# 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**

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

- ***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:

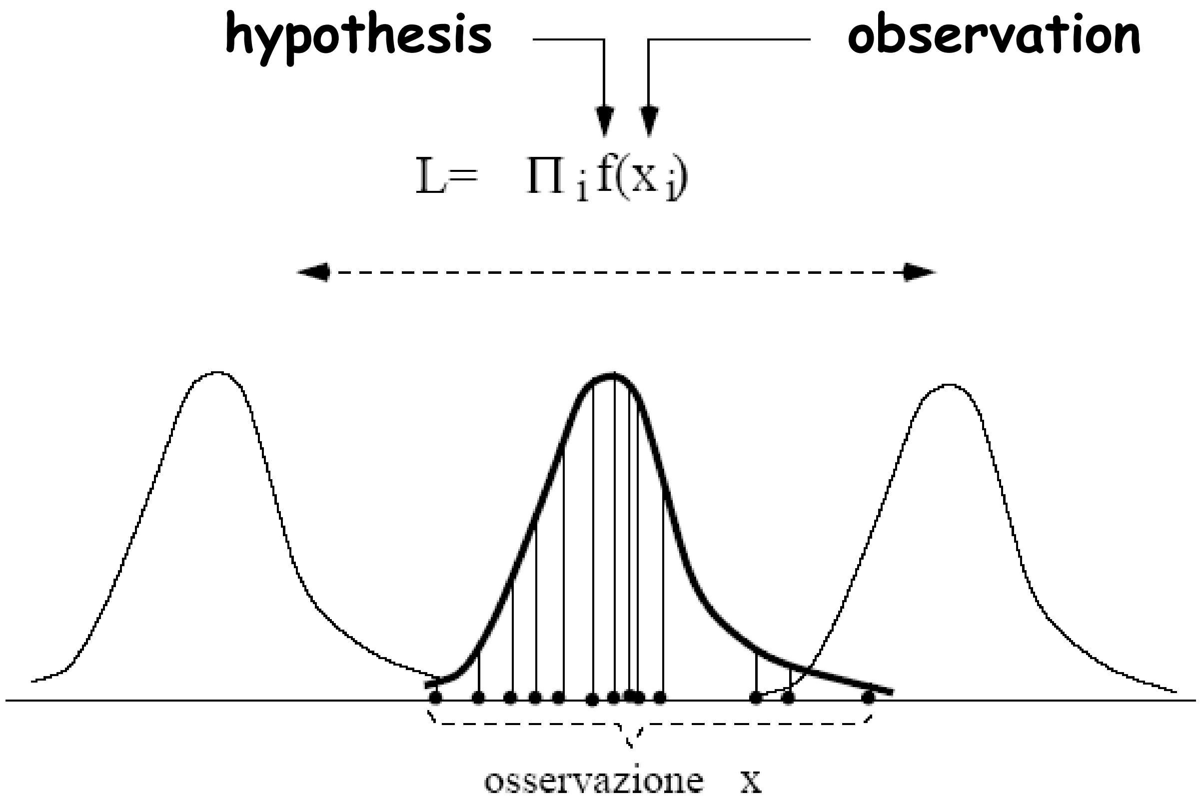
maximize the posterior probability

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

**Bayesians vs Frequentists**

- 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!

# 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**

$\bullet\bullet\circ$ $\quad$ $\circ\circ\bullet$

$p = 1/3$ $\quad$ $p = 2/3$

**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:**

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

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

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

## 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:**

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

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

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

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

# Estimators

- **Estimator of $\theta$**
  **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 $\theta$**

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

- **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 $\theta$

# Likelihood confidence intervals

$$\begin{aligned}\mathcal{L}(\theta) &\simeq \mathcal{L}(\hat\theta) + \mathcal{L}'(\hat\theta)(\theta-\hat\theta) + \frac{1}{2}\mathcal{L}''(\hat\theta)(\theta-\hat\theta)^2 \\ &\simeq \mathcal{L}(\hat\theta) + \frac{n}{2}\frac{\mathcal{L}''(\theta)}{n}(\theta-\hat\theta)^2 \\ &\simeq \mathcal{L}(\hat\theta) + \frac{nI(\theta)}{2}(\theta-\hat\theta)^2\end{aligned}$$

$$2[\ln L(\hat\theta) - \ln L(\theta)] \simeq nI(\theta)\,(\hat\theta-\theta)^2$$

**If one sets:**

$\eta(\theta)$ **for which** $nI_\eta(\eta) = 1$**:**

$$\ln L_\eta(\hat\eta) - \ln L_\eta(\eta) \simeq \frac{1}{2}(\hat\eta-\eta)^2 \ .$$

$$\ln L_\eta(\hat\eta) - \ln L_\eta(\eta) \simeq \frac{1}{2}(\hat\eta-\eta)^2 = \frac{1}{2}\chi^2_\alpha(1) \ , \ .$$

**Since** $\hat\eta \sim N(\eta, 1)$**:**

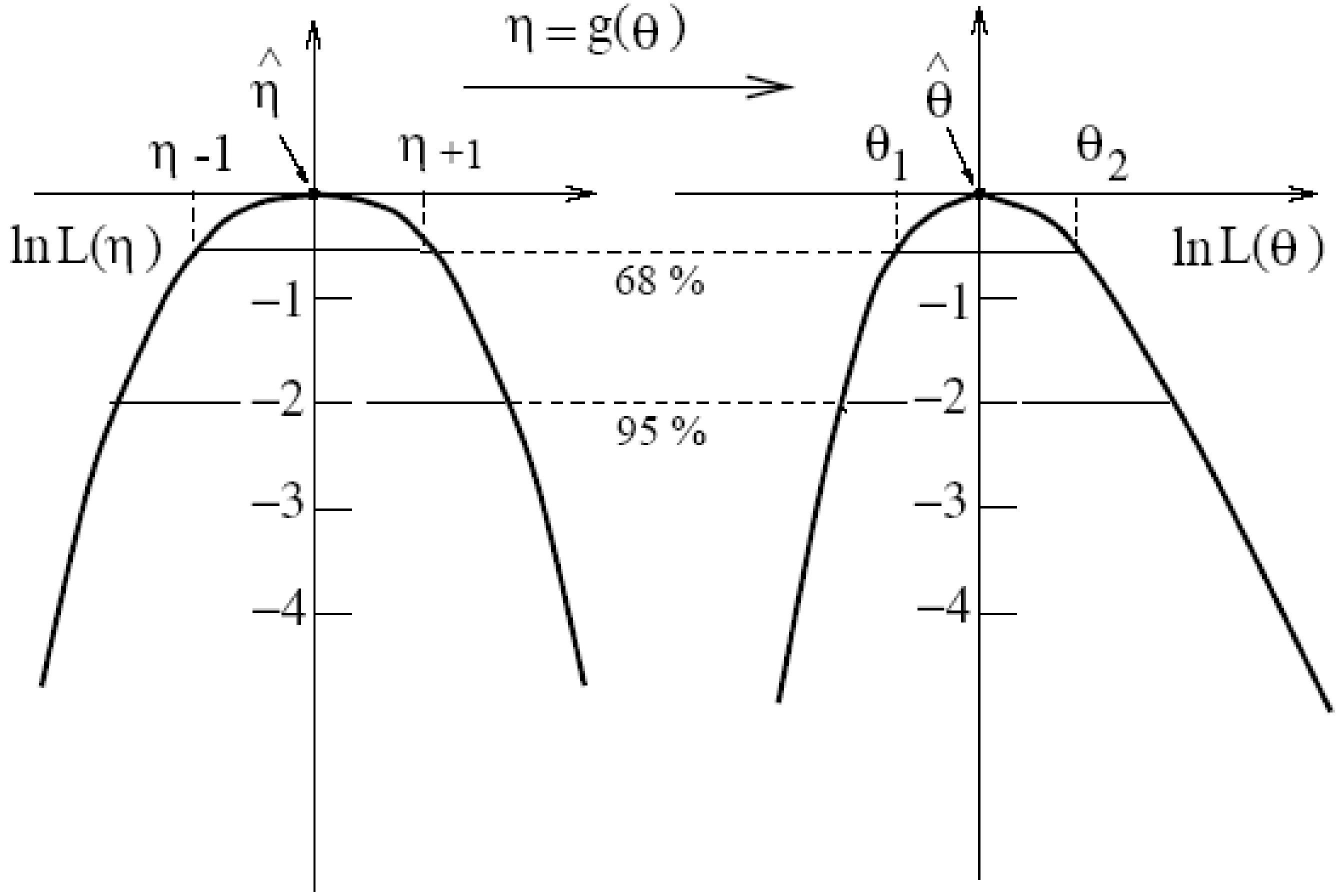
$$\begin{aligned}\ln L_\eta(\hat\eta) - \ln L_\eta(\eta) &= 0.5 \quad CL = 68.3\% \\ \ln L_\eta(\hat\eta) - \ln L_\eta(\eta) &= 2 \quad CL = 95.3\%\end{aligned}$$

**In general:**

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

**Multidimensional case:**

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



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

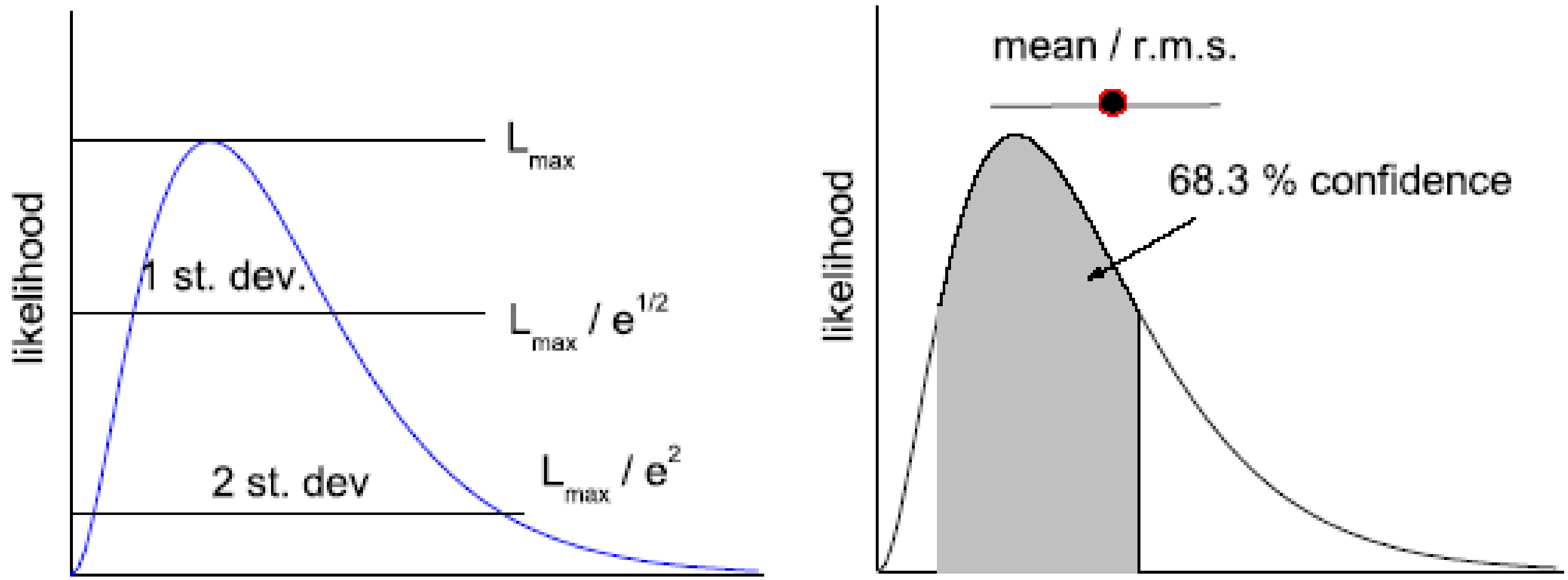



**Fig. 18.** Likelihood ratio limits (left) and Bayesian limits (right)

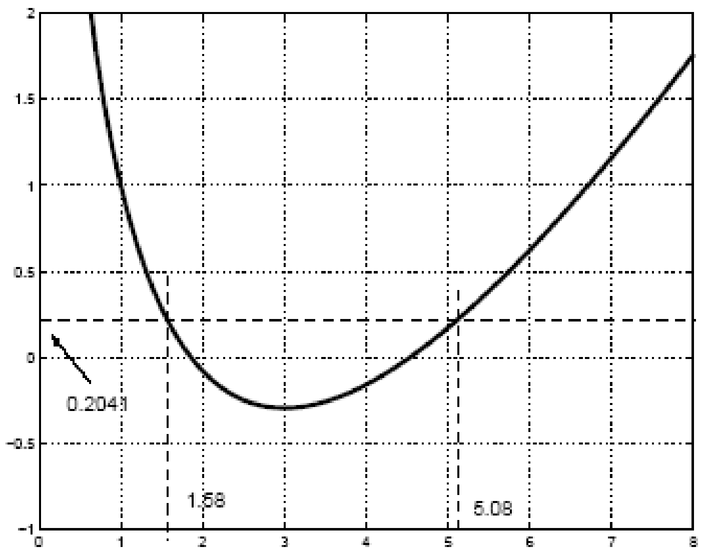
# The 3 event experiment (again)

**Likelihood:**

$$L(\mu; x) = \frac{\mu^x}{x!} e^{-\mu} \ , \qquad \hat{\mu} = x = 3 \ .$$

**$1\sigma$ interval:**

$$2[\ln L(\hat{\mu}; x) - L(\mu; x)] = 1$$
$$3 \ln 3 - 3 - 3 \ln \mu + \mu = 0.5$$
$$\mu - \ln \mu = 0.2041$$




$\mu = [1.58, 5.08]$

**Remember the 68% frequentist interval:**

$$\mu = [1.37, 5.92]$$

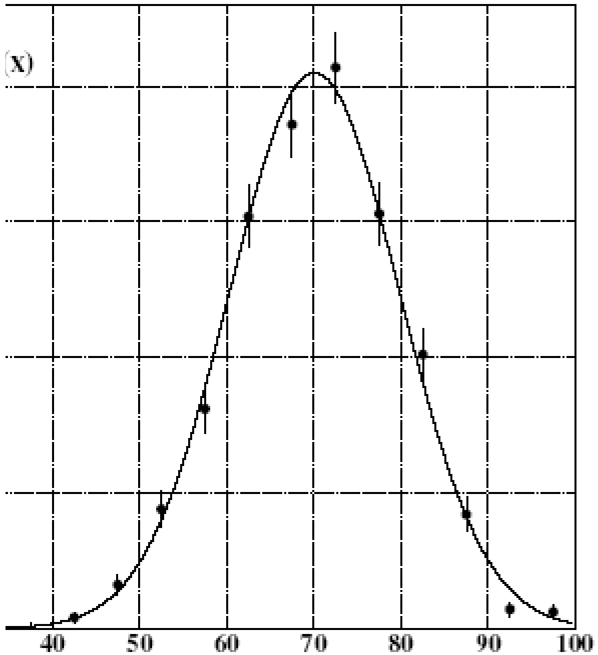
## Fit of Histograms

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-$\chi^2$ 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 $\sum_i p_i(\boldsymbol{\theta}) = 1$ implies $\sum_i \partial p_i(\boldsymbol{\theta})/\partial \theta_j = 0$.

The last member corresponds to the derivative of

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

This formula is from ML !!!!

with a constant denominator

# The extended likelihood

When the total number of events is a poissonian variable:

$$P\{I_i = n_i, \boldsymbol{N} = N\} = P\{I_i = n_i | \boldsymbol{N} = N\} P\{\boldsymbol{N} = N\}$$
$$= p(n_i, N) = \frac{N!}{n_i!(N-n_i)!} p_i^{n_i} (1-p_i)^{N-n_i} \frac{e^{-\lambda}\lambda^N}{N!} \; .$$

If $m_i = N - n_i$,

$$e^{-\lambda} = e^{-\lambda p_i} e^{-\lambda(1-p_i)} \; , \qquad \lambda^N = \lambda^{N-n_i} \lambda^{n_i} = \lambda^{m_i} \lambda^{n_i} \; ,$$

and

$$p(n_i, m_i) = \frac{e^{-\lambda p_i} (\lambda p_i)^{n_i}}{n_i!} \frac{e^{-\lambda(1-p_i)} [\lambda (1-p_i)]^{m_i}}{m_i!} \; ,$$

is the product of two poissonians, with averages $\lambda p_i$ and $\lambda(1-p_i)$

**Conclusions: the number of events in any channel follows the Poisson statistics**

# 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** $\mu_i = N\, p_i(\theta)$

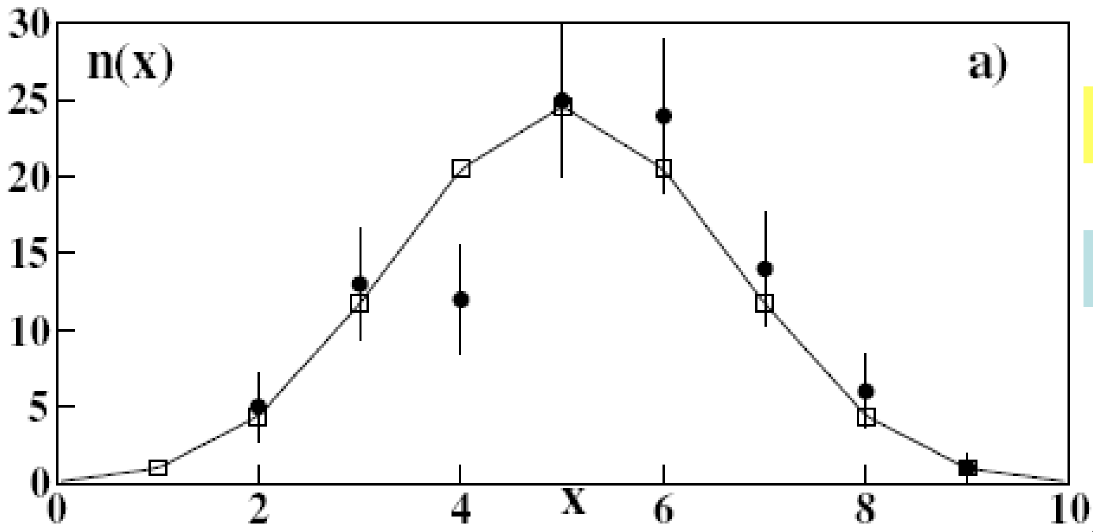
$$-\ln L(\theta, \underline{n}) = -\sum_{i=1}^{k} n_i \ln[p_i(\theta)] + N(\theta)$$

$N$ **is a function of** $\theta$ **as in the case of a detector efficiency,**

**If there is no functional relation between** $N$ **and** $\theta$

**the result is the same as for the non extended likelihood**

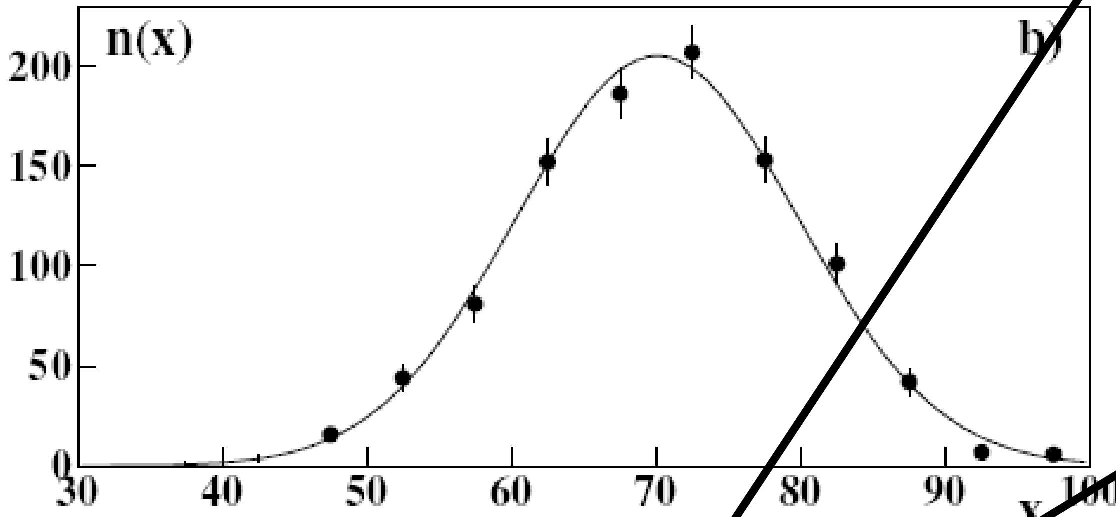
Binomial
p=0.5




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

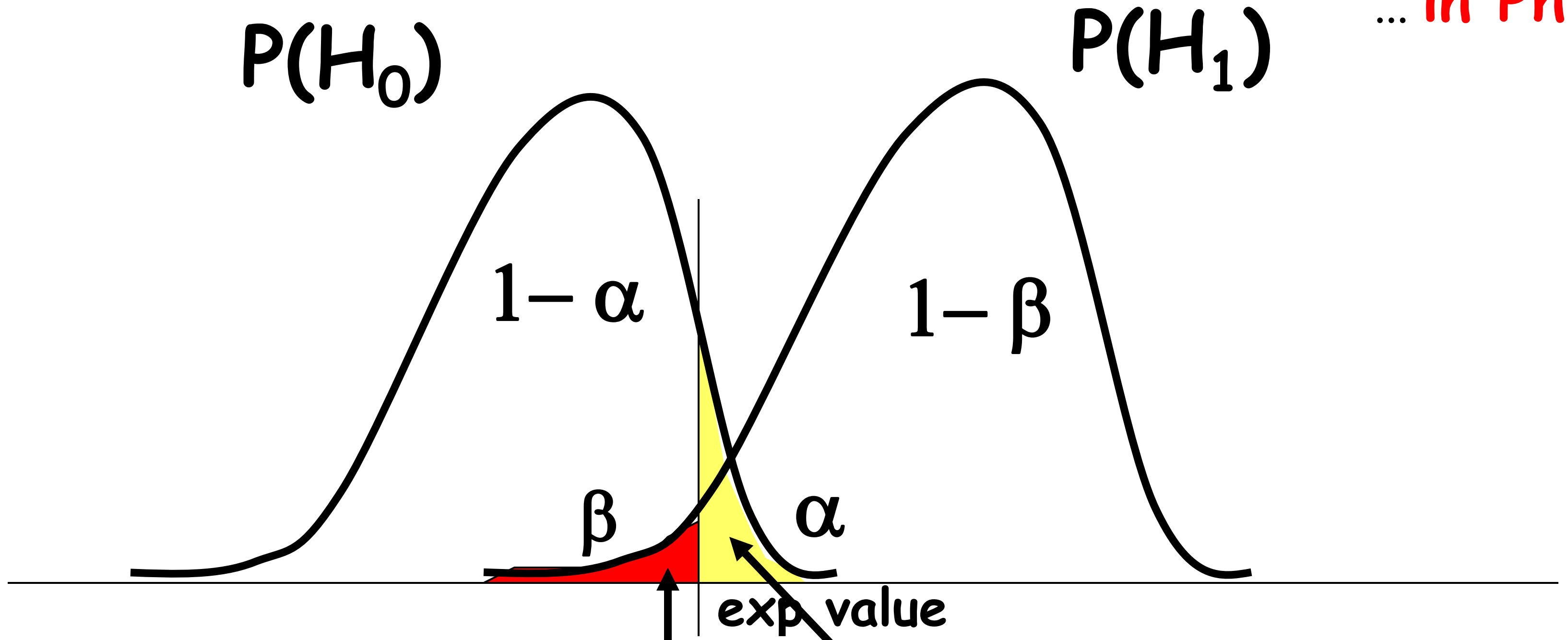
μ =69.97  0.31
σ =9.59  0.22

$\mathcal{L}$

$\chi^2$

## The other branch of Statistics: Hypothesis Testing

… **in Physics**




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

**power** ← $1-\beta$

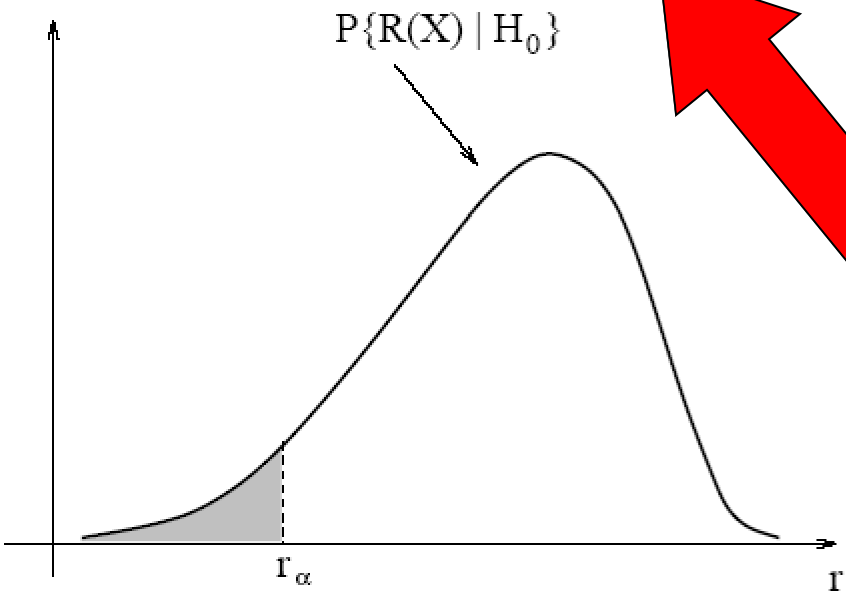
If $H_1$ is the discovery, the maximum power test maximizes the discovery probability, that is the good acceptance

When two simple hypotheses are given

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

the most powerful test, for $\alpha$ given, is

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




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

$$H_0 \;:\; \theta_1 = a \;, \quad \theta_2 = b$$
$$H_1 \;:\; \theta_1 \neq a \;, \quad \theta_2 \neq b$$

or

$$H_0 \;:\; \theta = a \;,$$
$$H_1 \;:\; \theta \geq a$$

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 average number of background events $N_B$ has been measured

A model predicts a number of signals $N_S$

The experiment measures the $x$ triggers:

$$\begin{array}{lll} H_0 & \text{no signal} & \mu = N_B \\ H_1 & \text{there is signal} & \mu = N_B + N_S \end{array}$$

**NP ratio:**

$$R = \frac{N_B^x \exp(-N_B)/x!}{(N_S + N_B)^x \exp[-(N_S + N_B)]/x!} = \left(\frac{N_B}{N_S + N_B}\right)^x \mathrm{e}^{-N_S} = ab^x = \psi(x)$$

**NP theorem application**

$$x = \ln R / \ln(ab)$$

**Conclusion:** the test statistics

$$S = \frac{X - N_B}{\sqrt{N_B}} \quad \text{or} \quad S = \frac{X - N_B}{\sqrt{X}}$$

has maximum power, because $R$ can be expressed as a function of $x$.

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 $2.9\sigma$ 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 $2\sigma$ [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.

N·H
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]

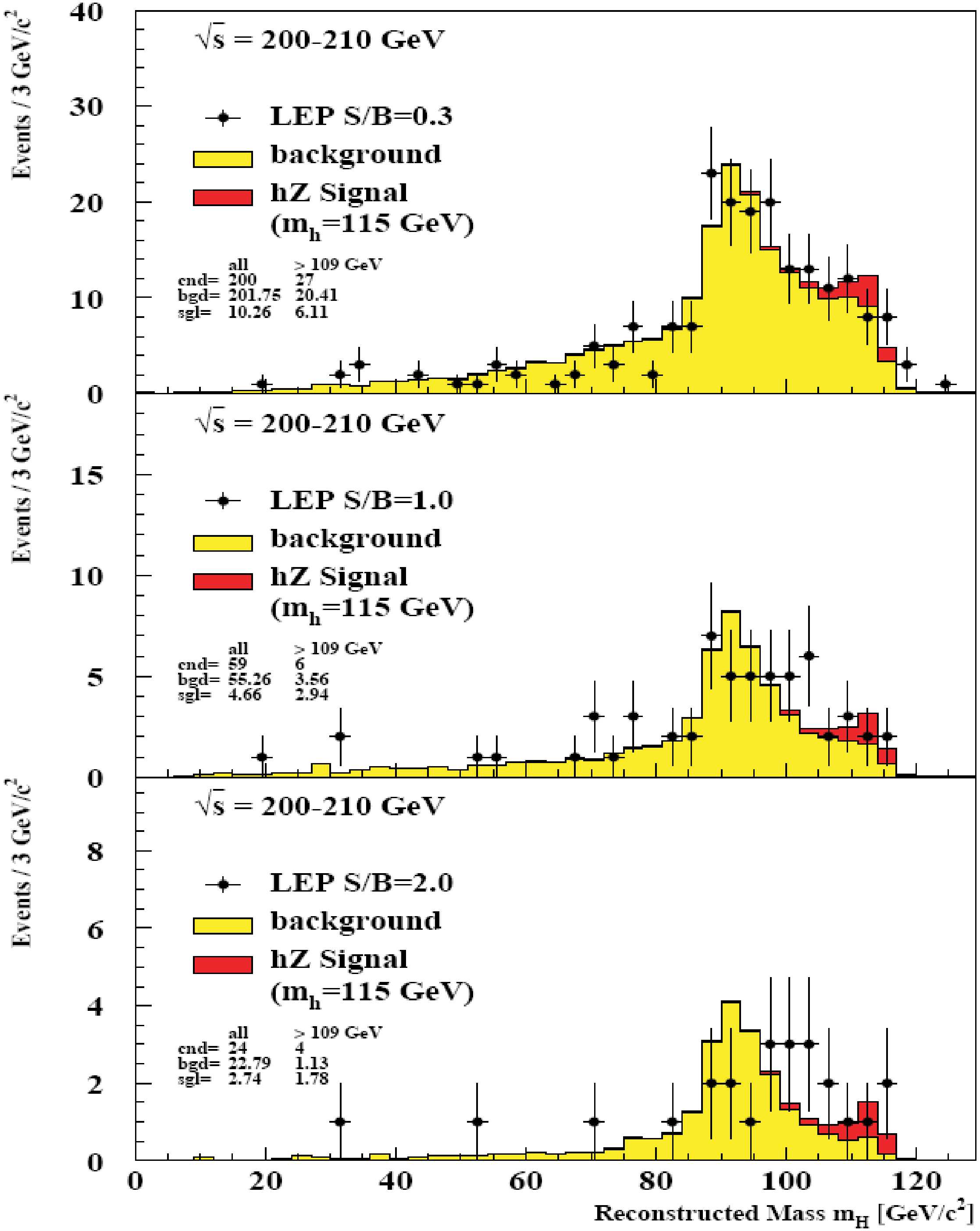
Received 7 March 2003; received in revised form 25 April 2003; accepted 28 April 2003

Editor: L. Rolandi

**Abstract**

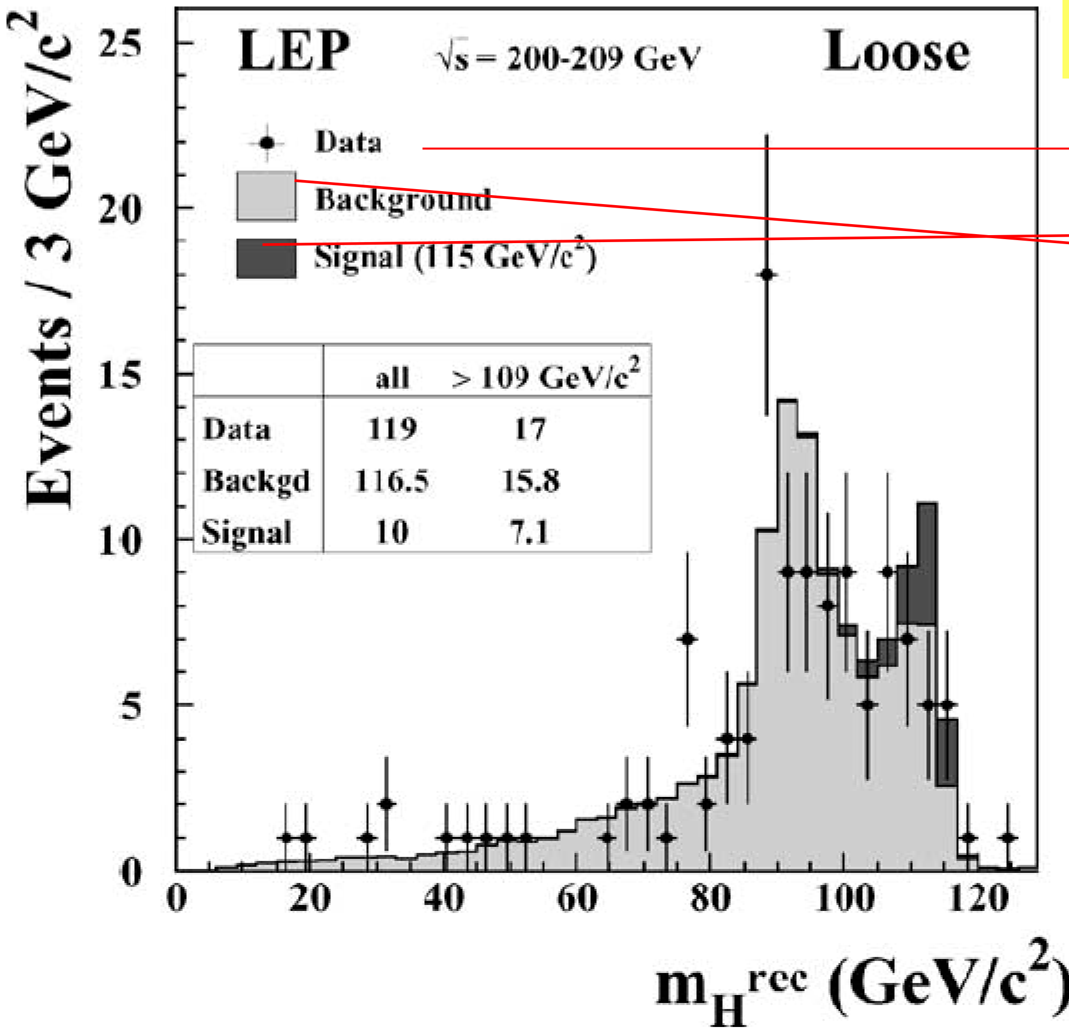
The four LEP Collaborations, ALEPH, DELPHI, L3 and OPAL, have collected a total of 2461 $pb^{-1}$ of $e^+e^-$ collision data at centre-of-mass energies between 189 and 209 GeV. The data are used to search for the Standard Model Higgs boson. The search results of the four Collaborations are combined and 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.

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

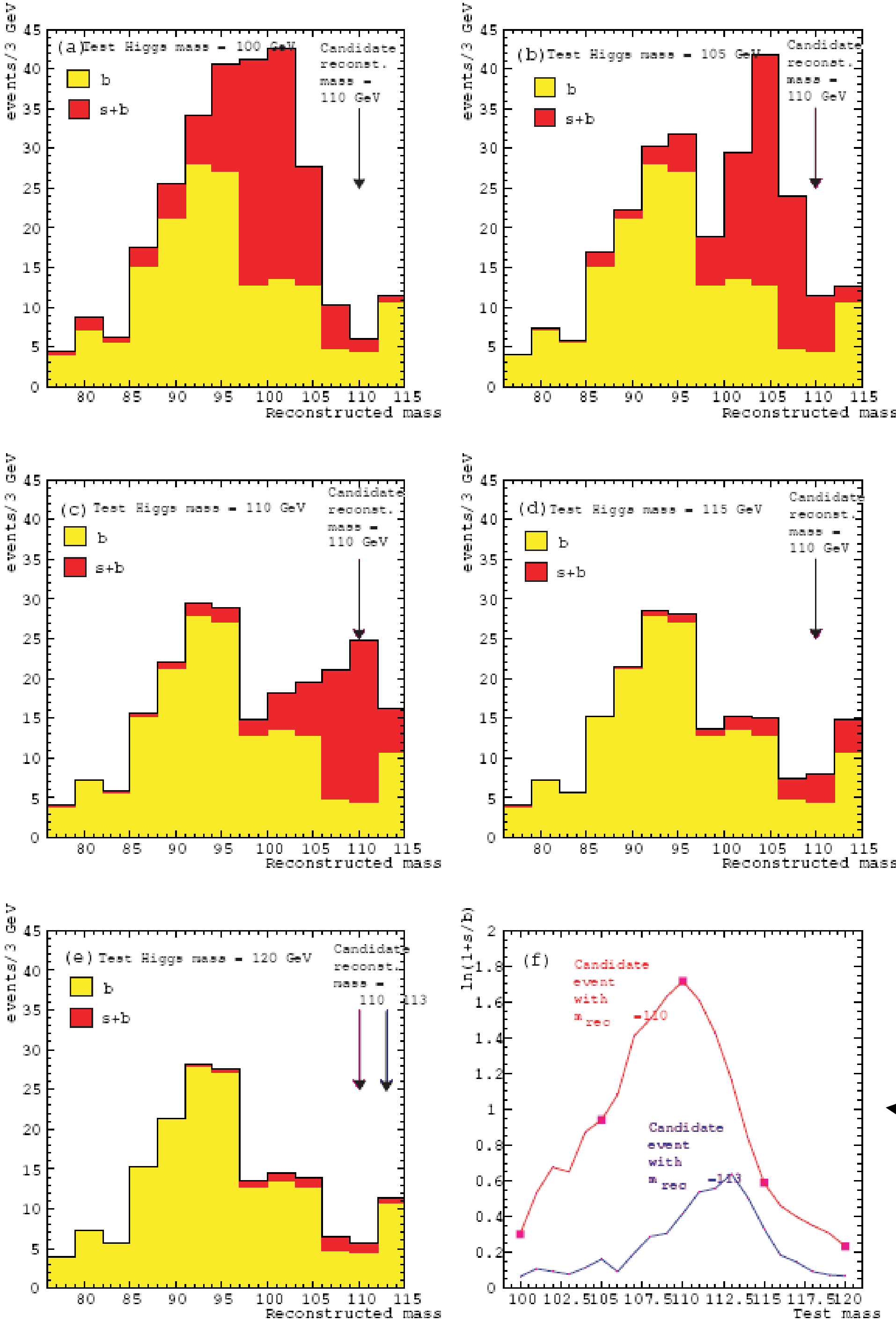
# ALEP, DELPHI, L3, OPAL, 2003




| | all | > 109 GeV/c² |
|---|---|---|
| Data | 119 | 17 |
| Backgd | 116.5 | 15.8 |
| Signal | 10 | 7.1 |

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

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 110-120 GeV

The weight plot was called **spaghetti plot**

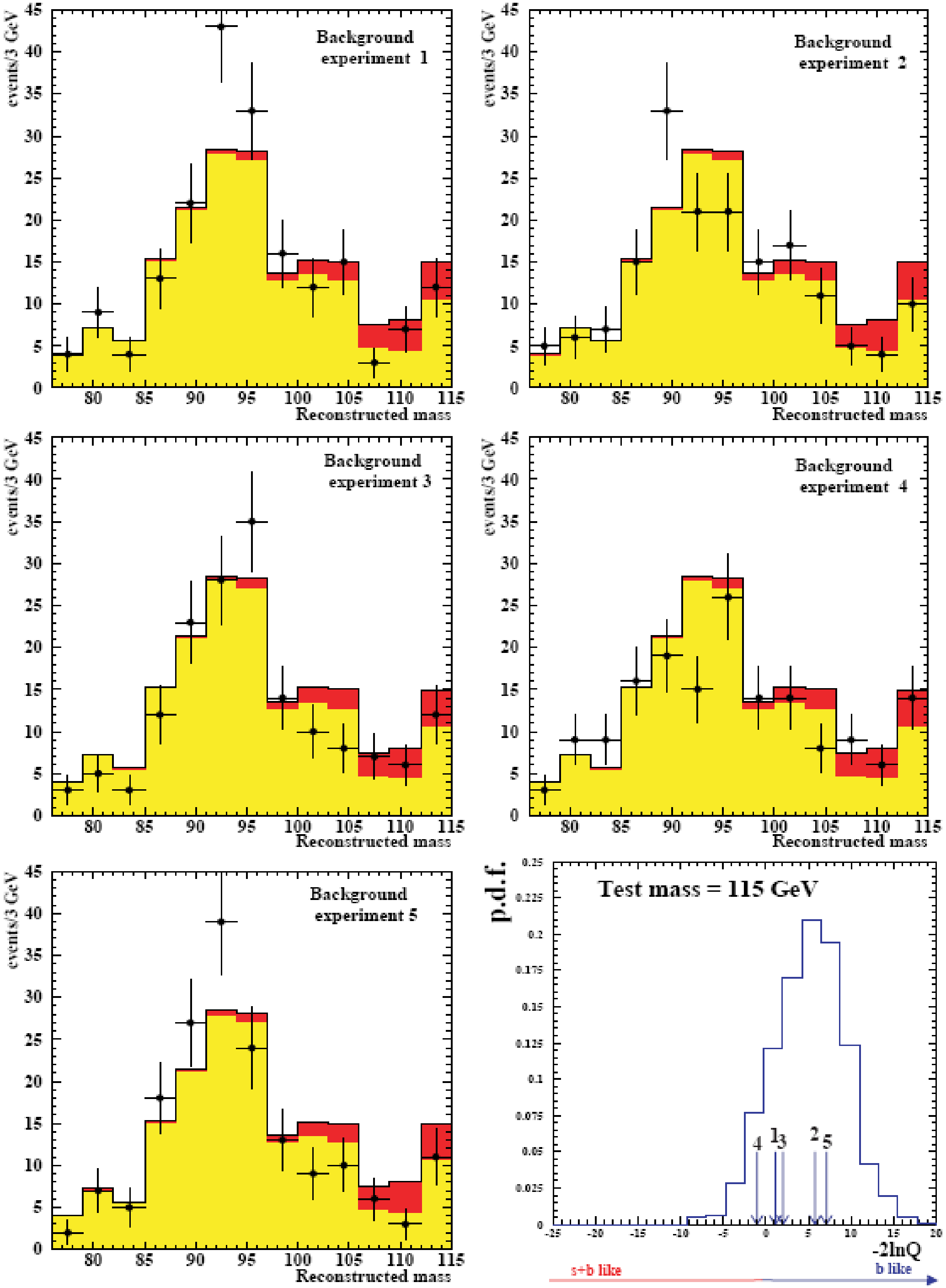
| | Expt | $E_{cm}$ | Decay channel | $m_{rec}$ (GeV) | $\ln(1+s/b)$ at 115 GeV |
|---|---|---|---|---|---|
| 1 | ALEPH | 206.6 | 4-jet | 114.1 | 1.76 |
| 2 | ALEPH | 206.6 | 4-jet | 114.4 | 1.44 |
| 3 | ALEPH | 206.4 | 4-jet | 109.9 | 0.59 |
| 4 | L3 | 206.4 | E-miss | 115.0 | 0.53 |
| 5 | ALEPH | 205.1 | 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

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

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



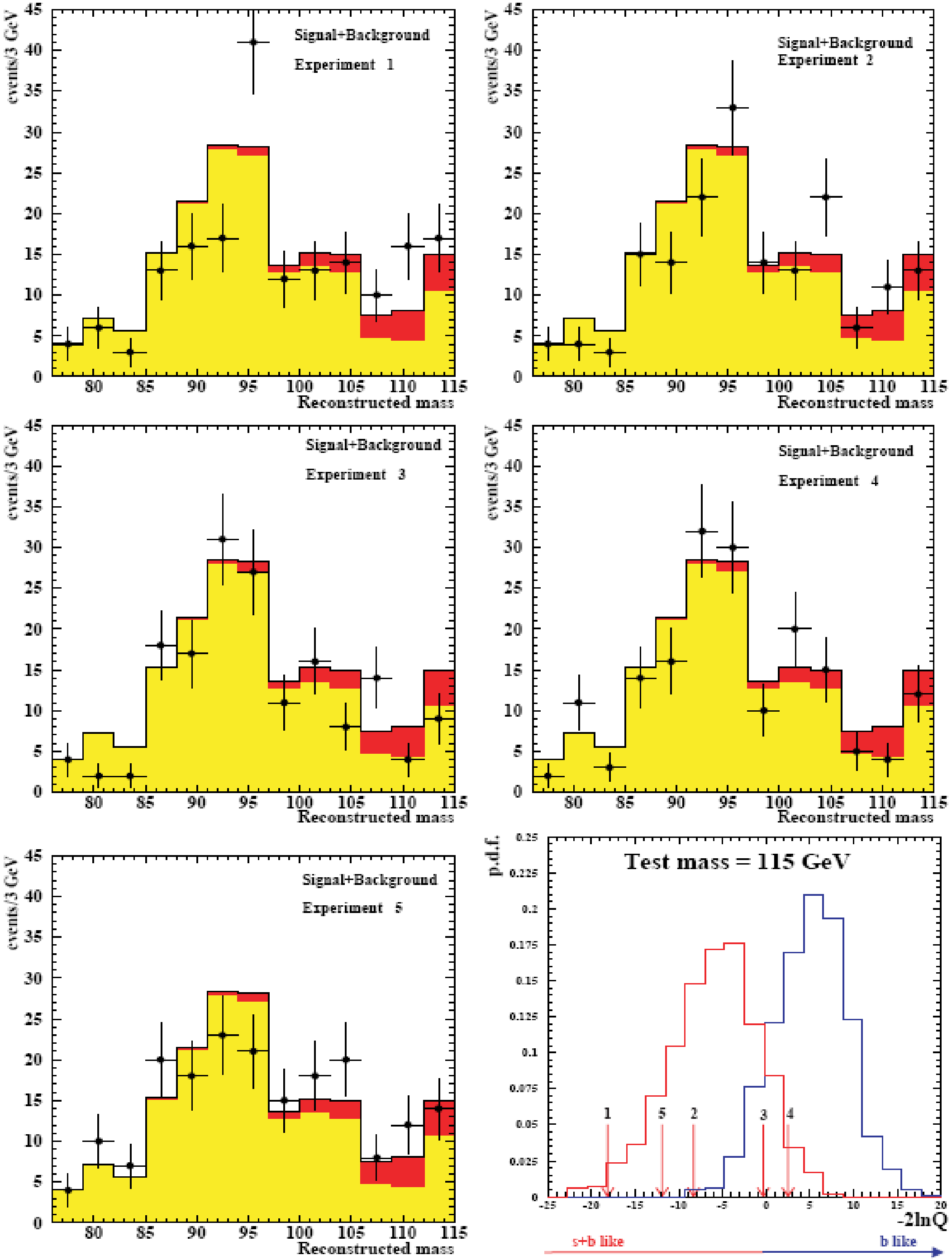

# MC toy model

$s_i$ red
$b_i$ yellow

Crosses: M data, Background only

ln(1+s/b) plot

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




# MC toy model

$s_i$ red
$b_i$ yellow

Crosses: M data,
Background + Signal

ln(1+s/b) plot

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

(in red is the previous one with background only)

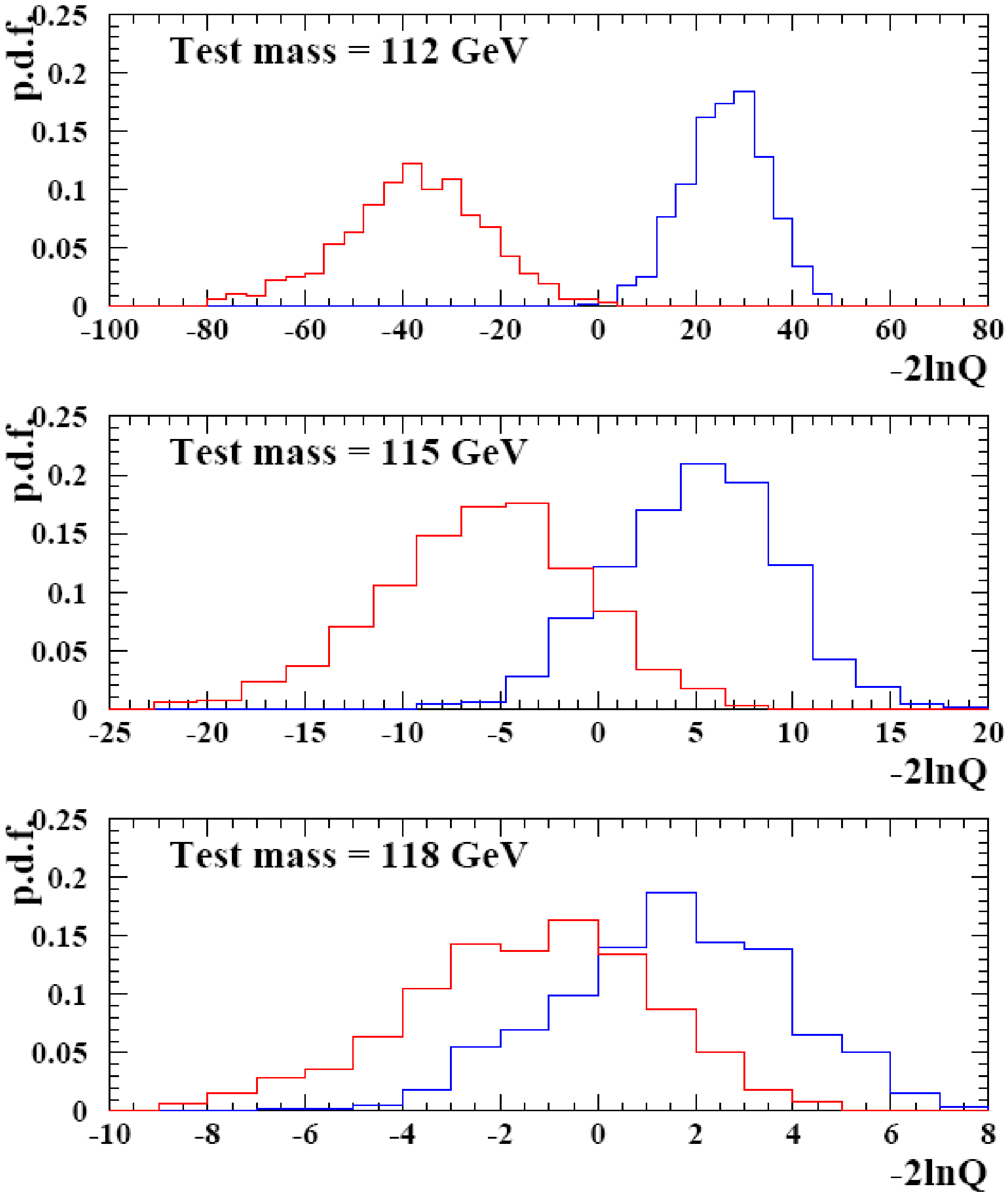
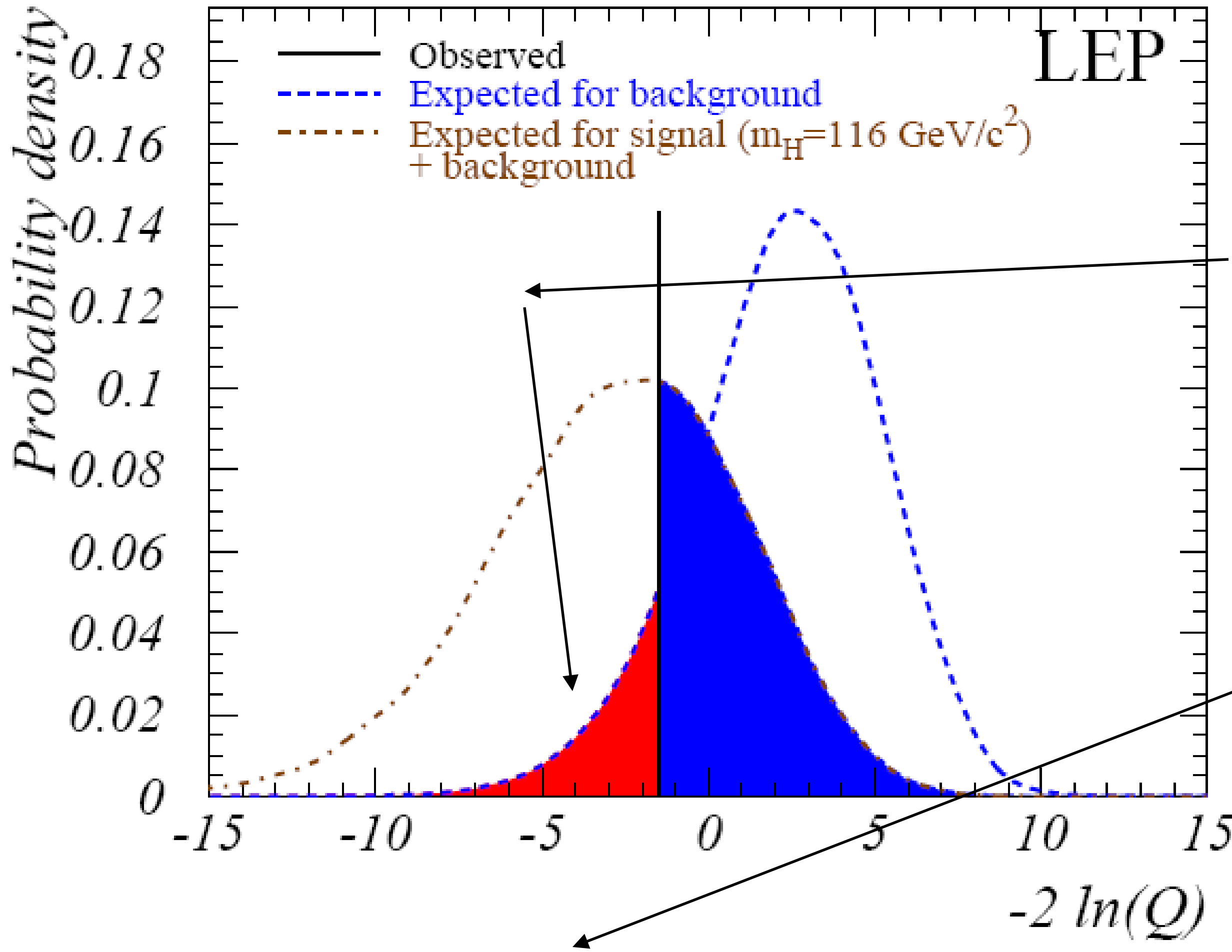
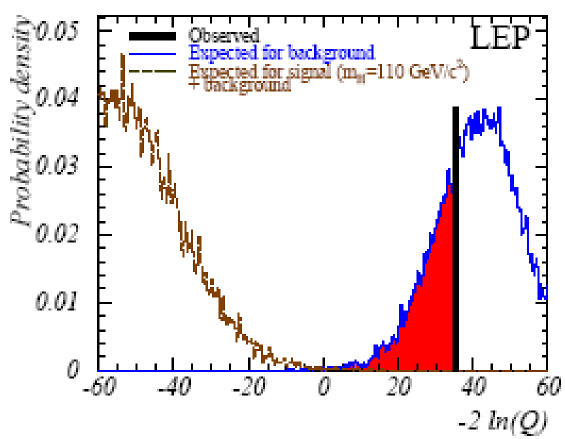


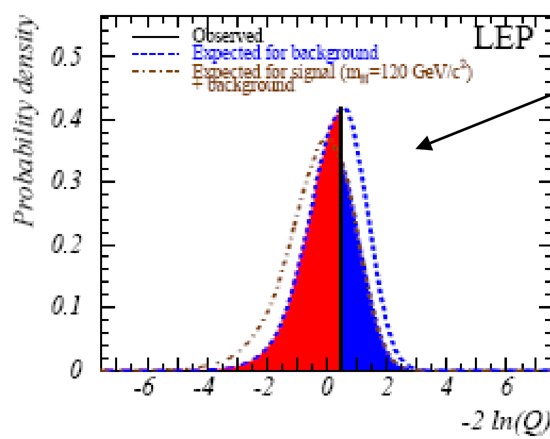

Figure 6: The separation between the Signal and the Background for various Higgs masses is shown by their likelihood p.d.f's.




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

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




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

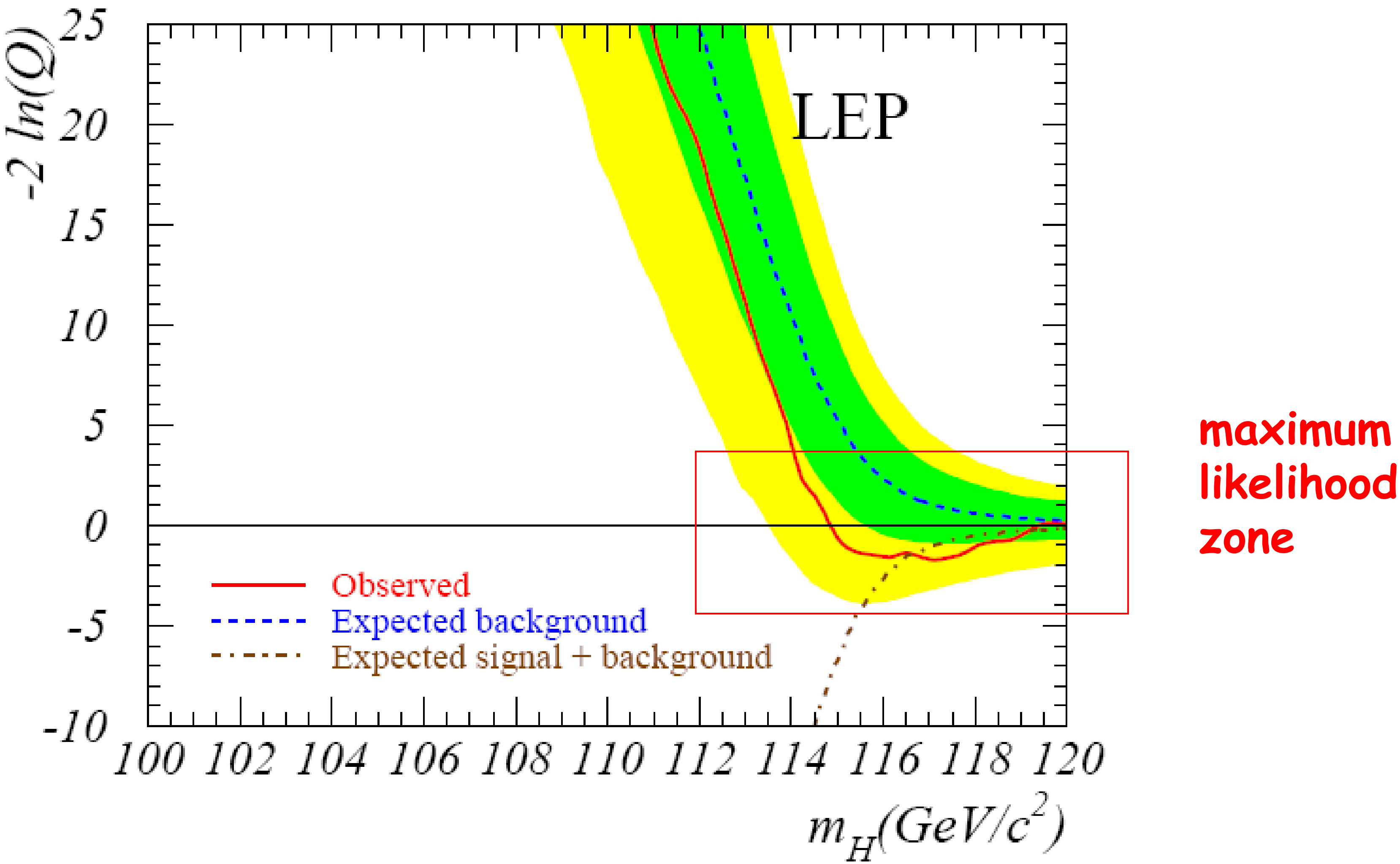



Figure 8: Observed and expected behavior of the likelihood $-2 \ln Q$ as a function of the test-mass $m_H$ for combined LEP experiments. The solid/red line represents the observation; the dashed/dash-dotted lines show the median background/signal+background expectations. The dark/green and light/yellow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

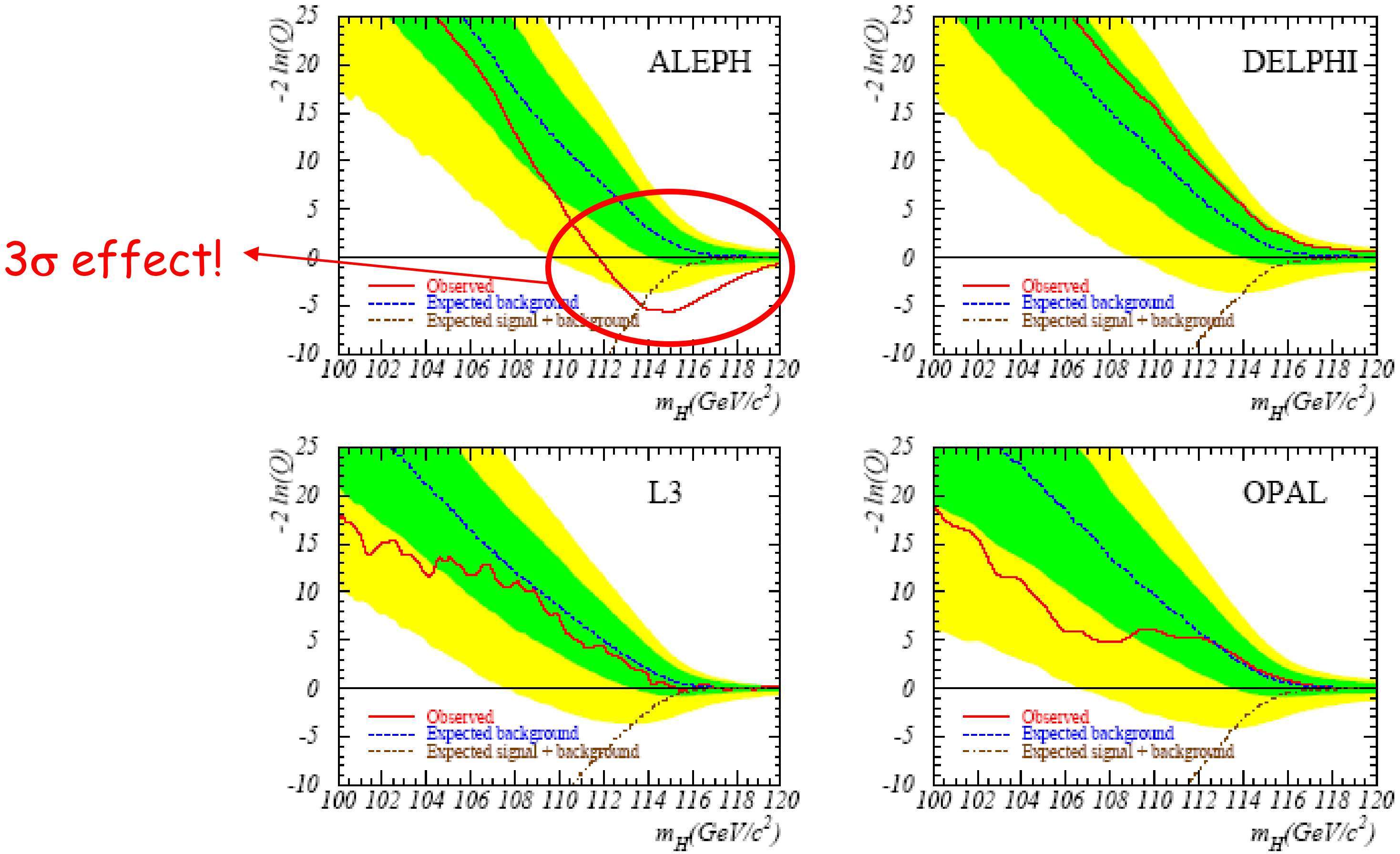
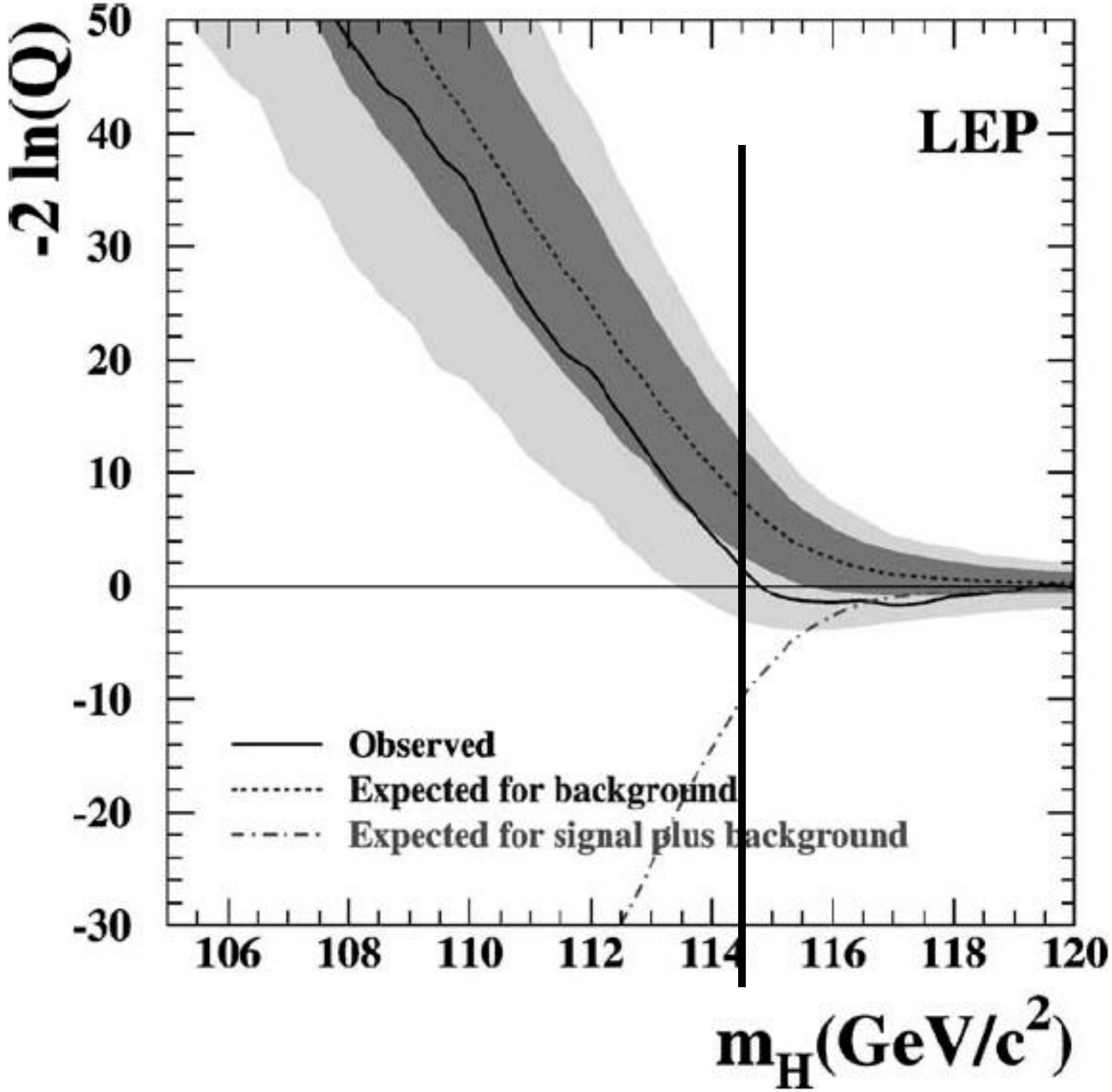
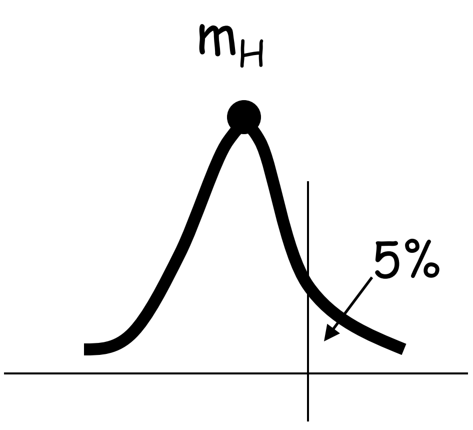



Figure 9: Observed and expected behavior of the likelihood $-2\ln Q$ as a function of the test-mass $m_H$ for the various experiments. The solid/red line represents the observation; the dashed/dash-dotted lines show the median background/signal+background expectations. The dark/green and light/yellow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




ALEPH
DELPHI
L3
OPAL
2003

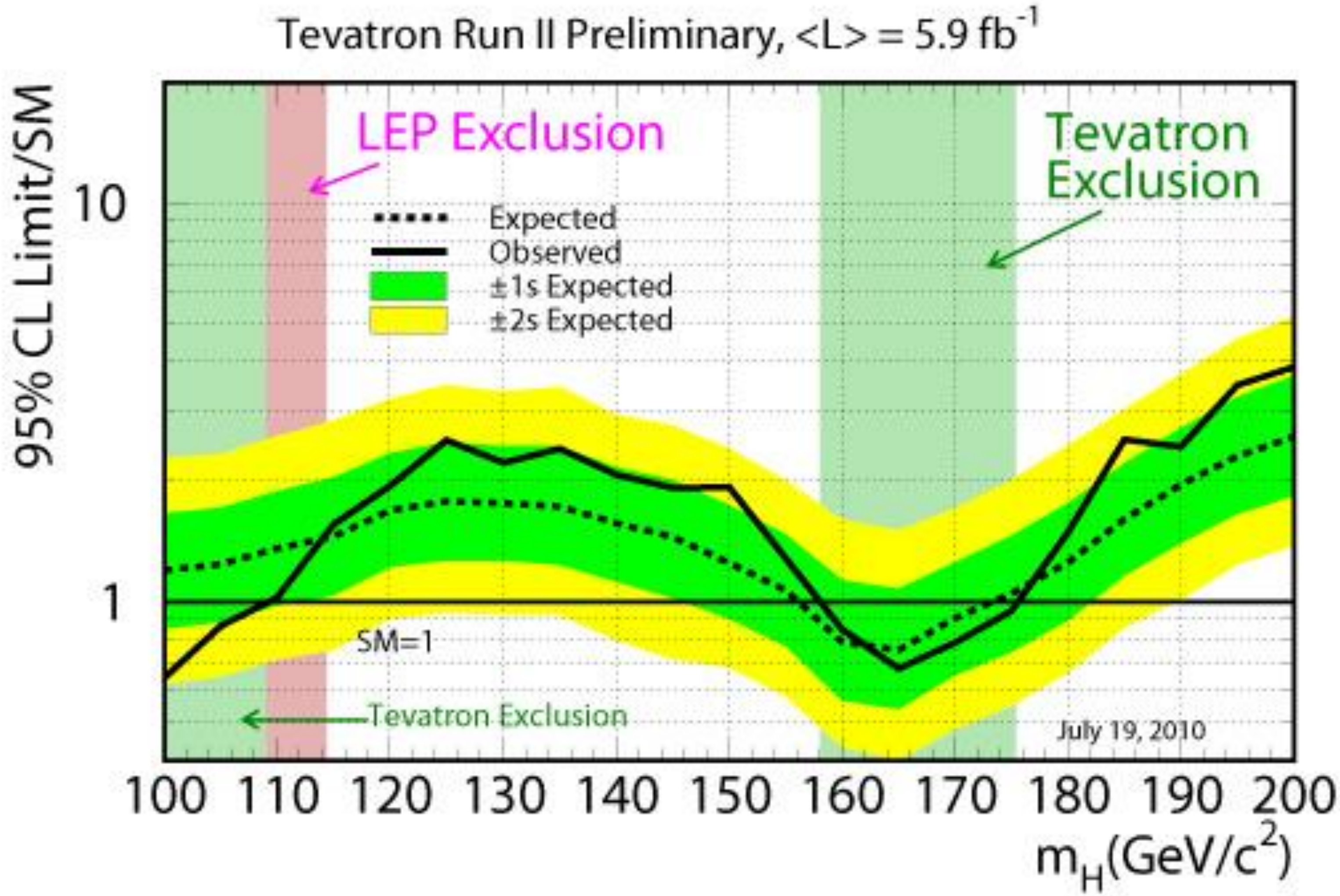



$m_H \geq 114.4\ GeV/c^2$ CL=95%

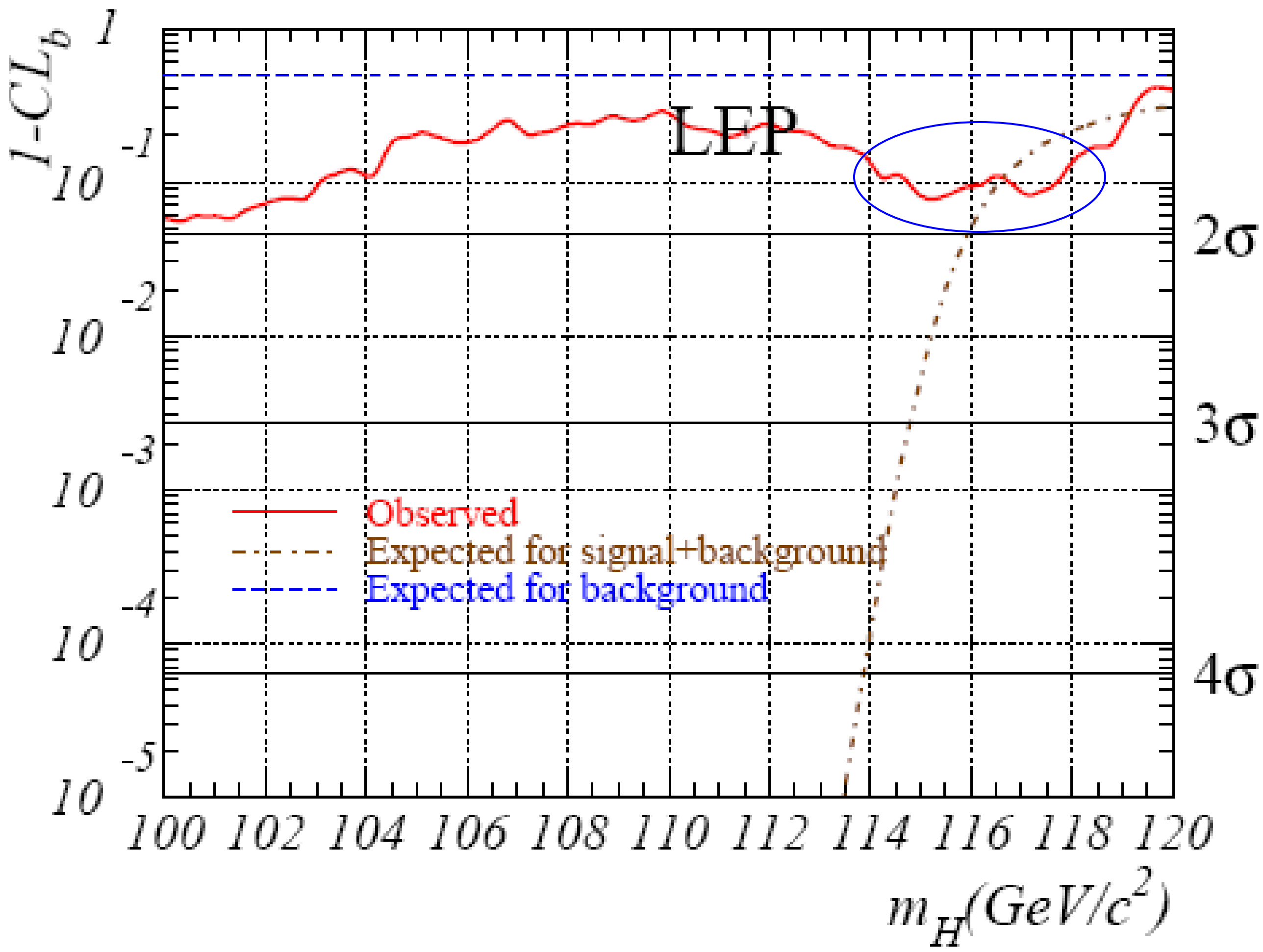
# Conclusions

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 $2\sigma$ level. When the LEP Higgs working group presented these results for the first time the significance was $2.9\sigma$ [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 $3\sigma$ 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.



Tevatron Run II Preliminary, <L> = 5.9 fb-1
LEP Exclusion
Tevatron Exclusion
Expected
Observed
±1s Expected
±2s Expected
SM=1
Tevatron Exclusion
July 19, 2010
95% CL Limit/SM
10
1
100 110 120 130 140 150 160 170 180 190 200
mH(GeV/c2)



1-$CL_b$
1
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
$10^{-5}$
LEP
2σ
3σ
4σ
Observed
Expected for signal+background
Expected for background
100 102 104 106 108 110 112 114 116 118 120
$m_H(GeV/c^2)$

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 |

# Bayesian Hypothesis test

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
- $\alpha$, $\beta$ , $1-\beta$ 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

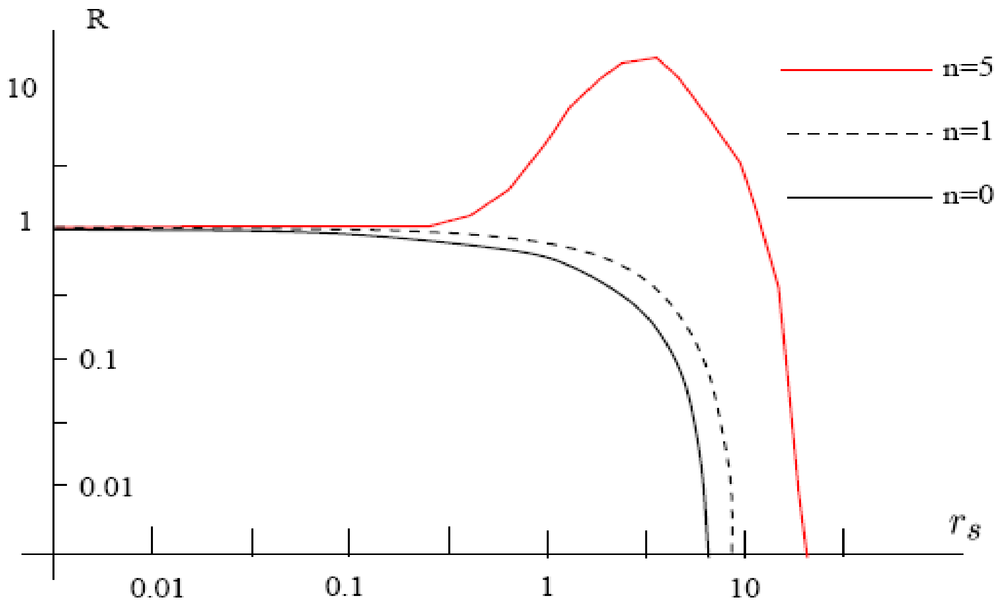



Figure 1: ratio $R$ for the poisson intensity parameter $r$ in units of events per month for an expected background rate $r_b = 1$ event/month and for $n = 0, 1, 5$ observed events

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

**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** $r_s = 4$

# Conclusions

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

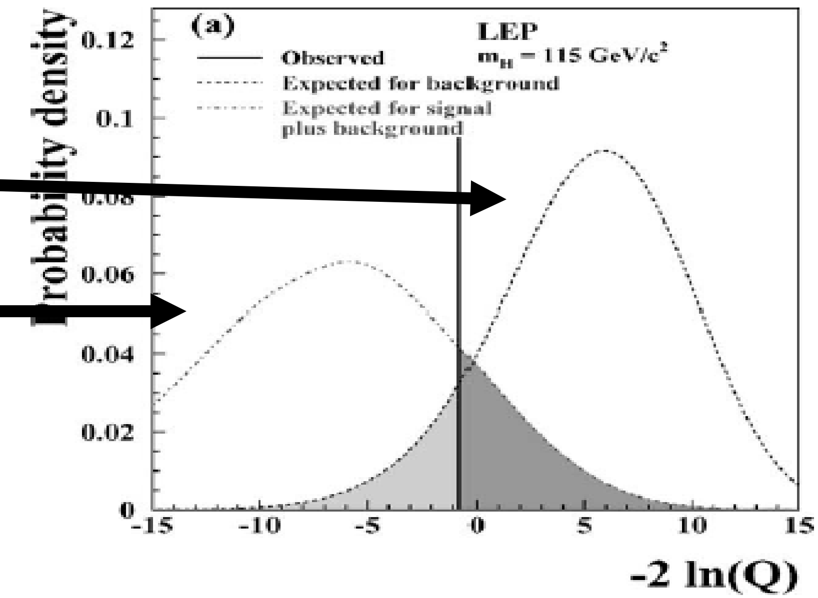
- The likelihood ratio is the maximum power test, that maximize the discovery potential

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

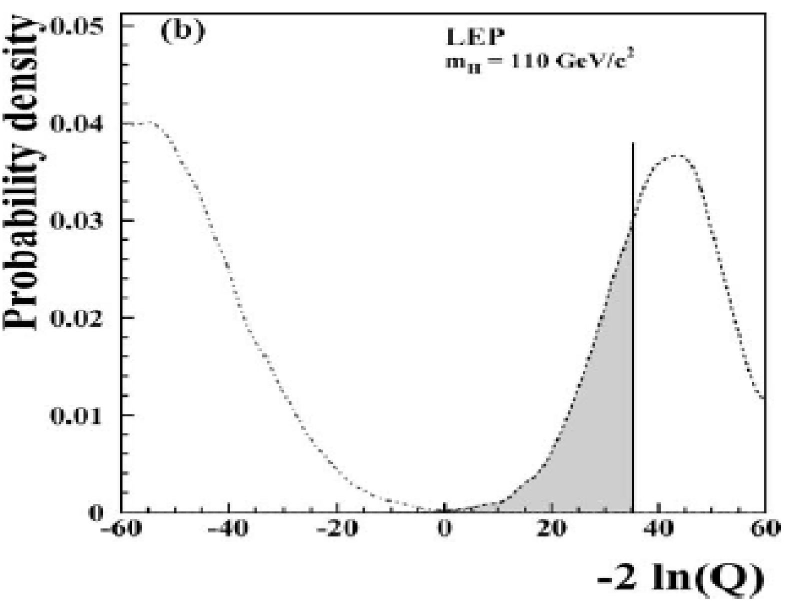
**MC samples**

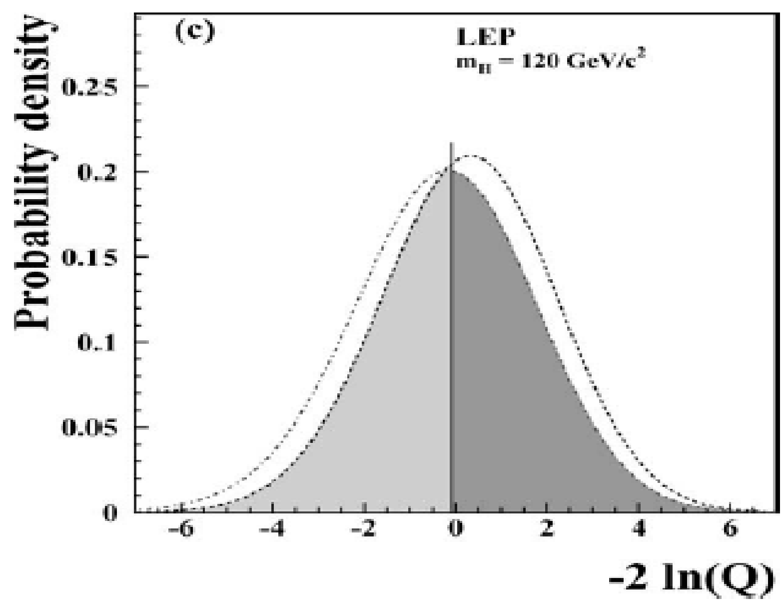
background

signal



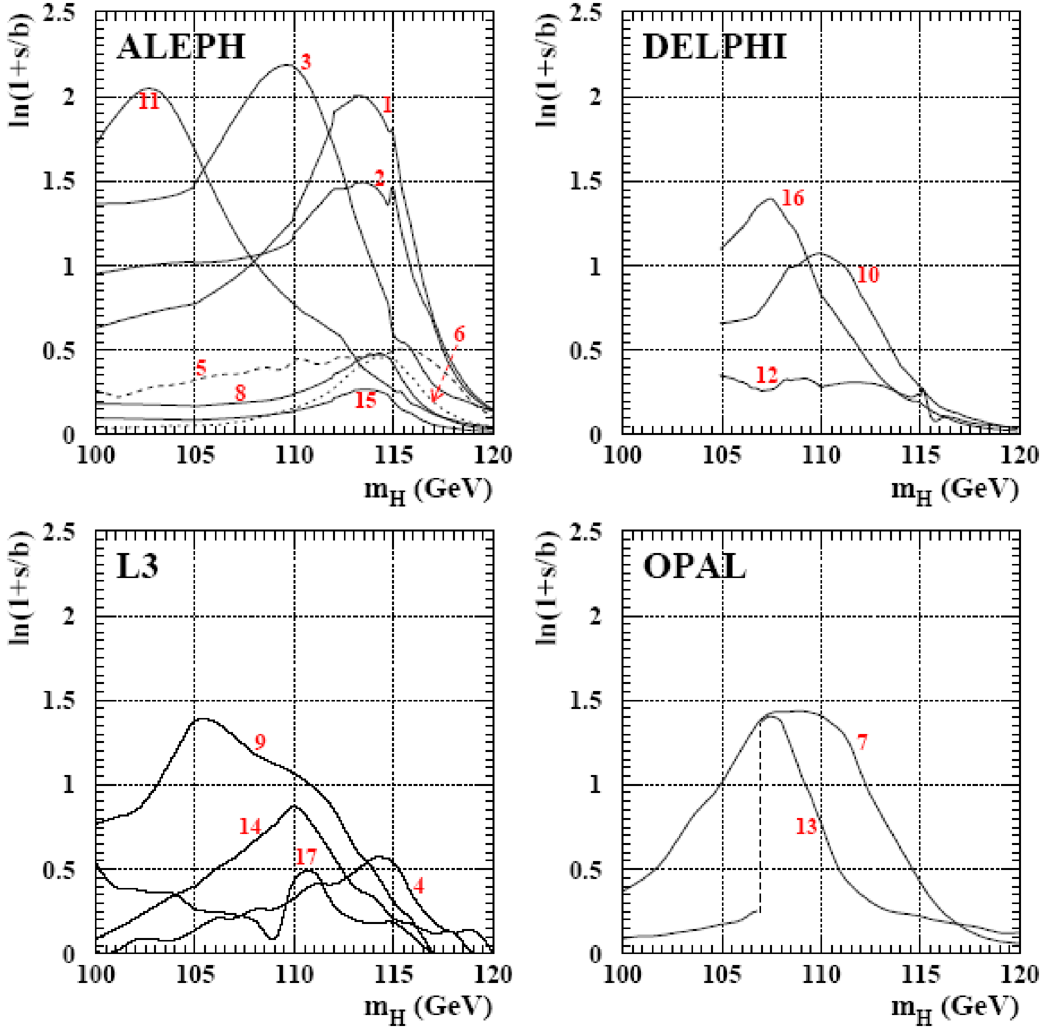

With a mass of 116 GeV 10% of the background only experiments give the observed signal




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




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].