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 $\hat{\theta}$ 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!

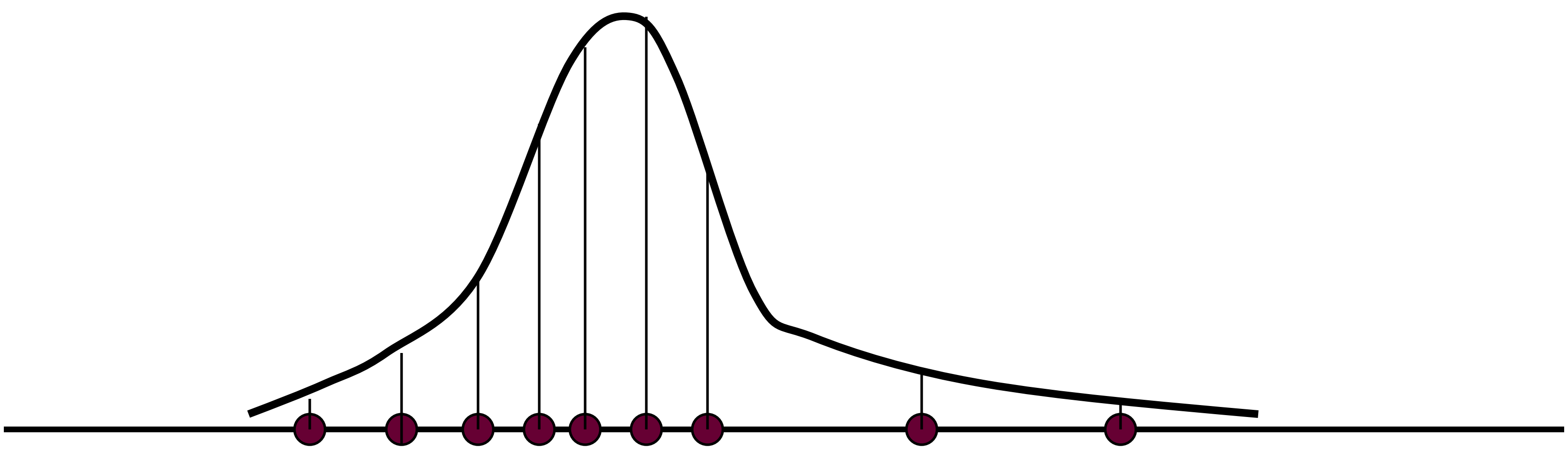
Bayesians vs Frequentists

Why ML does work?

hypothesis

$$L=\prod_k p(x_k;\theta)$$

observation



the p(x;θ) form is fitted to data by maximizing the ordinates of the observed data

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:

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

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

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

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

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

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

MINUIT/MINOS method

$$-2\ln \Delta L = -2[\ln L(\theta) - 2\ln L(\hat{\theta})] \cong \chi^2(\theta)$$

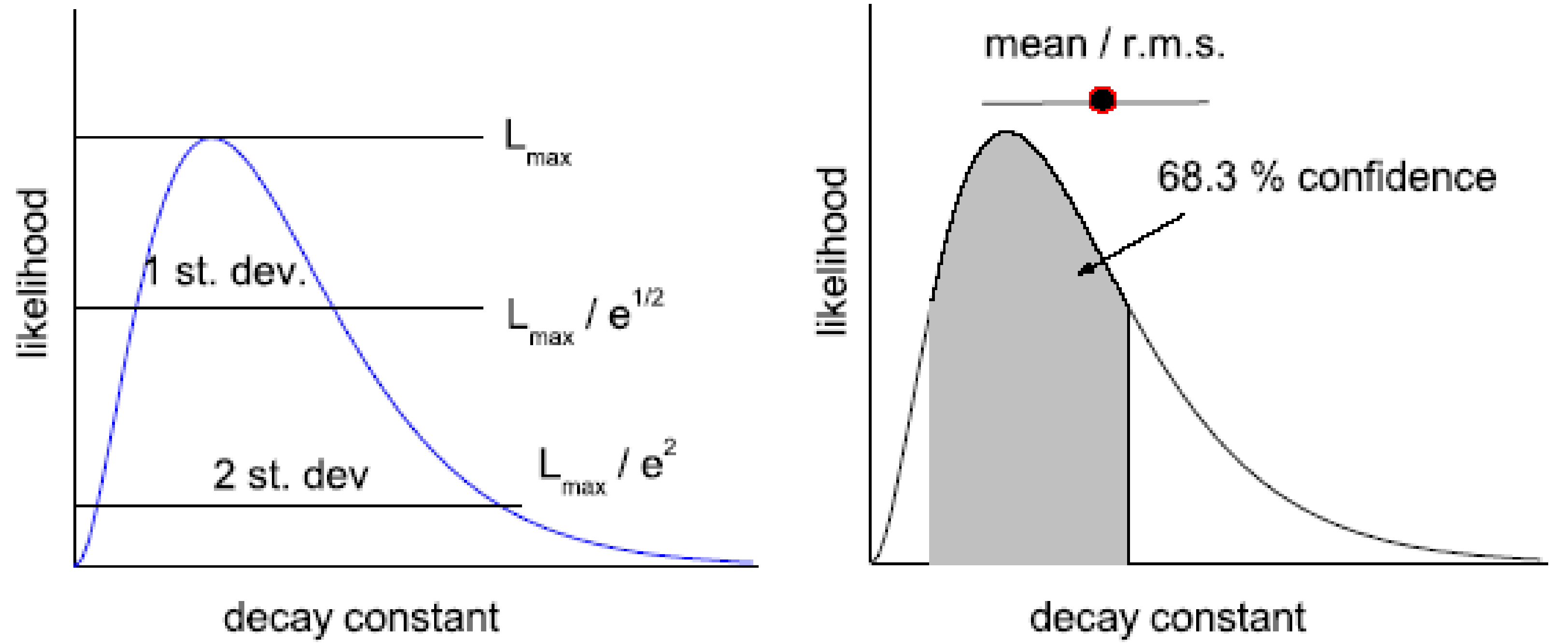
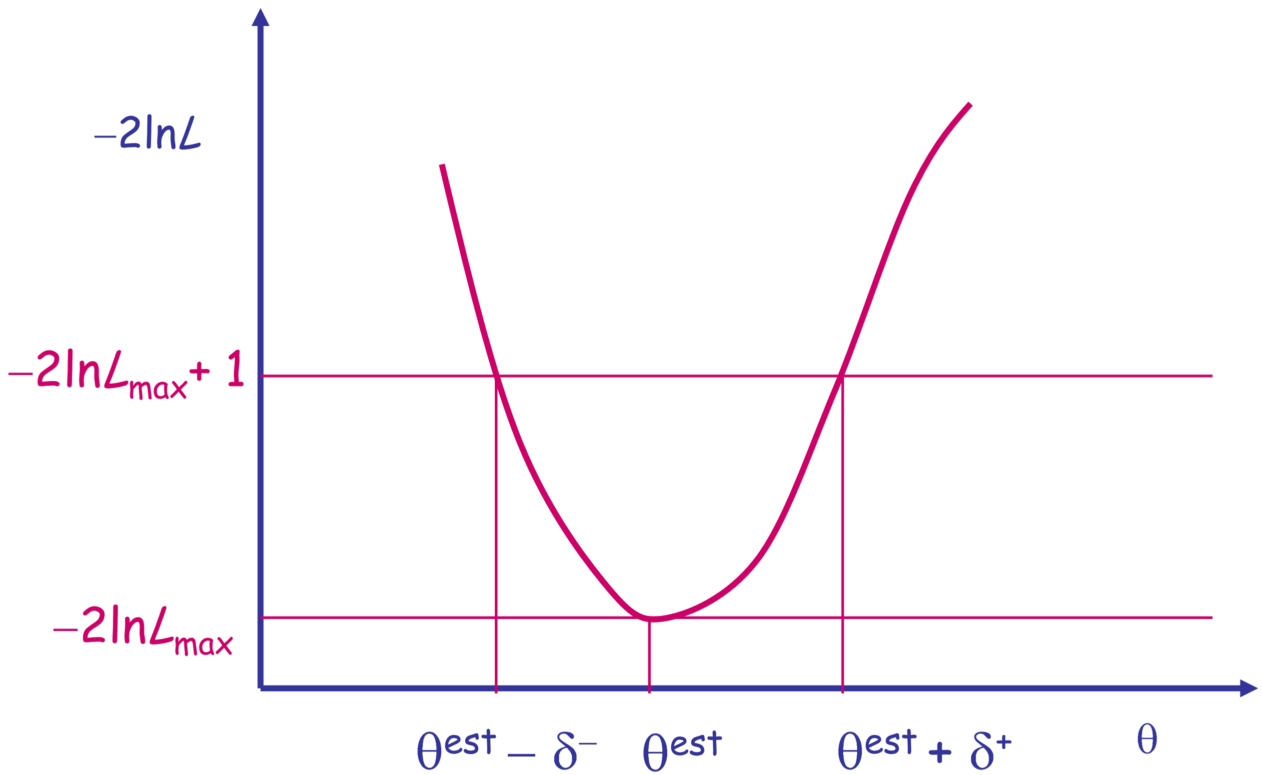


Fig. 18. Likelihood ratio limits (left) and Bayesian limits (right)

$$\ln e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma}} = -\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2} \Rightarrow -2\ln L(x;\theta) \approx \chi^2(\theta)$$

Error determination
MINUIT/MINOS method

- Error determined by the range around the Likelihood maximum for which $-2\ln L$ increases by *one*



- Errors can be asymmetric
  - Be careful about interpretation!
- Identical to PDF's σ for Gaussian models
- ML estimates are asymptotically Gaussian

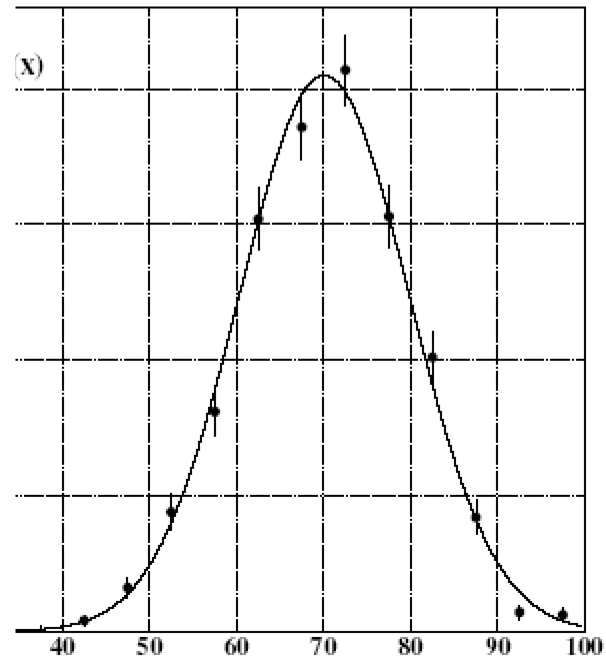
Fit of Histograms

The model is given by:

$$\mu_i(\boldsymbol{\theta}) = N \int_{\Delta_i} p(x;\boldsymbol{\theta})\,\mathrm{d}x \simeq N p(x_{0i};\boldsymbol{\theta})\Delta_i \equiv N p_i(\boldsymbol{\theta}) \ ,$$

$$L(\boldsymbol{\theta};\, \underline{n}) = \prod_{i=1}^{k} [p_i(\boldsymbol{\theta})]^{n_i} \ ,$$

$$\mathcal{L} = -\ln L(\boldsymbol{\theta};\, \underline{n}) = -\sum_{i=1}^{k} n_i \ln[p_i(\boldsymbol{\theta})] \ .$$



The second one correspond to the **pseudo-χ^2 minimization**. Indeed:

$$\sum_{i=1}^{k} \frac{n_i}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j} = \sum_{i=1}^{k} \frac{n_i - N p_i(\boldsymbol{\theta})}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j}$$

since $\Sigma_i\, p_i(\boldsymbol{\theta}) = 1$ implies $\Sigma_i\, \partial p_i(\boldsymbol{\theta})/\partial\theta_j = 0$.

This formula is from ML !!!!

The last member corresponds to the derivative of

$$\chi^2 = \sum_i \frac{(n_i - N p_i(\boldsymbol{\theta}))^2}{N p_i(\boldsymbol{\theta})} \simeq \sum_i \frac{(n_i - N p_i(\boldsymbol{\theta}))^2}{n_i} \ , \qquad (1)$$

with a **constant denominator**

The extended likelihood

$$L(\theta, \underline{n}) = \prod_i \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i}$$

$$-\ln L(\theta, \underline{n}) = -\sum_{i=1}^{k} n_i \ln[\mu_i(\theta)] + \sum_{i=1}^{k} \mu_i(\theta)$$

Since $\mu_i = N\, p_i(\theta)$

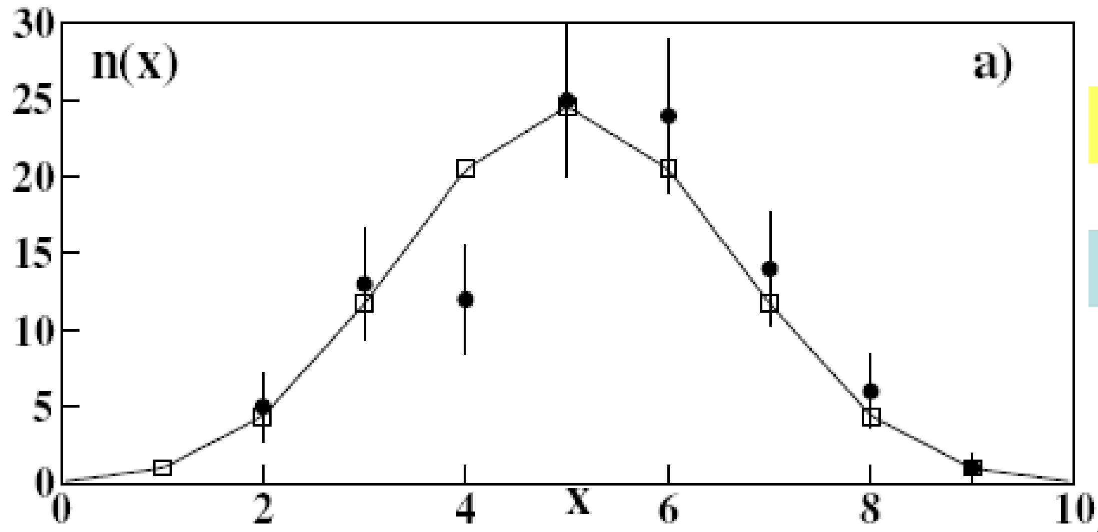
$$-\ln L(\theta, \underline{n}) = -\sum_{i=1}^{k} n_i \ln[p_i(\theta)] + N(\theta)$$

N **is a function of** θ **as in the case of a detector efficiency,**

If there is no functional relation between N **and** θ

the result is the same as for the non extended likelihood

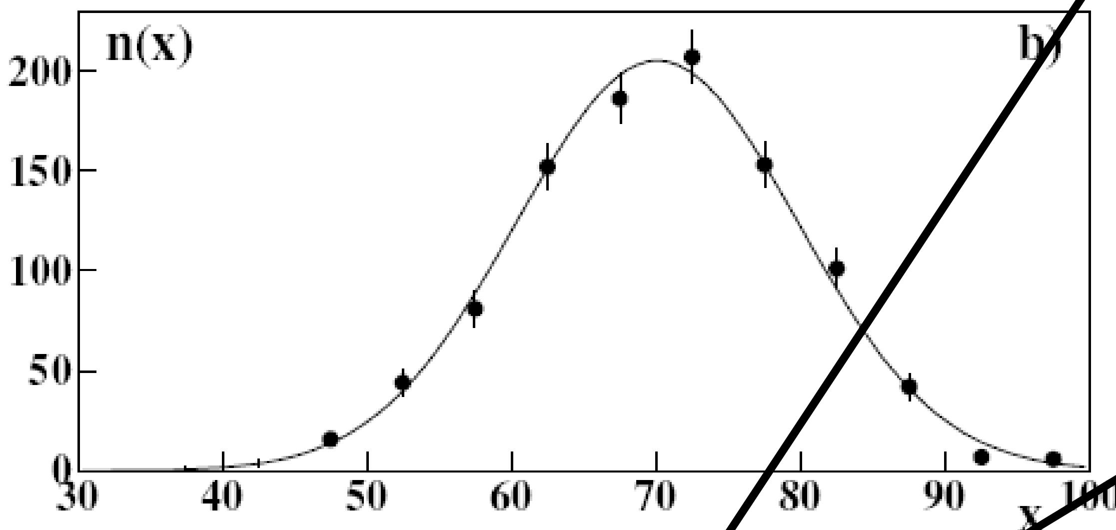
Binomial
p=0.5



p= 0.522 0.015

p= 0.528 0.017

Gaussian
μ=70
σ=10



μ =70.09 0.31
σ =9.73 0.22

μ =69.97 0.31
σ =9.59 0.22

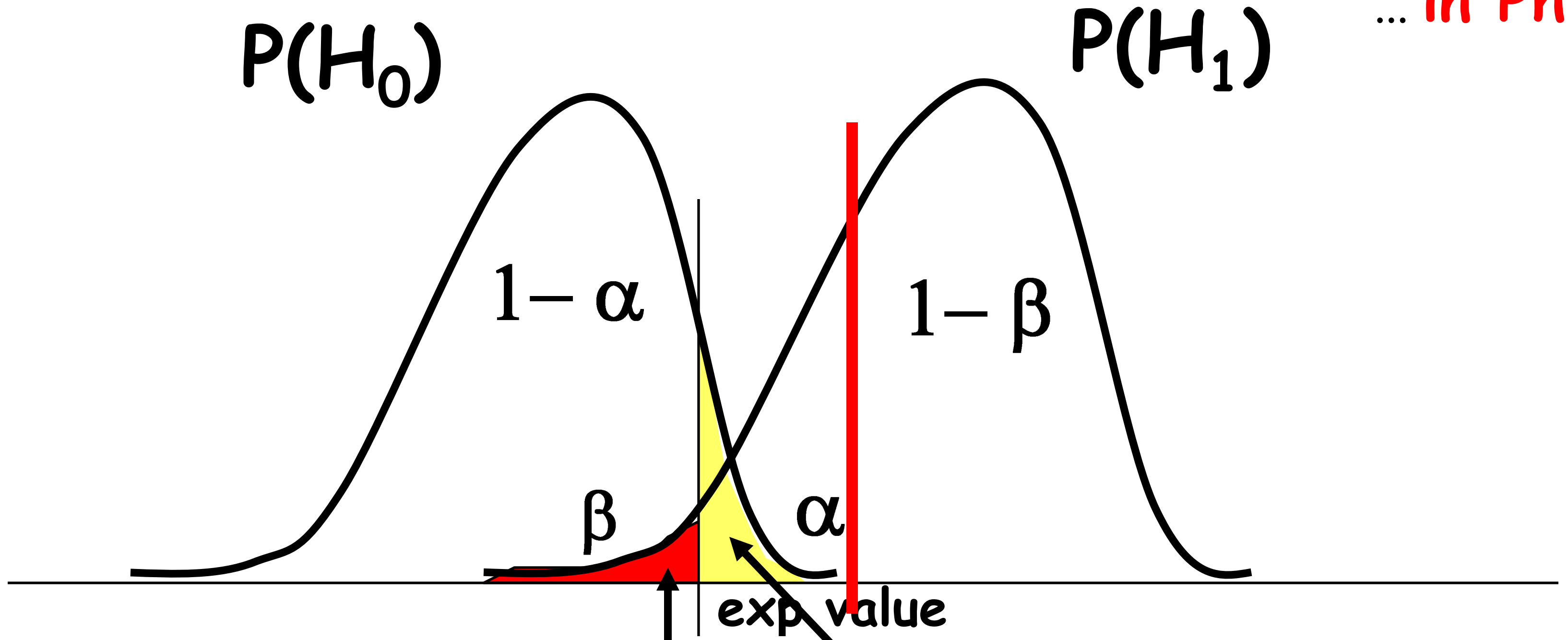
$\mathcal{L}$

χ^2

The other branch of Statistics: Hypothesis Testing

- A starting hypothesis (the null hypothesis) is defined:
  absence of the signal
- an observable must be defined:
  a trigger
- a test function, that is a random variable of known distribution, must be defined: *the number of trigger follows Poisson*
- at least one alternative hypothesis must be defined:
  the presence of the signal
- the rules for discriminating between the hypotheses must be defined:
  there are Bayesian and frequentist criteria!!

… **in Physics**



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

power

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

The connection between Hyp test and parameter estimation is the following one:

H_0 would be rejected at significance level α if the $(1-2\alpha) = CL$ confidence interval does not contain the value μ_s

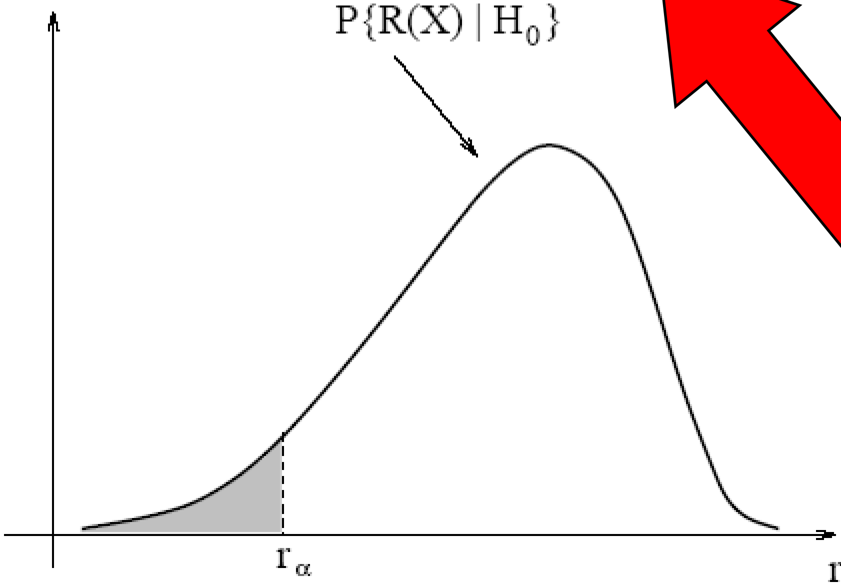
$$P = \left\{ \frac{x-\mu}{\sigma} \leq t \right\} = CL = 1 - 2\alpha$$

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 : $R = \psi(T)$.

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

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

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:

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

Usually one compare the quantity

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

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.

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

PHYSICS LETTERS B

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.

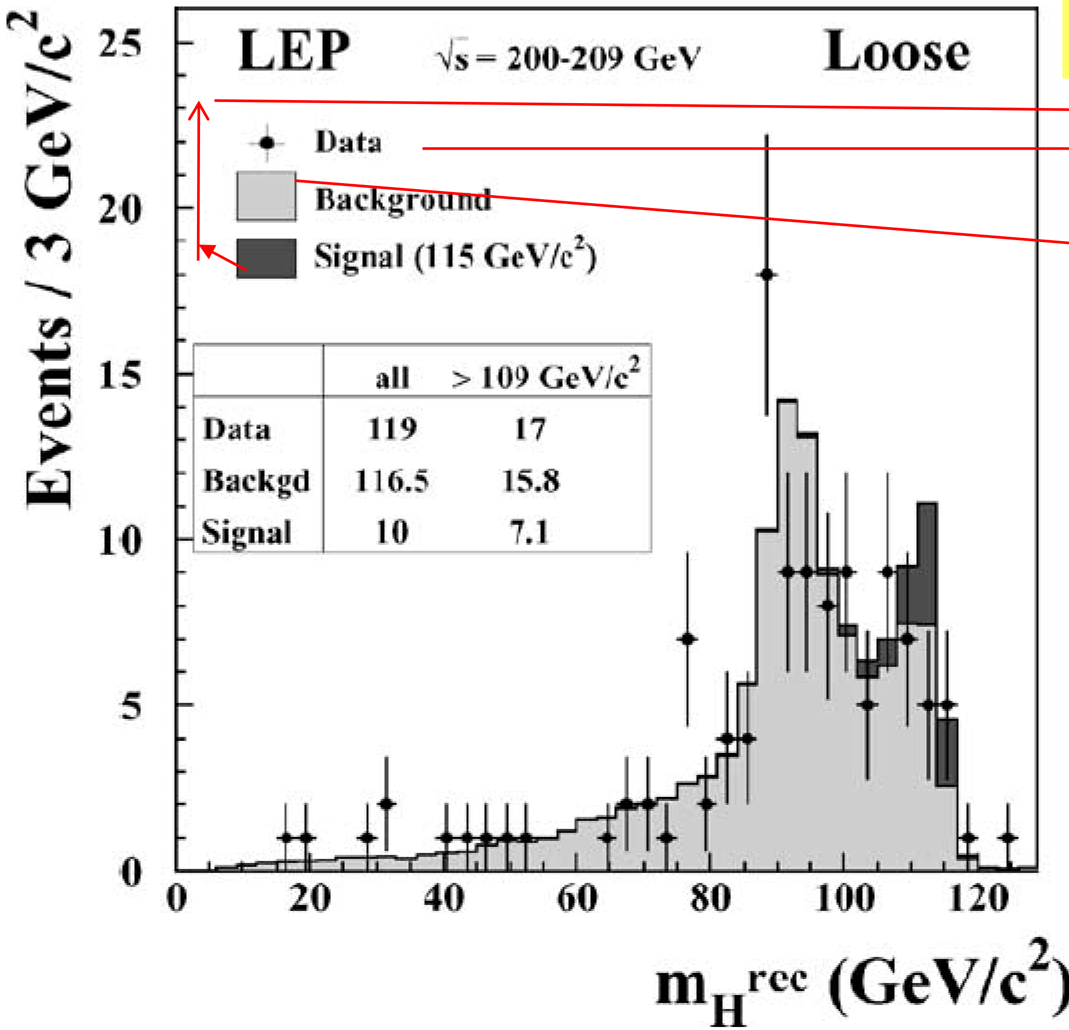
$e^+e^- \to HZ$

those of the associated Z boson. The searches at LEP encompass the four-jet final state $(H \to b\bar{b})(Z \to q\bar{q})$, the missing energy final state $(H \to b\bar{b})(Z \to \nu\bar{\nu})$, the leptonic final state $(H \to b\bar{b})(Z \to \ell^+\ell^-)$ where ℓ denotes an electron or a muon, and the tau lepton final states $(H \to b\bar{b})(Z \to \tau^+\tau^-)$ and $(H \to \tau^+\tau^-) \times (Z \to q\bar{q})$.

A preselection is applied by each experiment to reduce some of the main backgrounds, in particular, from two-photon processes and from the radiative return to the Z boson, $e^+e^- \to Z\gamma(\gamma)$. The remaining

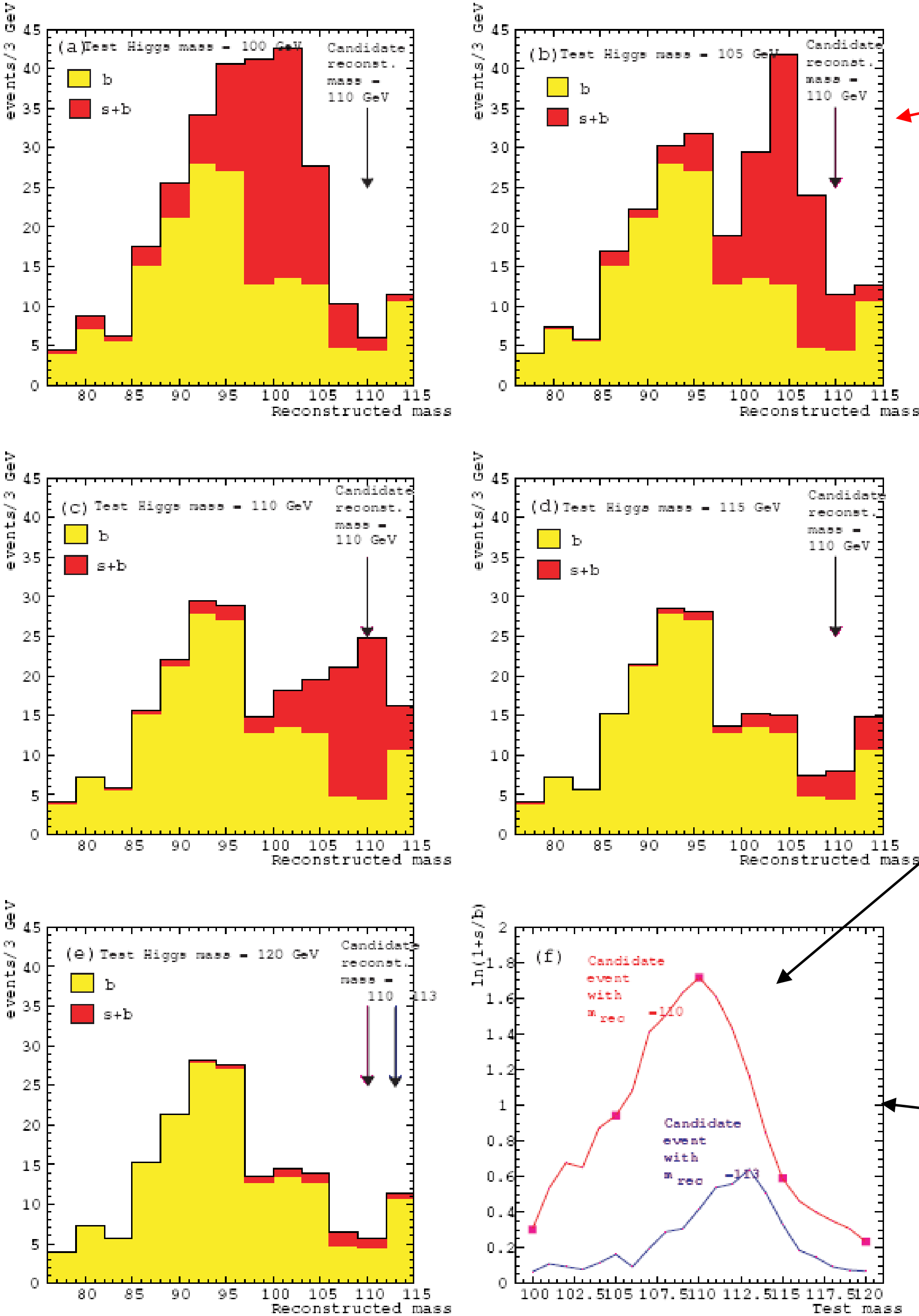
…hes / Physics Letters B 565 (2003) 61–75

ALEP, DELPHI, L3, OPAL, 2003



$$\ln Q = -S_{tot} + \sum_i n_i \ln(1 + \frac{S_i}{B_i})$$

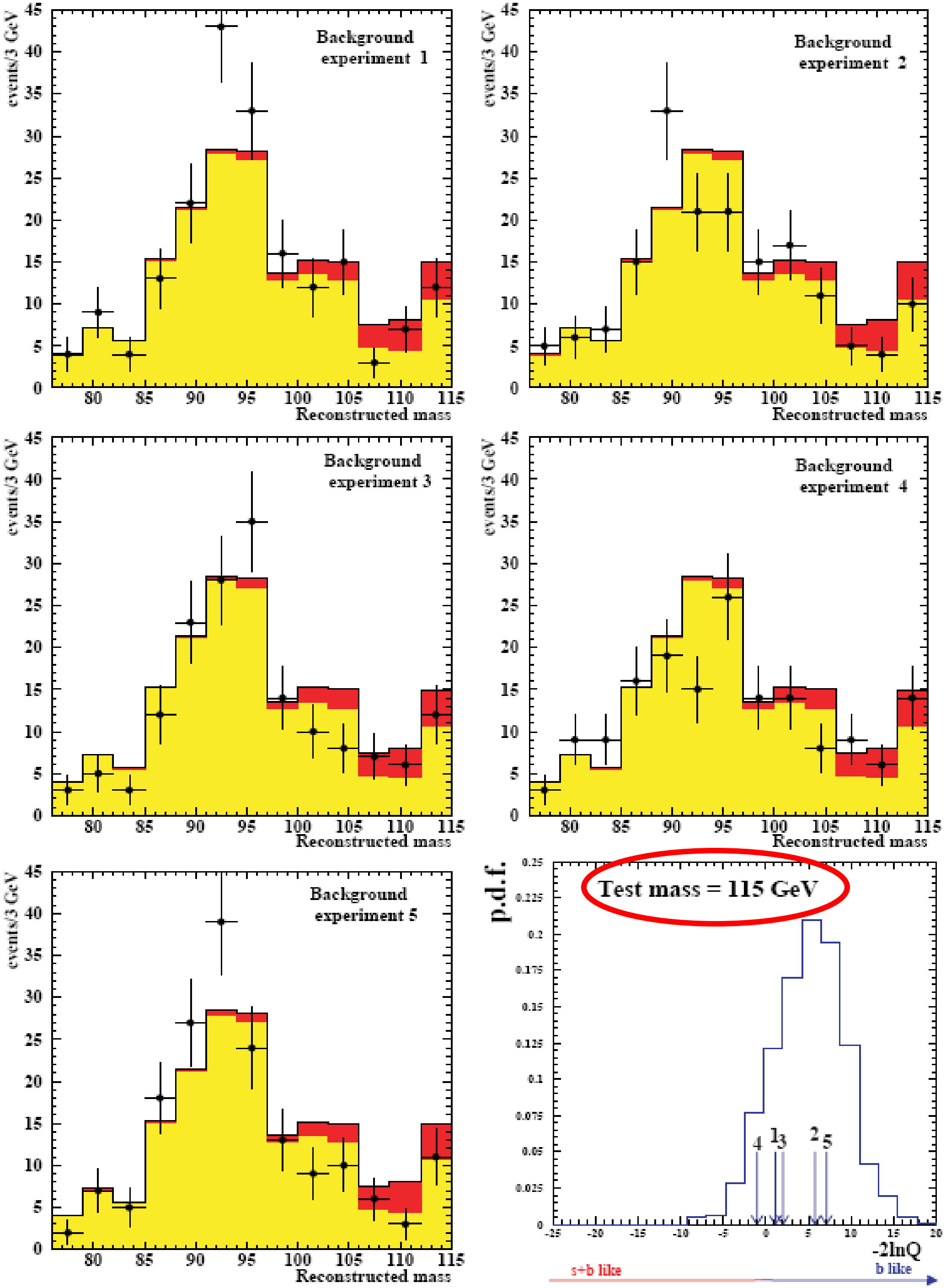
One can sum-up over the bins of histograms from different experiments and to construct a **GLOBAL** statistics!



MC toy model

First problem: due to detector efficiencies and to undetected neutrinos which accompain the Higgs decay products, the reconstructed mass could not coincide with the true mass
The figure shows the weight ln(1+s/b) when the reconstructed mass is 110 GeV and the weights are calculated for true Higgs masses bewtween 100-120 GeV

The weight plot was called spaghetti plot



MC toy model

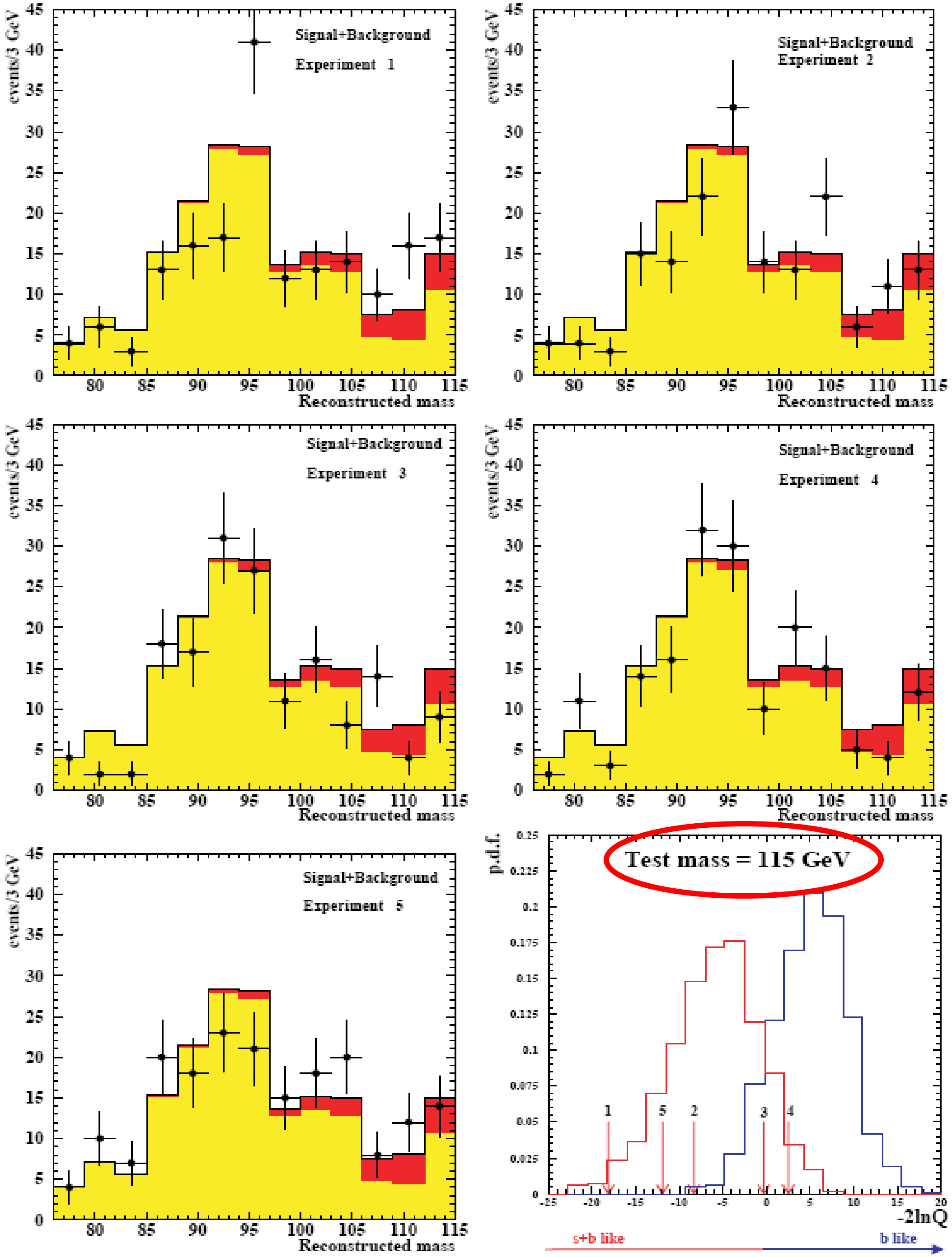
s_i red

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)

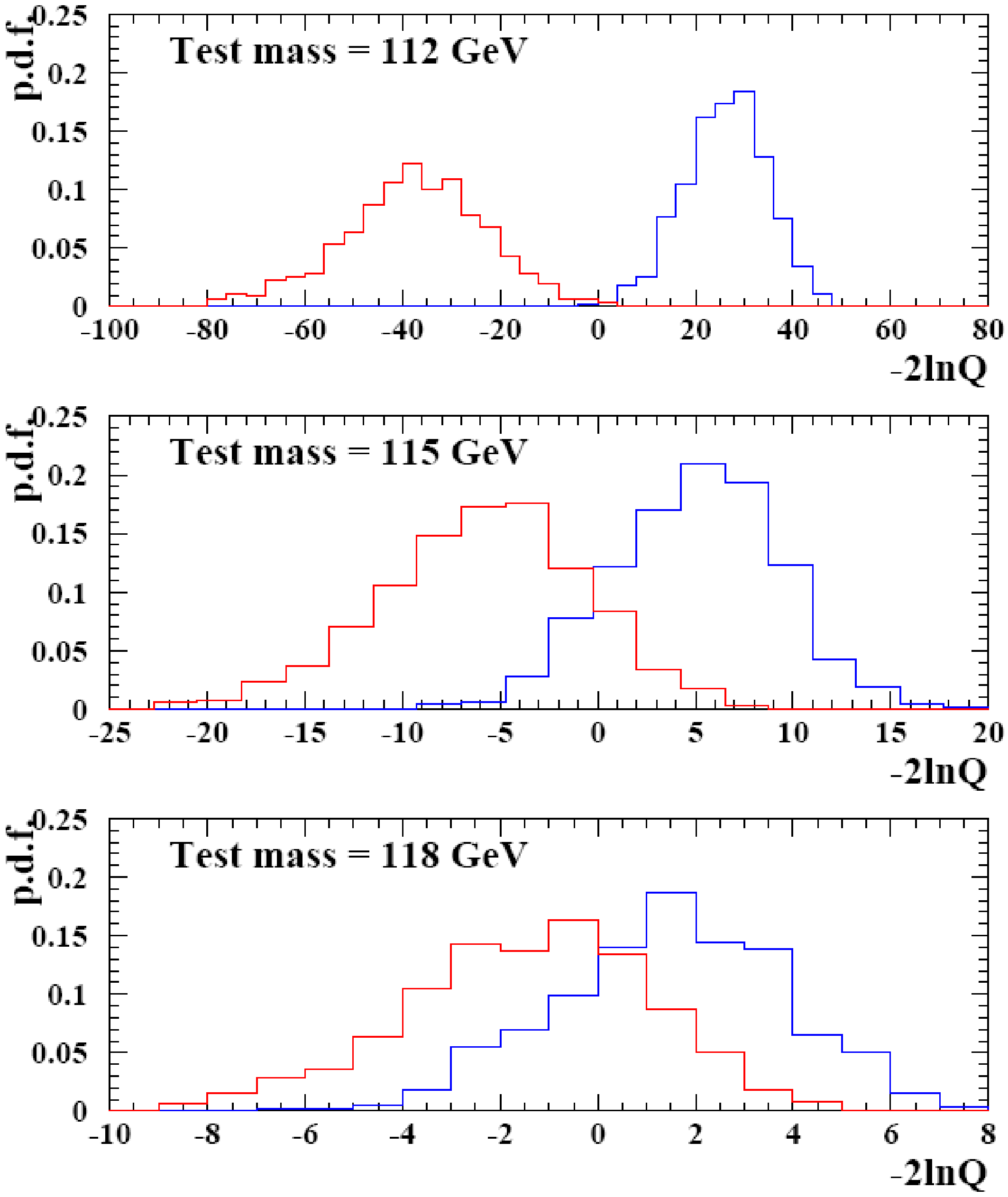
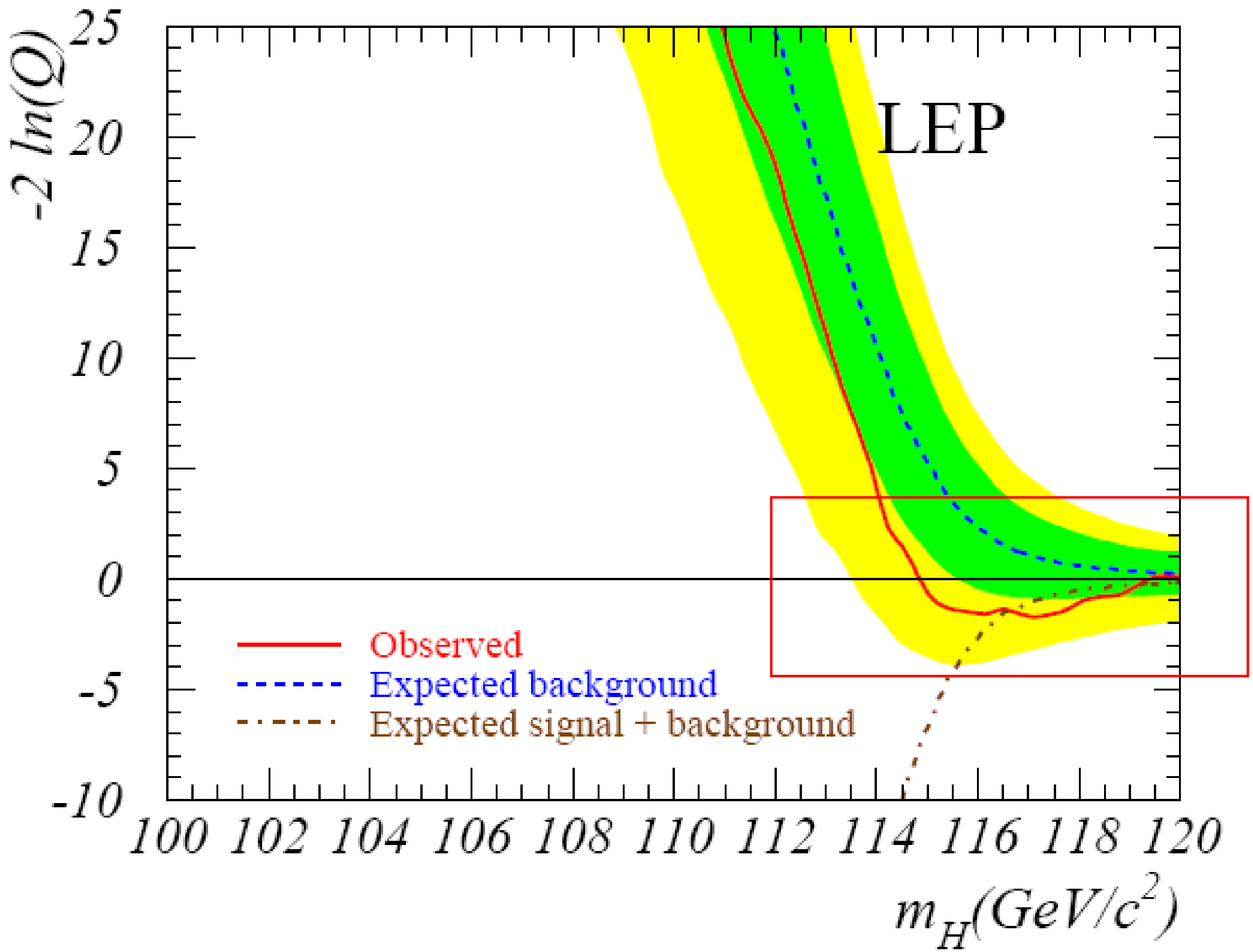


Figure 6: The separation between the Signal and the Background for various Higgs masses is shown by their likelihood p.d.f's.



maximum likelihood zone

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].

3σ effect!

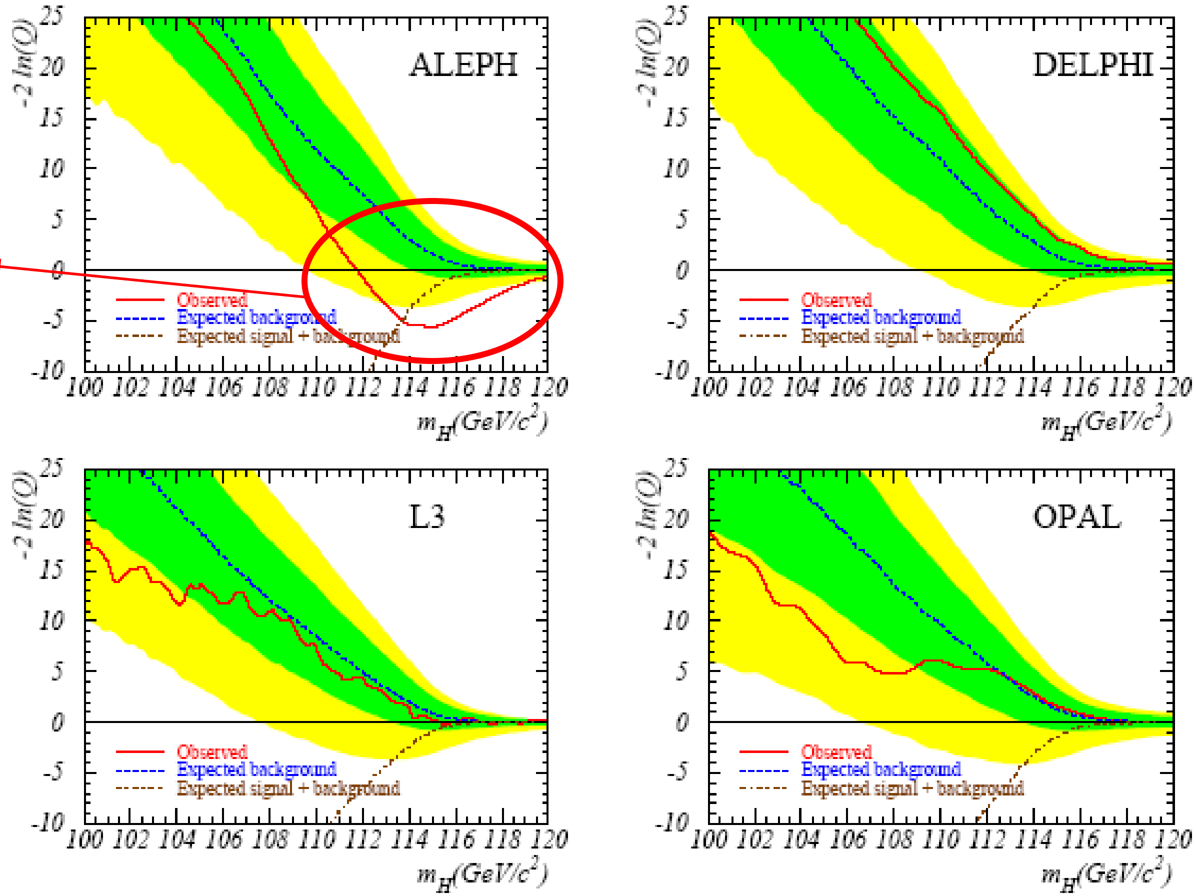
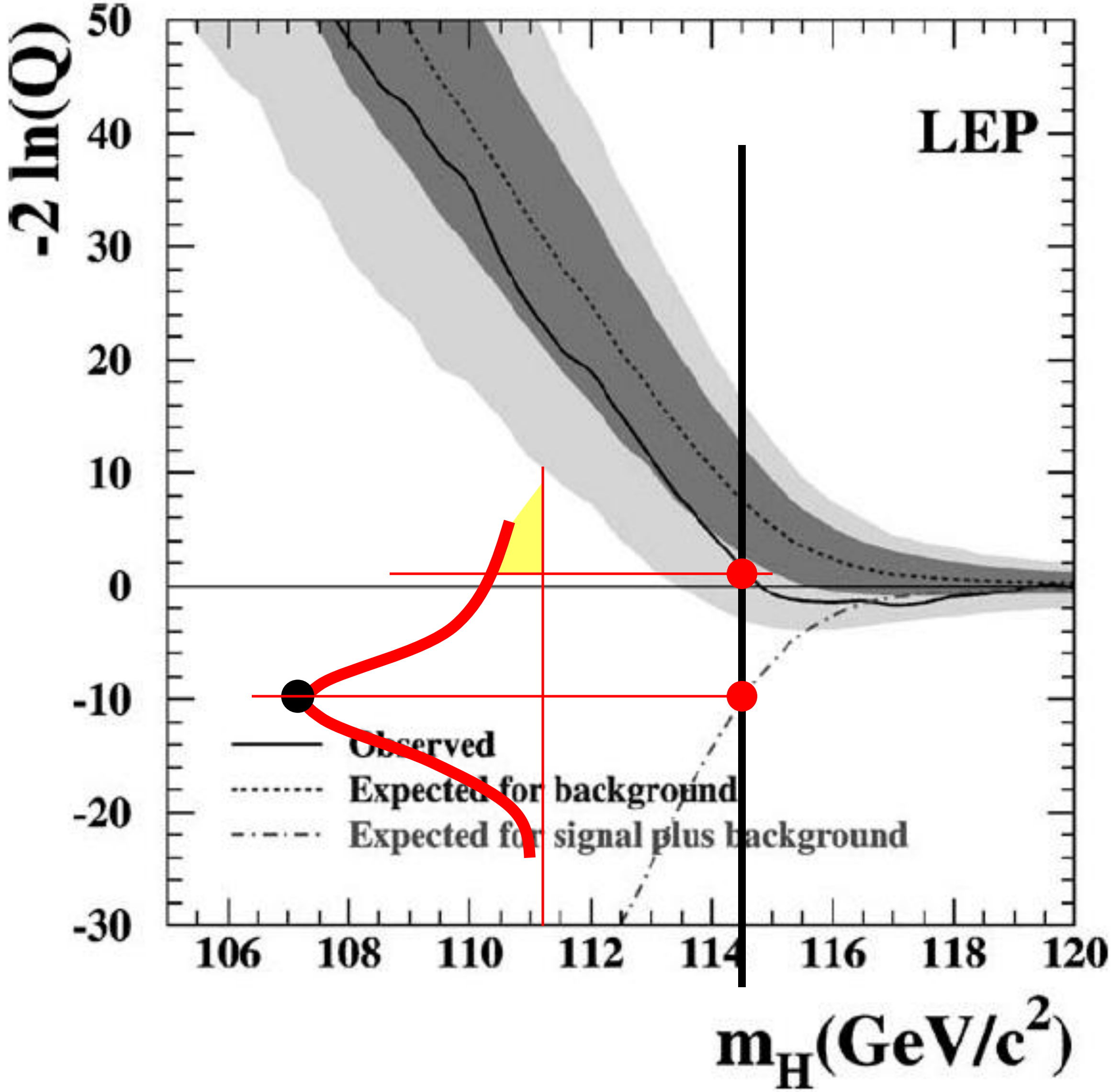
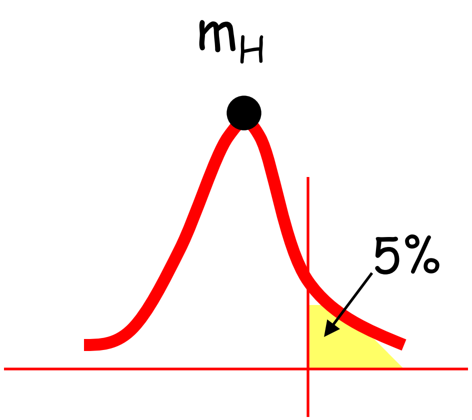


Figure 9: Observed and expected behavior of the likelihood $-2\ln Q$ as a function of the test-mass m_H for the various experiments. The solid/red line represents the observation; the dashed/dash-dotted lines show the median background/signal+background expectations. The dark/green and light/yellow shaded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



ALEPH
DELPHI
L3
OPAL
2003



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

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.

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

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

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

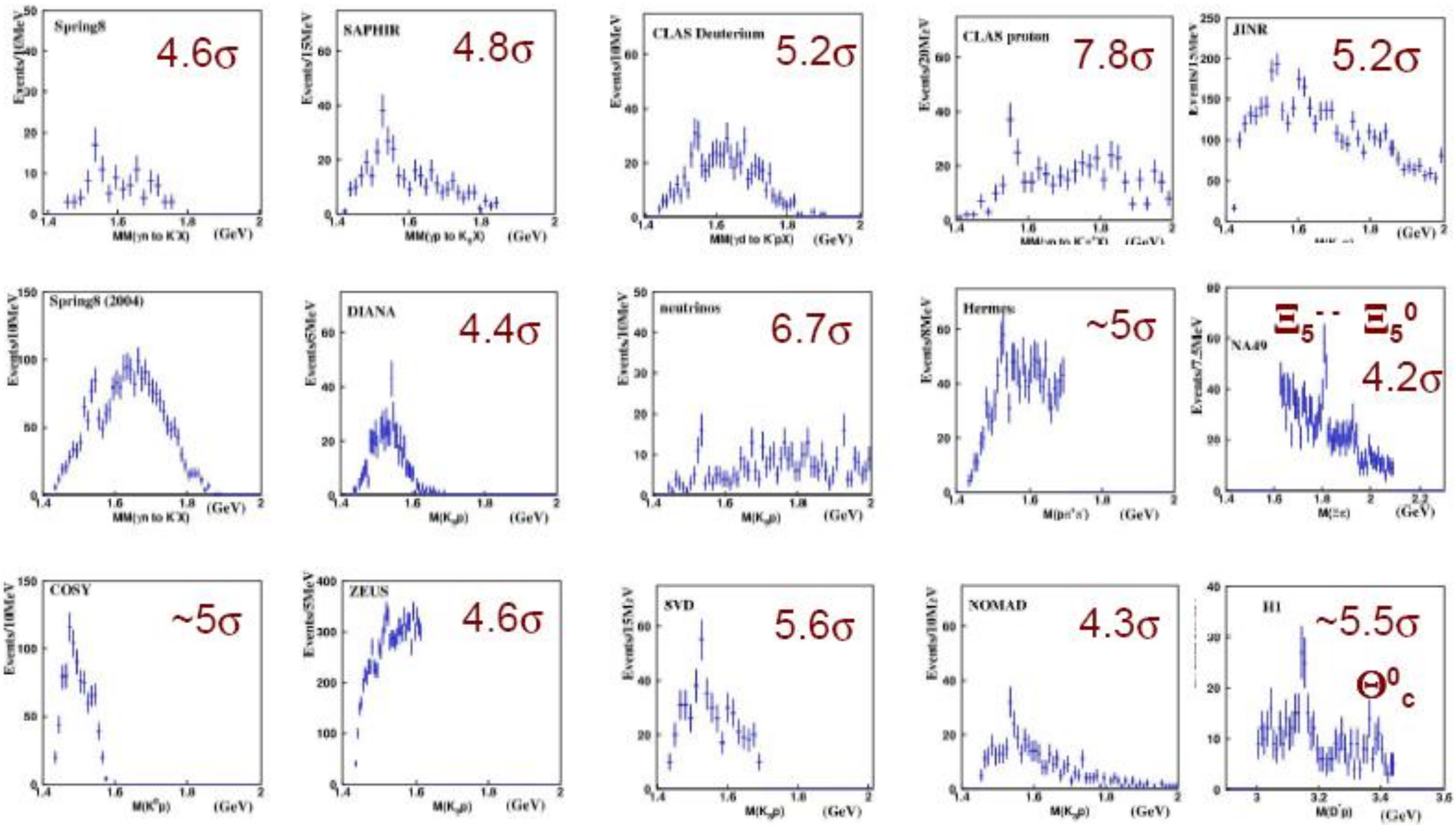
pentaquark, the unique baryon with positive strangeness.

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

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

.. **From the Curtis Meyer review (Miami 2004)**



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

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

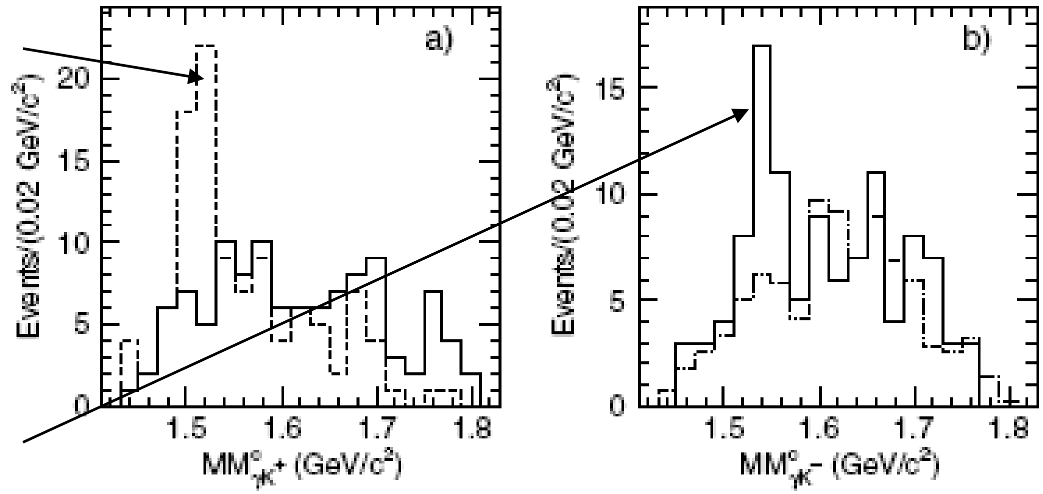


FIG. 3. (a) The $MM^c_{\gamma K^+}$ spectrum [Eq. (2)] for K^+K^- productions for the signal sample (solid histogram) and for events from the SC with a proton hit in the SSD (dashed histogram). (b) The $MM^c_{\gamma K^-}$ spectrum for the signal sample (solid histogram) and for events from the LH_2 (dotted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

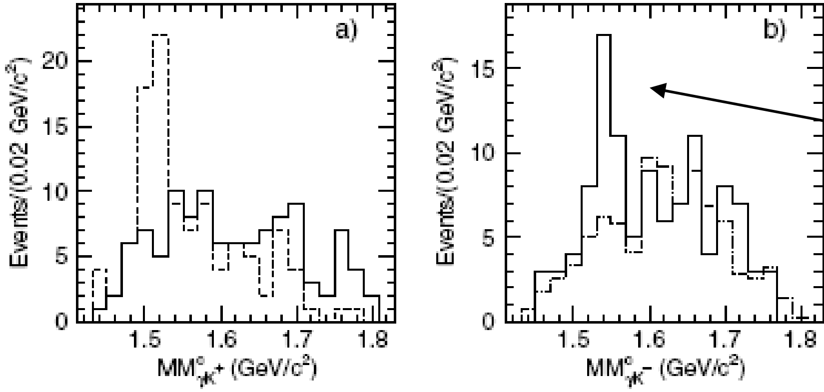


FIG. 3. (a) The $MM^c_{\gamma K^+}$ spectrum [Eq. (2)] for K^+K^- productions for the signal sample (solid histogram) and for events from the SC with a proton hit in the SSD (dashed histogram). (b) The $MM^c_{\gamma K^-}$ spectrum for the signal sample (solid histogram) and for events from the LH_2 (dotted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

4.6 sigma!

Is it convincing???

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

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

By using the new variables

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

$$m = N - n$$

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

one has

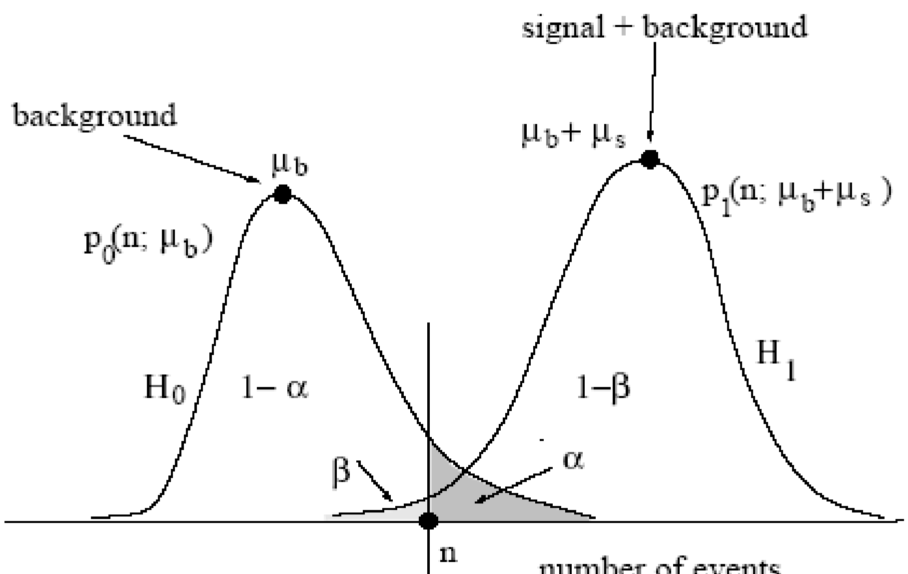
$$P(n|N)P(N) = \frac{\mathrm{e}^{-\lambda\varepsilon}(\lambda\varepsilon)^n}{n!}\,\frac{\mathrm{e}^{-\lambda(1-\varepsilon)}\lambda^m(1-\varepsilon)^m}{m!}$$

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.**

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

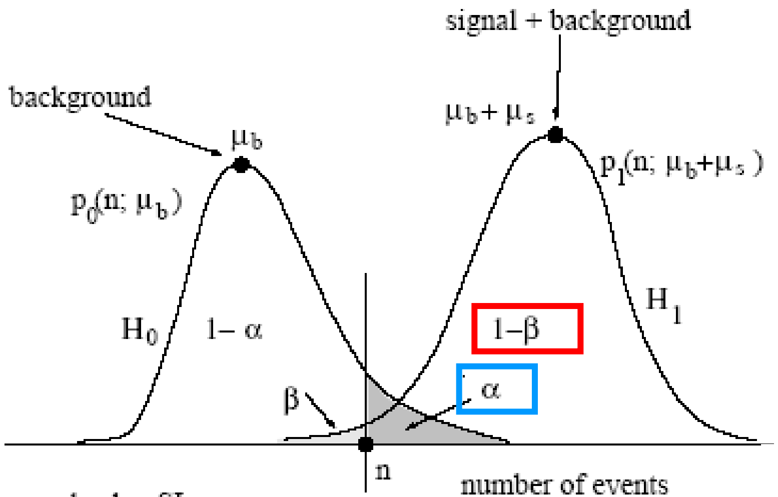
$$p(n, \mu_b + \mu_s) = \frac{(\mu_b + \mu_s)^n}{n!} e^{-\mu_b + \mu_s} \qquad (2)$$



α = backg. SL

β = signal 1− CL

1−β =signal CL or power of the test



α = backg. SL

β = signal 1− CL

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

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

$$1-\alpha = \Phi(Z) \quad \rightarrow \quad z = \Phi^{-1}(1-\alpha)$$

where

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z} \exp(-t^2/2)\,dt = \frac{1+\mathrm{erf}(Z/\sqrt{2})}{2}$$

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

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

Parameter estimation

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

Hypothesis test

This is the most common

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

WRONG

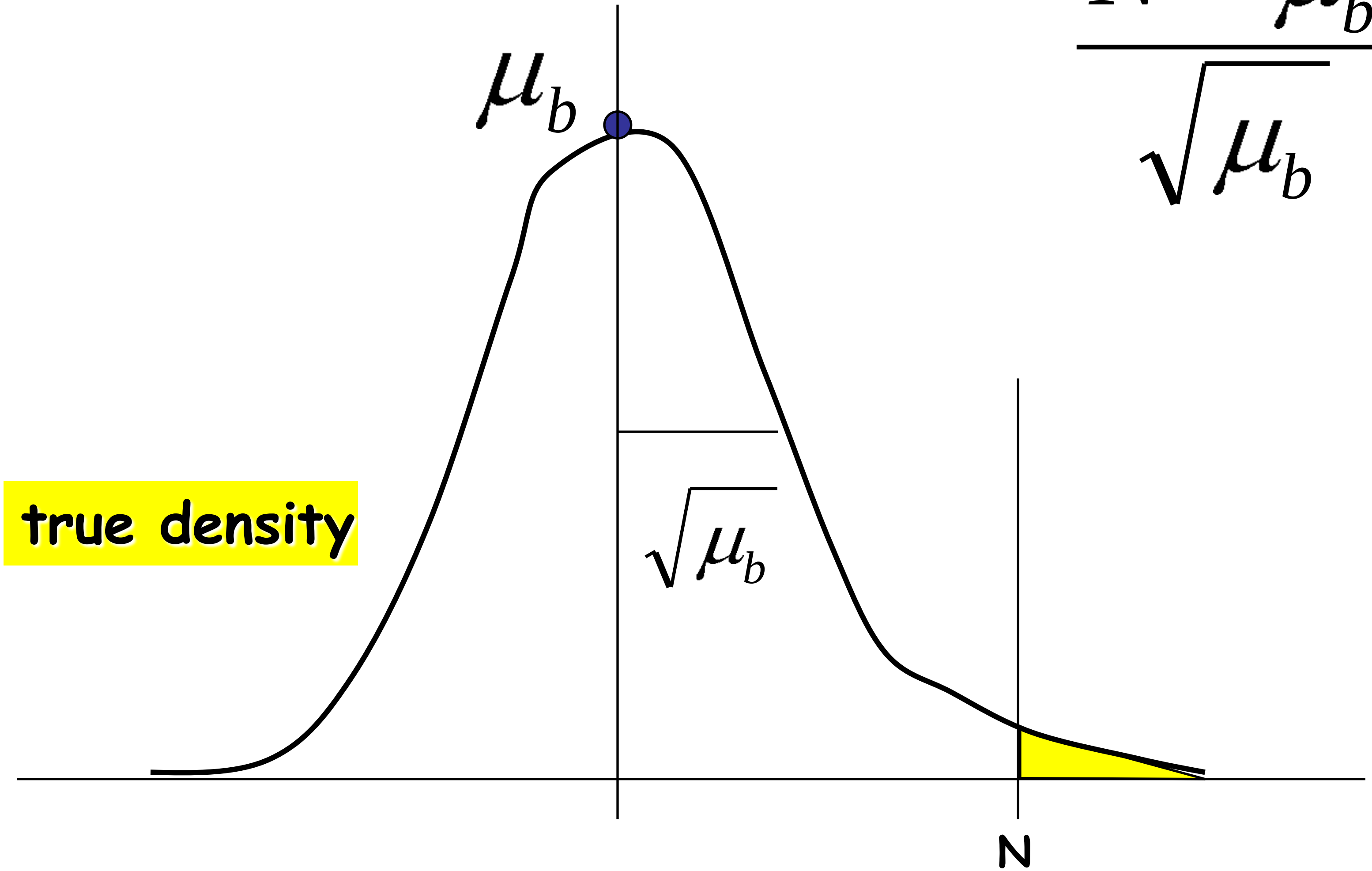
$$S_{sb} = \sqrt{N} - \sqrt{\mu_b} = \sqrt{N_s + N_b} - \sqrt{\mu_b}$$

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!!

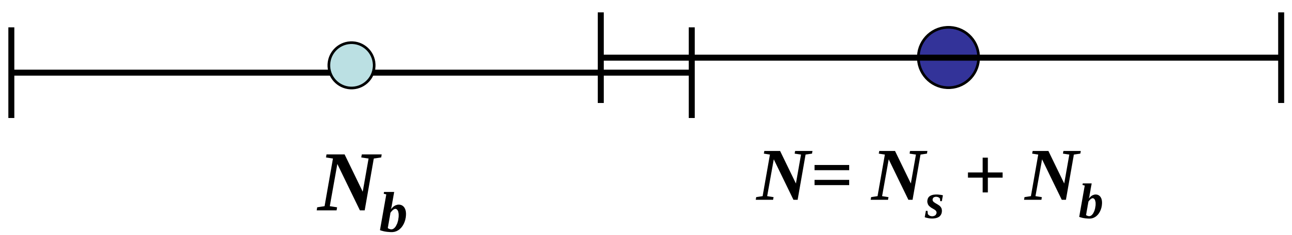
Hypothesis test I

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



Parameter estimation

$$N - N_b \pm \sqrt{N\ \ + N_b} \cong N_s \pm \sqrt{N_s + 2N_b}$$



$$\boxed{\frac{N_s}{\sqrt{N_s + 2N_b}}}$$

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

$$t = \frac{N - \mu_b}{\sqrt{\mu_b}} = \frac{\mu_s + \mu_b - N}{\sqrt{\mu_s + \mu_b}} \quad \Rightarrow \quad N = \sqrt{\mu_b\,(\mu_s + \mu_b)}\ , \quad t = \sqrt{\mu_s + \mu_b} - \sqrt{\mu_b}$$

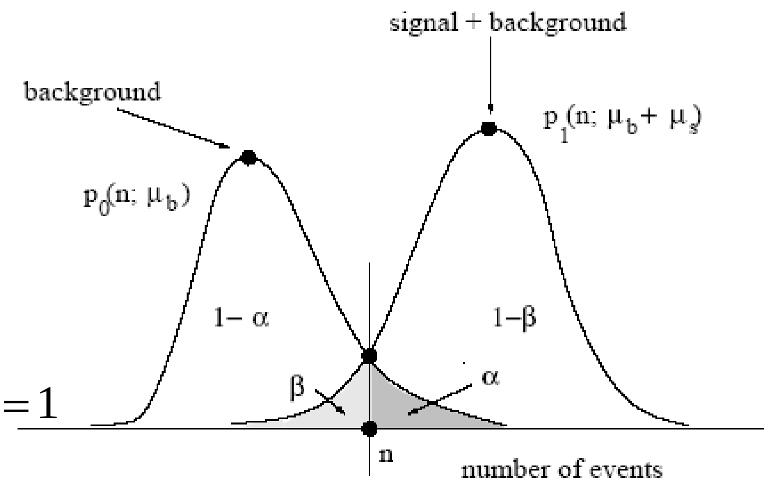
Therefore, one can define the statistic

$$S_{bs} = 2\,(\sqrt{N} - \sqrt{\mu_b}\,)$$

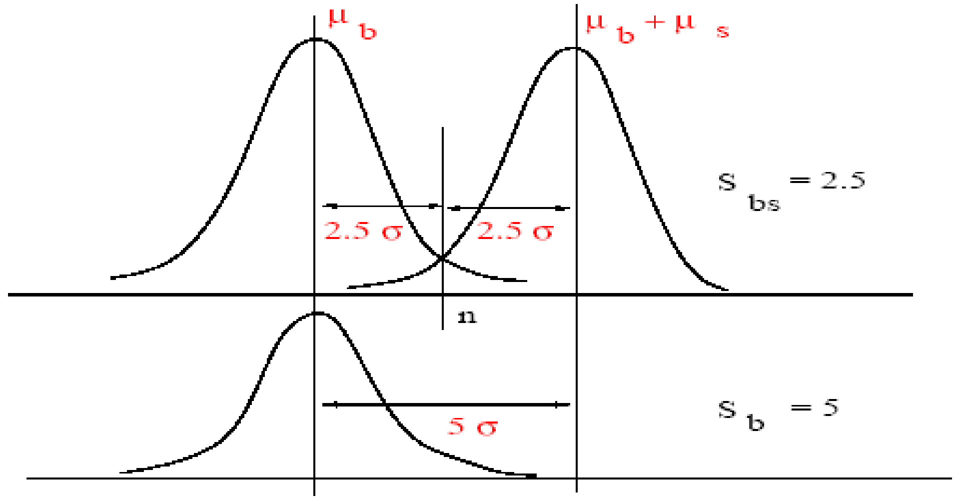
with expectation value

$$\langle S_{bs} \rangle = 2\left(\sqrt{\mu_b + \mu_s} - \sqrt{\mu_b}\right)$$

and unit variance: $\mathrm{Var}[S_{bs}] = 4\ \mathrm{Var}[\sqrt{N}\,] = 4\left(\frac{1}{2\sqrt{N}}\right)^2 N = 1$



Poissonian Signal detection



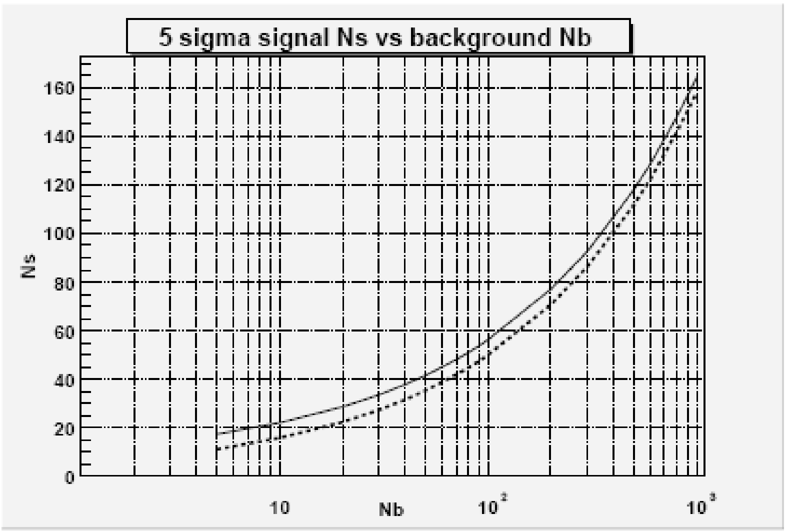


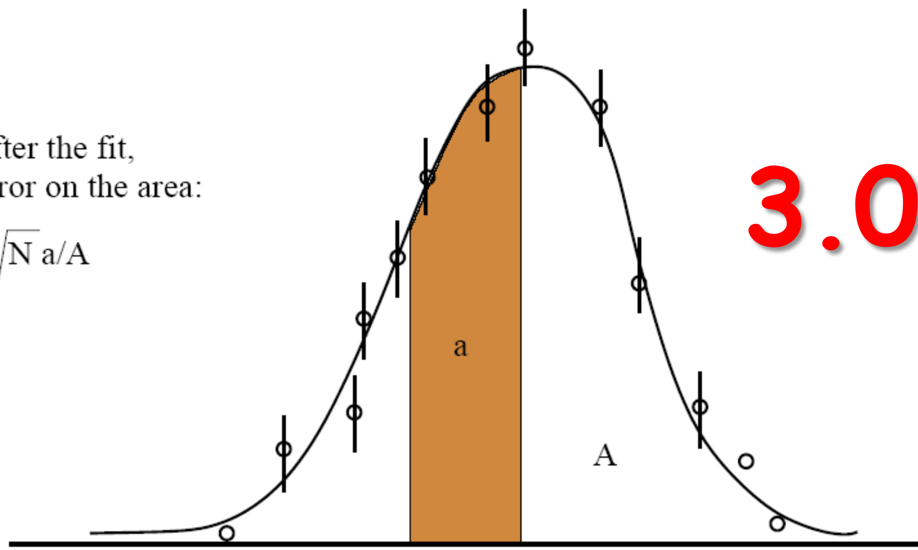
Figure 1: Number N_s of the signal events for $S_b = 5$ (dotted line) and $S_{bs} = 2.5$ (full line) versus the number N_b of background events.

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 $4.6^{+1.2}_{-1.0}\sigma$ $(19.0/\sqrt{17.0} = 4.6)$~~.

after the fit,
error on the area:

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



3.07

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

$$\frac{19}{\sqrt{17+2.8^2}} = 3.8$$

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. Bagdasaryan,[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)

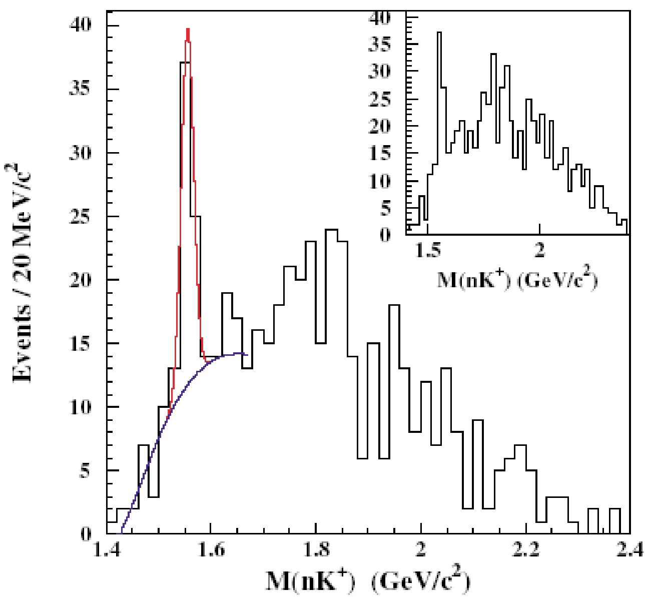


FIG. 4 (color online). The nK^+ invariant mass spectrum in the reaction $\gamma p \rightarrow \pi^+ K^- K^+ (n)$ with the cut $\cos\theta^*_{\pi^+} > 0.8$ and $\cos\theta^*_{K^+} < 0.6$. $\theta^*_{\pi^+}$ and $\theta^*_{K^+}$ are the angles between the π^+ and K^+ mesons and photon beam in the center-of-mass 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 $N_{\Theta^+} = 41 \pm 10$, $M = 1555 \pm 1\ \mathrm{MeV}/c^2$, and $\Gamma = 26 \pm 7\ \mathrm{MeV}/c^2$ (FWHM), where the errors are statistical. The systematic mass scale uncertainty is estimated to be $\pm 10\ \mathrm{MeV}/c^2$. 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 $40\ \mathrm{MeV}/c^2$ 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

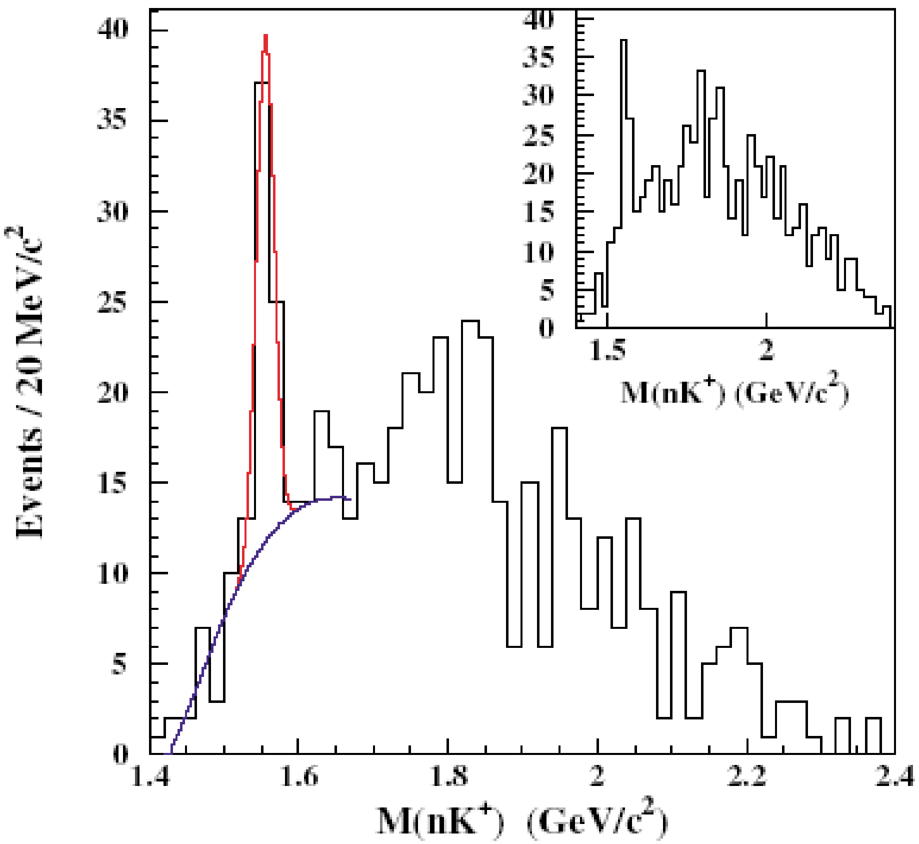


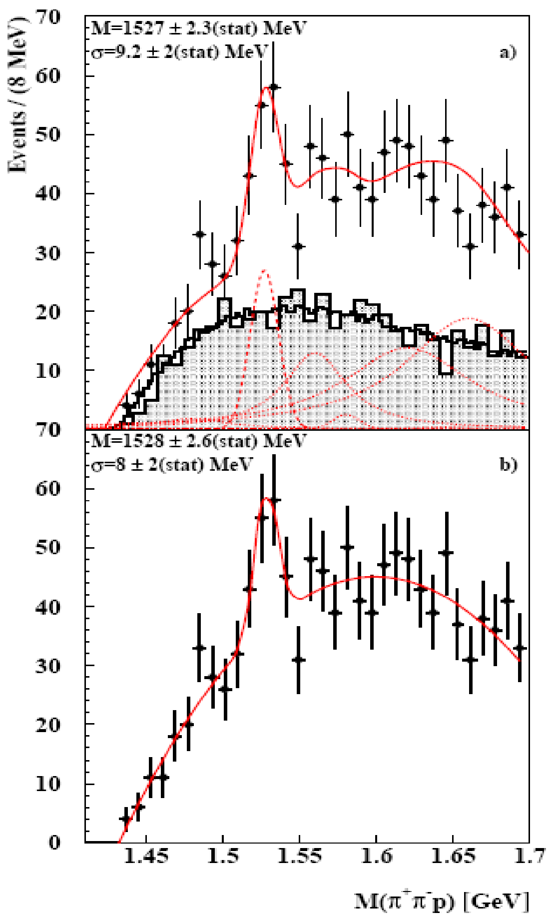
FIG. 4 (color online). The nK^+ invariant mass spectrum in the reaction $\gamma p \to \pi^+ K^- K^+ (n)$ with the cut $\cos\theta^*_{\pi^+} > 0.8$ and $\cos\theta^*_{K^+} < 0.6$. $\theta^*_{\pi^+}$ and $\theta^*_{K^+}$ are the angles between the π^+ and K^+ mesons and photon beam in the center-of-mass 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.

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

$$\frac{41}{\sqrt{41+27+27}} = 4.2$$

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

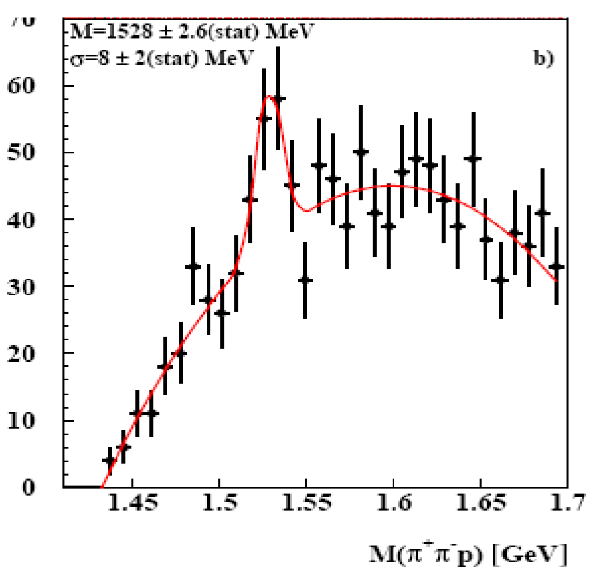
$$\Theta^+ \rightarrow pK^0_S \rightarrow p\pi^+\pi^-$$

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	$6.1\,\sigma$	78 ± 18	$4.3\,\sigma$
a')	$1527.0 \pm 2.5 \pm 2.1$	$24 \pm 5 \pm 2$	79	158	$6.3\,\sigma$	83 ± 20	$4.2\,\sigma$
b)	$1528.0 \pm 2.6 \pm 2.1$	$19 \pm 5 \pm 2$	56	144	$4.7\,\sigma$	59 ± 16	$3.7\,\sigma$
b')	$1527.8 \pm 3.0 \pm 2.1$	$20 \pm 5 \pm 2$	52	155	$4.2\,\sigma$	54 ± 16	$3.4\,\sigma$



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

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

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

Statistics

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

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

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

Experiment	Signal	Background	Publ.	ξ_1	ξ_2	ξ_3
				Significance		
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!!

All these methods estimate true values through measured quantities .....

but ...

Consider N_{on} and $N_{\text{off}} \approx \text{Pois}(\lambda \mu_b)$ with λ known

$$N_{\text{on}} = N_s + N_b$$

$$N_b \cong \frac{N_{\text{off}}}{\lambda} \quad \lambda = \frac{N_{\text{off}}}{N_b} \qquad \sigma_b = \frac{\sqrt{N_{off}}}{\lambda} \quad \lambda = \frac{N_b}{\sigma_b^2}$$

A first rigorous solution

R. Cousins et al, NIM A 595(2008)480

The joint probability of observing $n_{\rm on}$ and $n_{\rm off}$ is the product of Poisson probabilities for $n_{\rm on}$ and $n_{\rm off}$, and can be rewritten as the product of a single Poisson probability with mean $\mu_{\rm tot} = \mu_{\rm on} + \mu_{\rm off}$ for the total number of events $n_{\rm tot}$, and the binomial probability that this total is divided as observed if the binomial parameter ρ is $\rho = \mu_{\rm on}/\mu_{\rm tot} = 1/(1+\lambda)$:

$$\lambda = \mu_{\rm off} / \mu_{\rm on} \xrightarrow{H_0} \mu_{\rm off} / \mu_{\rm b}$$

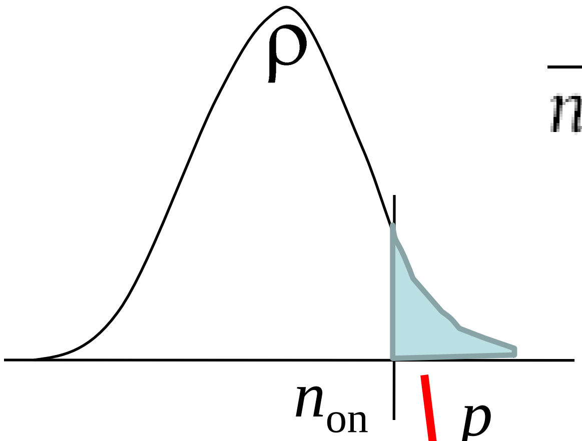
$$P(n_{\rm on}, n_{\rm off}) = \frac{e^{-\mu_{\rm on}} \mu_{\rm on}^{n_{\rm on}}}{n_{\rm on}!} \times \frac{e^{-\mu_{\rm off}} \mu_{\rm off}^{n_{\rm off}}}{n_{\rm off}!}$$

$$= \frac{e^{-(\mu_{\rm on}+\mu_{\rm off})}(\mu_{\rm on}+\mu_{\rm off})^{n_{\rm tot}}}{n_{\rm tot}!} \tag{9}$$

$$\times \frac{n_{\rm tot}!}{n_{\rm on}!(n_{\rm tot}-n_{\rm on})!} \rho^{n_{\rm on}} (1-\rho)^{(n_{\rm tot}-n_{\rm on})}. \tag{10}$$

λ is the known normalization constant supposing that the **on** measurement does not contain the signal **(H_0 hyp.)**

A rigorous solution



$$\frac{n_{\text{tot}}!}{n_{\text{on}}!(n_{\text{tot}} - n_{\text{on}})!} \rho^{n_{\text{on}}} (1-\rho)^{(n_{\text{tot}}-n_{\text{on}})}$$

$$\rho = \mu_{\text{on}}/\mu_{\text{tot}} = 1/(1+\lambda)$$

$$\lambda = \mu_{\text{off}} / \mu_{\text{on}}$$

$$p_{\text{Bi}} = \sum_{j=n_{\text{on}}}^{n_{\text{tot}}} P(j|n_{\text{tot}};\rho) \;\; = B(\rho, n_{\text{on}}, 1+n_{\text{off}})/B(n_{\text{on}}, 1+n_{\text{off}})$$

$$Z = \Phi^{-1}(1-p) = -\Phi^{-1}(p)$$

$$\boxed{Z = \sqrt{2}\,\text{erf}^{-1}(1-2p)}$$

where

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} \exp(-t^2/2)\,\text{d}t = \frac{1+\text{erf}(Z/\sqrt{2})}{2} \quad = 1-p$$

For the simple on/off problem with $n_{\text{on}} = 140$, $n_{\text{off}} = 100$, and $\tau = 1.2$, the ROOT commands are:

```
double n_on  = 140.
double n_off = 100.
double tau = 1.2
double P_Bi  = TMath::BetaIncomplete(1./(1.+tau),n_on,n_off+1)
double Z_Bi  = sqrt(2)*TMath::ErfInverse(1 - 2*P_Bi)
```

Pentaquark: n_on=36, n_off= 17*2.17 = 36.7,

$\tau = \lambda = 17/2.8^2 = 2.17$, Z=3.07

A 2nd rigorous solution

R. Cousins et al, NIM A 595(2008)480

$$\mathscr{L}_{\mathrm{P}} = \frac{(\mu_{\mathrm{s}} + \mu_{\mathrm{b}})^{n_{\mathrm{on}}}}{n_{\mathrm{on}}!} \mathrm{e}^{-(\mu_{\mathrm{s}}+\mu_{\mathrm{b}})} \frac{(\tau\mu_{\mathrm{b}})^{n_{\mathrm{off}}}}{n_{\mathrm{off}}!} \mathrm{e}^{-\tau\mu_{\mathrm{b}}} \tag{20}$$

while for the Gaussian-mean background problem with either absolute or relative σ_{b}, it is

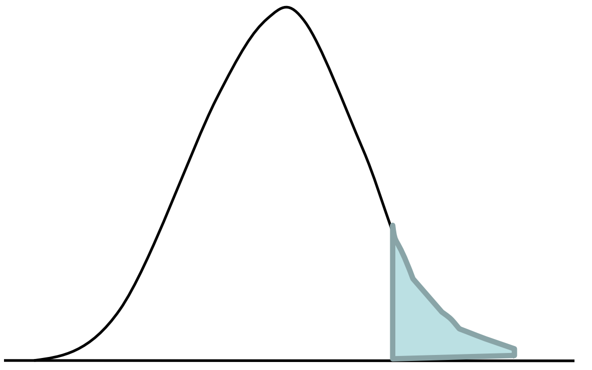
$$\mathscr{L}_{\mathrm{G}} = \frac{(\mu_{\mathrm{s}} + \mu_{\mathrm{b}})^{n_{\mathrm{on}}}}{n_{\mathrm{on}}!} \mathrm{e}^{-(\mu_{\mathrm{s}}+\mu_{\mathrm{b}})} \frac{1}{\sqrt{2\pi\sigma_{\mathrm{b}}^2}} \exp\left(-\frac{(\hat{\mu}_{\mathrm{b}} - \mu_{\mathrm{b}})^2}{2\sigma_{\mathrm{b}}^2}\right) \tag{21}$$

where as discussed below we have explored the effect of truncating the Gaussian pdf in $\hat{\mu}_{\mathrm{b}}$ and renormalizing prior to forming $\mathscr{L}_{\mathrm{G}}$.

Using either $\mathscr{L}_{\mathrm{P}}$ or $\mathscr{L}_{\mathrm{G}}$, one obtains the log-likelihood ratio

$$\Lambda(\mu_{\mathrm{s}}) = \frac{\mathscr{L}(\mu_{\mathrm{s}}, \tilde{\tilde{\mu}}_{\mathrm{b}}(\mu_{\mathrm{s}}))}{\mathscr{L}(\tilde{\mu}_{\mathrm{s}}, \tilde{\mu}_{\mathrm{b}})} \qquad -2\ln\Lambda(\mu_{\mathrm{s}}) < F^{-1}_{\chi^2_1}(1-2\alpha) \tag{22}$$

A 2nd rigourous solution



$$Z_{PL} = \sqrt{-2\ln\Lambda(\mu_s = 0)} \tag{24}$$

where the likelihood ratio is computed using $\mathscr{L}_P$ or $\mathscr{L}_G$, as appropriate for the problem.

For the on/off problem and $\mathscr{L}_P$, the explicit result obtained from Eq. (24) was given by Li and Ma (their Eq. 17) [8]:

$$Z_{PL} = \sqrt{2}\left(n_{on}\ln\frac{n_{on}(1+\tau)}{n_{tot}} + n_{off}\ln\frac{n_{off}(1+\tau)}{n_{tot}\tau}\right)^{1/2}. \tag{25}$$

Pentaquark: n_on=36, n_off= 17*2.17 = 36.7,

$\tau = \lambda = 17/2.8^2 = 2.17$, Z=3.25

Table 1
Test cases and significance results

Reference	[40]	[41]	[42]	[43]	[44]	[44]	[45]	[46]	[47]	[48]
n_{on}	**4**	**6**	**9**	**17**	**50**	**67**	**200**	**523**	**498 426**	**2 119 449**
n_{off}	**5**	18.78	17.83	40.11	**55**	**15**	**10**	**2327**	**493 434**	23 650 096
τ	**5.0**	14.44	4.69	10.56	**2.0**	**0.5**	**0.1**	**5.99**	**1.0**	**11.21**
$\hat{\mu}_{\text{b}}$	1.0	**1.3**	**3.8**	**3.8**	27.5	30.0	100.0	388.6	493 434	**2 109 732**
$s = n_{\text{on}} - \hat{\mu}_{\text{b}}$	3.0	4.7	5.2	13.2	22.5	37	100	**134**	**4992**	**9717**
σ_{b}	0.447	**0.3**	**0.9**	**0.6**	3.71	7.75	31.6	8.1	702.4	433.8
$f = \sigma_{\text{b}}/\hat{\mu}_{\text{b}}$	0.447	0.231	0.237	0.158	0.135	0.258	0.316	0.0207	0.00142	0.000206
Reported p		**0.003**	**0.027**	**2E−06**						
Reported Z		2.7	1.9	**4.6**				**5.9**	**5.0**	**6.4**
See conclusion										
$Z_{\text{Bi}} = Z_{\Gamma}$ binomial	**1.66**	**2.63**	**1.82**	**4.46**	**2.93**	**2.89**	**2.20**	**5.93**	**5.01**	**6.40**
Z_{N} Bayes Gaussian	1.88	2.71	1.94	4.55	3.08	*3.44*	*2.90*	**5.93**	**5.02**	**6.40**
Z_{PL} profile likelihood	1.95	2.81	1.99	4.57	3.02	3.04	2.38	**5.93**	**5.01**	**6.41**
Z_{ZR} variance stabilization	1.93	2.66	1.98	4.22	3.00	3.07	2.39	5.86	**5.01**	**6.40**
Not recommended										
$Z_{\text{BiN}} = s/\sqrt{n_{\text{tot}}/\tau}$	*2.24*	*3.59*	2.17	*5.67*	3.11	**2.89**	**2.18**	6.16	**5.01**	**6.41**
$Z_{\text{nn}} = s/\sqrt{n_{\text{on}} + n_{\text{off}}/\tau^2}$	1.46	*1.90*	1.66	*3.17*	2.82	3.28	*2.89*	*5.54*	**5.01**	**6.40**
$Z_{\text{ssb}} = s/\sqrt{\hat{\mu}_{\text{b}} + s}$	1.50	*1.92*	1.73	*3.20*	3.18	*4.52*	*7.07*	5.88	*7.07*	6.67
$Z_{\text{bo}} = s/\sqrt{n_{\text{off}}(1+\tau)/\tau^2}$	*2.74*	*3.99*	*2.42*	*6.47*	*3.50*	*3.90*	*3.02*	6.31	**5.03**	**6.41**
Ignore σ_{b}										
Z_{P} Poisson: ignore σ_{b}	2.08	2.84	2.14	4.87	*3.80*	*5.76*	*8.76*	*6.44*	*7.09*	6.69
$Z_{\text{sb}} = s/\sqrt{\hat{\mu}_{\text{b}}}$	*3.00*	*4.12*	*2.67*	*6.77*	*4.29*	*6.76*	*10.00*	*6.82*	*7.11*	6.69
Unsuccessful ad hockery										
Poisson: $\mu_{\text{b}} \to \hat{\mu}_{\text{b}} + \sigma_{\text{b}}$	1.56	2.51	1.64	**4.47**	3.04	*4.24*	*5.51*	6.01	*6.09*	**6.39**
$s/\sqrt{\hat{\mu}_{\text{b}} + \sigma_{\text{b}}}$	*2.49*	*3.72*	*2.40*	*6.29*	*4.03*	*6.02*	*8.72*	*6.75*	*7.10*	6.69

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:

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

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

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 ε, $b \ll 1$,

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

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

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

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

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:

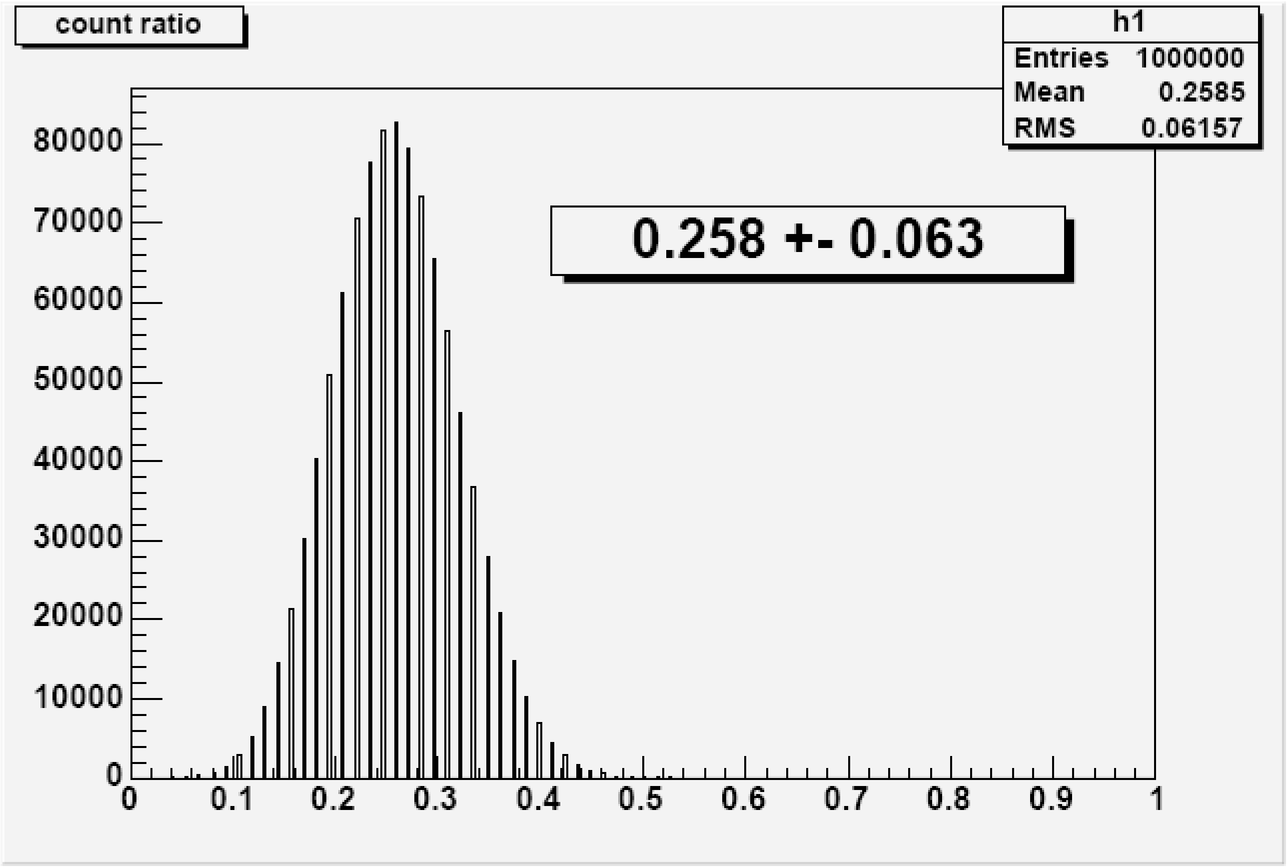
- **SVX tagging: b jets identification by searching for secondary vertices in the Silicon Vertex detector;**
- **SLT tagging: to search for an additional soft lepton from semileptonic b decay**

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

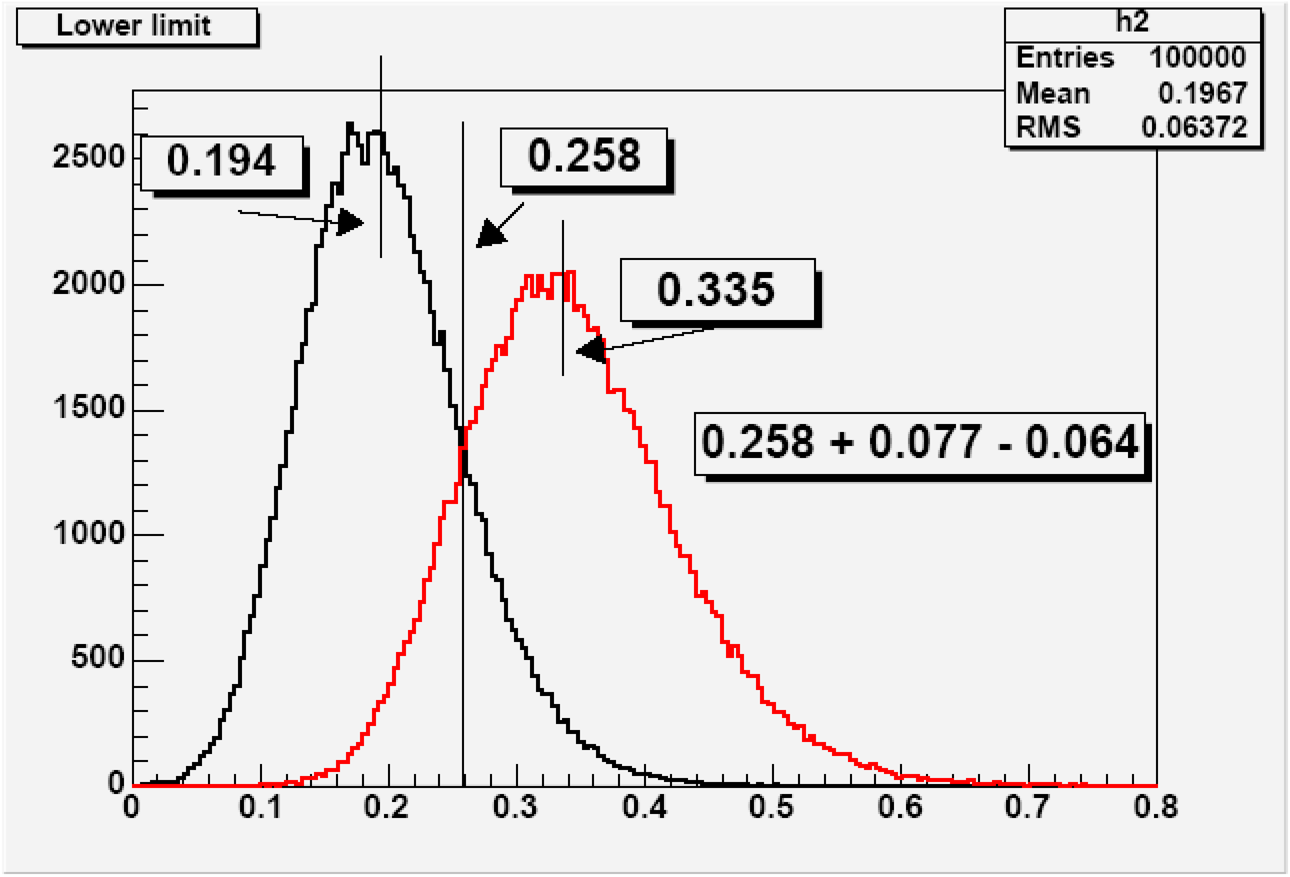
The bootstrap method for confidence levels

With fixed efficiencies we have the binomial/gaussian distribution



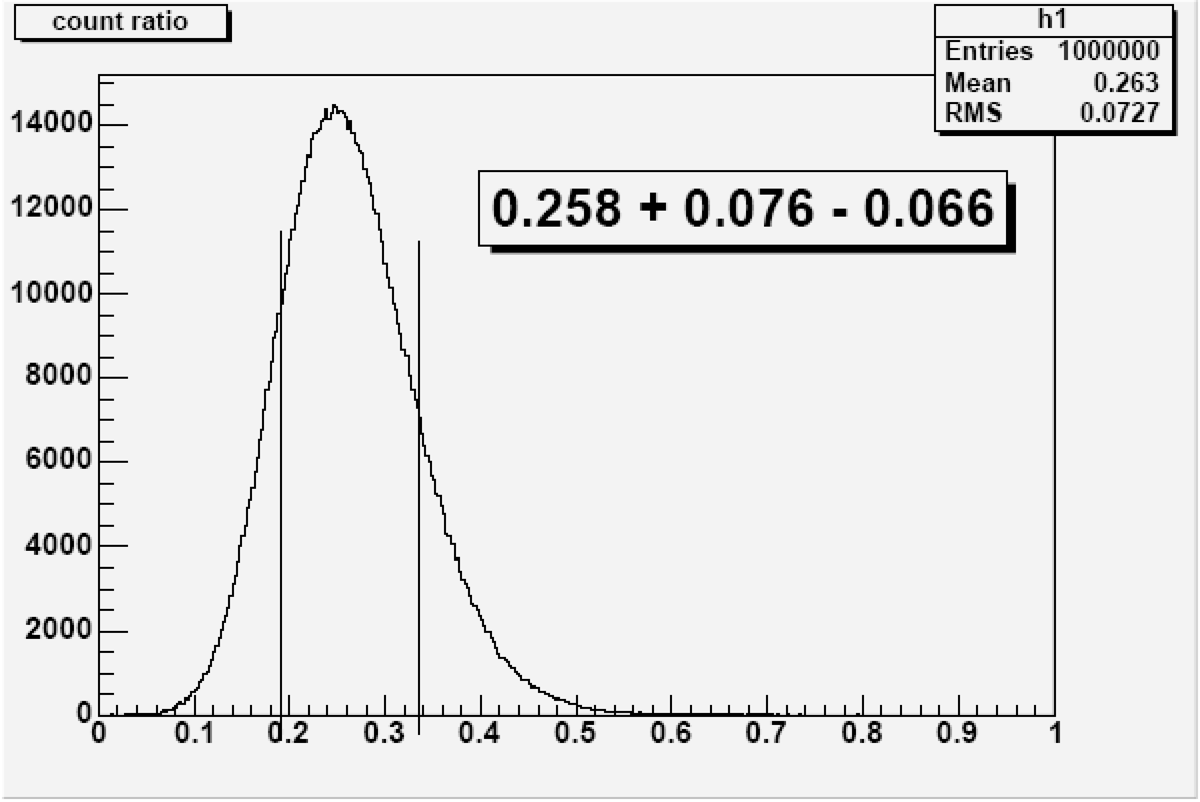
The grid method for confidence levels

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 non parametric Bootstrap

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:

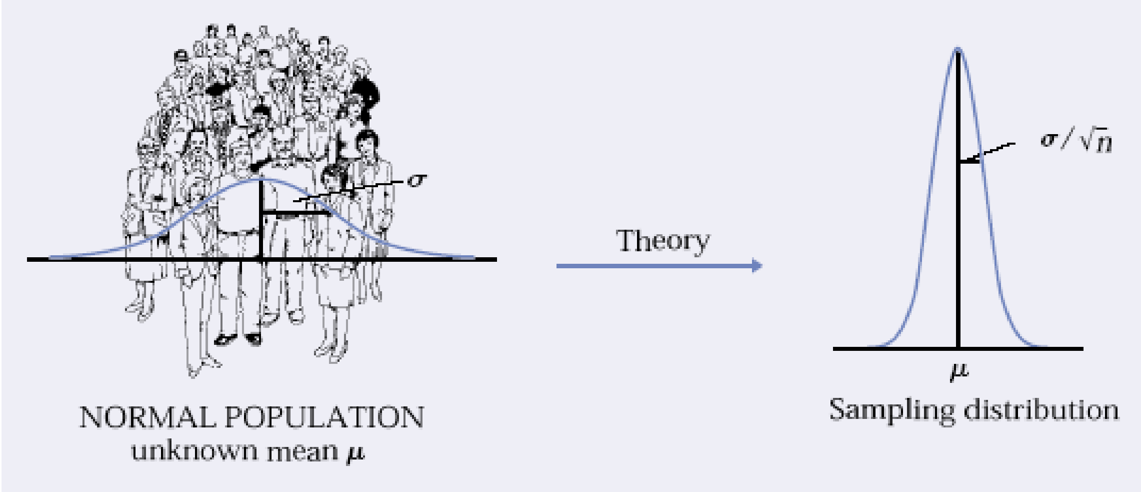
The non parametric Sampling methods

- **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

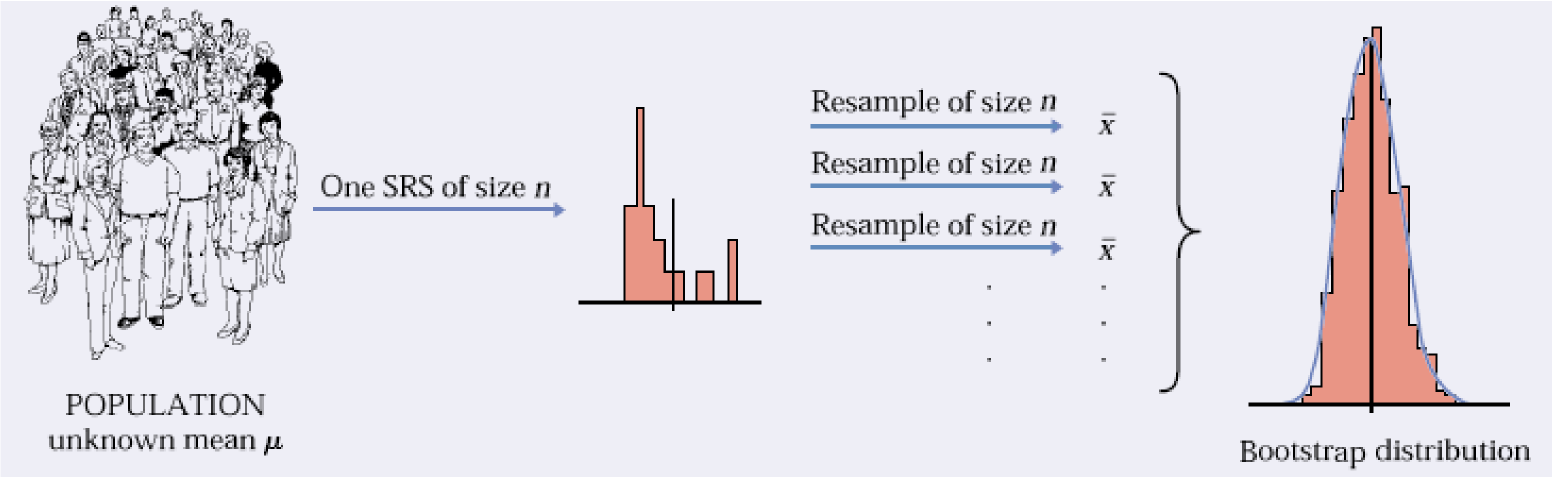
The best one !!!

These methods, familiar among statisticians, are practically not (yet) used by physicists (only 3 papers with non parametric Bootstrap!) ← **Up to 2006**

Non parametric Bootstrap



(b)



(c)

- **Center:** A statistic is biased as an estimate of the parameter if its sampling distribution is not centered at the true value of the parameter. We can check bias by seeing whether the bootstrap distribution of the statistic is centered at the value of the statistic for the original sample.

  bias

  More precisely, the **bias** of a statistic is the difference between the mean of its sampling distribution and the true value of the parameter. The **bootstrap estimate of bias** is the difference between the mean of the bootstrap distribution and the value of the statistic in the original sample.

  bootstrap estimate of bias

- **Spread:** The bootstrap standard error of a statistic is the standard deviation of its bootstrap distribution. The bootstrap standard error estimates the standard deviation of the sampling distribution of the statistic.

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 $s_{\rm res}$ of the (re)samples:**

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

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

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

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

- **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**

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

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}

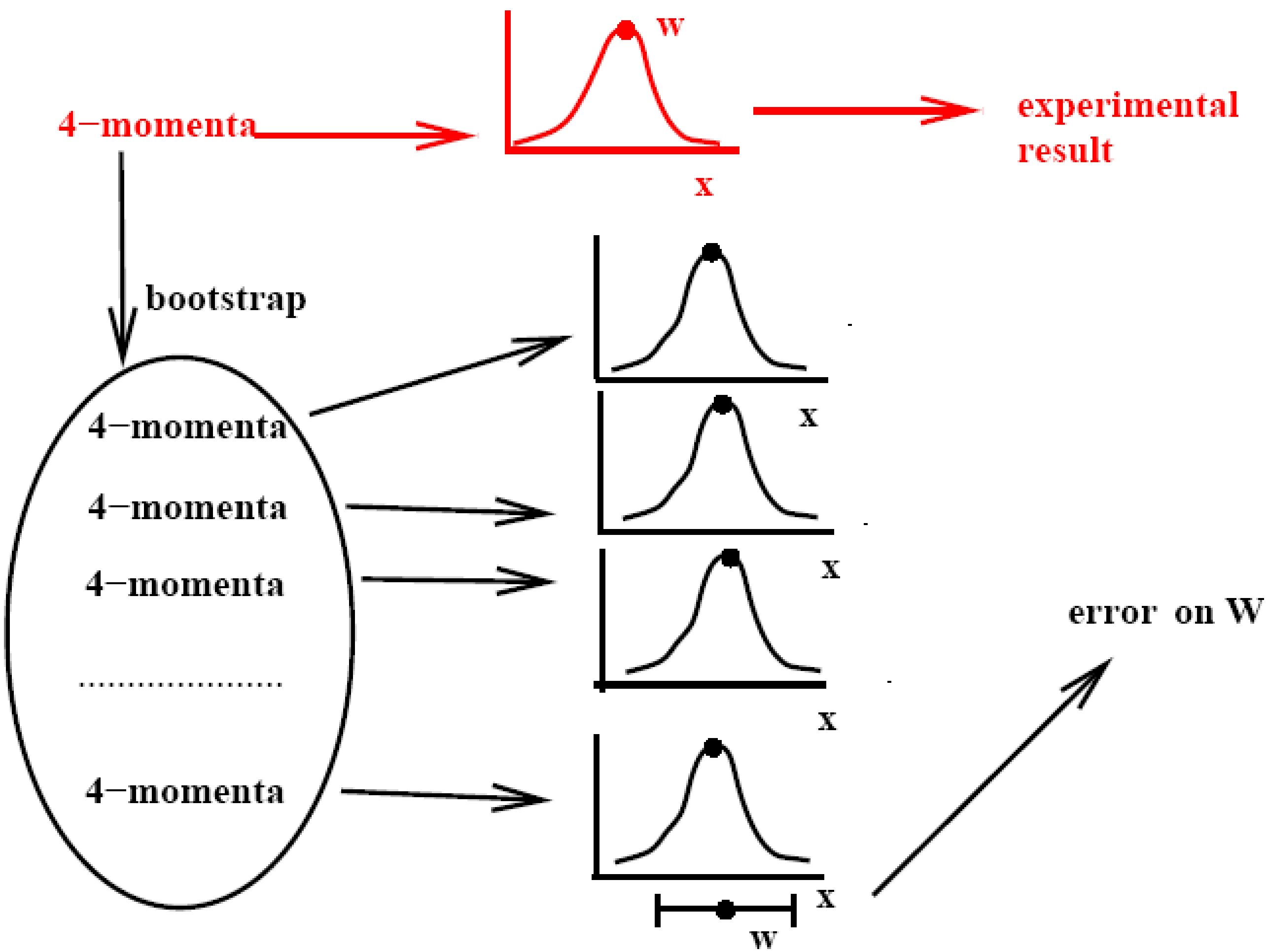
- make different estimates with different bootstrap samples (with replacement) $N_B \leq N$ and verify that the variances scales as $1/N_B$. This verify the condition

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

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

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:

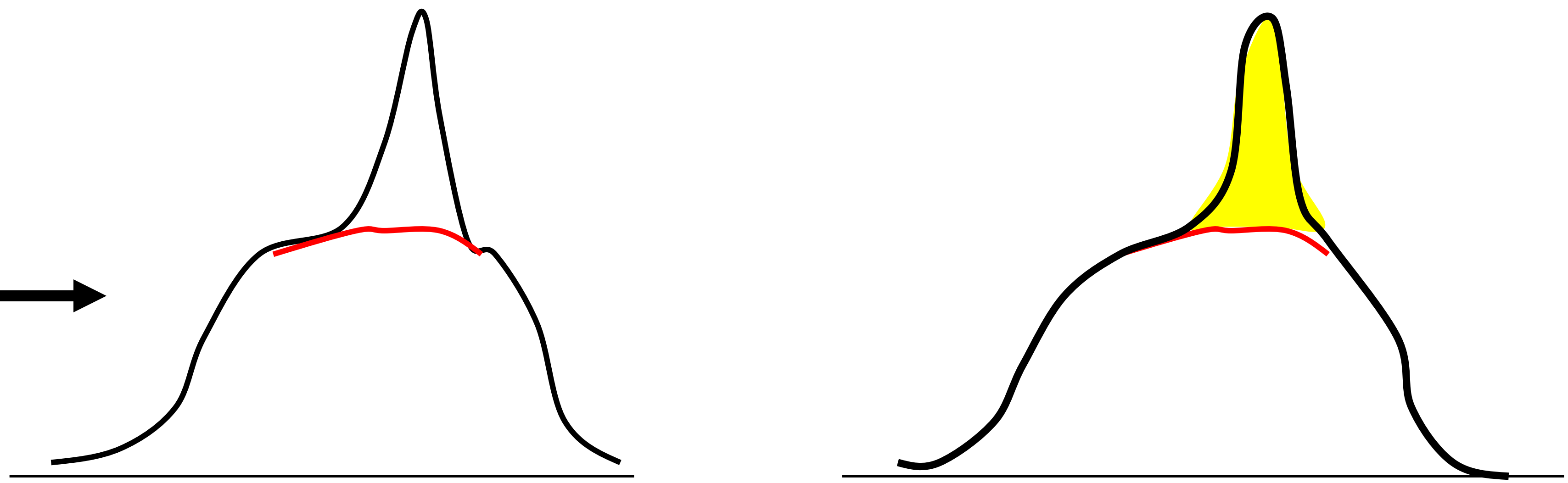
1. Draw a resample of size n with replacement from the first sample and a separate resample of size m from the second sample. Compute a statistic that compares the two groups, such as the difference between the two sample means.

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. Use the Binomial formula!

- **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 end

- Back-up slides

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 $x = 2$

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

Bayesian Hypothesis test

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

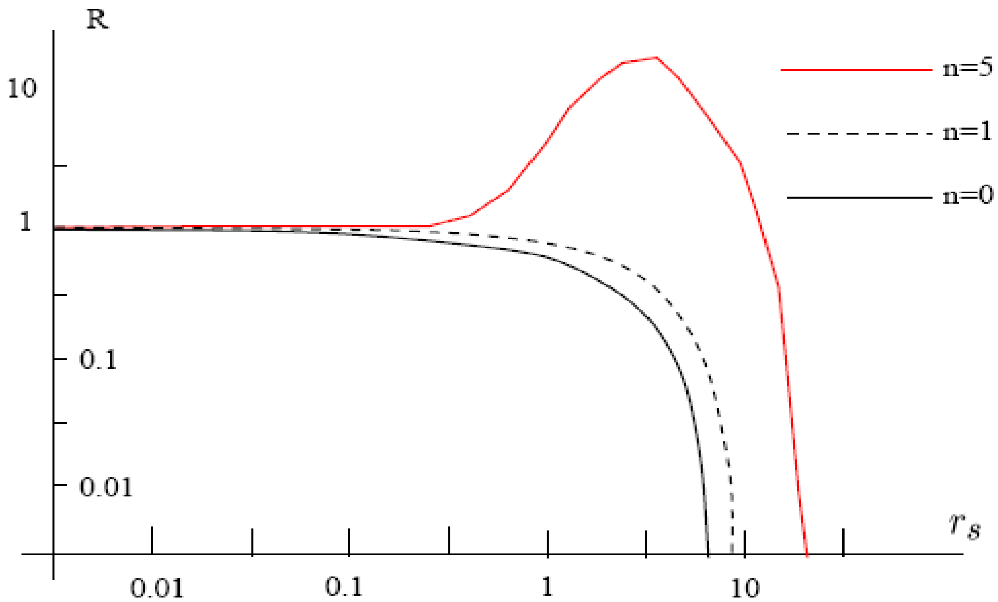
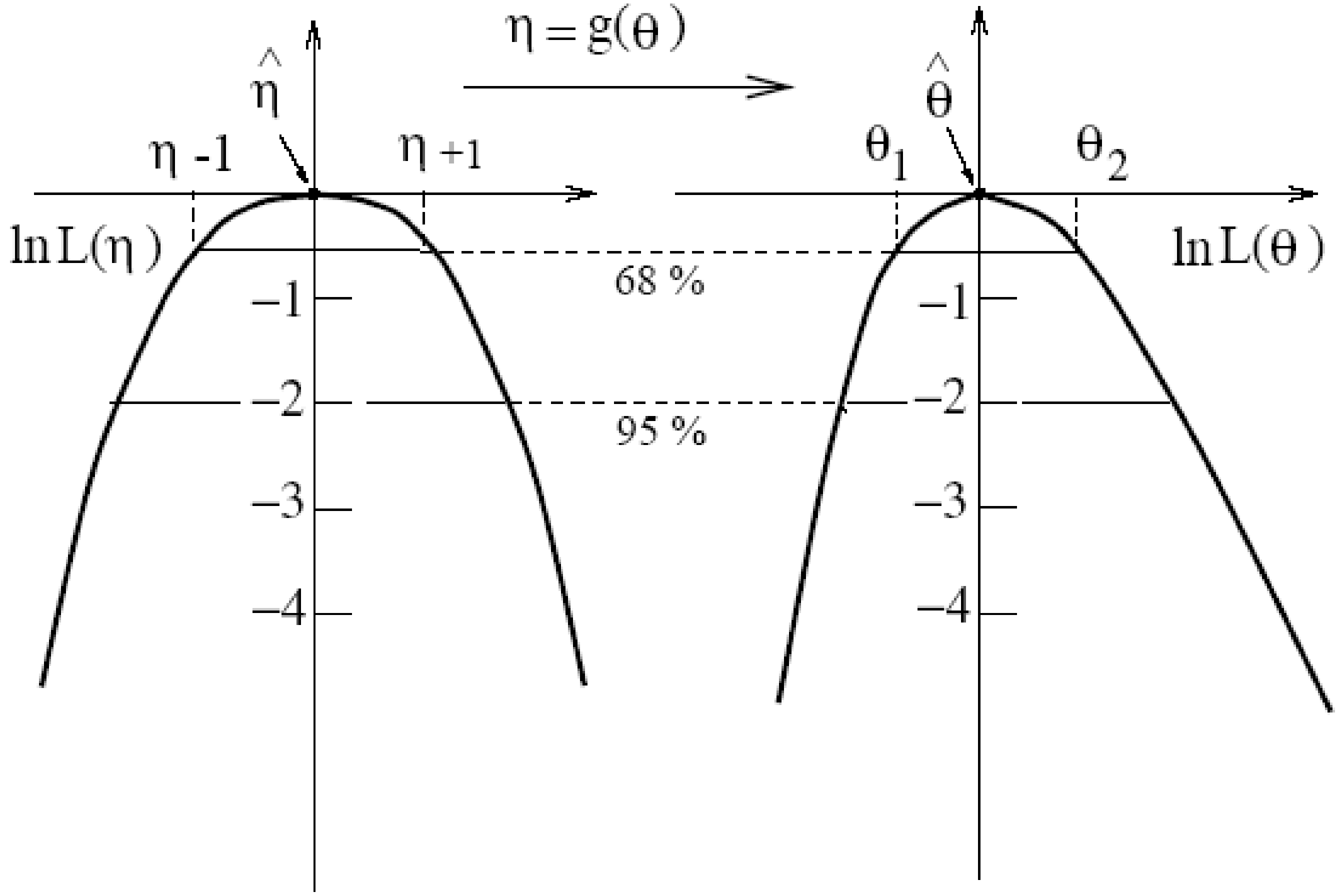


Figure 1: ratio R for the poisson intensity parameter r in units of events per month for an expected background rate $r_b = 1$ event/month and for $n = 0, 1, 5$ observed events

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

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** $r_s = 4$



$\eta = g(\theta)$
$\hat{\eta}$
$\hat{\theta}$
η -1
η +1
θ_1
θ_2
$\ln L(\eta)$
$\ln L(\theta)$
68 %
95 %
−1
−2
−3
−4

LEP real data

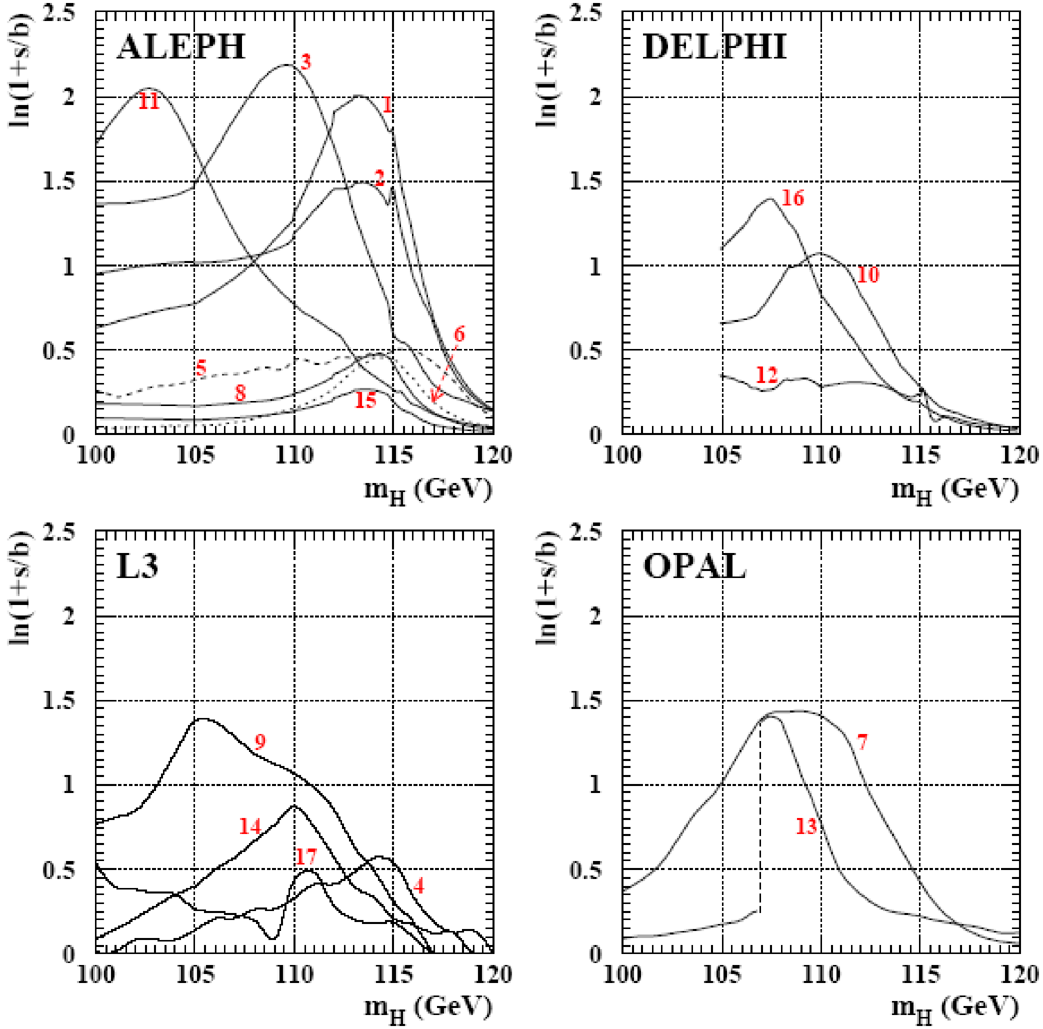


Figure 3: Evolution of the event weight $\ln(1+s/b)$ with test-mass m_H for the events with the largest weight at $m_H = 115$ GeV. The labels correspond to the candidate numbers in the first column of Table 1. The sudden increase in the weight of the OPAL missing-energy candidate labeled "13" at $m_H = 107$ 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].

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:

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

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

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

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:

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

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

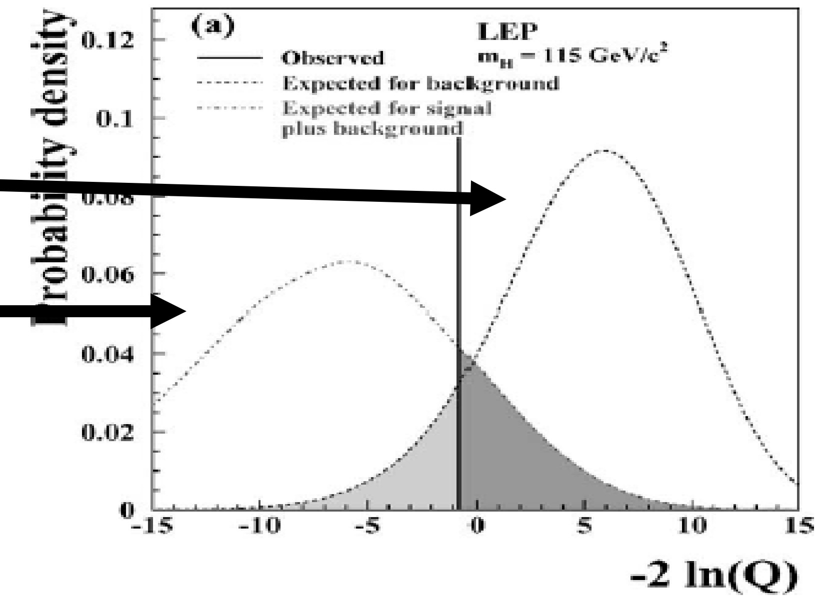
$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

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

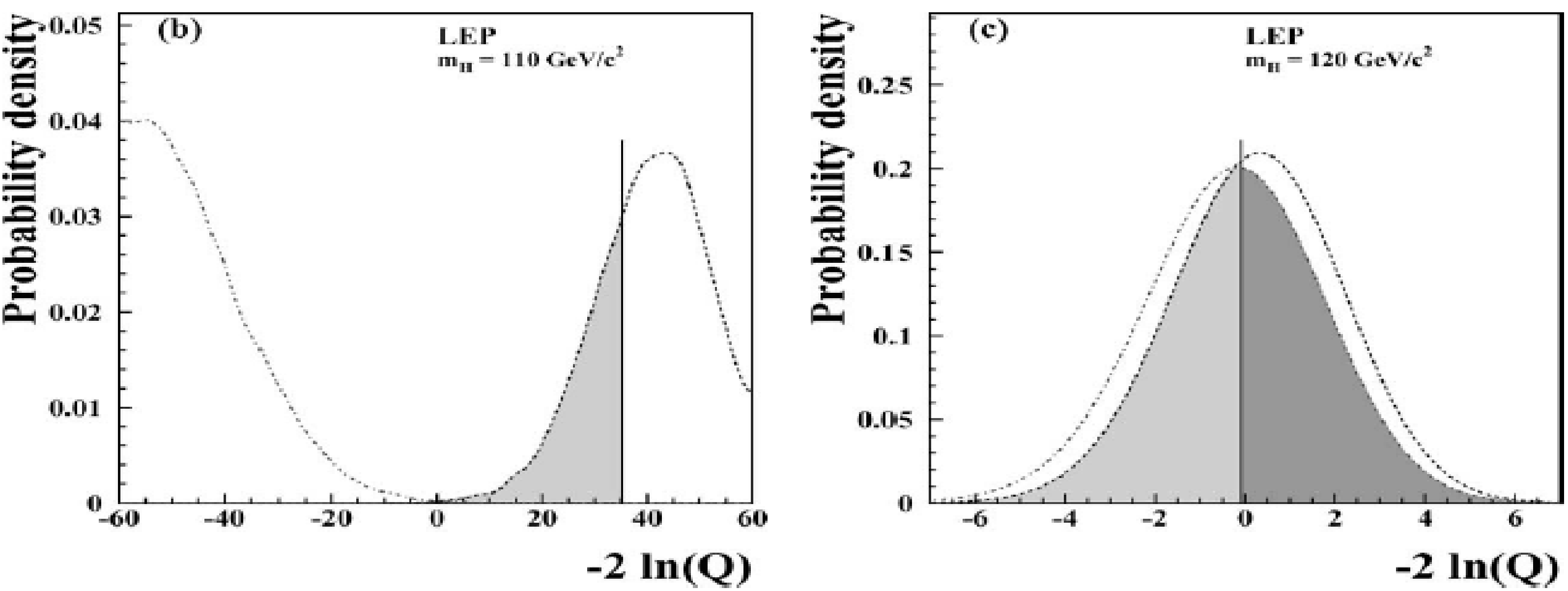
MC samples

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

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

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

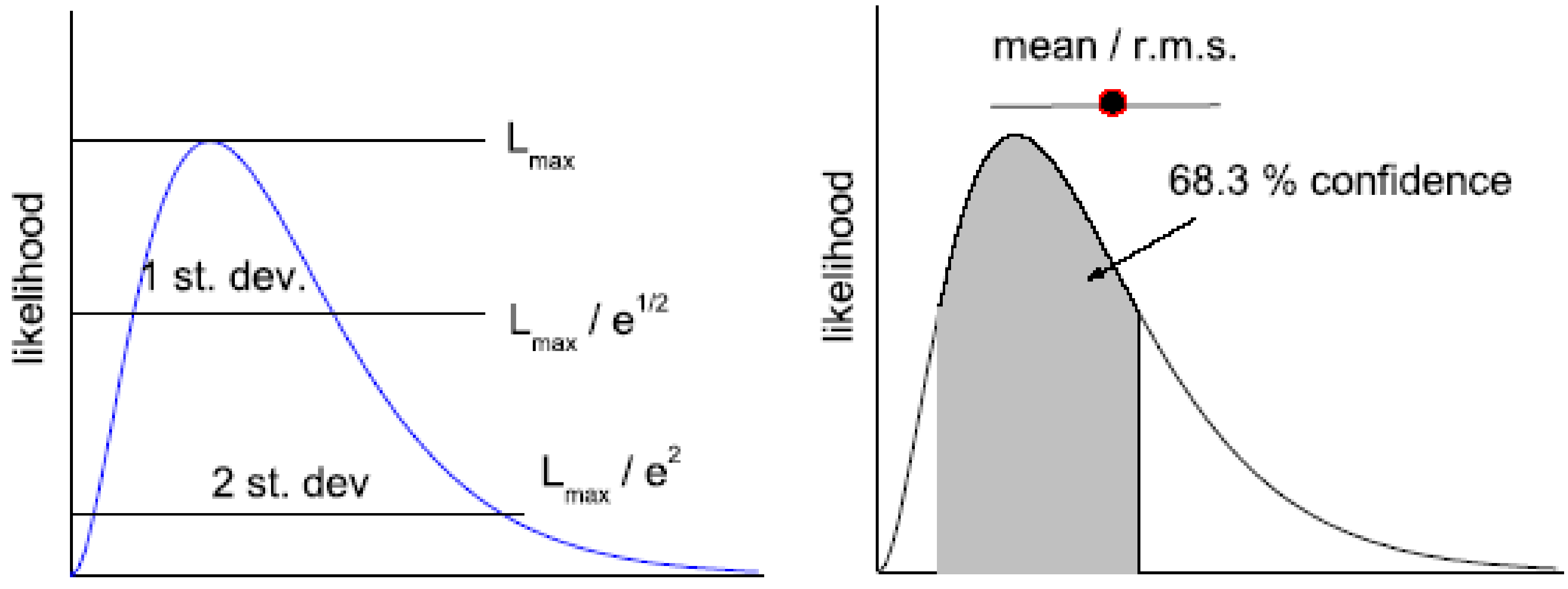
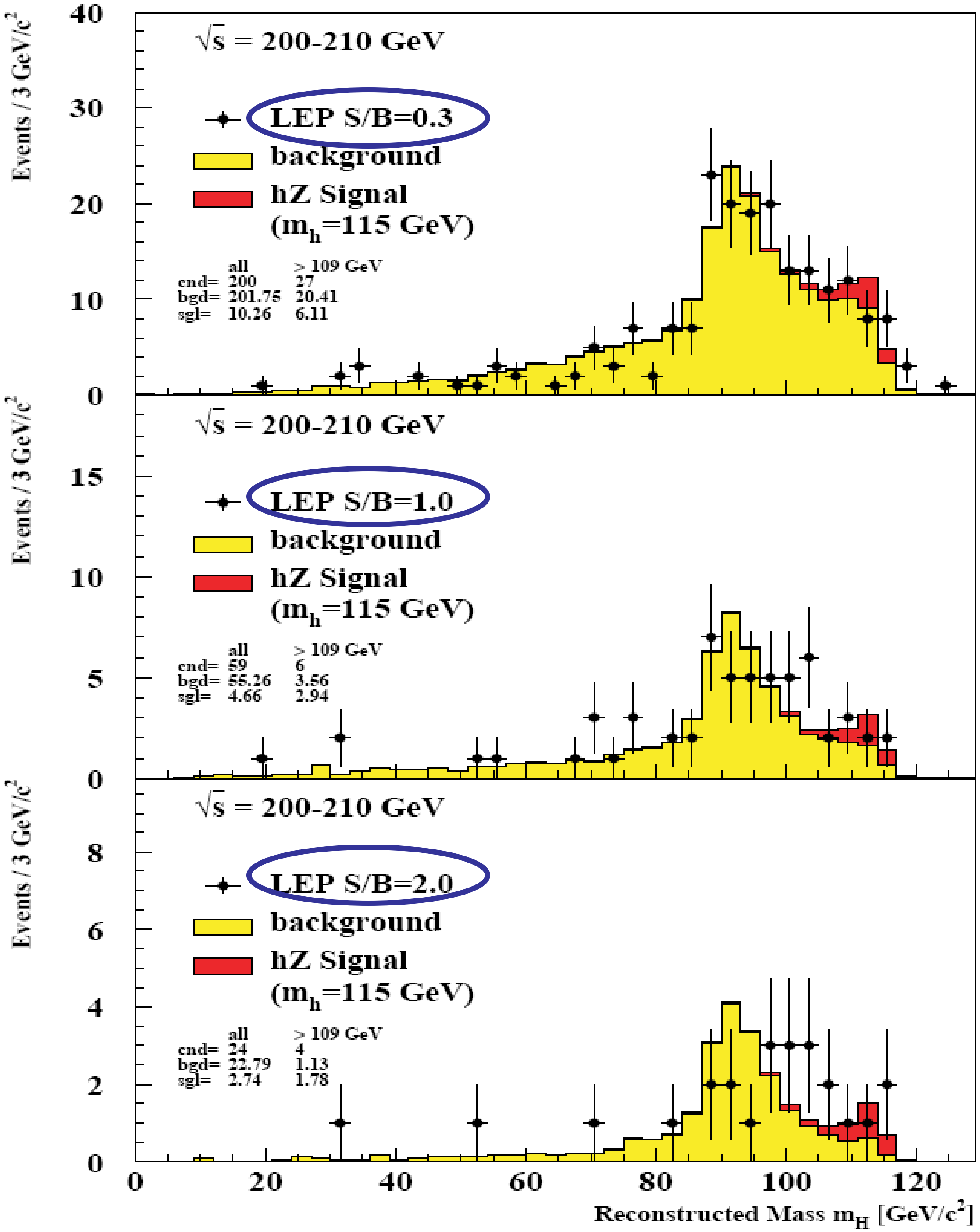


Fig. 18. Likelihood ratio limits (left) and Bayesian limits (right)

	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].

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

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)

Likelihood function

- Given a sample $(x_1, \ldots, x_n)$, the likelihood function expresses the probability density of the sample, as a function of the unknown parameters

$$L = \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_n)$$

- Sometimes the used notation for parameters is the same as for conditional probability:

$$f(x_i | \theta_1, \cdots, \theta_n)$$

- If the size n of the sample is also a random variable, the extended likelihood function is also used:

$$L = p(n; \theta_1, \cdots, \theta_n) \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_n)$$

  Where p is most of the times a Poisson distribution whose average is a function of the unknown parameters

- In many cases it is convenient to use **$-\ln L$** or **$-2\ln L$**
  - $\Pi_i \rightarrow \Sigma_i$

We have been using other estimators:

$N_S/\sqrt{N_B}$,

$N_S/\sqrt{N_B + N_S}$,

$2(\sqrt{N_B + N_S} - \sqrt{N_B})$

Finaly, we calculate the signal statistical significance as:

$S_L = \sqrt{2\left(\ln L_{B+S} - \ln L_B\right)}$

where compute two likelihoods:

$$\ln L_B = \sum_{i=1}^{10} \left(-b_i + n_i \cdot \ln b_i\right)$$

$$\ln L_{B+S} = \sum_{i=1}^{10} \left(-b_i - s_i + n_i \cdot \ln\left(b_i + s_i\right)\right)$$

b_i, s_i, n_i are the number of predicted background and signal events and observed data events in the i-th bin