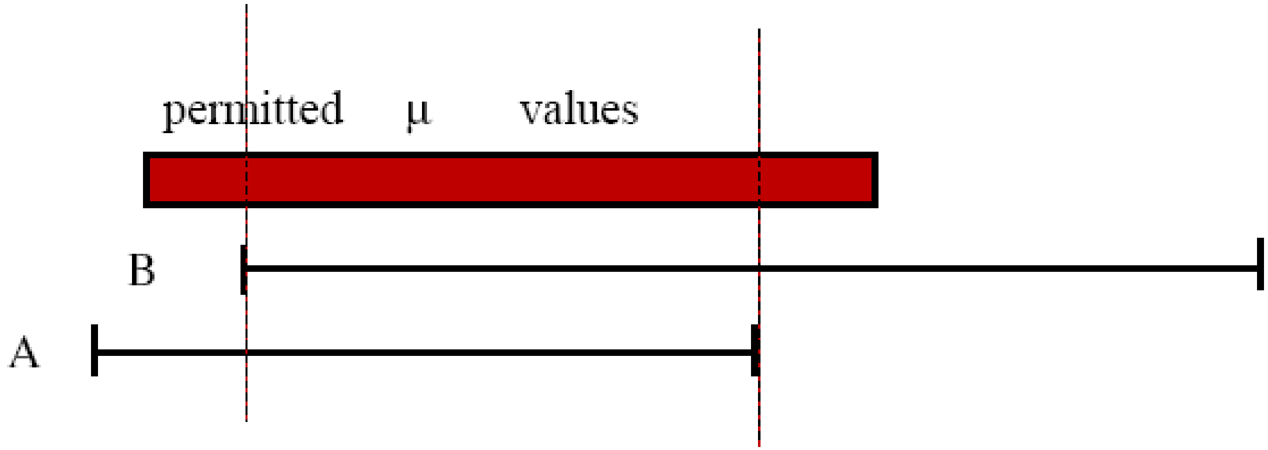
Here the coverage is correct



BUT, when $x < -1.96\,\sigma$ we have $\mu = 0$

Are we Happy?

Not only the COVERAGE!!



A and B have the same coverage but

A minimizes the probability to contain wrong μ values

The A interval is the result of a more powerful estimator
New concepts are necessary

Some frequentist problems - II

A counting experiment with background

$$P(n;\mu,b) = \frac{(n+b)^n}{n!}\,\mathrm{e}^{-(\mu+b)}$$

The background b is known, the counts n are measured.

Find the upper limit for the μ parameter with a fixed CL, for example 90%

b/n	0	1	2	3	4
0	2.30	3.89	5.32	6.68	7.99
1	1.30	2.89	4.32	5.58	6.99
2	0.30	1.89	3.32	4.68	5.99
3	−0.79	0.89	2.32	3.68	4.99

When no events are counted,
an experiment with an expected background of
3 events measures
a negative m upper limit for the expected value
of the counts!!

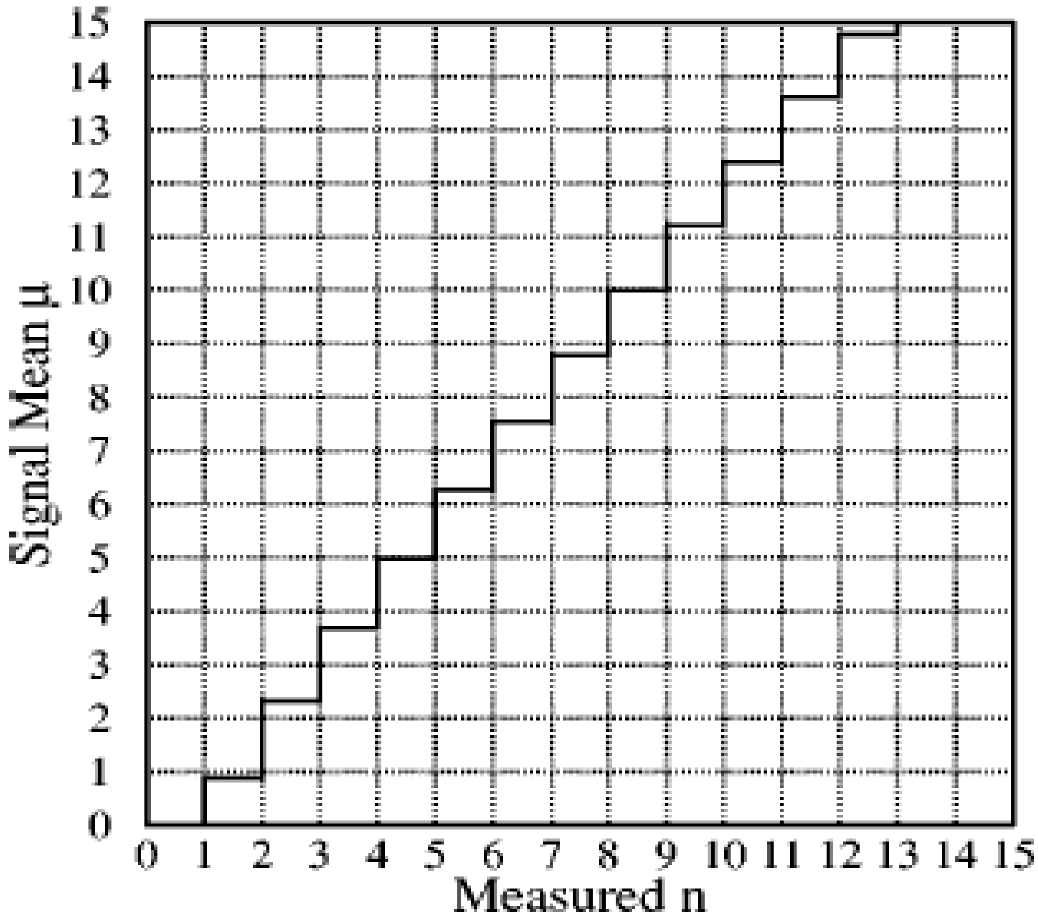


FIG. 5. Standard confidence belt for 90% C.L. upper limits, for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b=3.0$. The second line in the belt is at $n=+\infty$.

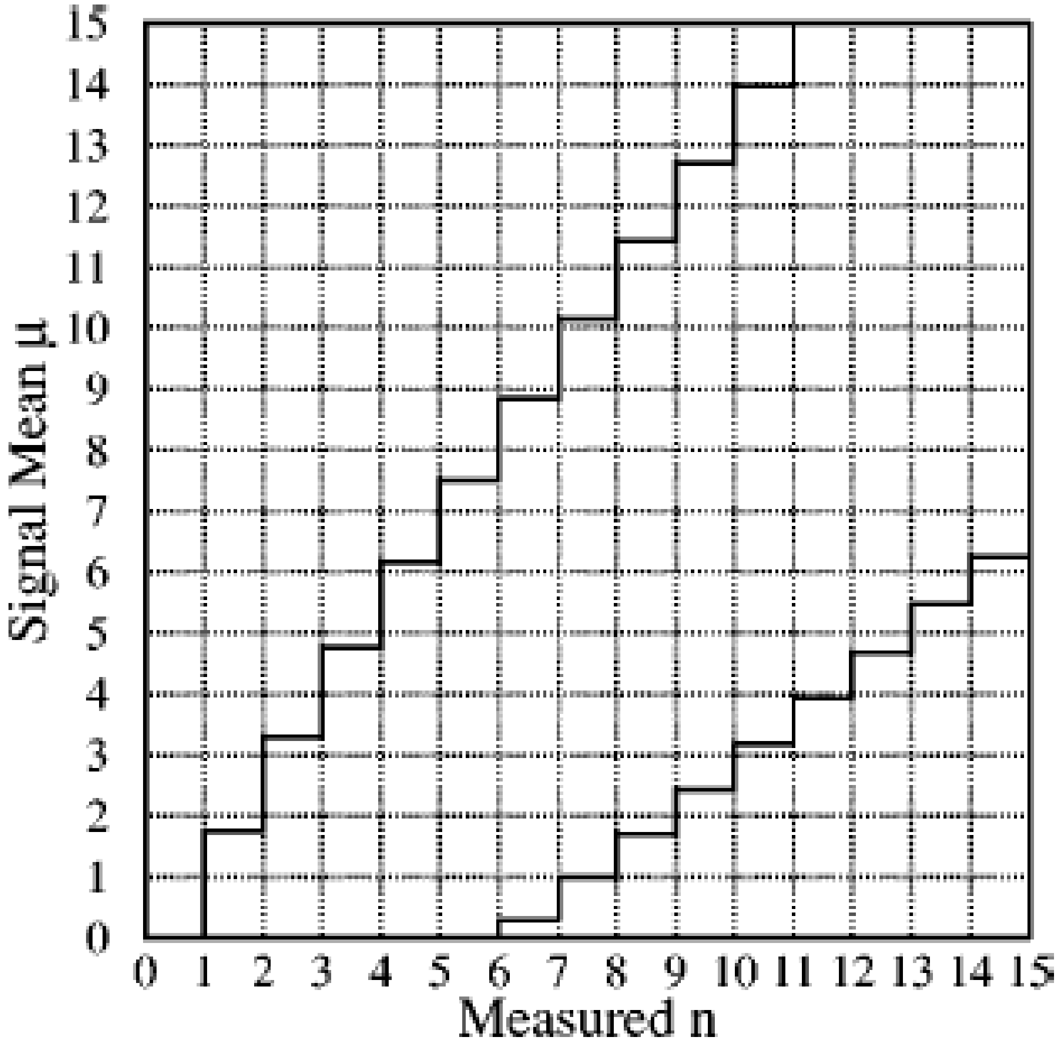


FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b = 3.0$.

PHYSICAL REVIEW D VOLUME 57, NUMBER 7 1 APRIL 1998

Unified approach to the classical statistical analysis of small signals

Gary J. Feldman*
Department of Physics, Harvard University, Cambridge, Massachusetts 02138

Robert D. Cousins†
Department of Physics and Astronomy, University of California, Los Angeles, California 90…

(Received 21 November 1997; published 6 March 1998)

We give a classical confidence belt construction which unifies the treatment of upper confidence limits for null results and two-sided confidence intervals for non-null results. The unified treatment solves a problem (apparently not previously recognized) that the choice of upper limit or two-sided intervals leads to intervals which are not confidence intervals if the choice is based on the data. We apply the construction to two related problems which have recently been a battleground between classical and Bayesian statistics: Poisson processes with background and Gaussian errors with a bounded physical region. In contrast with the usual classical construction for upper limits, our construction avoids unphysical confidence intervals. In contrast with some popular Bayesian intervals, our intervals eliminate conservatism (frequentist coverage greater than the stated confidence) in the Gaussian case and reduce it to a level dictated by discreteness in the Poisson case. We generalize the method in order to apply it to analysis of experiments searching for neutrino oscillations. We show that this technique both gives correct coverage and is powerful, while other classical techniques that have been used by neutrino oscillation search experiments fail one or both of these criteria.

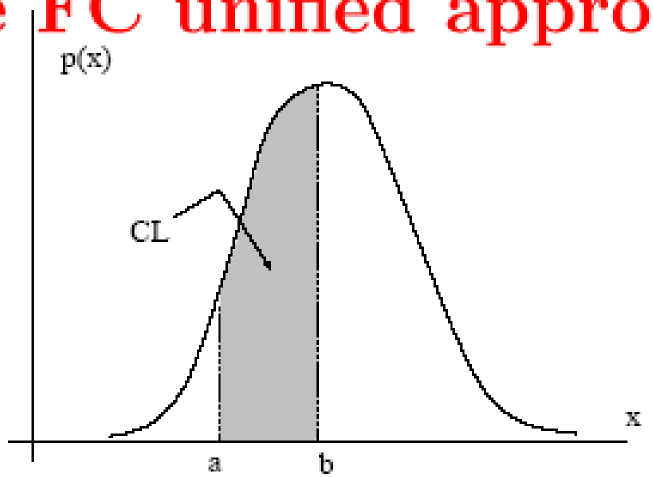
[S0556-2821(98)00109-X]

PACS number(s): 06.20.Dk, 14.60.Pq

Bayesians recall the date

topcite

The FC unified approach



$$R = \frac{p(x;\mu)}{\max\left[p(x;\mu)\right]} = \frac{p(x;\mu)}{p(x;\hat{\mu})}$$

$$CL = \sum_R p(x;\mu) \rightarrow \int_R p(x;\mu)\,\mathrm{d}x$$

This frequentist approach is called unified:

- **the technique for the one-sided and the two-sided intervals is the same (by FC)**
- **the approach merges the parameter estimate and the hypothesis testing techniques**

For a given CL and μ, there are many (a, b) choices. The FC ordering Principle (FCOP), inspired to the NP theorem, is to choose the extremes a and b that cover the points ordered for decreasing values of

The FC Neyman curve

The idea is to use the likelihood ratio

$$R = \frac{p(x;\mu)}{p(x;\hat{\mu})} = \frac{L(\mu;x)}{L(\hat{\mu};x)}$$

- $R \simeq 1 \rightarrow \mu \simeq \hat{\mu}$
- $R \simeq 0 \rightarrow \mu$ far from the ML estimate
- constraints on μ are fully considered
- $P\{\mu \in I_{\rm FC}\}$ minimum when $\mu \neq \mu_{\rm true}$

The method is officially recommanded from PDG 1998

The FC idea

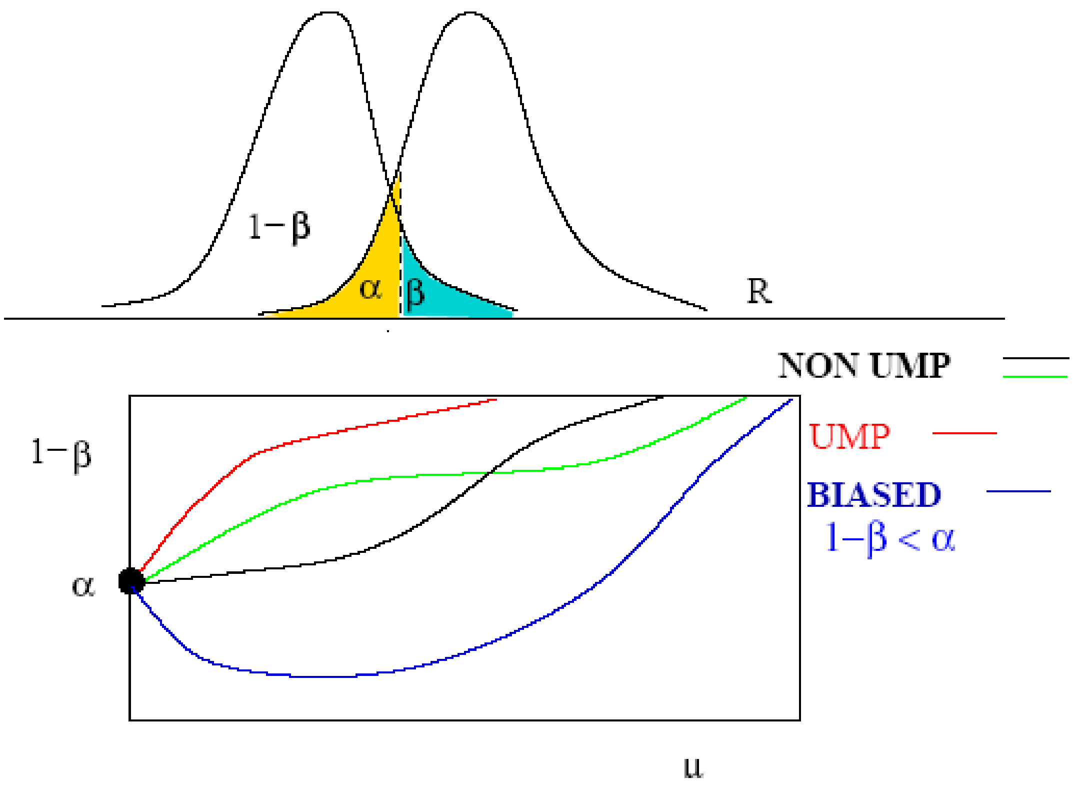
The ratio

$$R = \frac{L(\mu; x)}{L(\hat{\mu}; x)}$$

minimizes β and maximizes $(1 - \beta)$

The R test tends to be Uniformly Most Powerful.

The probability that I contains false μ values is minimized



Poisson with background
The FC method

$$P(n;\mu,b) = \frac{(n+b)^n}{n!}\,\mathrm{e}^{-(\mu+b)}$$

$$\frac{\partial}{\partial\mu}\left\{(\mu+b)^n \mathrm{e}^{-(\mu+b)}\right\} =$$

$$\mathrm{e}^{-(\mu+b)}\left[n(\mu+b)^{n-1} - (\mu+b)^n\right] = 0$$

Hence

$$\hat{\mu} = \begin{cases} n-b & \text{if} \quad n > b \\ 0 & \text{otherwise} \end{cases} \quad \rightarrow \quad \hat{\mu} = \max(0, n-b)$$

The method when $\mu = 0.5$, $b = 3$, $CL = 90\%$

This is only ONE POINT on the Neyman curve

n	$P(n;\mu,b)$	$\hat{\mu}$	$P(n;\hat{\mu},b)$	R	score 90%
0	0.030	0	0.050	0.607	6
1	0.106	0	0.149	0.708	3
2	0.185	0	0.224	0.826	3
3	0.216	0	0.224	0.963	2
4	0.189	1	0.195	0.966	1
5	0.132	2	0.175	0.753	4
6	0.077	3	0.161	0.480	7
7	0.039	4	0.149	0.259	
8	0.017	5	0.140	0.121	
9	0.007	6	0.125	0.018	
10	0.002	7	0.125	0.006	

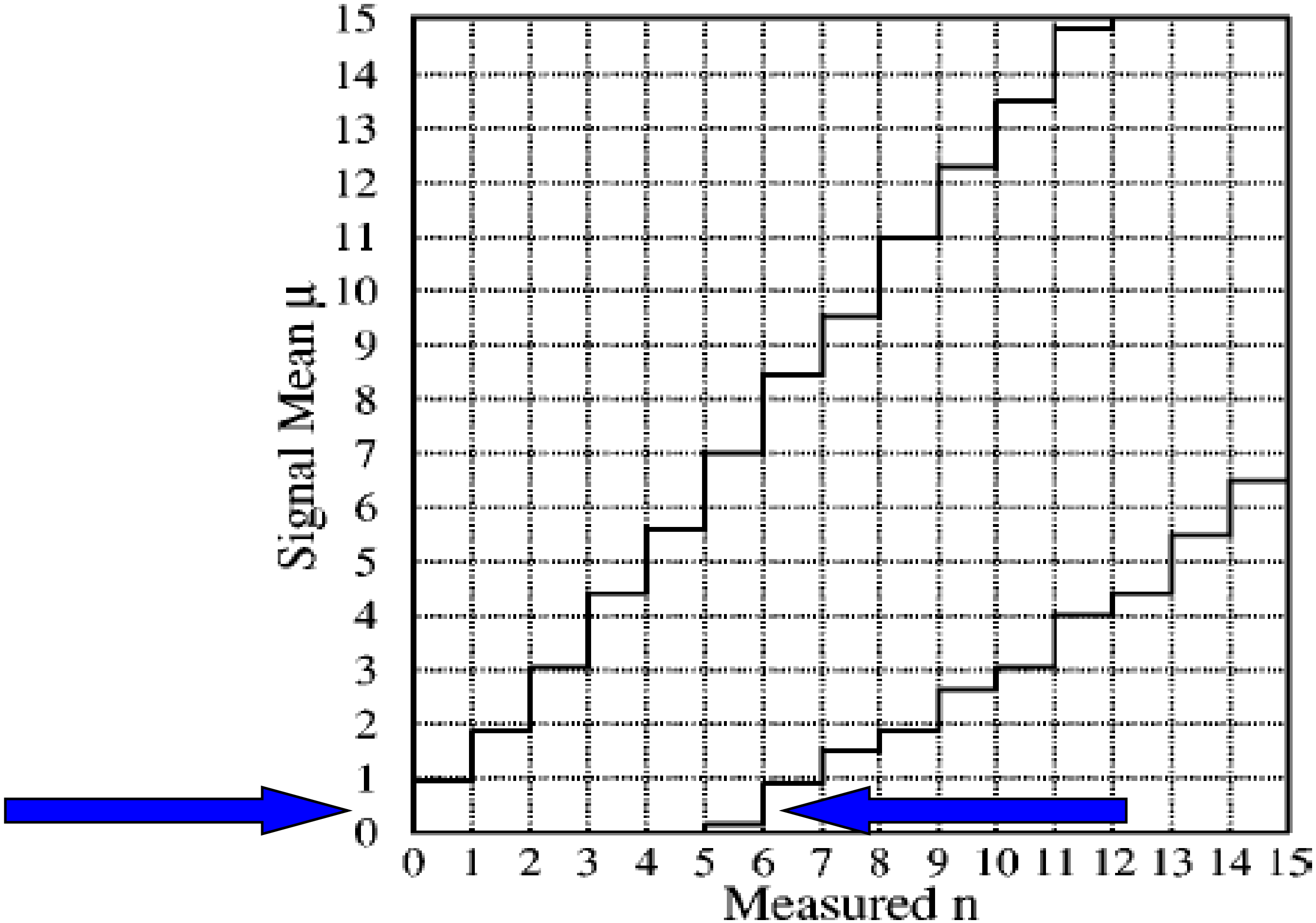


FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b = 3.0$.

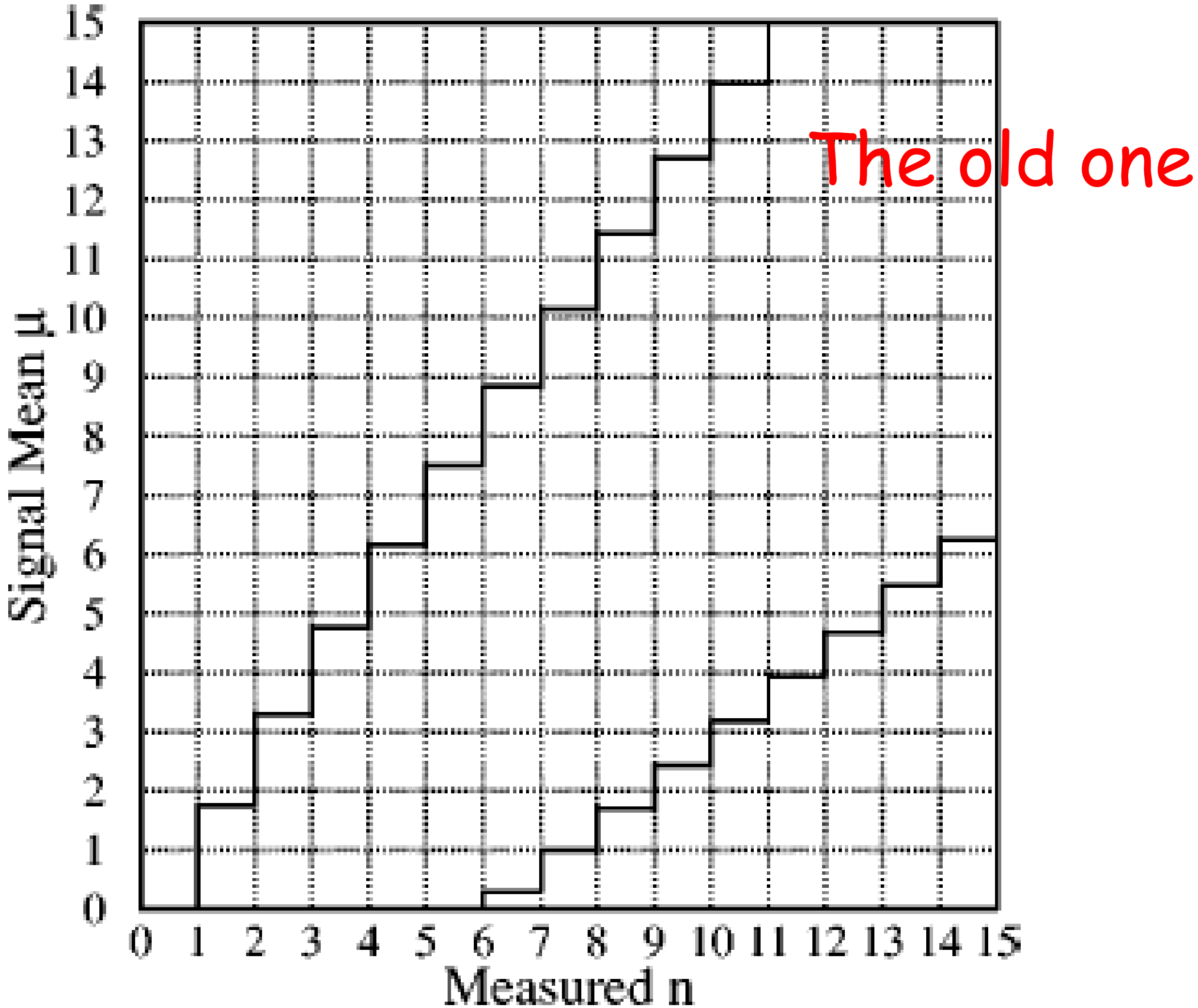


FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b=3.0$.

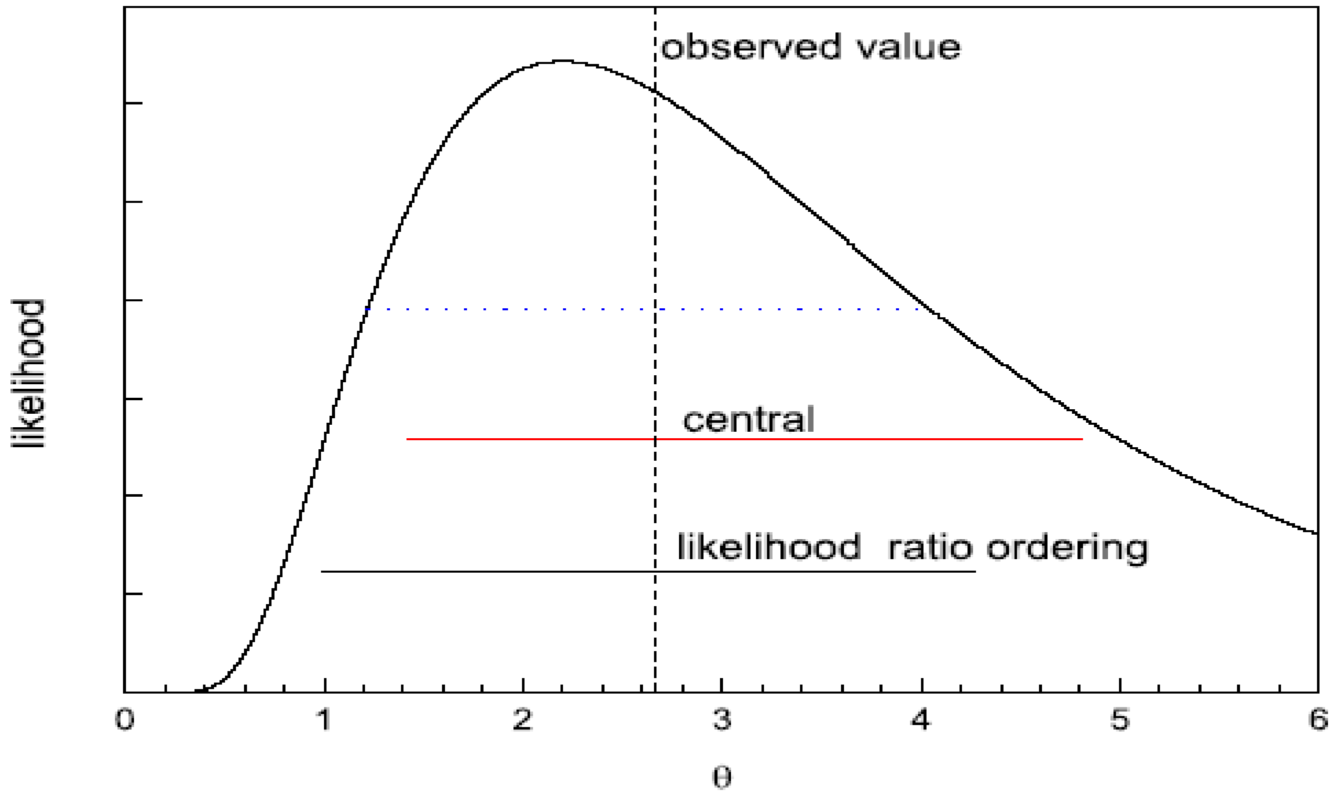


Fig. 4. Position measurement from drift time. The error is due to diffusion. Classical confidence intervals are shown together with the likelihood function

Gaussian with constraints
The FC method

$$\frac{\partial}{\partial \mu}\left[\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]\right] = 0 \to \hat{\mu} = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x \leq 0 \end{cases} \to \max(0, x)$$

$$R = \frac{g(x;\mu)}{g(x;\hat{\mu})}, \quad \text{where } g(x;\hat{\mu}) = \begin{cases} \frac{1}{\sqrt{2\pi}} & x \geq 0 \\ \frac{\exp(-x^2/2)}{\sqrt{2\pi}} & x < 0 \end{cases}$$

$$R = \begin{cases} \mathrm{e}^{-(x-\mu)^2/2} & x \geq 0 \\ \mathrm{e}^{(x\mu-\mu^2/2)} & x < 0 \end{cases}$$

m > 0

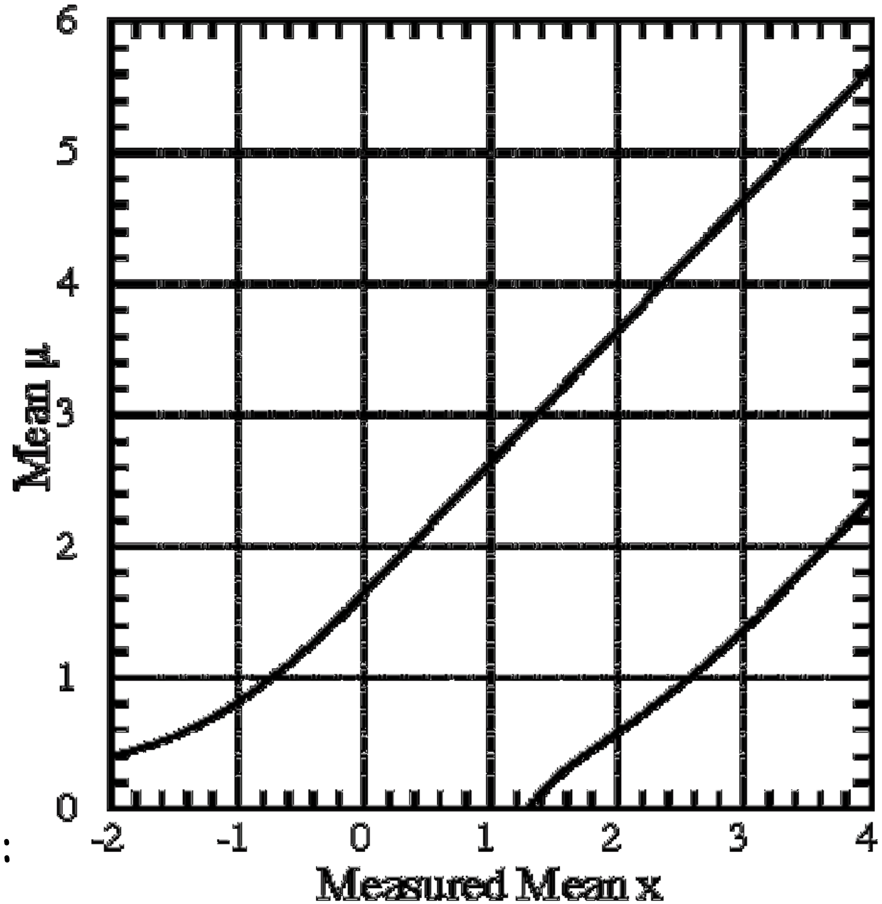
Then, the Neyman construction:

One determines the order in which the values of x are integrated.

That is, for any value of μ, one determines the interval $[x_1, x_2]$ such that $R(x_1) = R(x_2)$ and that

$$\int_{x_1}^{x_2} g(x;\mu)\,\mathrm{d}x = CL$$

From Feldman notes:



(1) This approaches the central limits for x >>1.

(2) The upper limit for x = 0 is 1.64, the two-sided rather than the one-sided limit.

(3) From the defining 1937 paper of Neyman, this is the only valid confidence belt, since there are 4 requirements for a valid belt:

  (a) It must cover.

  (b) For every x, there must be at least one μ.

  (c) No holes (only valid for single μ).

  (d) Every limit must include its end points.

A famous Paradox

An experiment that measures less events than the expected background, will report

a better (lower) upper limit

than an identical experiment which measures a number of events equal to the background

If no events are detected, the experiment with expected background will find

a better upper limit

than the experiment with no background.

A famous Paradox

Standard frequentist result:
counts with background, $CL = 90\%$, the upper limits improves when μ_b increases

μ_b \ $n\rightarrow$	0	1	2	3	4
0.0	2.30	3.89	5.32	6.68	7.99
0.5	1.80	3.39	4.82	6.18	7.49
1.0	1.30	2.89	4.32	5.58	6.99
2.0	0.30	1.89	3.32	4.58	5.99

A paradox?

YES from the Bayesian point of view

$$P(H|\mathrm{data})$$

NO from the frequentist point of view

$$P(\mathrm{data}|H)$$

It implies simply that the upper limit "improves", but in a small number of experiments, when $n < \mu_b$

A frequentist remedy: the Sensitivity

- **a μ_b value is fixed**
- **some Poisson data are generated via MC**

$$p(x;\mu_b) = \frac{\mu_b^x}{x!}\,\mathrm{e}^{-\mu_b}$$

- **the FC average upper limit is found for a given CL**

The sensitivity

Our suggestion for doing this is that in cases in which the measurement is less than the estimated background, the experiment reports both the upper limit and the "sensitivity" of the experiment, where the "sensitivity" is defined as the average upper limit that would be obtained by an ensemble of experiments with the expected background and no true signal. Table XII gives these values, for the case of a measurement of a Poisson variable.

TABLE XII. Experimental sensitivity (defined as the average upper limit that would be obtained by an ensemble of experiments with the expected background and no true signal), as a function of the expected background, for the case of a measurement of a Poisson variable.

b	68.27% C.L.	90% C.L.	95% C.L.	99% C.L.
0.0	1.29	2.44	3.09	4.74
0.5	1.52	2.86	3.59	5.28
1.0	1.82	3.28	4.05	5.79
1.5	2.07	3.62	4.43	6.27
2.0	2.29	3.94	4.76	6.69
2.5	2.45	4.20	5.08	7.11
3.0	2.62	4.42	5.36	7.49
3.5	2.78	4.63	5.62	7.87
4.0	2.91	4.83	5.86	8.18
5.0	3.18	5.18	6.32	8.76
6.0	3.43	5.53	6.75	9.35
7.0	3.63	5.90	7.14	9.82
8.0	3.86	6.18	7.49	10.27
9.0	4.03	6.49	7.81	10.69
10.0	4.20	6.76	8.13	11.09
11.0	4.42	7.02	8.45	11.46
12.0	4.56	7.28	8.72	11.83
13.0	4.71	7.51	9.01	12.22
14.0	4.87	7.75	9.27	12.56
15.0	5.03	7.99	9.54	12.90

The Unified Approach in two (or more) dimensions

One has to estimate the (x, y) CL regions and sometimes the z confidence region

$$z = f(x, y)$$

having measured x_m and y_m.

- the Likelihood Ratio is chosen as:

$$R = \frac{L(x, y; \mu)}{L(x, y; \hat{\mu})} \quad \rightarrow$$

$$-2 \ln R = -2[\ln L(x, y; \mu) - \ln L(x, y; \hat{\mu})] \simeq \chi^2(\hat{\mu})$$

- for a given CL a μ_x and μ_y map is made, calculating at each point the quantile

$$R \leq R_{1-CL}$$

where

$$\int_0^{R_{1-CL}} g(R)\, \mathrm{d}R = 1 - CL$$

When the R-space is constrained or complicated MC integration must be used.

- when we have to find $z = f(x, y)$ a second map must be generated (usually via MC) for each (μ_x, μ_y) with the values

$$z_{1-CL/2} \quad , \qquad z_{1-(1-CL)/2}$$

The Unified Approach in two (or more) dimensions

- **Now the**

$$(\mu_x, \mu_y)$$

plane is mapped with three numbers:

$$R_{1-CL} \quad , \quad z_{1-CL/2} \quad , \quad z_{1-(1-CL)/2}$$

- **the acceptance region** $(\mu_x, \mu_y)_{CL}$ **is defined by**

$$R(x_m, y_m) < R_{1-CL}$$

$$z_{1-CL/2} \leq z_m = f(x_m, y_m) \leq z_{1-(1-CL)/2}$$

The Unified Approach in two (or more) dimensions

When we have x_i gaussian variables $\sim N(\mu_i, \sigma_i^2)$

$$-2\ln R = \Delta\chi^2 = \sum_i \left[\frac{(x_i-\mu_i)^2}{\sigma_i^2} - \frac{(x_i-\hat{\mu}_i)^2}{\sigma_i^2}\right]$$

When there are no cuts or constraints

$$\hat{\mu}_i = x_i$$

and we reobtain the usual χ^2.
In the case of Poisson variables we have

$$-2\ln\left[\frac{\Pi_i\, \mu_i^{x_i} e^{-\mu_i}}{\Pi_i\, \hat{\mu}_i^{x_i} e^{-\hat{\mu}_i}}\right] = 2\sum_i [\mu_i - x_i \ln\mu_i - \hat{\mu}_i + x_i \ln\hat{\mu}_i]$$

When there are no cuts or constraints:

$$-2\ln R = 2\sum_i [\mu_i - x_i \ln\mu_i - x_i + x_i \ln x_i]$$

Example

Ratio of two gaussians $z = x/y$, comparison of three methods

- standard

$$\sigma[z] = \frac{x}{y}\sqrt{\frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}}$$

- bootstrap:
  sample from two gaussian of average x and y (measured value), take the ratio, make the histogram and find the two quantiles $z_{1-CL/2}$ and $z_{1-(1-CL)/2}$.

- unified method: find

$$R(x_m, y_m) < R_{1-CL}$$

$$z_{1-CL/2} \leq z_m = f(x_m, y_m) \leq z_{1-(1-CL)/2}$$

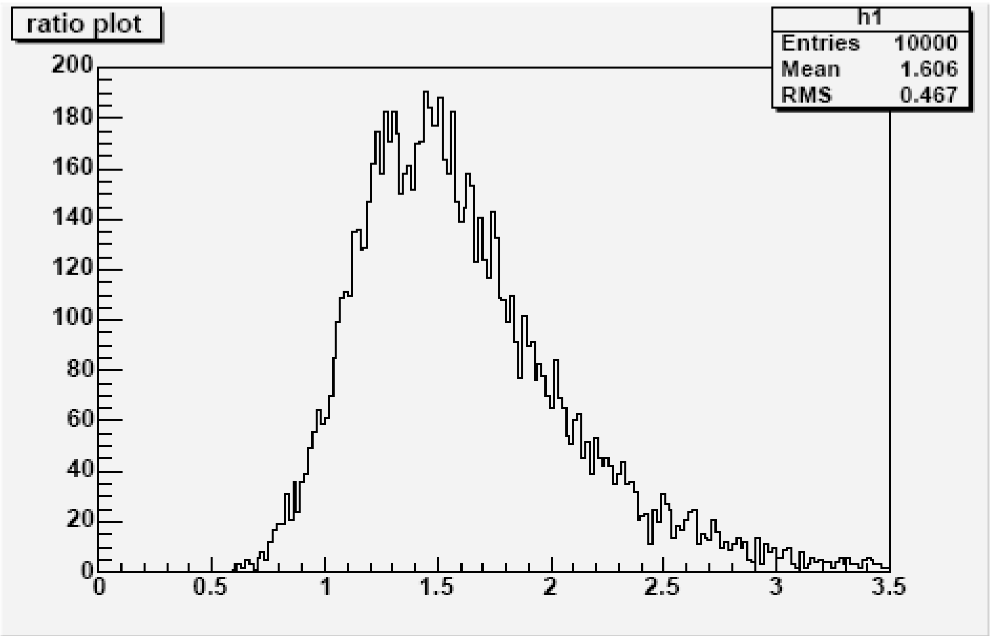
Example

Measured

$$x = 20 \qquad y = 13 \qquad \sigma_x = \sigma_y = 3$$

$z = x/y$ **interval with** $CL = 0.683$**;**
coverage calculated with MC:

- **standard:** $z = 0.730 \pm 0.002$**;**
- **bootstrap** $z = 0.700 \pm 0.002$**;**
- **unified:** $z = 0.696 \pm 0.002$**;**



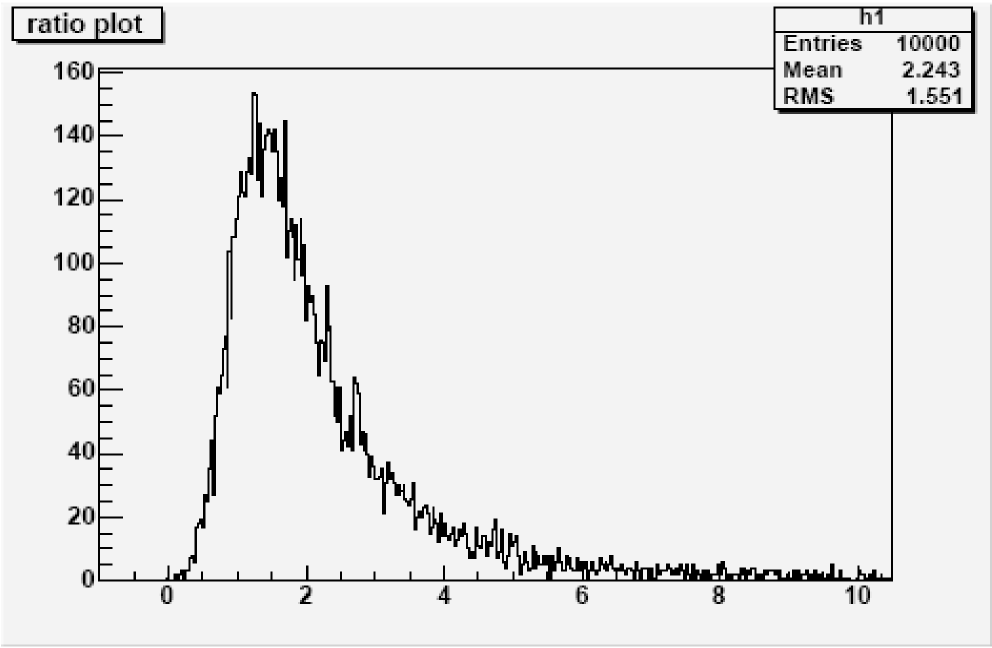
Example

Measured

$$x = 11 \qquad y = 6 \qquad \sigma_x = \sigma_y = 3$$

$z = x/y$ **interval with** $CL = 0.683$
with the cut $y > 0$**:**

- **standard:** $z = 0.774 \pm 0.002$;
- **bootstrap** $z = 0.736 \pm 0.002$;
- **unified:** $z = 0.686 \pm 0.002$;



Neutrino Oscillations

The QM probability is:

$$P(\nu_\mu \to \nu_e) = \sin^2(2\theta)\,\sin^2\left(\frac{1.27\Delta m^2 L}{E}\right)$$

L is the distance (km), E is the energy in GeV and $\Delta m^2 = |m_1^2 - m_2^2|$ is $(\mathrm{eV/c^2})^2$.

The result is plotted in the Δm^2 vs. $\sin^2(2\theta)$ plane

The goal is the search for an ν_e excess (the signal) over the normal ν_e background.
The data are usually a bin content, the signal n_i a and the backg. b_i.
The expected number of event is:

$$\mu_{true} = F\left(\sin^2(2\theta)\,\sin^2\left(\frac{1.27\Delta m^2 L}{E}\right) \right)$$

The frequentist test is

$$\chi^2 = \sum_i \frac{(n_i - b_i - \mu_i)^2}{\sigma_i^2}$$

Frequentist neutrino Monte Carlo

The starting point is the key formula:

$$2\Delta[\ln L] \equiv 2\left(\ln L(\hat{\theta}) - \ln L(\theta)\right) = \chi^2_\alpha(1) \ ,$$

Recall that from probability calculus

$$\sum_i^n \frac{(X_i - \mu_{\rm true})^2}{\sigma^2} \sim \chi^2(n)$$

and asymptotically, in statistics

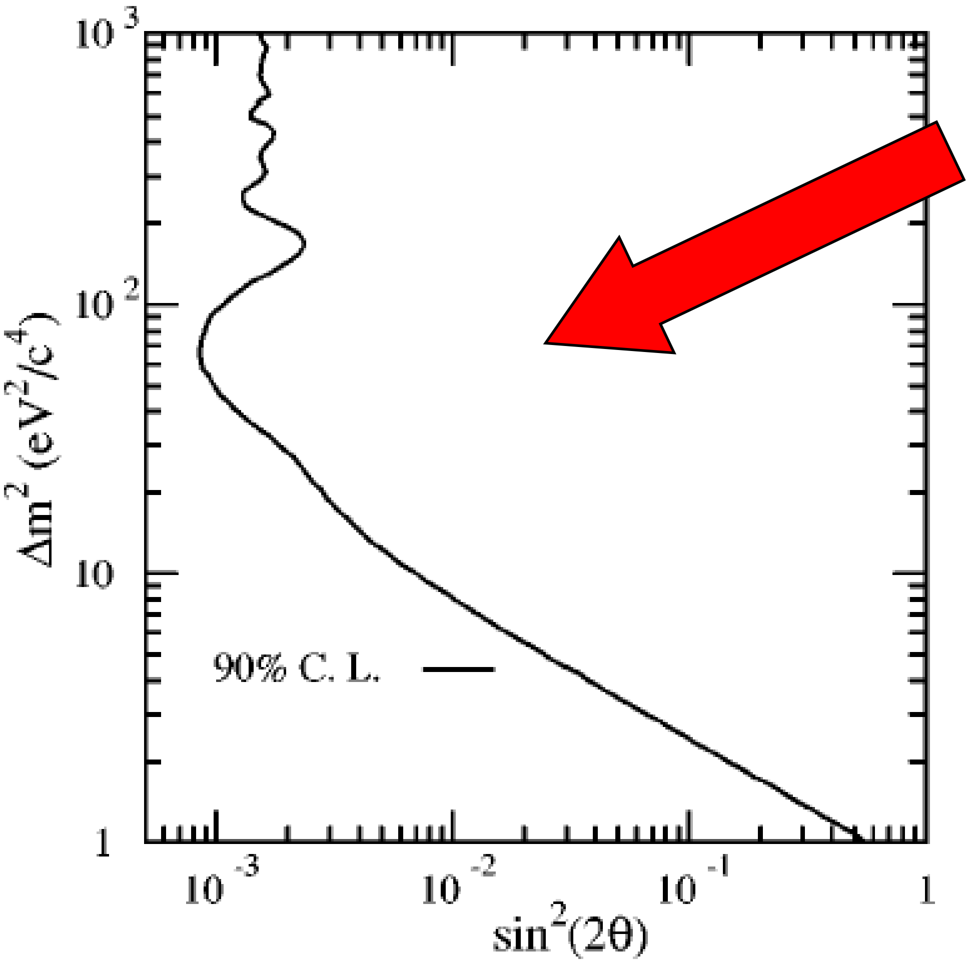
$$\underbrace{\sum_i^n \frac{(X_i - \mu(p)_{\rm true})^2}{\sigma^2}}_{n \ \text{d.o.f.}} - \underbrace{\sum_i^n \frac{(X_i - \hat{\mu}(p))^2}{\sigma^2}}_{(n-p) \ \text{d.o.f.}} \sim \underbrace{\chi^2(p)}_{p \ \text{d.o.f.}}$$

For $CL = 90\%$ and two parameters

$$\chi^2(\sin^2(2\theta), \Delta m^2) - \chi^2(\text{best}) \le 4.6$$

The region is given by the points with χ^2 within 4.6 of the minimum (acceptances neglected)

The acceptance zone for no oscillations (Hypothesis)



if the experiment falls here, we can reject the hypothesis with 90% CL

FIG. 11. Calculation of the confidence region for an example of the toy model in which $\sin^2(2\theta)=0$. The 90% confidence region is the area to the left of the curve.

Frequentist neutrino Monte Carlo

Another method is the Raster Scan
A grid is made for each Δm^2 value:

$$\Delta\chi^2 = \sum_i^n \frac{(n_i - b_i - \mu_{i\ \mathrm{true}}(\sin^2(2\theta))^2}{\sigma_i^2}$$

$$- \sum_i^n \frac{(n_i - b_i - \hat{\mu}_i(\sin^2(2\theta))^2}{\sigma_i^2} \leq (1.65)^2$$

χ^2 is calculated as a function of $\sin^2(2\theta)$
Raster Scan gives an exact coverage but does not give the optimum coverage.
The point is that $\hat{\mu}$ is found on one parameter only, without taking into account the compound probability of both $(\sin^2(2\theta), \Delta m^2)$, as required by the R test

Unified Approach (FC)

Monte Carlo

The correct method
applies the FC ordering principle

$$2\ln R = \Delta\chi^2 = \sum_i^n \frac{(n_i - b_i - \mu_i)^2}{\sigma_i^2} - \sum_i^n \frac{(n_i - b_i - \hat{\mu}_i)^2}{\sigma_i^2}$$

- $\forall[\sin^2(2\theta), \Delta m^2]$ a set of n_i, b_i values is simulated;
- for each point, the true value

$$\mu_i \rightarrow \sin^2(2\theta)\sin^2\left(\frac{1.27\Delta m^2 L}{E}\right)$$

is calculated with the model,

here the cuts are Taken into account

$$\mu_{\text{best}} \equiv \hat{\mu}$$

gives the highest probability for the physically allowed values

- $\forall[\sin^2(2\theta), \Delta m^2]$ the quantile χ^2_{CL} is calculated

$$P\{\Delta\chi^2 \leq \Delta\chi^2_{CL}\} = CL$$

- when one experiment gives $\{n_i, b_i\}$, the acceptance region is

$$\Delta\chi^2(n_i, b_i; \sin^2(2\theta), \Delta m^2) \leq \chi^2_{CL}$$

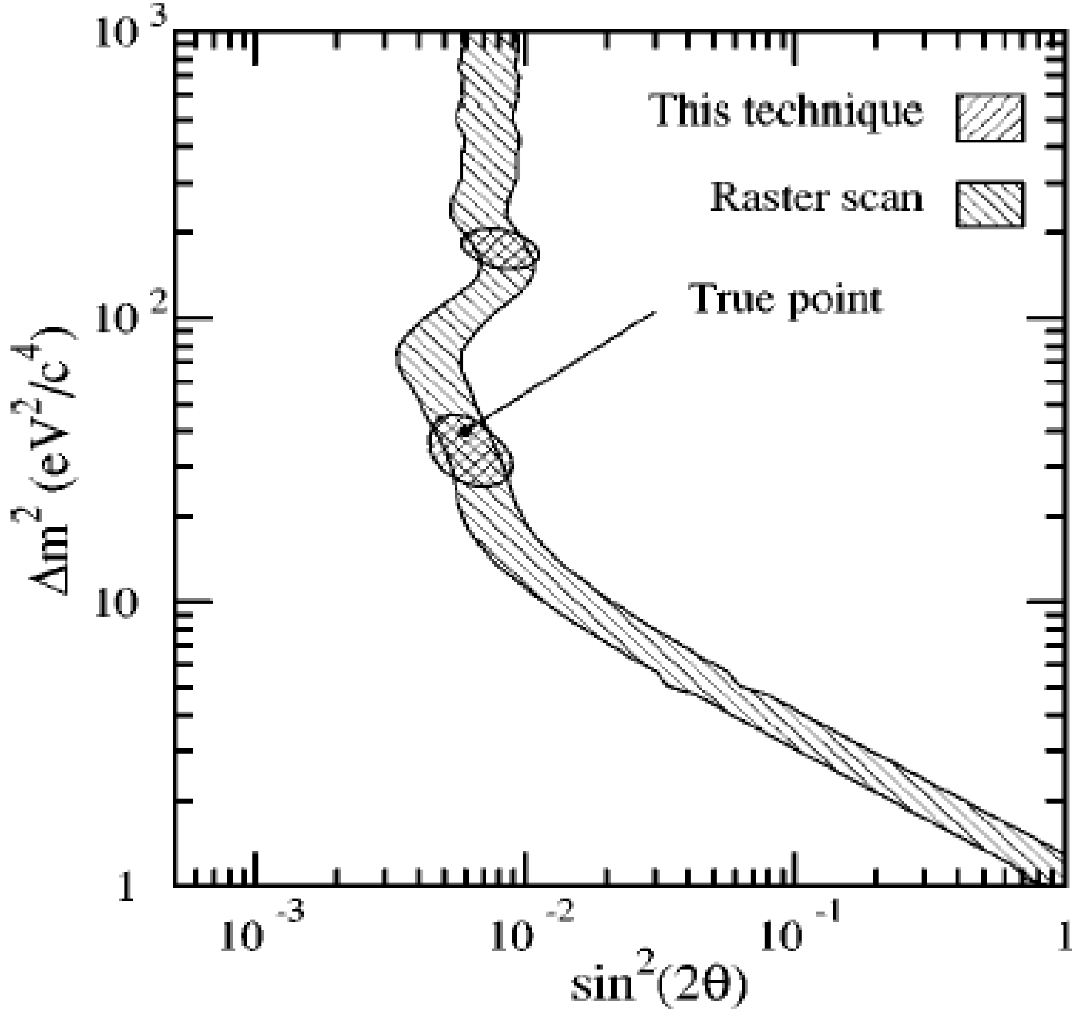


FIG. 12. Calculation of the confidence regions for an example of the toy model in which $\Delta m^2 = 40\ (\mathrm{eV}/c^2)^2$ and $\sin^2(2\theta) = 0.006$, as evaluated by the proposed technique and the raster scan.

This is a really a good result!

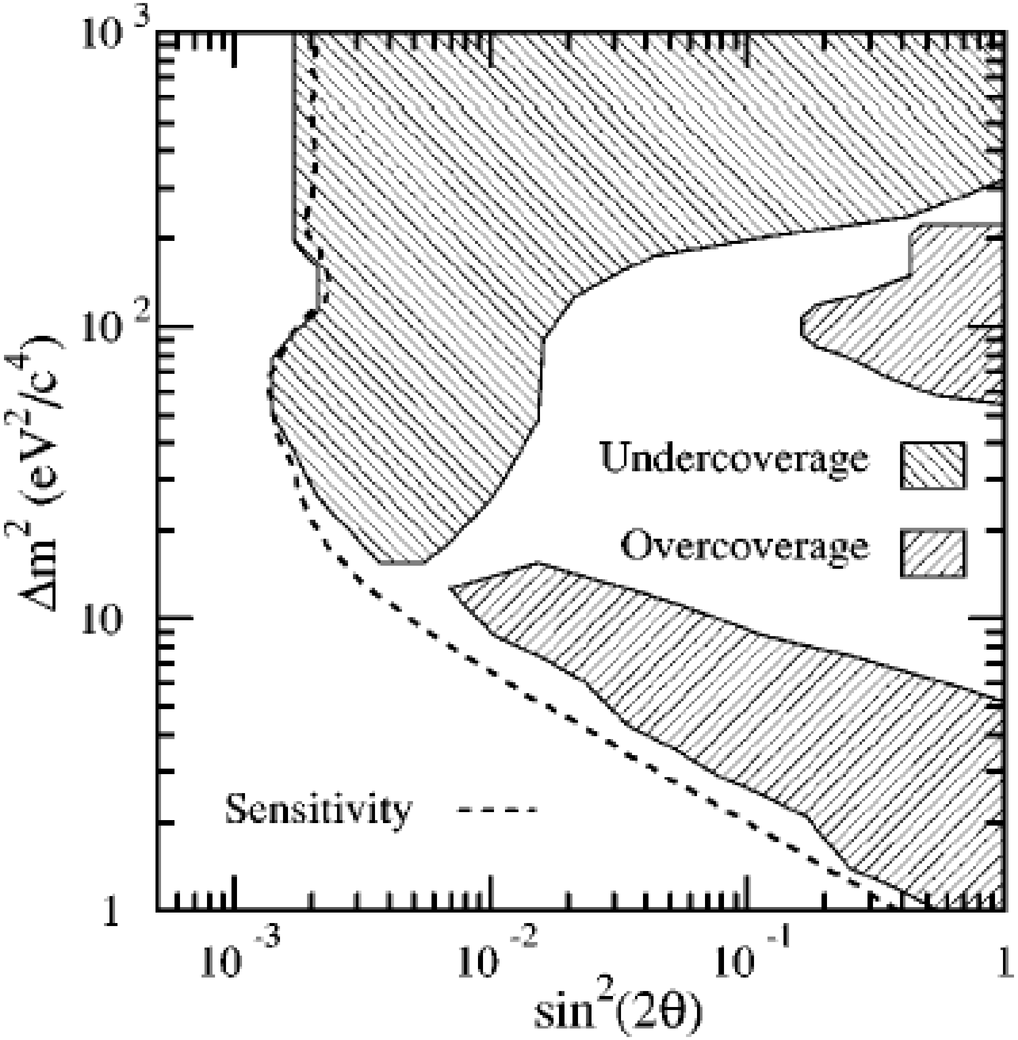


FIG. 14. Regions of significant under- and overcoverage for the global scan.

TABLE XI. Properties of the proposed technique for setting confidence regions in neutrino oscillation search experiments and three alternative classical techniques defined in the text.

Technique	Always gives useful results	Gives proper coverage	Is powerful
Raster scan		√	
Flip-flop raster scan			
Global scan			√
Proposed technique	√	√	√

Some frequentist problems III

A lifetime measurement gives

$$t = 1 \text{ s}$$

Evaluate the lifetime λ with $\alpha \equiv CL = 0.683$.
We use the method of the pivot quantity
(which is in this case $q = \lambda t$):

$$p(t;\lambda)\,\mathrm{d}t = \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \mathrm{e}^{-q}\,\mathrm{d}q$$

Hence we use the pivotal equality

$$P\{q_1 < q(t;\lambda) < q_2\} = P\{\lambda_1(q_1) < \lambda < \lambda_2(q_2)\} = \alpha = CL$$

Then, the interval is

$$\int_{q_1}^{q_2} \mathrm{e}^{-q}\,\mathrm{d}q = \mathrm{e}^{-q_1} - \mathrm{e}^{-q2} = \alpha \equiv CL$$

$$\begin{aligned} q_2 &= -\ln(\mathrm{e}^{-q_1} - \alpha) \\ \frac{q_1}{t} &< \lambda < \frac{q_2}{t} = -\frac{\ln(\mathrm{e}^{-q_1} - \alpha)}{t} \end{aligned}$$

$\partial/\partial q_1 = 0$ **gives the minimum interval as**

$$0 < \lambda \leq -\frac{1}{t}\ln(1 - CL)$$

$$0 < \lambda \leq 1.15 \quad \text{s}^{-1}$$

A Paradox: the lifetime measurement II

$$0 < \lambda \le -\frac{1}{t}\ln(1-CL)$$

$$0 < \lambda \le 1.15 \quad \mathrm{s}^{-1}$$

If one works in terms of $\tau = 1/\lambda$ and repeatsthe same procedure, finds:

$$\frac{t}{-\ln(\mathrm{e}^{-q_1}-\alpha)} \le \tau \le \frac{t}{q_1}$$

$\partial/\partial q_1 = 0$ **gives the minimum interval as**

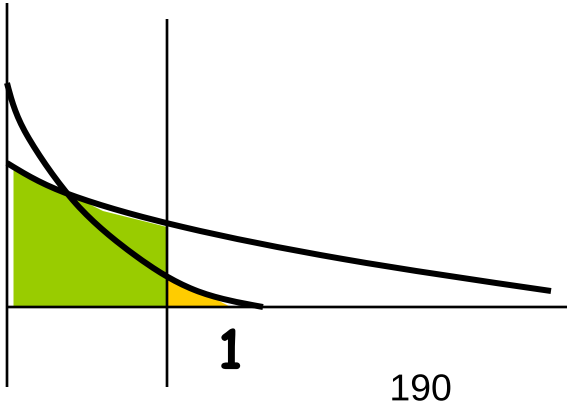
$$0.15 < \lambda \le 2.65 \quad \mathrm{s}$$

corresponding to

$$0.377 < \lambda \le 5.88 \quad \mathrm{s}^{-1}$$

Note that both the λ intervals give the right CL!!

$$\int_0^1 1.15\,\mathrm{e}^{-1.15t}\,\mathrm{d}t = 0.683$$



$$\int_1^\infty 5.88\,\mathrm{e}^{-5.88t}\,\mathrm{d}t = 0.0027 \;, \quad \int_0^1 0.377\,\mathrm{e}^{-0.377t}\,\mathrm{d}t = 0.3140$$

$$1 - 0.3140 - 0.0027 = 0.683$$

A Paradox: the lifetime measurement III

The likelihood ratio is invariant w.r.t. any changement of variable (fundamental theorem).
In this case, for a 1σ interval

$$\ln \hat{L} - \ln L = 0.5$$

we have to find the $\hat{\lambda}$ value

$$\frac{\partial}{\partial \lambda} \lambda \, \mathrm{e}^{-\lambda} = 0 \quad \rightarrow \quad 1 - \lambda^2 = 0 \quad \rightarrow \hat{\lambda} = 1$$

The confidence interval is the solution of the equation

$$\ln \mathrm{e}^{-1} - \ln \lambda \mathrm{e}^{-\lambda} = 0.5 \;\; \rightarrow \quad \lambda - \ln \lambda = 1.5$$

The solution is

$$\boxed{0.301 \leq \lambda \leq 2.358 \quad \mathrm{s}^{-1}}$$

A Paradox: the lifetime measurement IV

$$0.301 \le \lambda \le 2.358 \quad \mathrm{s}^{-1}$$

$$\int_1^\infty 2.358\,\mathrm{e}^{-2.358t}\,\mathrm{d}t = 0.0946 \;, \quad \int_0^1 0.301\,\mathrm{e}^{-0.301t}\,\mathrm{d}t = 0.2599$$

$$1 - 0.2599 - 0.0946 = 0.645$$

The coverage is wrong

Some physicists use the LR technique instead of the frequentist or Bayesian approaches (see G. Zech, EPJ direct C12, 1-81 (2002))

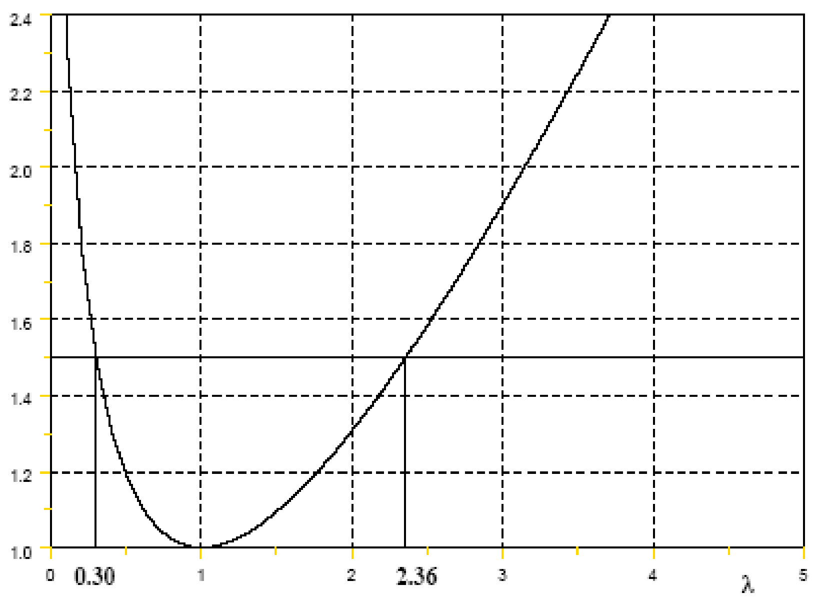


Figure 1: $\lambda - \ln\lambda = 1.5$

Frequentism, Bayes or likelihood ratio??

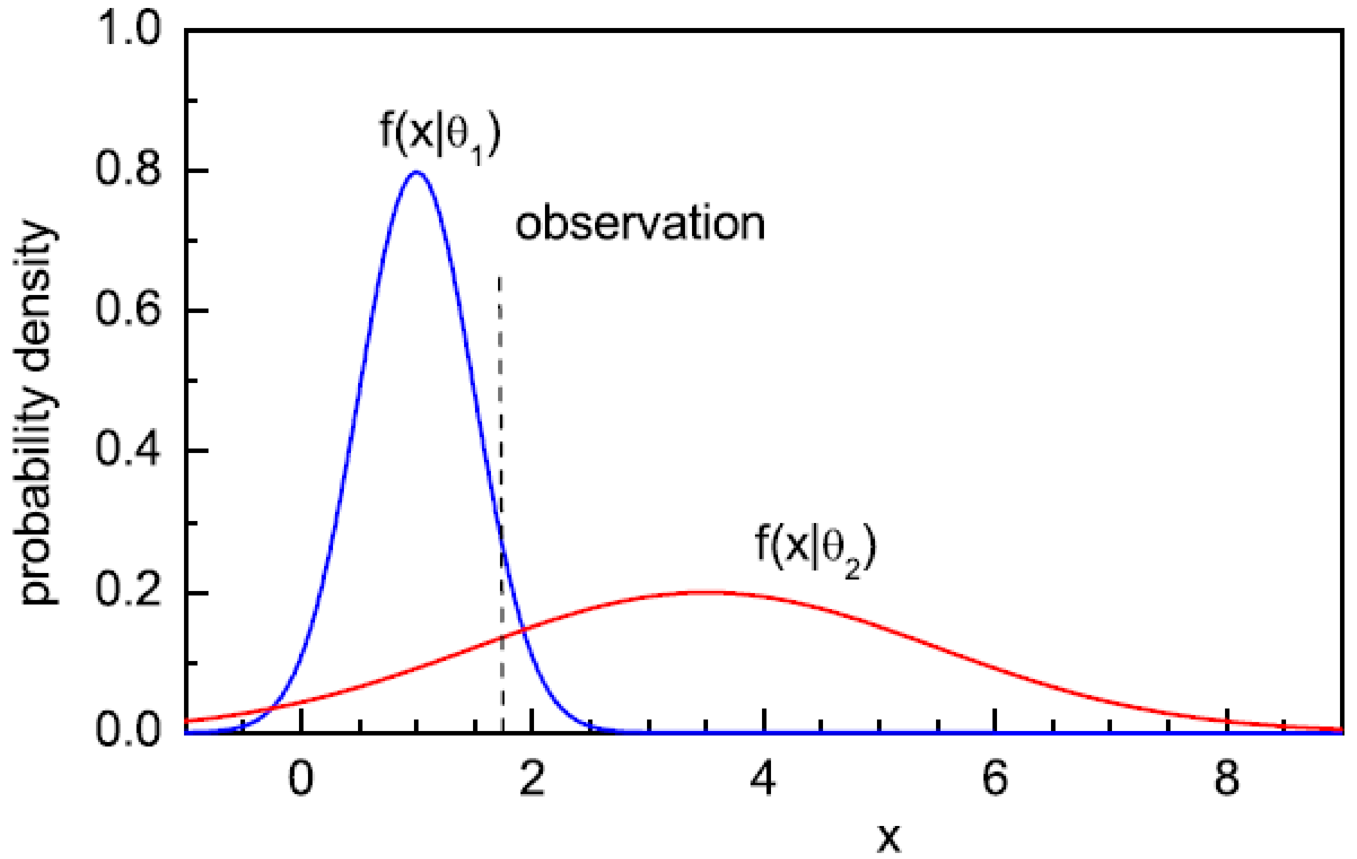
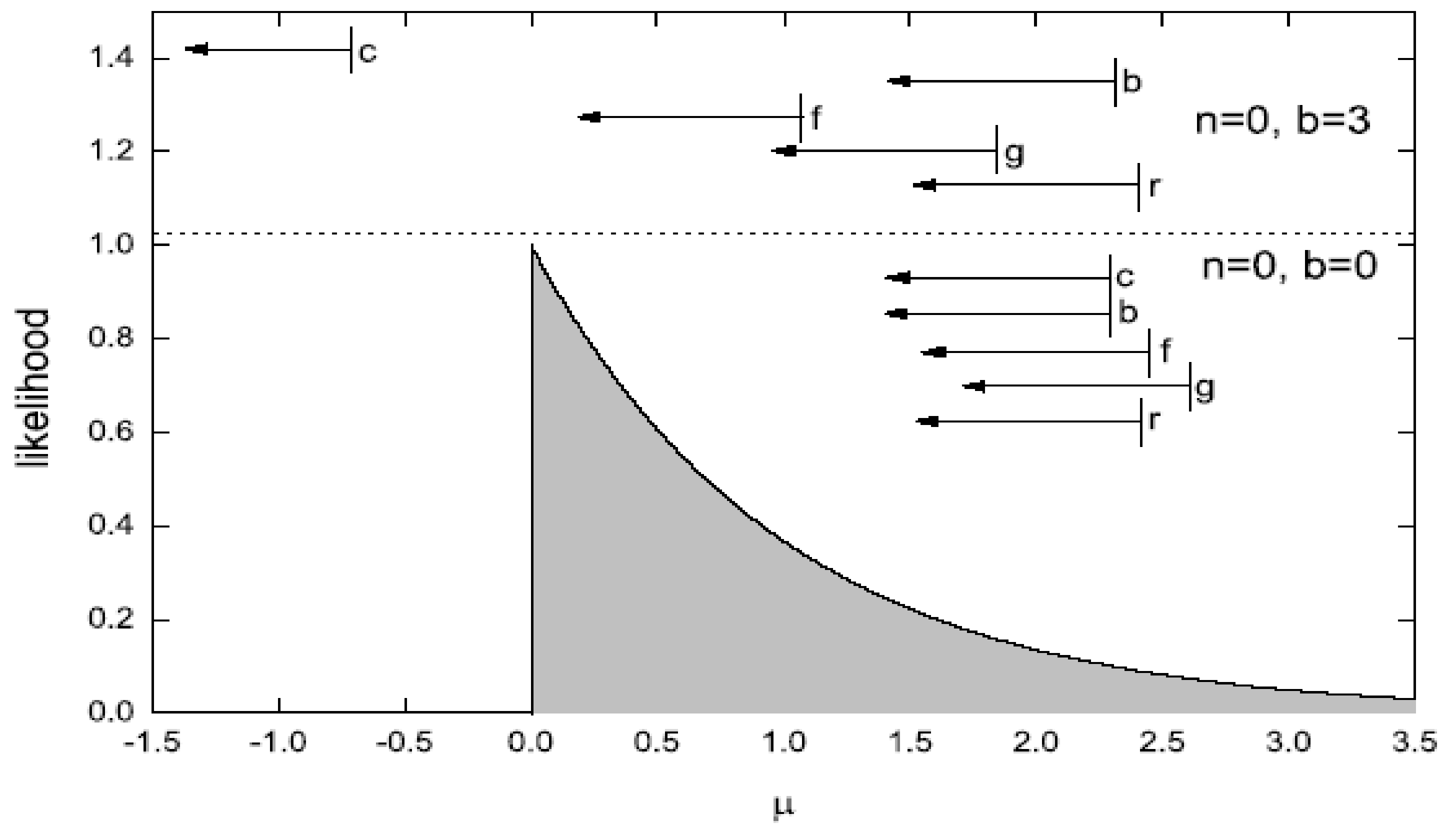


Fig. 1. The likelihood is larger for parameter θ_1, but the observation is less then 1 st. dev. off θ_2. Classical approaches include θ_2 and exclude θ_1 within a 68.3% confidence interval

Draw the conclusions by yourself .....

Table 4. Comparison of different approaches to define error intervals, see text

method:	classical	unified	likelihood	Bayesian u.p.	Bayesian a.p.
consistency	- -	- -	++	+	+
precision	- -	-	+	+	+
universality	- -	- -	-	+	++
simplicity	-	- -	++	+	+
variable transform.	-	++	++	- -	- -
nuisance parameter	-	-	-	+	+
error propagation	-	-	+	+	+
combining data	-	-	++	+	-
coverage	+	++	- -	- -	- -
objectivity	-	-	++	+	-
discrete hypothesis	-	-	+	+	+



FC unified

Fig. 16. Likelihood function for zero observed events and 90% confidence upper limits with and without background expectation. The labels refer to [9] (f), Bayesian (b), [41] (g) and [42] (r) classical (c)

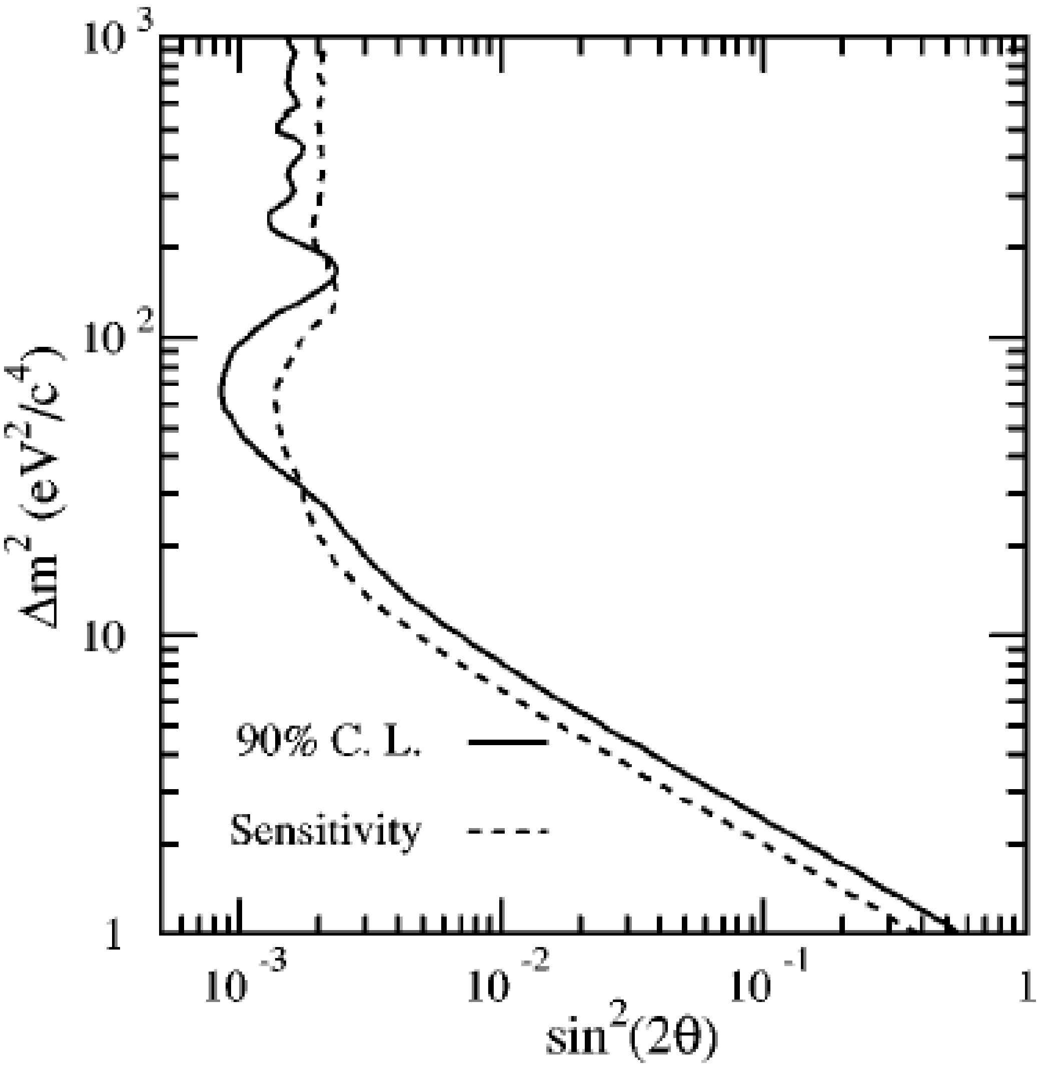


FIG. 15. Comparison of the confidence region for an example of the toy model in which $\sin^2(2\theta)=0$ and the sensitivity of the experiment, as defined in the text.

1. Calcolo delle probabilità
2a. Statistica frequentista
2b. Statistica bayesiana
3. Likelihood
4a. Fondo e segnale
4b. Metodi Bootstrap
5. Approccio Unificato
6. ***Unfolding***

Alberto Rotondi
Università di Pavia

Folding theorem

$$Z = f(X_1, X_2) \ .$$

$Z_1 \equiv Z$ e $Z_2 = X_2$ auxiliary variable

$$Z_1 = f(X_1, X_2) \ , \qquad Z_2 = X_2 \ .$$

The jacobian is

$$|J| = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial z_1} & \frac{\partial f_1^{-1}}{\partial z_2} \\ 0 & 1 \end{vmatrix} = \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| \ ,$$

From the general theorem one obtains

$$p_{\mathbf{Z}}(z_1, z_2) = p_{\mathbf{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| \ .$$

by integrating on the auxiliary variable :

$$p_{Z_1}(z_1) = \int p_{\mathbf{Z}}(z_1, z_2)\, \mathrm{d}z_2 \ .$$

hence

$$\begin{aligned} p_Z(z) &= \int p_{\mathbf{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z} \right| \mathrm{d}x_2 \\ &= \int p_{\mathbf{X}}\left(f_1^{-1}(z, x_2), x_2\right) \left| \frac{\partial f_1^{-1}}{\partial z} \right| \mathrm{d}x_2 \ , \end{aligned}$$

which is the probability density

Convolution theorem

For independent variables:

$$p_Z(z) = \int p_{X_1}\left(f_1^{-1}(z, x_2)\right) \, p_{X_2}(x_2) \frac{\partial f_1^{-1}}{\partial z} \, \mathrm{d}x_2 \; .$$

when Z is given by the sum

$$Z = X_1 + X_2 \; ,$$

we have

$$X_1 = f_1^{-1}(Z, X_2) = Z - X_2 \; , \qquad \frac{\partial f_1^{-1}}{\partial z} = 1 \; ,$$

and we obtain

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{\boldsymbol{X}}(z - x_2, x_2) \, \mathrm{d}x_2 \; .$$

When X_1 and X_2 are independent, we obtain the convolution integral

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{X_1}(z - x_2) \, p_{X_2}(x_2) \, \mathrm{d}x_2 \; ,$$

In physics

(δ instrument function, f signal) :

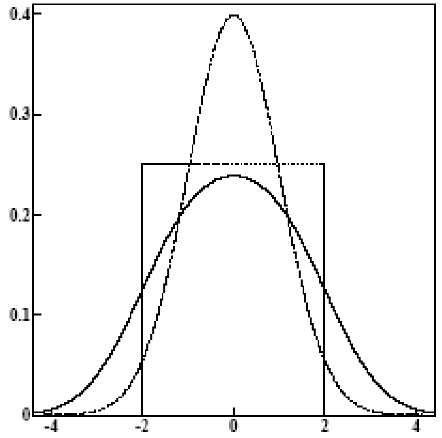
$$g(y) = \int_{-\infty}^{+\infty} f(y - x) \, \delta(x) \, \mathrm{d}x \; ,$$

Uniform*Gaussian

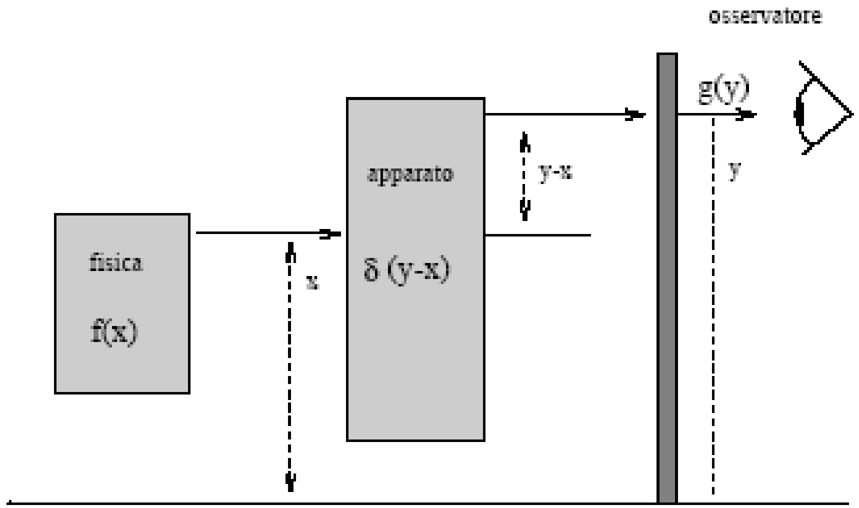
When $Z = X + Y$ **where**
$X \sim N(\mu, \sigma^2)$ e $Y \sim U(a, b)$.
one has immediately

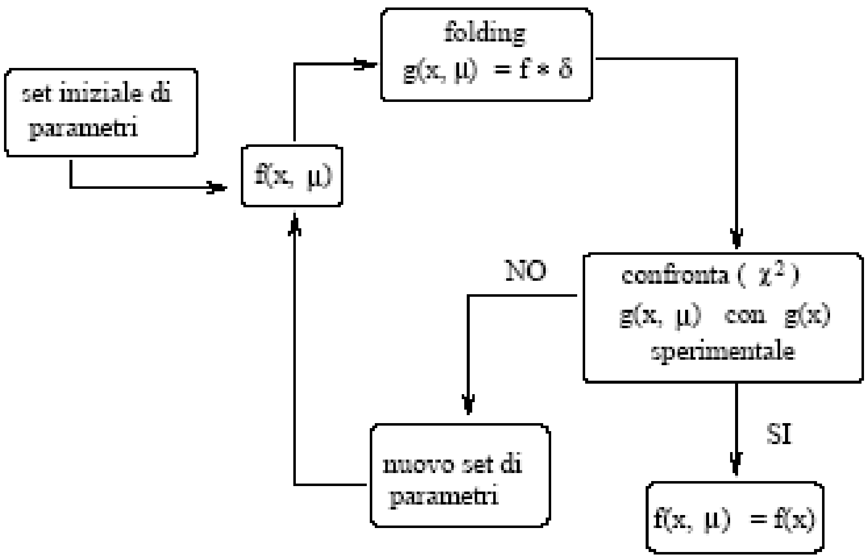
$$\begin{aligned} p_Z(z) &= \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z-y-\mu)^2}{2\sigma^2}\right] dy \\ &= \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[y-(z-\mu)]^2}{2\sigma^2}\right] dy \ . \end{aligned}$$

$$p_Z(z) = \frac{1}{b-a}\left[\Phi\left(\frac{b-(z-\mu)}{\sigma}\right) - \Phi\left(\frac{a-(z-\mu)}{\sigma}\right)\right]$$



$$g(y) = \int f(x)\, \delta(y - x)\, \mathrm{d}x \ ,$$





)1

The 1D problem

In the reconstruction of an histogram,

- the true histogram (image) where the bin contents are the expected values

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N) \ , \quad \mu_j = \mu_{\rm tot} p_j = \mu_{\rm tot} \int_{\rm bin\ i} f_t(y)\, {\rm d}y$$

The problem in 2-D

A picture in a $x - y$ plane is the result of a double dimensional folding, where the true points are smeared out by detector effects.

$$N = \sum_{ij=1}^{n_c} N_{ij}(\mathrm{exp}) \ , \tag{93}$$

N is the total number of events and $N_{ij}(\mathrm{exp})$ is the recorded number of event in the pixel placed at the ith-row and jth-column.

The observed $N_{ij}(\mathrm{exp})$ events have to be compared with the expected values $N_{ij}(\mathrm{th})$ predicted by a model.

$$N_{ij}(\mathrm{th}) = N P_{ij}(\mathrm{obs}) = N \sum_{i'j'} P_{i'j'}(\mathrm{true}) \, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) \ , \tag{94}$$

that is, the number of events observed in the ijth-cell is due to the presence into the $i'j'$th-cell, times the probability P_v that the PSF shifts the point from the $i'j'$ to the ij-cell. One has to sum on all the cells near the $ij-$one.

In the case of a two dimensional Gaussian point spread function PSF:

$$P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) = \frac{1}{2\pi \, \sigma_x \sigma_y} \exp\left[-\frac{(x_{ij} - x_{i'j'})^2}{2\,\sigma_x^2} - \frac{(y_{ij} - y_{i'j'})^2}{2\,\sigma_y^2} \right] , \tag{96}$$

Fourier techniques

$$f(x) = \int F(t)\ \mathrm{e}^{2\pi i\,xt}\ \mathrm{d}t$$

Convolution:

$$f(x) = \int g(y)\delta(x-y)\,\mathrm{d}y$$

$$\int F(t)\,\mathrm{e}^{2\pi ixt}\ \mathrm{d}t = \int G(t)\ \mathrm{e}^{2\pi i\,yt}\ \Delta(t)\,\mathrm{e}^{2\pi i\,(x-y)t}\ \mathrm{d}t$$

$$\int F(t)\,\mathrm{e}^{2\pi i\,xt}\ \mathrm{d}t = \int G(t)\ \Delta(t)\,\mathrm{e}^{2\pi i\,xt}\ \mathrm{d}t \to F(t) = G(t)\,\Delta(t)$$

Correlation

$$\mathrm{Corr}(g,\delta) \equiv \int g(x+y)\,\delta(y)\,\mathrm{d}y \to G(t)\Delta^*(t)$$

if the functions are real

$$G(t) = G(-t)^*\ , \quad \mathrm{Corr}(g,\delta) \to G(t)\Delta(-t)$$

Autocorrelation (Wiener theorem)

$$\mathrm{Corr}(g,g) \to |G(t)|^2$$

Total Power:

$$P(f) \equiv \int |f(x)|^2\,\mathrm{d}x = \int |F(t)|^2\,\mathrm{d}t$$

Power Spectral Density (in the Fourier space):

$$PSD(f) \equiv |F(t)|^2 + |F(-t)|^2 \quad \stackrel{\mathrm{f(x)\ real}}{\longrightarrow} 2|F(t)|^2 \quad 0 \le t \le \infty$$

Image Deconvolution

$$D(\boldsymbol{x}) = \int \mathrm{d}\boldsymbol{y} I(\boldsymbol{y})\, \delta(|\boldsymbol{x} - \boldsymbol{y}|)$$

In the absence of noise

$$I = F^{-1}\left[\frac{F(D)}{F(\delta)}\right]$$

where F is the Fourier transform.

For a real image $I(n_1, n_2)$ the Fourier transform is:

$$F(k_1, k_2) = \sum_{n_2=0}^{N_2-1} \sum_{n_1=0}^{N_1-1} \mathrm{e}^{2\pi i k_2 n_2/N_2}\, \mathrm{e}^{2\pi i k_1 n_1/N_1}\; I(n_1, n_2)$$

$$F(k_1, k_2) = FFT_2[FFT_1[I(n_1, n_2)]]$$

For the routines see for example *Numerical Recipes*

The problem with fluctuations

original

Gaussian smearing

Poisson statistics

Fourier (un)restored

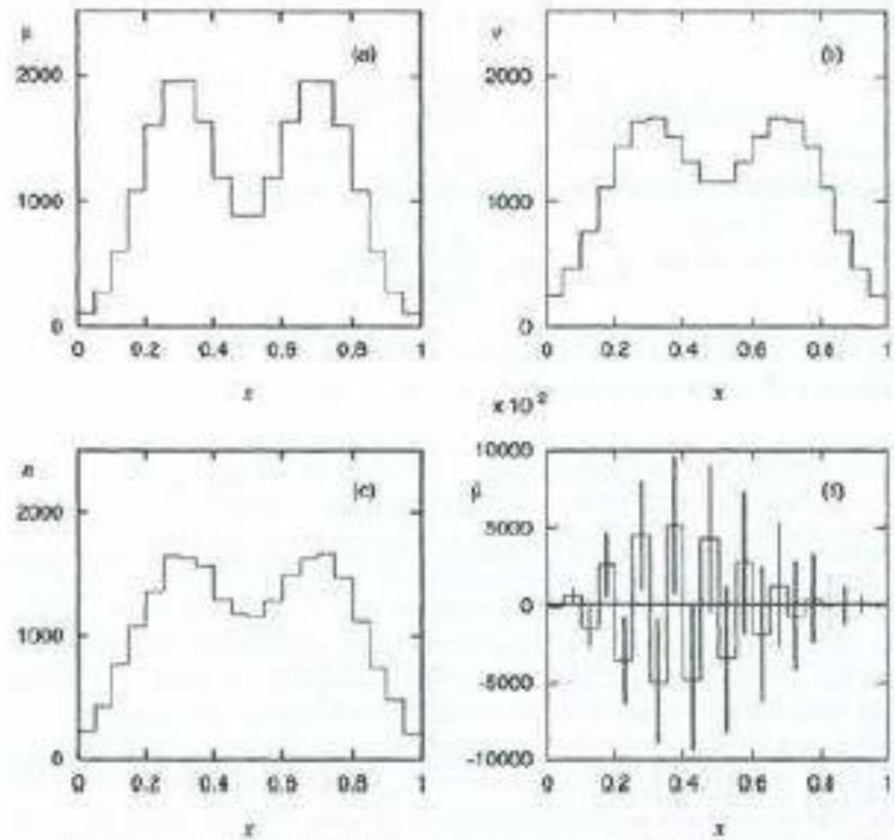


Fig. 11.1 (a) A hypothetical true histogram μ, (b) the histogram of expectation values $\nu = R\mu$, (c) the histogram of observed data n, and (d) the estimators $\hat{\mu}$ obtained from inversion of the response matrix.

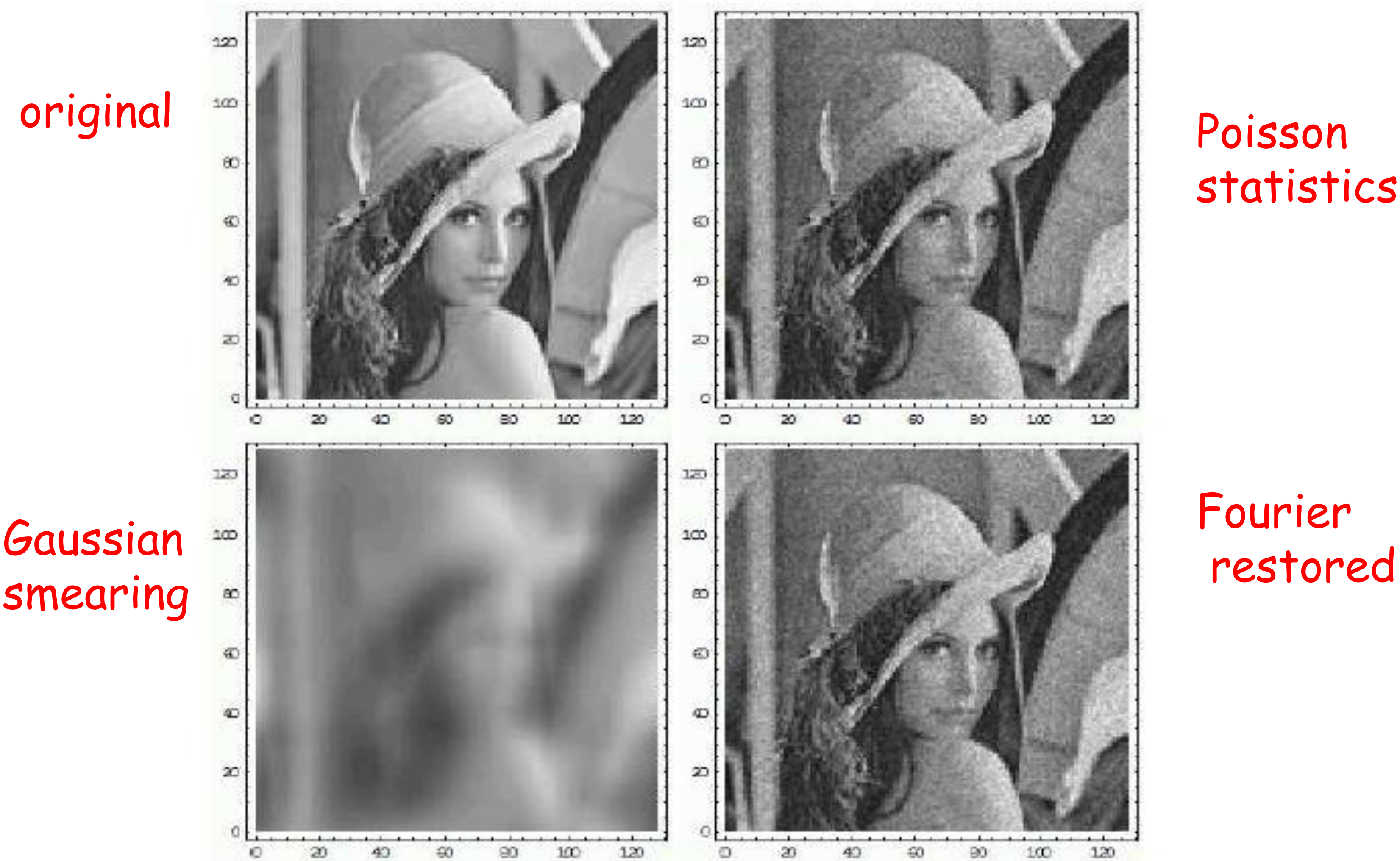


Figure 11: **Lena restored by FFT: The original image (top left) is sampled with Poisson statistics (top right) and smeared with a 2D 10-bins Gaussian PSF (bottom left): the Fourier restored image (bottom right) is similar to the Poisson sampled image. In this case the noise term N is neglected.**

original

Gaussian smearing

Poisson statistics

Fourier (un)restored

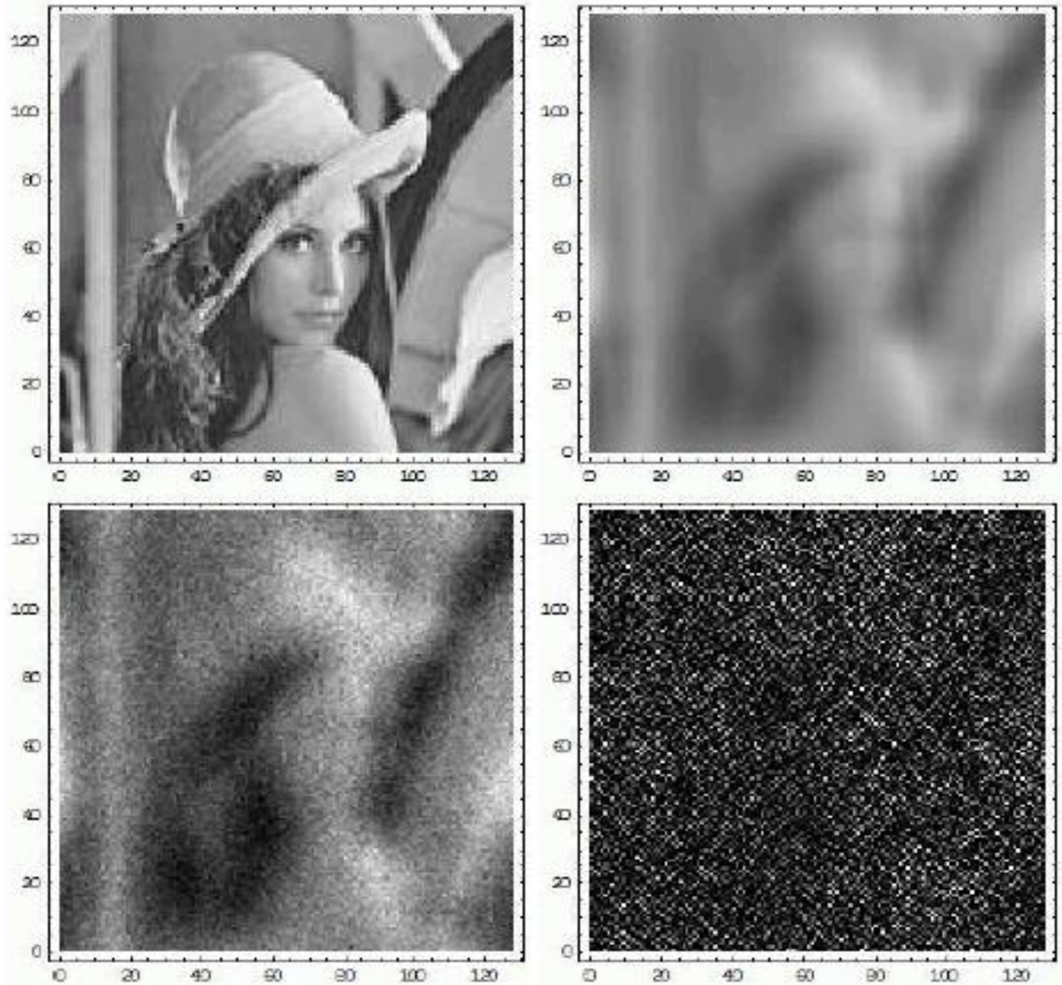


Figure 12: **Lena not restored by FFT: In this case the noise term N is not ignored: the original image (top left) is smeared with a 2D 10-bins Gaussian PSF (top right) and the result is sampled with Poisson statistics (bottom left): the Fourier restored image (bottom right) cannot recover the information lost in the noise. Another approach, statistical in nature, is required.**

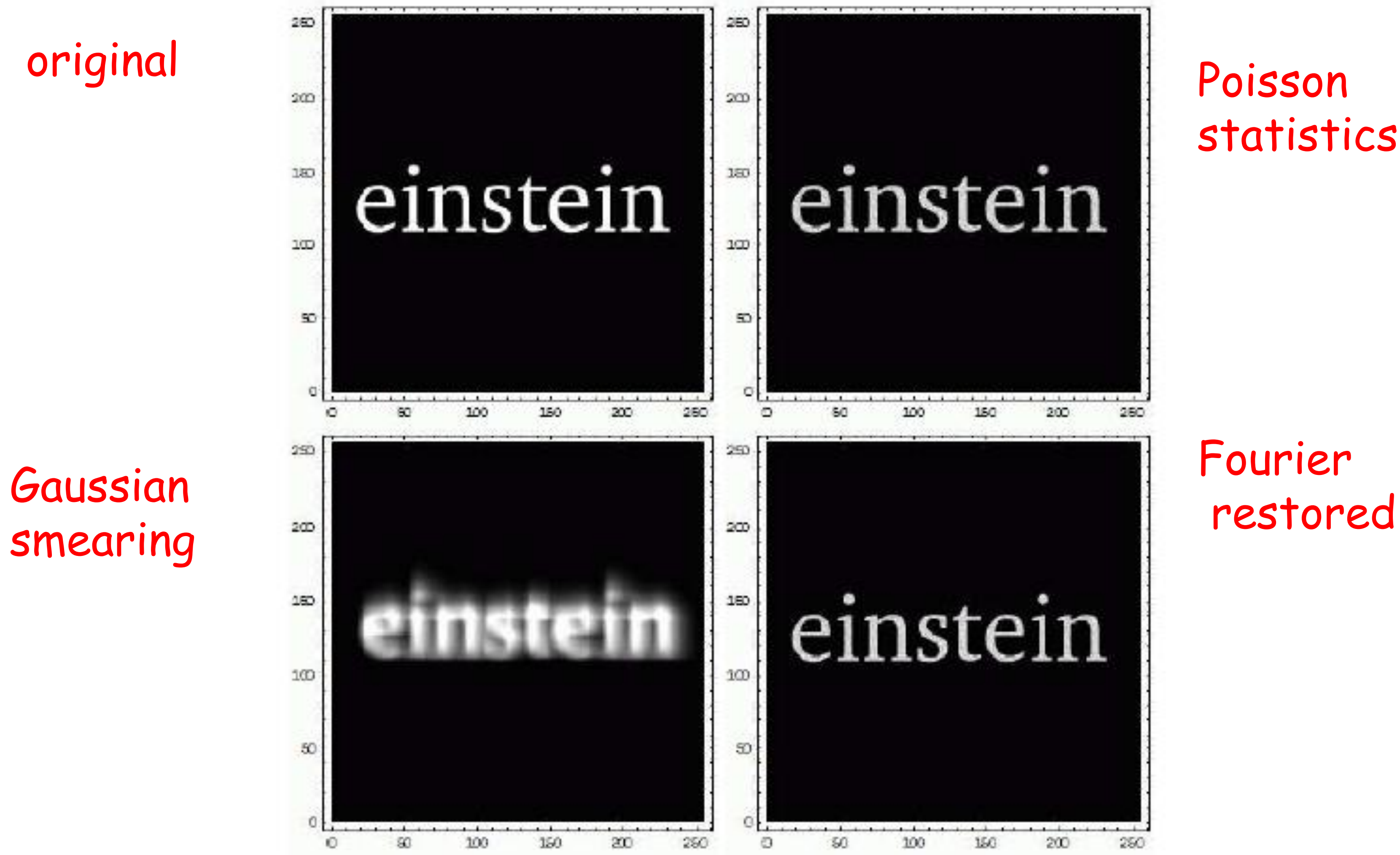


Figure 13: **Einstein restored by FFT: explanation as in Figure 1.**

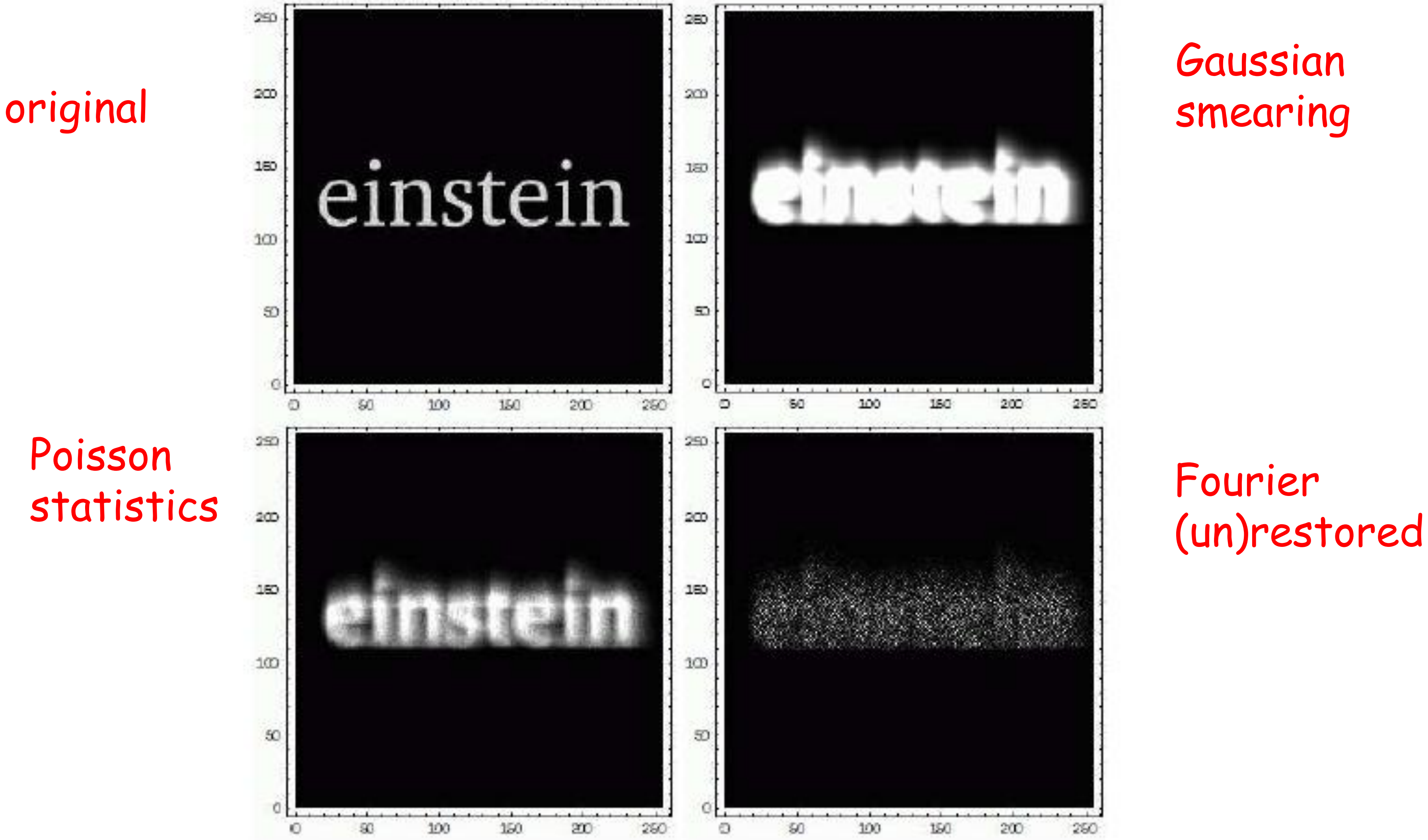


Figure 14: **Einstein not restored by FFT: explanation as in Figure 2.**

When true $\rightarrow$ obs $\equiv \mu \rightarrow \nu$ deterministic methods (FFT) can be used

$$\boldsymbol{\nu} = R * \boldsymbol{\mu} \tag{19}$$

Image restoration

When true $\rightarrow$ smeared $\rightarrow$ obs $\equiv \mu \rightarrow \nu \rightarrow n$ statistical methods must be used

$$\boldsymbol{n} = \boldsymbol{\nu} + \boldsymbol{\rho} = R * \boldsymbol{\mu} + \boldsymbol{\rho}$$

In the poissonian or binomial case we have to minimize:

$$-\ln L(\boldsymbol{\mu}) = -\sum_i \ln P(n_i, \nu_i)$$

In the gaussian case we must minimize

$$\chi^2(\boldsymbol{\mu}) = \sum_{ij} (\nu_i - n_i)(V^{-1})_{ij}(\nu_j - n_j)$$

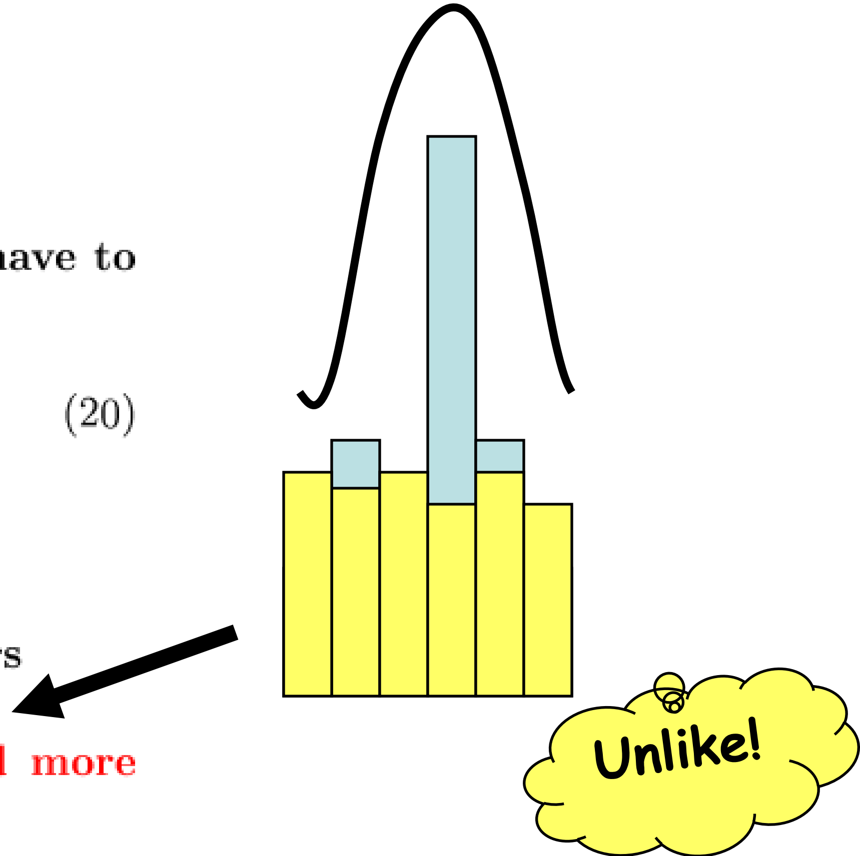
In 2-D, when $N_{ij}(\exp)$ contains fluctuations, we have to minimize:

$$-2 \ln L(\boldsymbol{n}|\boldsymbol{\nu}, \boldsymbol{\mu}) \simeq \chi^2 = \sum_{ij} \frac{[N_{ij}(\exp) - N P_{ij}(\text{obs})]^2}{N P_{ij}(\text{obs})} \tag{20}$$

where

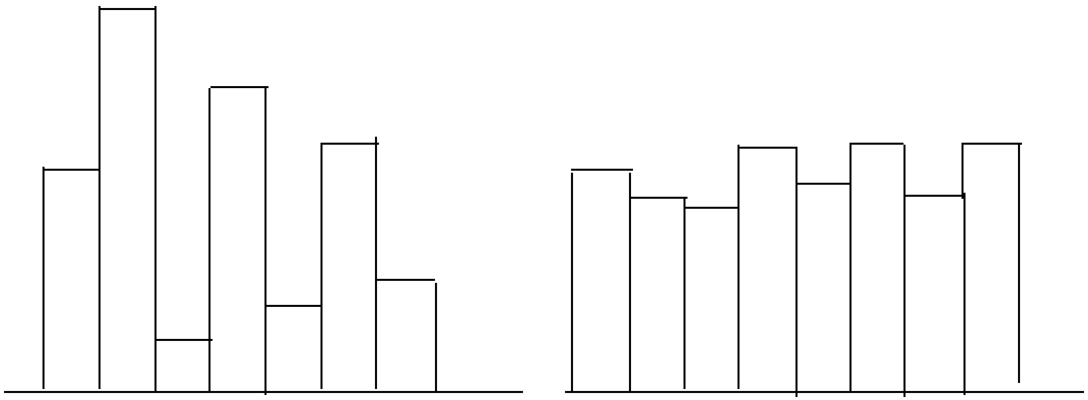
$$P(\text{obs}) = P(\nu|\mu)\, P(\mu)$$

If all the pixel contents $\boldsymbol{\mu}$ are the free parameters to be determined the problem has zero DoF



This is a ILL-POSED problem with many (and more probable) unrealistic solution!!

Explanation:



spike solution
HIGLY PROBABLE

smooth solution
UNLIKE

The smeared distributions of two input distributions cannot be distinguished if they agree on a large scale of x but differ by oscillations on a "microscopic" scale much smaller than the experimental resolution

many solutions give a good χ^2

the spike ones are more probable!

Cure: to add to χ^2 an empirical regularization term $C[p]$.

$$\chi^2 \rightarrow \alpha\,\chi^2 + C[P(\text{true})]$$

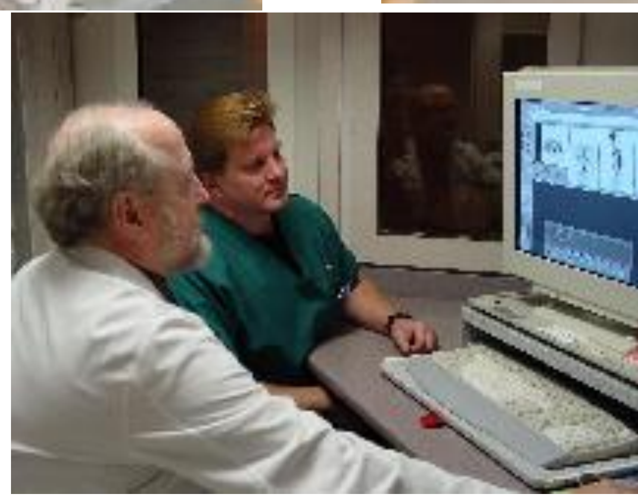
or

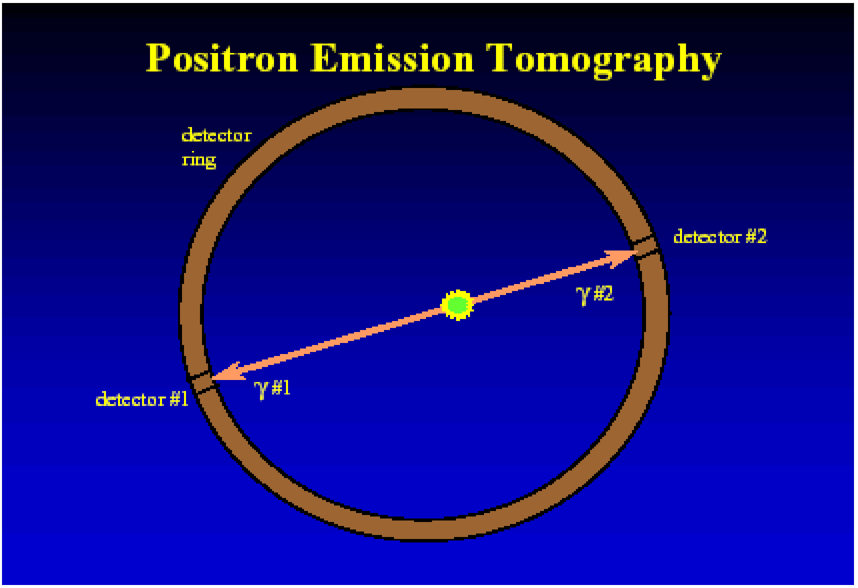
$$\chi^2 \rightarrow \chi^2 + \alpha C[P(\text{true})]$$

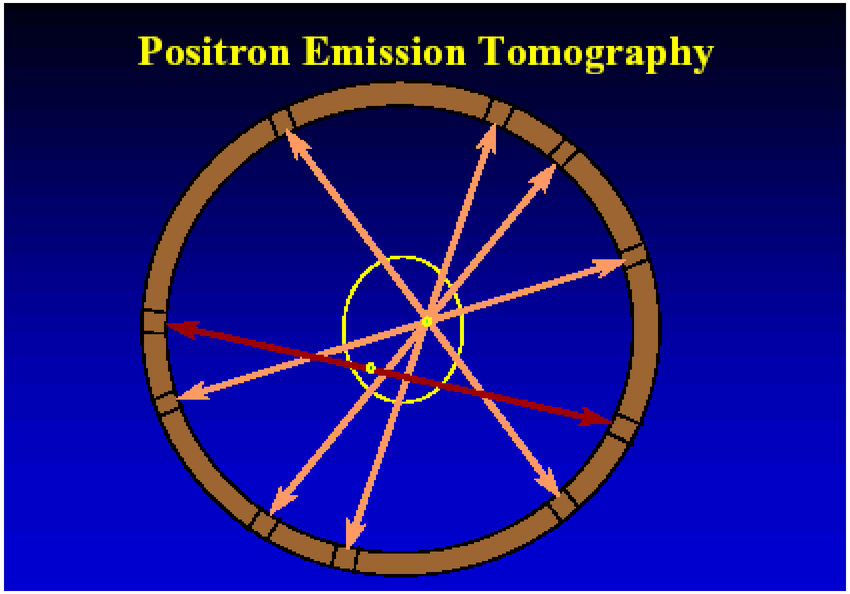
or
to increase the DoF by using a parametric model

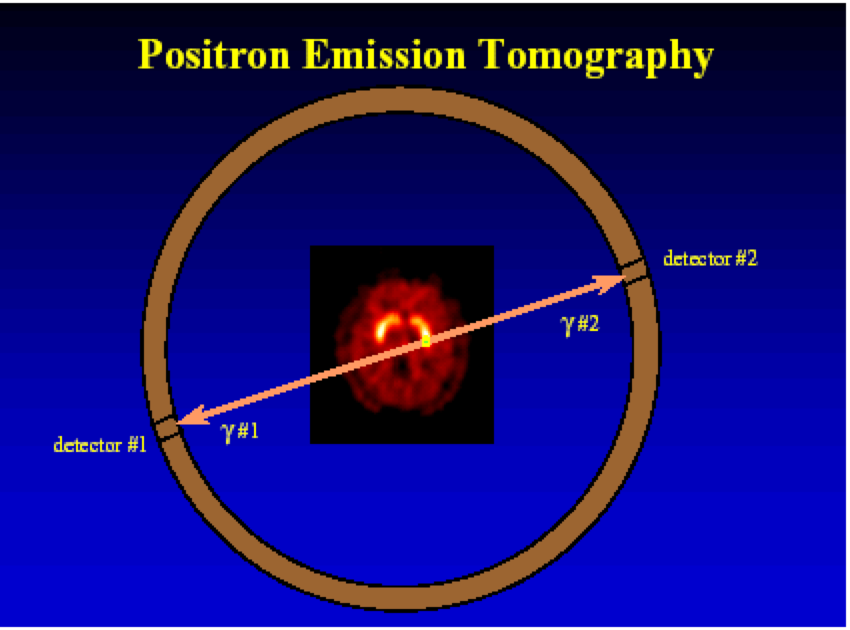
$$P(\nu\,|\,\mu)P(\mu) \rightarrow P(\nu\,|\,\mu')$$

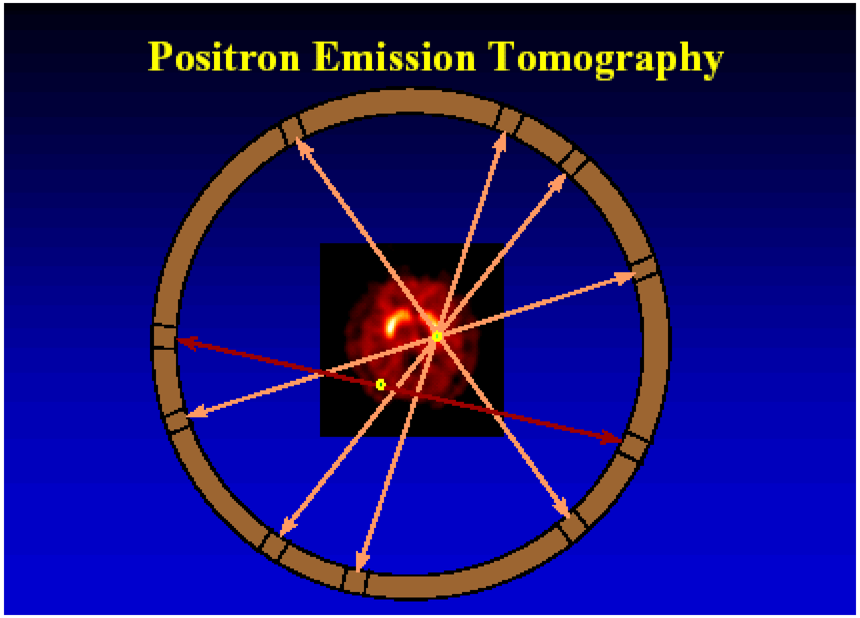
PET: positron emission thomography




Positron Emission Tomography
detector
ring
detector #2
γ #2
detector #1
γ #1


Positron Emission Tomography


Positron Emission Tomography
detector #2
γ #2
detector #1
γ #1


Positron Emission Tomography

Remember

$$\mu_i \xrightarrow{PSF} \nu_i \xrightarrow{random} n_i$$

and consider (95) as a form of the Bayes theorem

$$P(\text{true}_{ij}|\text{obs}) \propto \sum_{i',j'} P_{i'j'}(\text{true})\, P_v(\text{obs}_{ij}|\text{true}_{i'j'}) = L(\boldsymbol{n}|\boldsymbol{\mu})P(\boldsymbol{\mu})$$

Bayesians say: posterior = likelihood × prior One maximizes $P(\text{true}_{ij}|\text{obs}) \equiv F(\boldsymbol{\mu})$ (or minimize $-F(\boldsymbol{\mu})$):

$$F(\boldsymbol{\mu}) = \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + \ln P(\boldsymbol{\mu}) \tag{99}$$

following the Maximum Likelihood (ML) principle. The practical (no Bayesian) experimentalist introduces an empirical regularization parameter α and considers the prior $P(\boldsymbol{\mu})$ as a regularization function $C(\boldsymbol{\mu})$:

The frequentist assumes

$$P(\mu) = 1$$

$$F(\boldsymbol{\mu}) = \alpha \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + C(\boldsymbol{\mu}) \tag{100}$$

By keeping fixed the normalization:

$$\nu_T = \sum_i \sum_j R_{ij}\hat{\mu}_j + \rho_i = n_T$$

the objective function is

$$F(\boldsymbol{\mu}) = \alpha \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \tag{101}$$

where λ is a Lagrange multiplier

$$\frac{\partial F}{\partial \lambda} = 0 \rightarrow \sum_i n_i = n_T$$

The objective function to be minimized is

$$-F(\boldsymbol{\mu}) = -2\ln L(\boldsymbol{n}|\boldsymbol{\mu}) - \alpha\, C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \quad (23)$$

$$\mu_j = \mu_{\rm tot} p_j = \mu_{\rm tot} \int_{\rm bin\, j} f_t(y)\, {\rm d}y$$

where $\alpha > 0$. Some regularization terms:

Regularization terms

- minimum second derivative (Tichonov)

$$C(\boldsymbol{\mu}) = -\int [f_t''(y)]^2\, {\rm d}y \simeq -\sum_{i=1}^{M-2} [-\mu_i + 2\mu_{i+1} - \mu_{i+2}]^2$$

- minimum variance:

$$C(\boldsymbol{\mu}) = -{\rm Var}[\boldsymbol{\mu}] \equiv ||C\boldsymbol{\mu}||^2 = -\sum_i \mu_i^2$$

- maximum entropy (MaxEnt)

$$C(\boldsymbol{\mu}) = -\sum_i p_i \ln p_i = -\sum_i \frac{\mu_i}{\mu_T} \ln \frac{\mu_i}{\mu_T}$$

- cross-entropy

$$C(\boldsymbol{\mu}) = -\sum_i p_i \ln \frac{p_i}{q_i} = -\sum_i \frac{\mu_i}{\mu_T} \ln \frac{\mu_i}{\mu_T\, q_i}$$

where $\boldsymbol{q} = (q_1, q_2, \ldots, q_n)$ is the most likely a priori shape for the true distribution μ_i.

The objective function to be minimized is

$$-F(\boldsymbol{\mu}) = -2\ln L(\boldsymbol{n}|\boldsymbol{\mu}) - \alpha\, C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \quad (24)$$

Some choices of $\alpha > 0$ are:

- Bayesian

$$\alpha = \frac{1}{\mu_T}$$

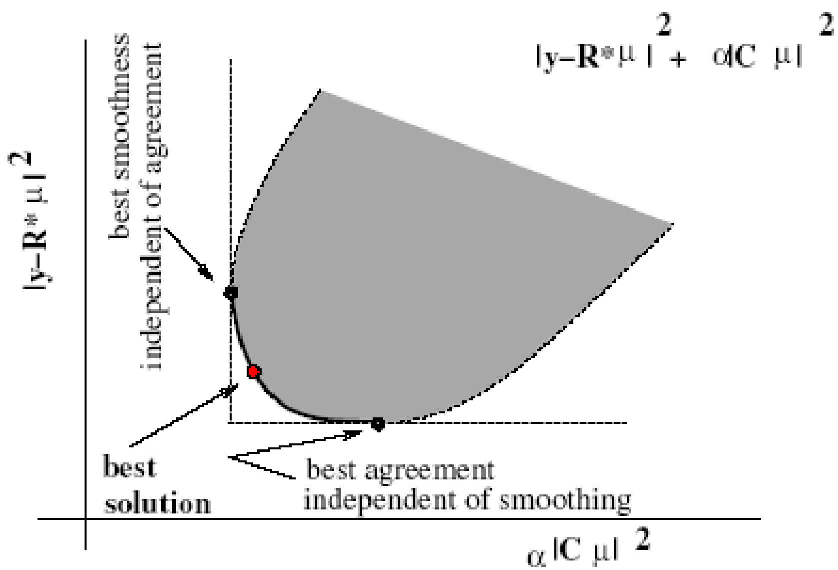
usually too much smoothing

- $A = \chi^2 + \alpha ||C\mu||^2 > 0$

we can regularize the solution by choosing $\chi^2 \simeq DoF \equiv$ number of pixel N, with the condition:

$$\alpha = \frac{A - N}{||C\mu||^2}$$

Regularization parameter



Regularization parameter

Two-peak deconvolution with the regularized ML methods (Glen Cowan, Statistical Data Analysis, Oxford (2000))

Examples of unfolding 181

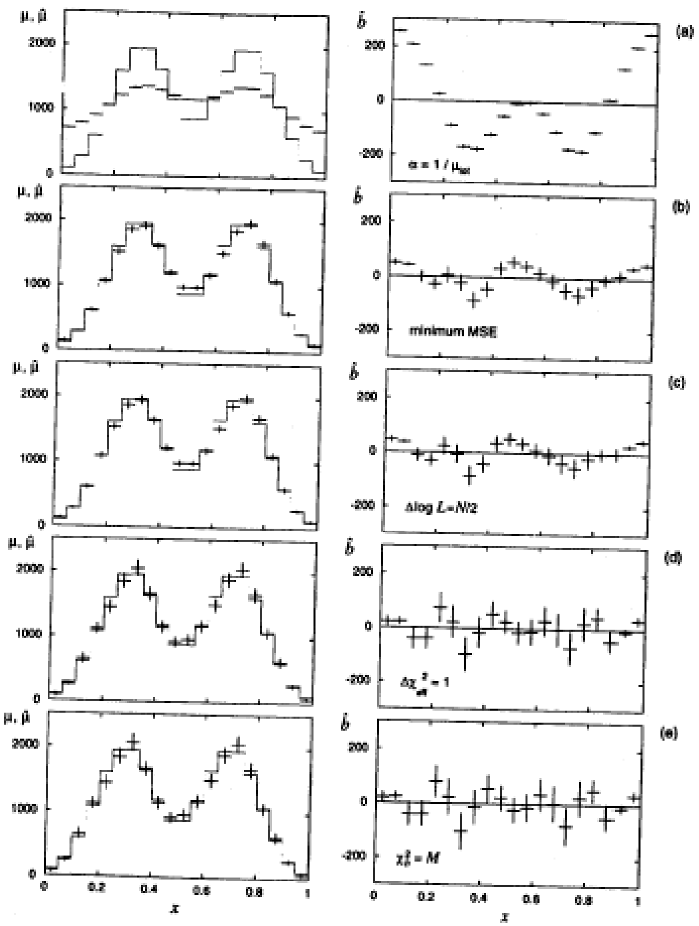
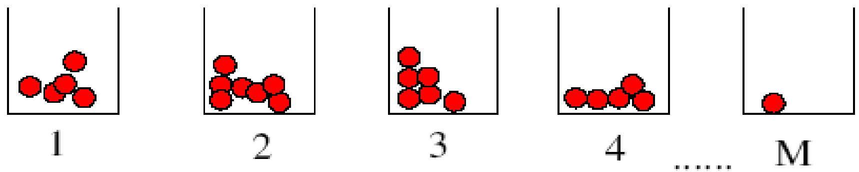


Fig. 11.8 MaxEnt unfolded distributions shown as points with the true distribution shown as a histogram (left) and the estimated biases (right) for different values of the regularization parameter α. The examples correspond to (a) the Bayesian prescription …

We have M boxes and a monkey that throws N ball randomly into them.
What is the box-balls configuration of highest probability? Probability of a configuration:

$$p = \frac{1}{M^N}\frac{N!}{n_1!n_2!\ldots n_N!} = \mathrm{e}^{\ln p}$$

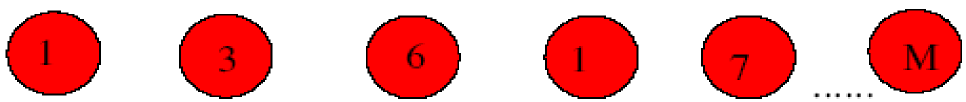
What is MaxEnt ???



M boxes and N balls

equal to

N balls labelled randomly from 1 to M



$$\ln p = -N\ln M + \ln N! - \sum_i \ln(n_i!)$$

Stirling formula: $n! = \sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \rightarrow \ln n! \simeq n\ln n - n$

$$\ln p = -N\ln M + N\ln N - N + \sum n_i - \sum_i n_i \ln n_i$$

$$\sum p_i = 1 \ , \qquad p_i = \frac{n_i}{N} \ , \ \ln p = -N\ln M - N\sum_i p_i \ln p_i$$

The most probable configuration means to maximize

$$\ln p(\boldsymbol{\mu}) = S = -\sum_i p_i \ln p_i$$

$$N_{ij}(\text{th}) = N P_{ij}(\text{obs}) = N \sum_{i'j'} P_{i'j'}(\text{true})\, P_v(\text{obs}_{ij}|\text{true}_{i'j'}) \quad (25)$$

We write this equation considering the operator R:

$$n = R * \mu$$

The iterative method (Van Cittert 1930) adds with a weight β the residual r_k to the current solution

$$\mu_{k+1} = \mu_k + \beta[n - R * \mu_k] \quad (26)$$

The method is based on the known equation

$$\sum_{i=0}^{k} q^n = \frac{1 - q^{k+1}}{1 - q} \quad (27)$$

for $k \to \infty$ the series converges if $|q < 1|$.
By applying iteratively (26)

$$\begin{aligned}
\mu_{k+1} &= \beta n + (1 - \beta R)\mu_k = \beta n + (1 - \beta R)(\beta n + (1 - \beta R)\mu_{k-1}) \\
&= \beta n + \beta(1 - \beta R)n + (1 - \beta R)^2 \mu_{k-1} \\
&= \beta n + \beta(1 - \beta R)n + \beta(1 - \beta R)^2 n + (1 - \beta R)^3 \mu_{k-2} \ldots \\
&= \sum_{i=0}^{k} \beta(1 - \beta R)^i n \ .
\end{aligned}$$

From (27):

$$\mu_{k+1} = \frac{I - (I - \beta R)^{k+1}}{\beta R} \beta n \to R^{-1} n = \mu \ , \quad \text{for} \quad k \to \infty \ .$$

if $|I - \beta R| < 1$

The iterative principle

The iterative principle

In summary, we use

$$\mu_{k+1} = \mu_k + \beta[n - R * \mu_k] \tag{28}$$

with the initial condition

$$\mu_0 = n \ .$$

The convergence is assured if

$$|I - \beta R| < 1$$

Since $|1 - \beta x| < 1$ implies $0 < \beta < 2/x$, in the case of the operator R, which can be transformed in a square matrix

$$R' = (R * \mu)\, \mu^{-1}$$

we obtain the condition:

$$0 < \beta < \frac{2}{\text{max eigenvalue of R}'}$$

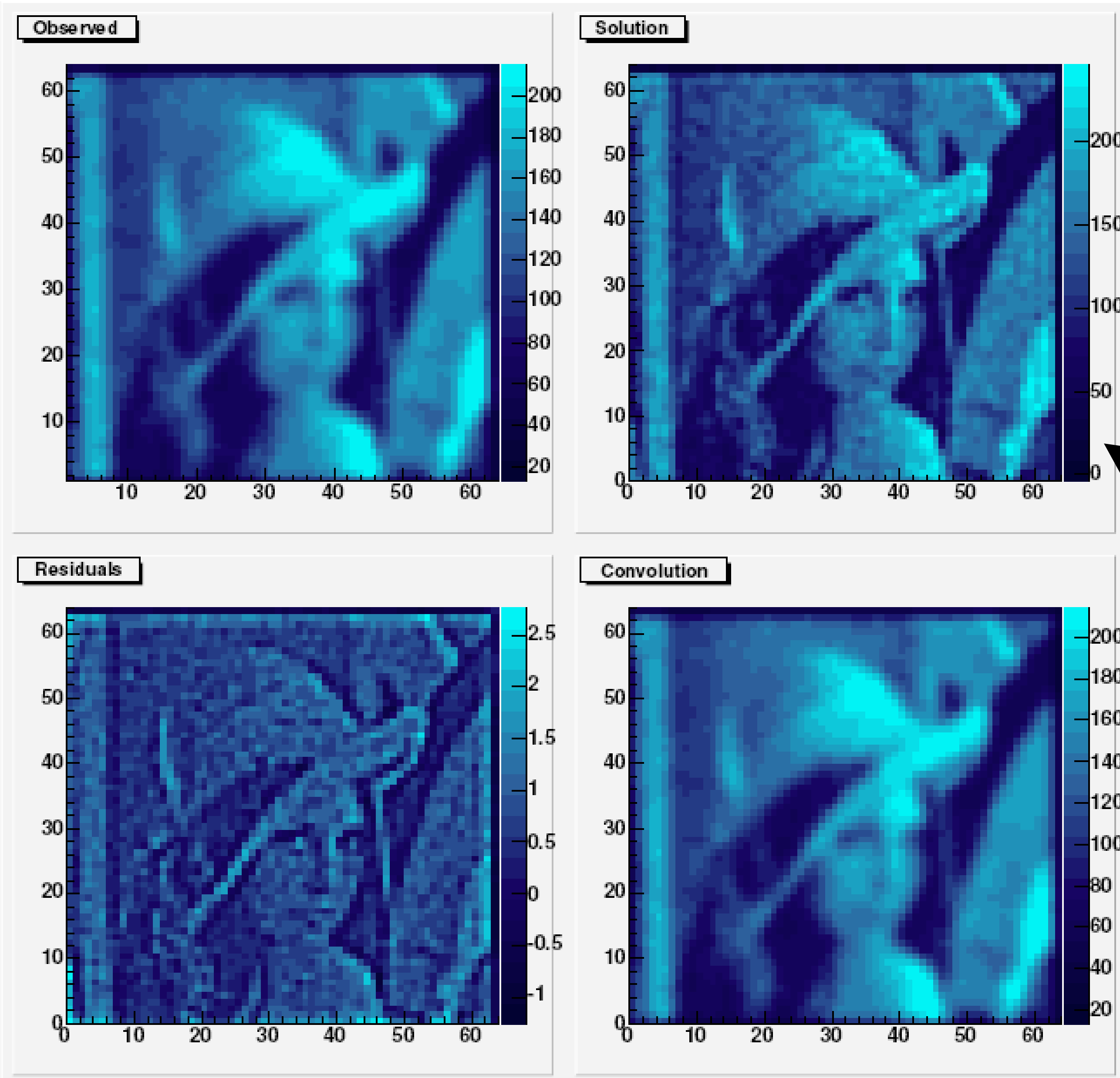
Note that we works always with square matrices $R * \mu$, μ, μ^{-1} and R'.

However, this step must be repeated at each iteration

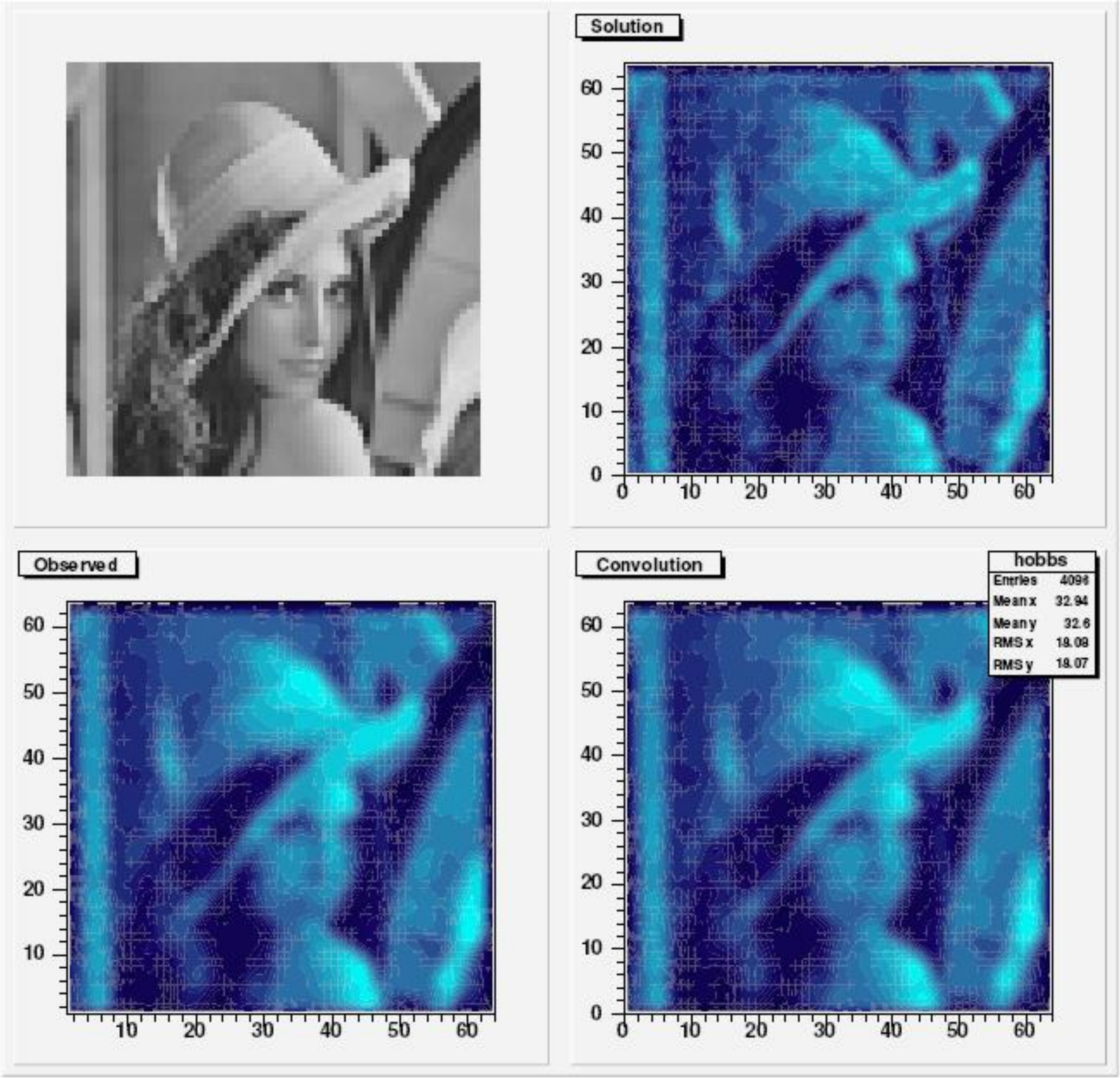
This method sometimes gives spectacular result!

However, often it gives irregular solutions.

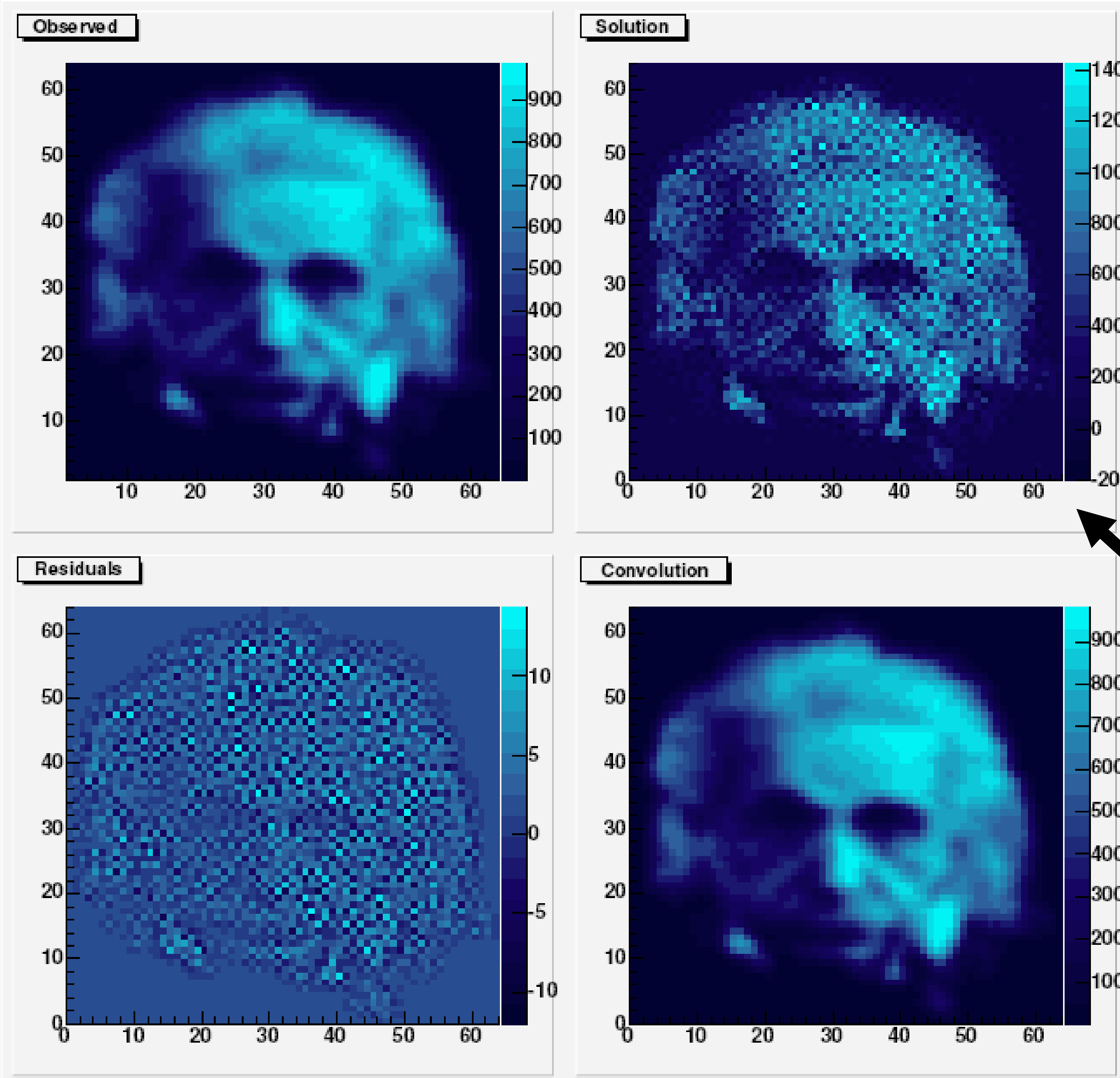
Without β $\quad \mu_{k+1} = \mu_k + [n - R * \mu_k]$



The iterative Principle without best fit



The iterative Principle without best fit + smoothing



The iterative Principle without best fit

Bad!

Consider the case with statistical fluctuations

$$\mu_i \overset{PSF}{\rightarrow} \nu_i \overset{random}{\rightarrow} n_i$$

$$\mathbf{n} = \mathbf{R} * \mu + \mathbf{r}$$

To have regular solutions and to make the method robust, we must search for an iterative solution which minimize the χ^2:

$$\chi^2 = ||\mathbf{R} * \mu - \mathbf{n}||^2 = \frac{1}{2} \sum_{\mathbf{ik}} \left(\sum_{\mathbf{mn}} \mathbf{R}_{\mathbf{i-m,k-n}} \, \mu_{\mathbf{mn}} - \mathbf{n}_{\mathbf{ik}} \right)^2 .$$

Note that is the case in which
the PSF depends on the pixel difference only
(translational invariance)
In this case we consider R as an operator and we can work with symmetric $M \times N$ matrices.
Minimum χ^2 w.r.t μ_{ik} gives the equations:

$$\frac{\partial \chi^2}{\partial \mu_{mn}} = \sum_{ik} \left(\sum_{rs} R_{i-r,k-s} \, \mu_{rs} - n_{ik} \right) R_{i-m,k-n} = 0$$

$$0 < m < M \; , \quad 0 < n < N \; .$$

Hence :

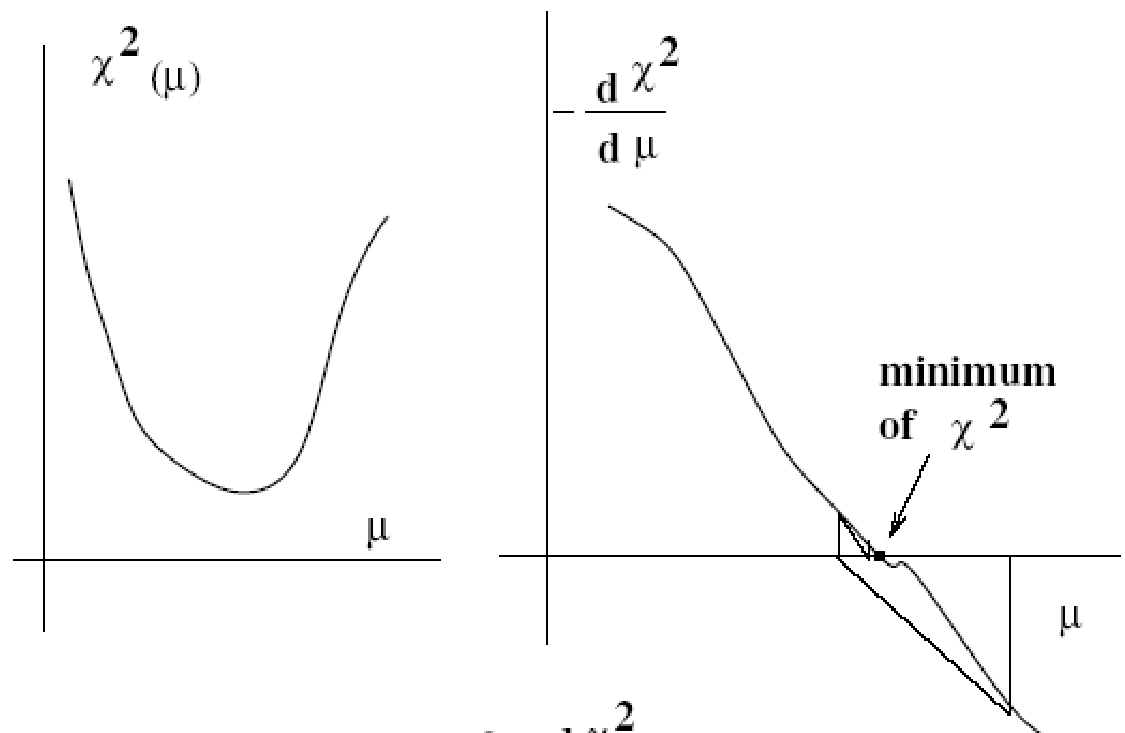
$$\frac{\partial \chi^2}{\partial \mu} = R * [R * \mu - n] = 0 \tag{29}$$

The iterative solution toward the μ's that minimize χ^2 is obtained with the Robbins-Monro algorithm (1951).

The iterative algorithm + best fit

It belongs to the family of the Newton-Raphson methods From (29):

The iterative algorithm + Best fit



$$\mu_{k+1} = \mu_k - \beta_k \frac{d\chi^2}{d\mu}$$

$$\mu_{k+1} = \mu_k + \beta_k R * [n - R * \mu] \qquad (30)$$

This formula minimizes χ^2. For example when $\beta_k = \beta$:

$$\begin{aligned} \mu_{k+1} &= \mu_k + \beta R * [n - R * \mu_k] \\ &= \beta R * n + (I - \beta R^2) * \mu_k \\ &= \beta R * n + \beta(I - \beta R^2) * R * n + (I - \beta R^2)^2 * \mu_{k-1} \\ &= \sum_{i=0}^{k} \beta(I - \beta R^2)^i * R * n = \frac{I - (I - \beta R^2)^{k+1}}{\beta R^2} \beta R * n \to R^{-1} n \end{aligned}$$

The iterative algorithm + best fit + regularization

$$\mu_{k+1} = \mu_k + \beta_k R * [n - R * \mu_k] \qquad (31)$$

Previous method converges if

$$||I - \beta R * R|| < 1$$

when β is independent of k. In this case

$$0 < \beta < \frac{2}{\text{max eigenvalue of } (R * R * \mu) * \mu^{-1}}$$

When β_k depends on k convergence is assured if (Robbins and Munro 1951)

$$\lim_{N\to\infty} \beta_N = 0 \ , \quad \sum_{N=1}^{\infty} \beta_N = \infty \ , \quad \sum_{N=1}^{\infty} \beta_N^2 < \infty \ ,$$

Next, the method must be regularized by adding a term to the χ^2:

$$\chi^2 = ||R * \mu - n||^2 + \alpha ||C * \mu||^2 \qquad (32)$$

For example,

$$||C * \mu||^2 = \sum_{ik} \mu_{ik}^2$$

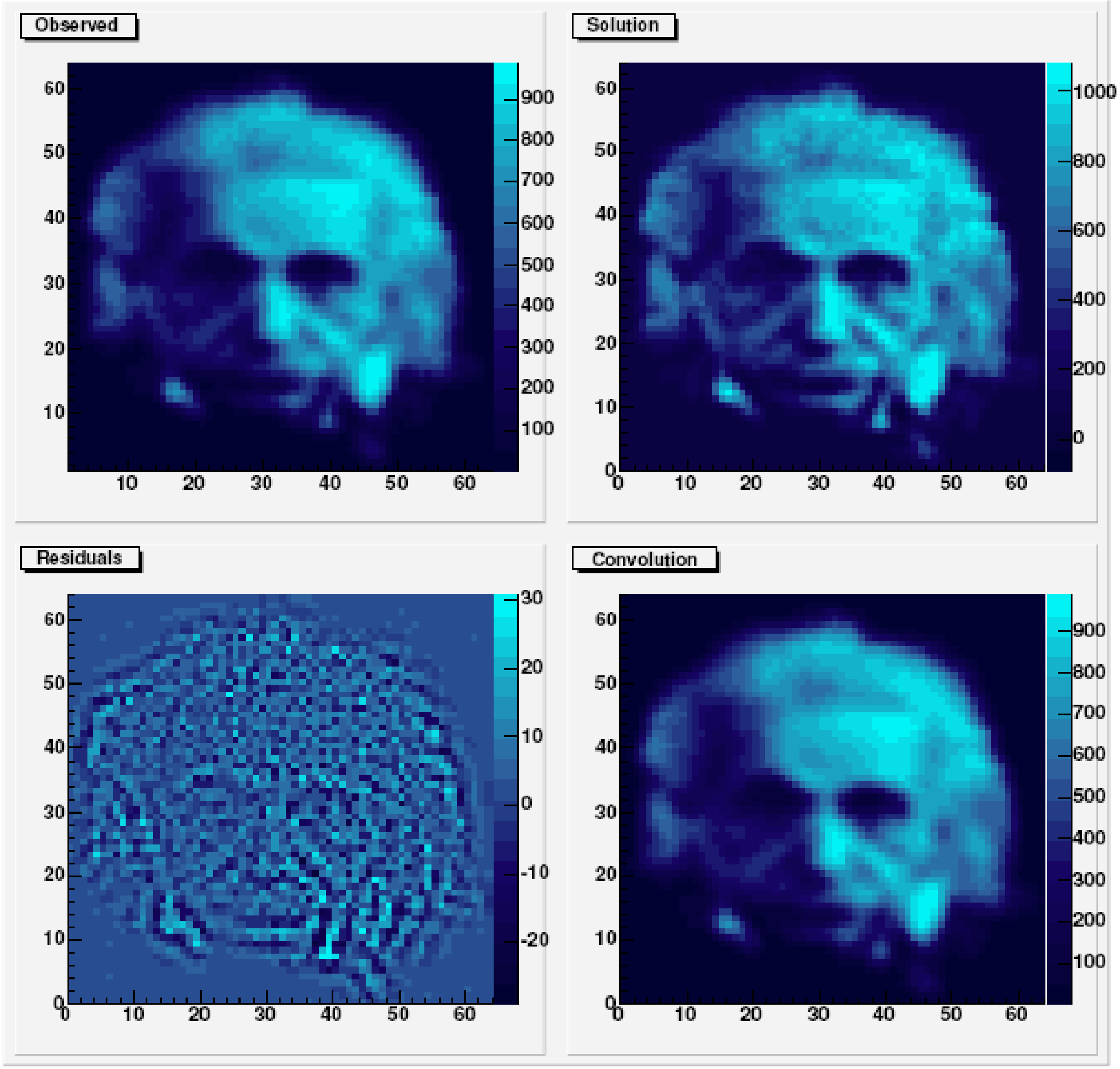
$$||C * \mu||^2 = |\sum_{ik} \mu_{ik} \ln \mu_{ik}/\mu_{\text{tot}}| \ , \qquad \alpha < 0$$

The iterative solution becomes

$$\mu_{k+1} = \mu_k + \beta_k [R * n - (R * R + \alpha C * C)\mu_k] \qquad (33)$$

$$\mu_{k+1} = \mu_k + \beta_k[R * n - [R * R * \mu_k + \alpha(\ln \mu_k/\mu_T + I)]$$

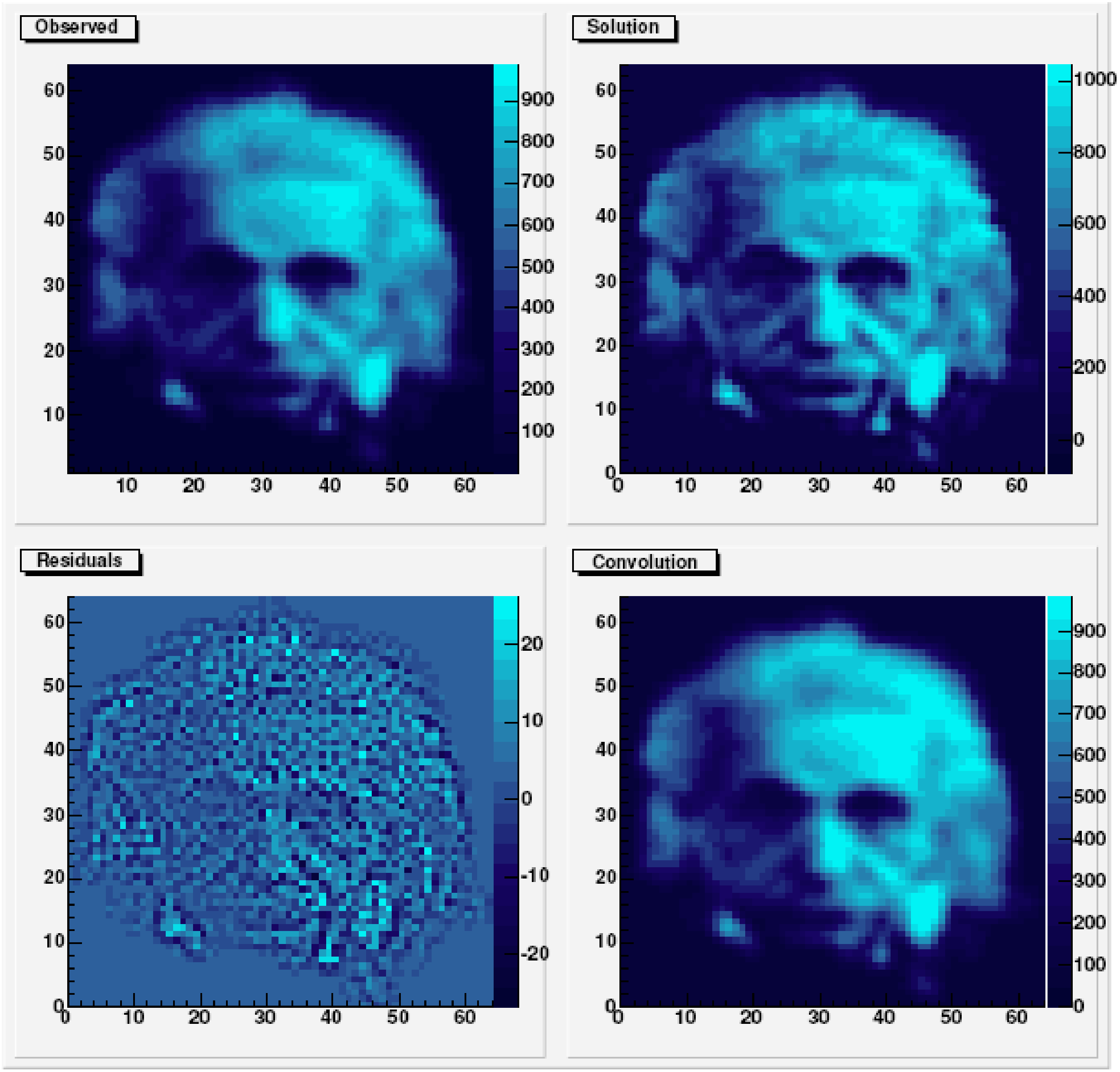
About 40 iterations, regularized with Maximum entropy



The iterative algorithm + best fit + MaxEnt regularization

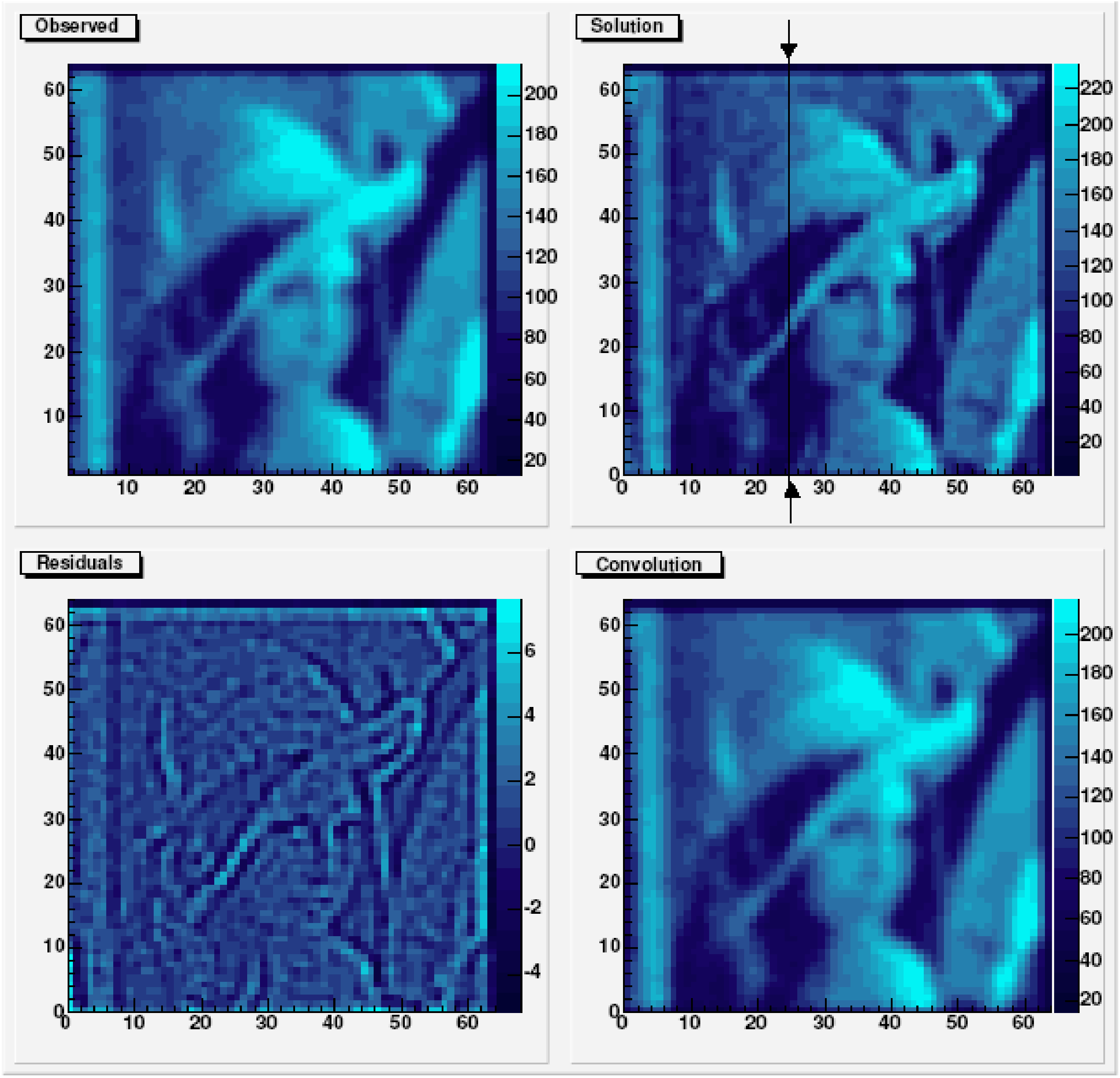
$$\mu_{k+1} = \mu_k + \beta_k[R * n - (R * R + \alpha I) * \mu_k]$$

About 100 iterations, regularized with the sum of squares



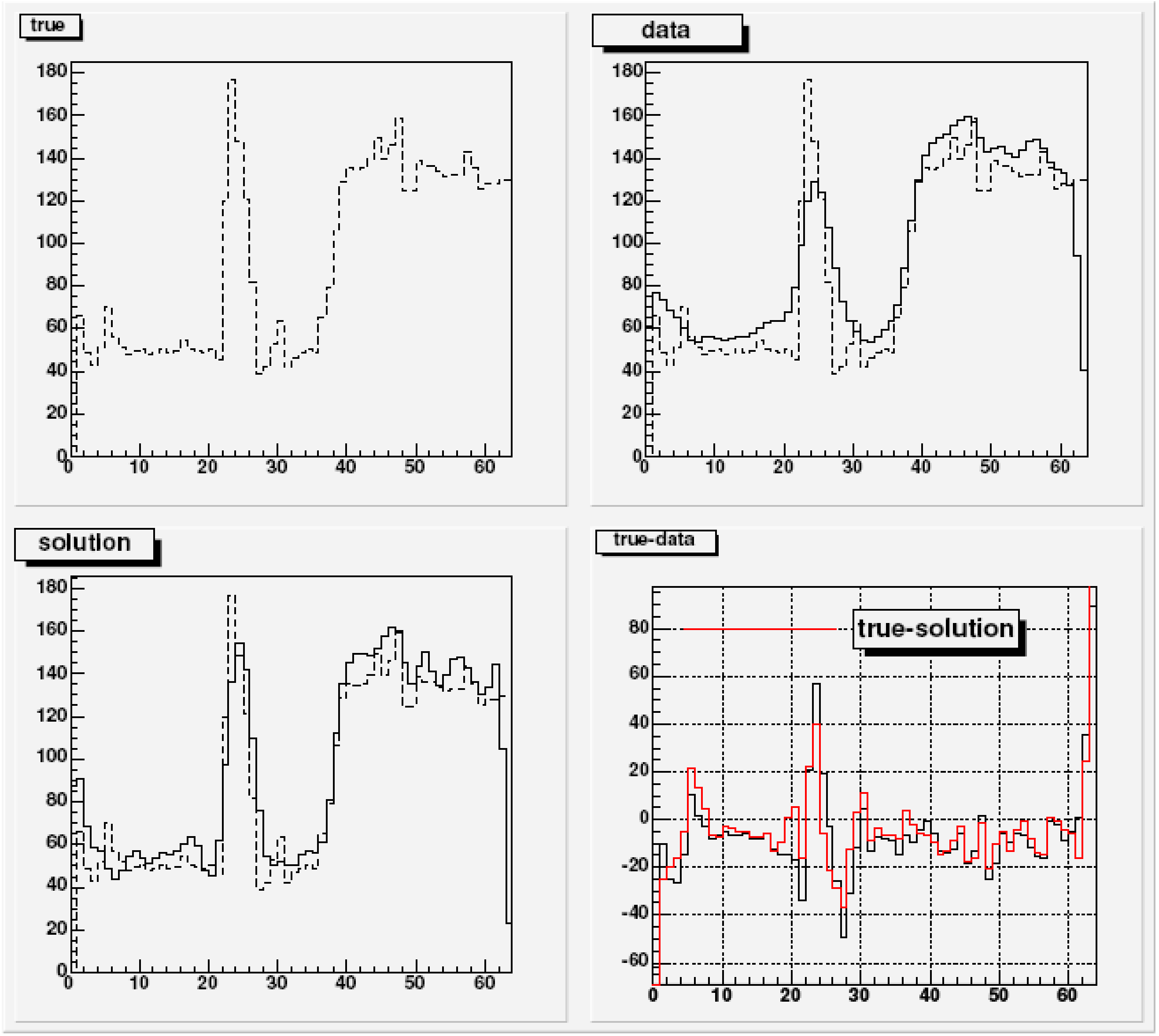
The iterative algorithm + best fit + Tichonov regularization

$$\mu_{k+1} = \mu_k + \beta_k[R * n - (R * R + \alpha I) * \mu_k]$$



The iterative algorithm + best fit + Tichonov regularization

$$\mu_{k+1} = \mu_k + \beta_k[R * n - (R * R + \alpha I) * \mu_k]$$



The iterative algorithm + best fit + Tichonov regularization

Stopping Rules

- χ^2 variation

$$\chi^2_k = ||y - R * \mu_k||^2 = \sum_i^{M\times N} \frac{(y_i - \Sigma_{ij} R_{ij}\mu_{j(k)})^2}{\Sigma_{ij} R_{ij}\mu_{j(k)}}$$

$$\frac{\chi^2_k - \chi^2_{k-1}}{\chi^2_{k-1}} < 10^{-6}$$

If the regularization is good, one has $\chi^2_k \simeq$ DoF.

- Signal to noise ratio **(usually measured in decibel)**

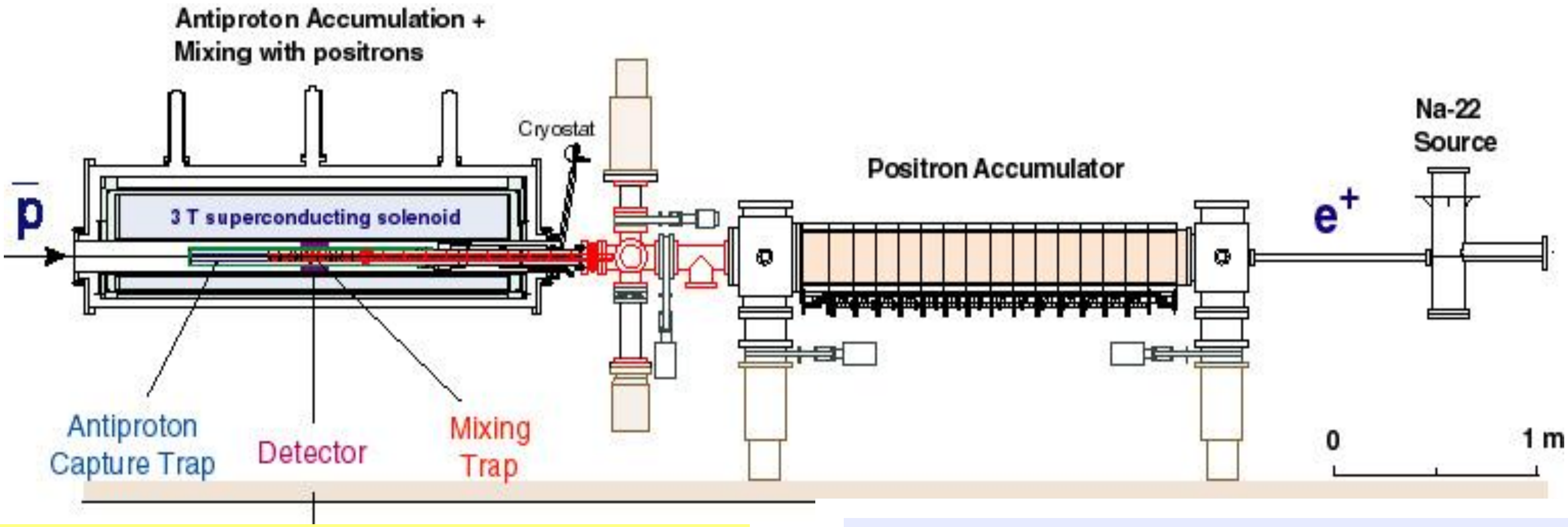
$$SNR = 10\log_{10}\left[\frac{\Sigma_i(\mu_i - y_i)^2}{\Sigma_i(\mu_i - \mu_{true\ i})^2}\right]$$

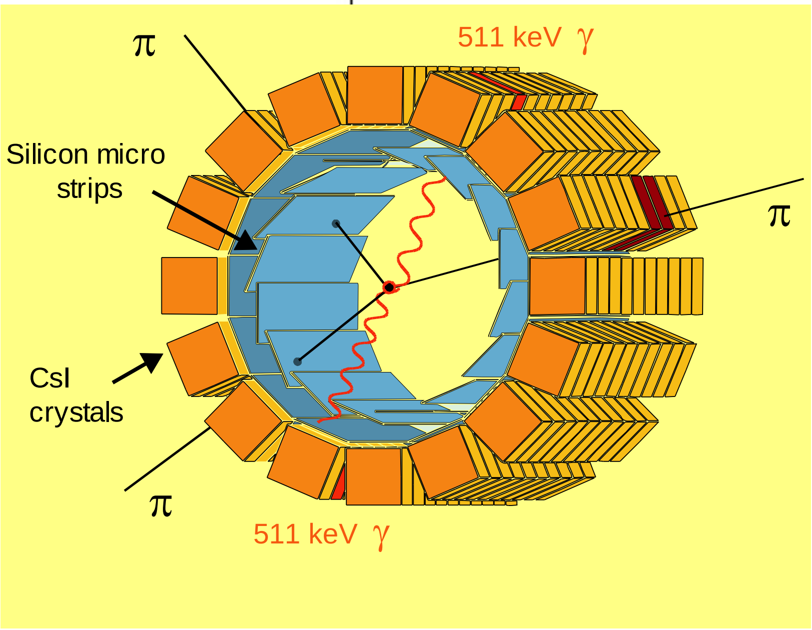
where μ_{true} is the true image. This quantity is used in the MC simulations during when the true image is known.

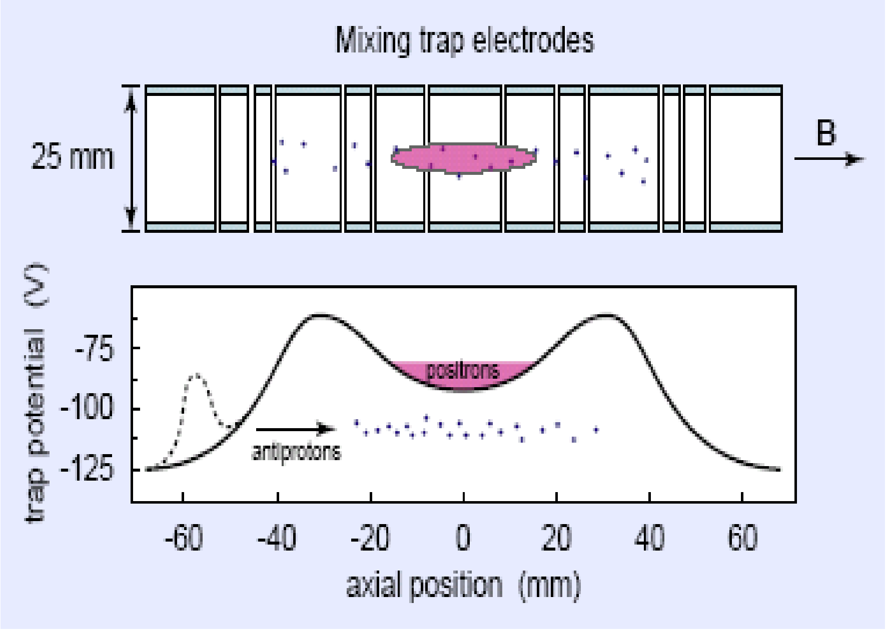
- convergence of the solution

$$\frac{||\mu_k - \mu_{k-1}||^2}{||\mu_{k-1}||^2} < 10^{-6}$$

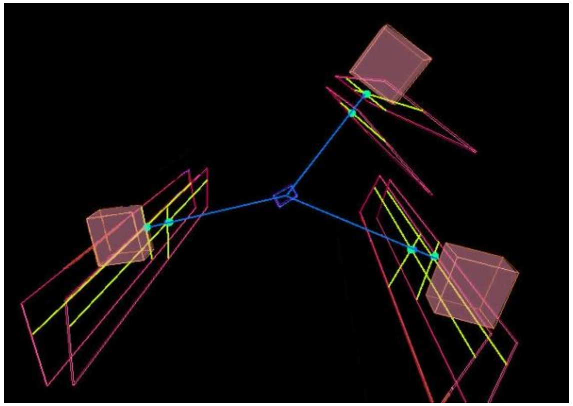
ATHENA apparatus

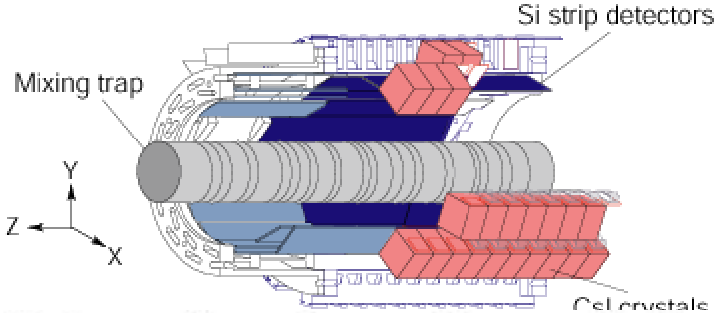






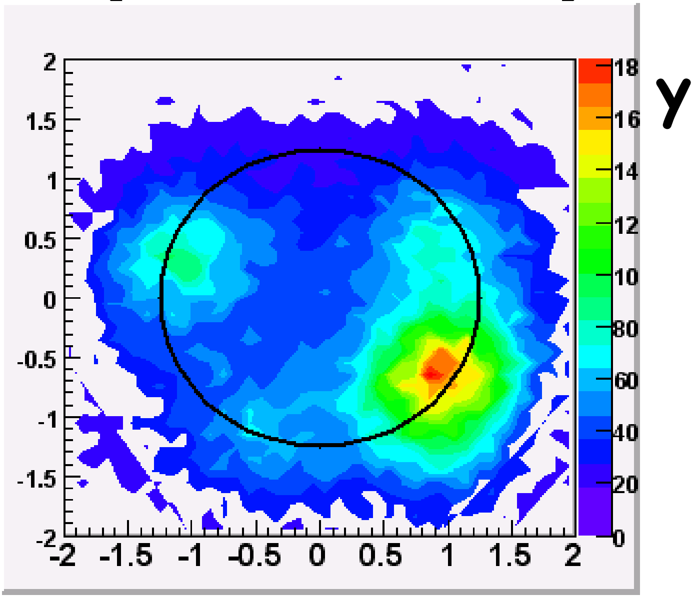
From the ATHENA detector



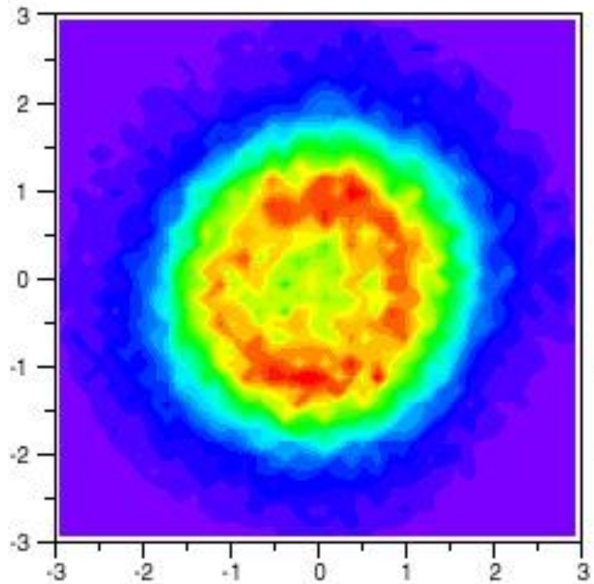


Distribution of annihilation vertices when antiprotons are mixed with ...

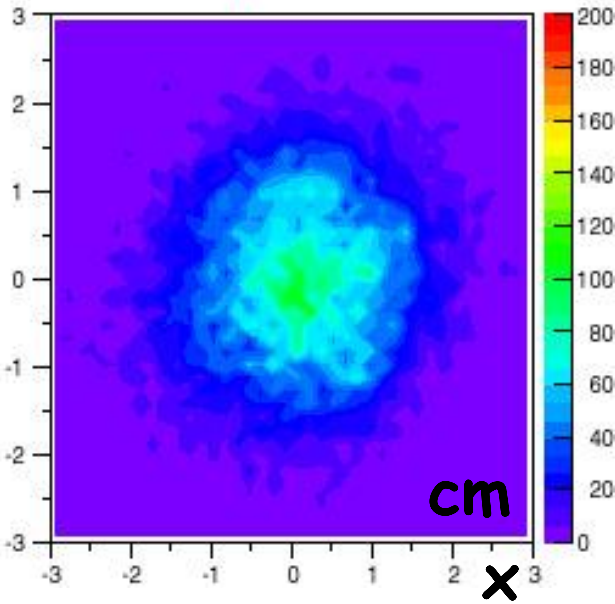
Pbar-only (with electrons)



cold positrons

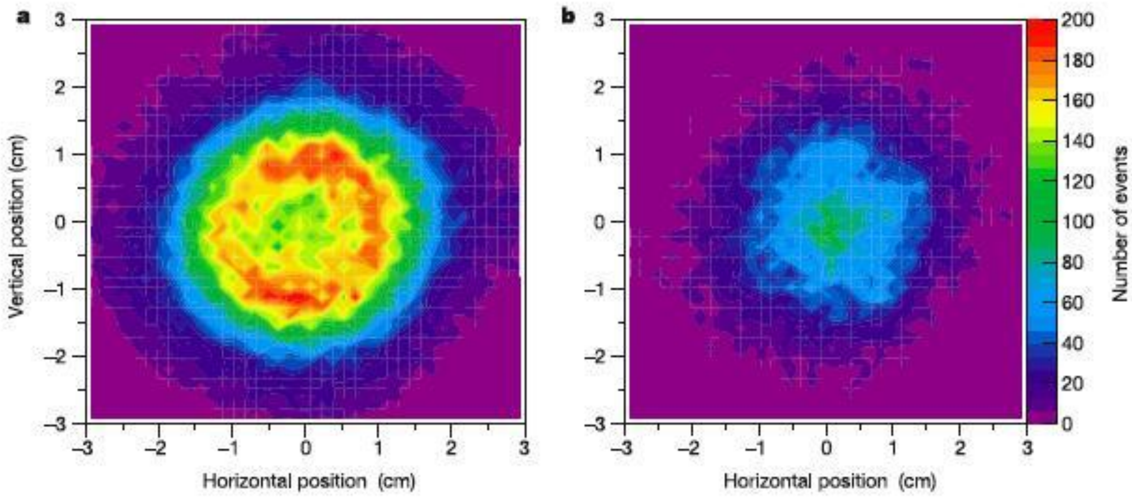


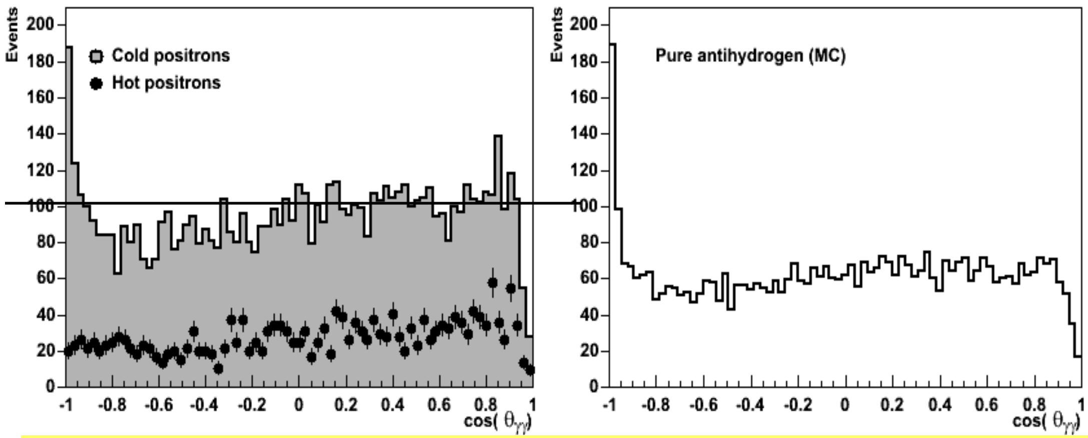
hot positrons



antihydrogen !!!!!!!!!!

FIRST COLD ANTIHYDROGEN PRODUCTION & DETECTION (2002)
M. Amoretti et al., Nature 419 (2002) 456
M. Amoretti et al., Phys. Lett. B 578 (2004) 23





SIGNAL ANALYSIS:

opening angle
xy vertex distribution
radial vertex distribution

65 % +/- 10% of annihilations are due to antihydrogen

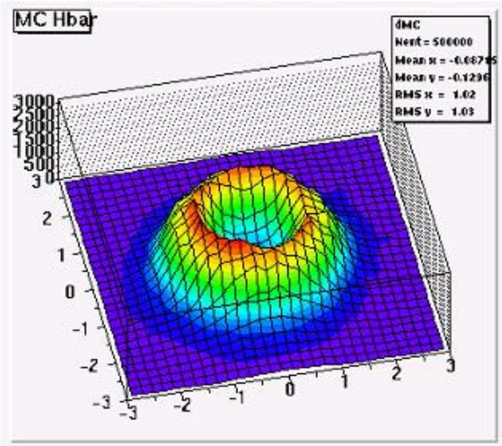
between 2002 & 2004 **more than 2 millions** antihydrogen atoms have been produced

that's about 2 x 10^{-15} mg
.. or .. 1000 Giga years for a gram

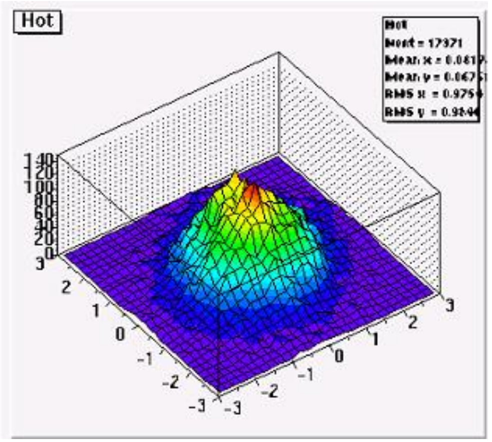
$$\frac{80}{\sqrt{190+110}} = 4.7; \quad \frac{80}{\sqrt{190}} = 6.5; \quad \frac{80}{\sqrt{110}} = 8$$

Annihilation vertex in the trap x-y plane

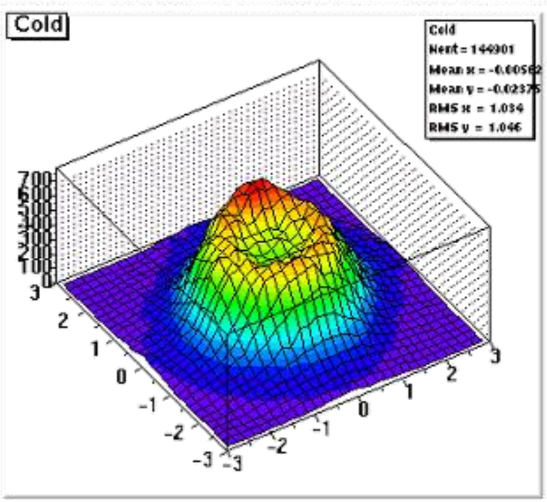
Hbar (MC)



BCKG (HotMixData)



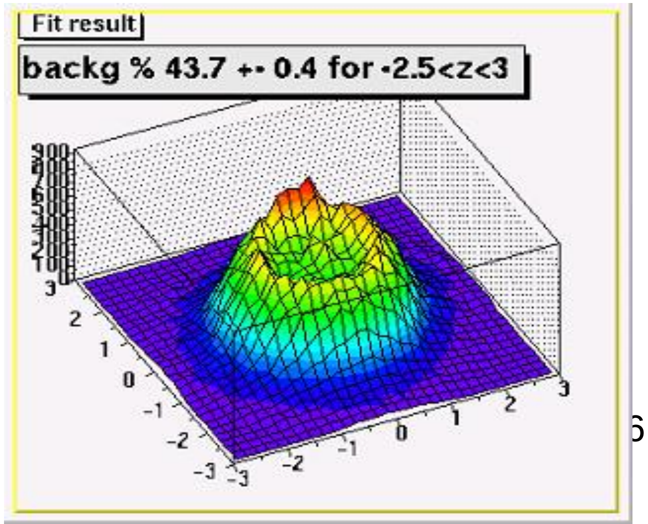
Cold Mix data



Pbar vertex XY projection (cm)

x Hbar + (1-**x**) BCKG =

ML Fit Result

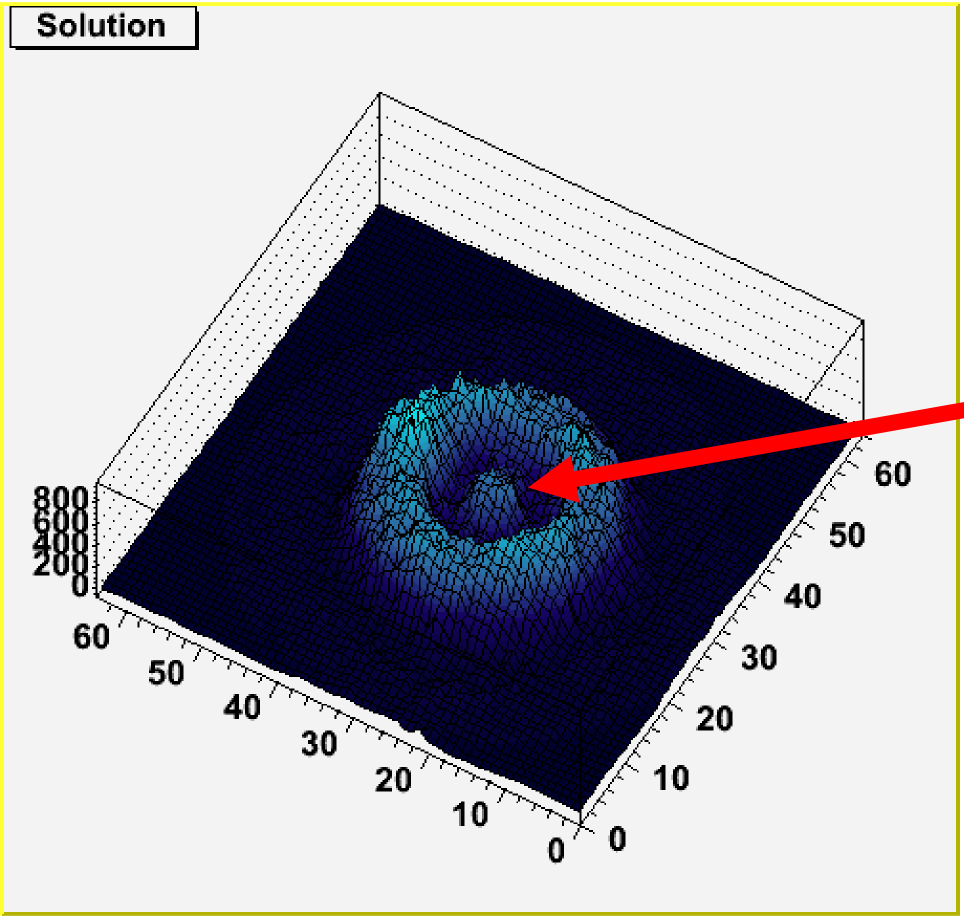


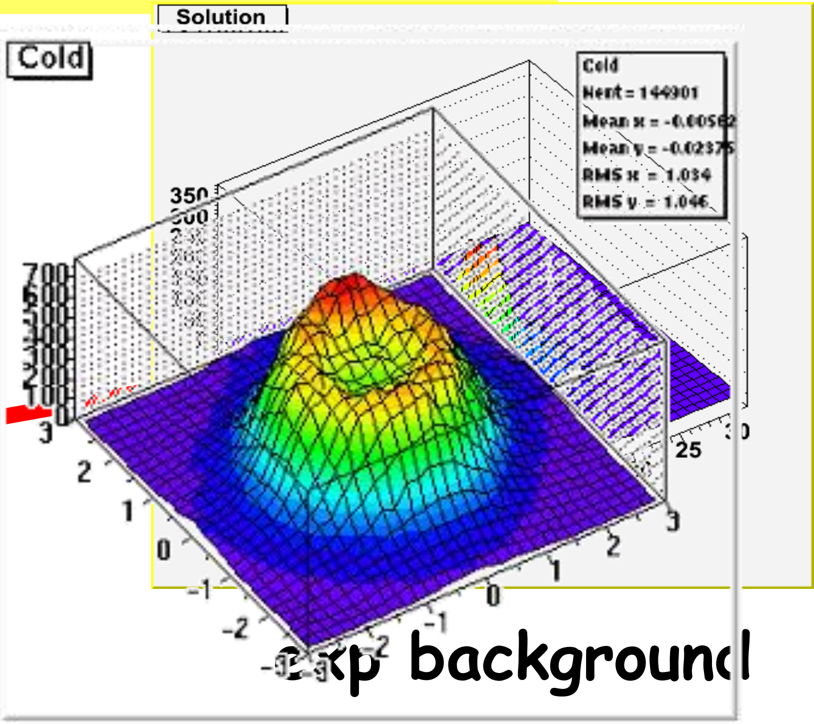
Hbar percentage

x = 0.65 ± 0.05

Iteratve best fit method

Cold Mix data





Cold Mix

The vertex algorithm resolution function is gaussian with $\sigma \cong 3$ mm

The 2D deconvolution reveals two different annihilation modes

The iterative algorithms + best fit + regularization

- iterative algorithms are used in unfolding (ill posed) problems
- they need a Bayesian regularization term
- when there are degrees of freedom, one can use a best fit of a signal+background function to the data
- in this case there are no Bayesian terms (pure frequentist approach)

Conclusions

- don't be dogmatic

- use Bayes to parametrize the a priori knowledge if any, not the ignorance

- in the case of poor a priori knowledge, use the frequentist methods