Confidence Limits and Intervals 3: Various other topics

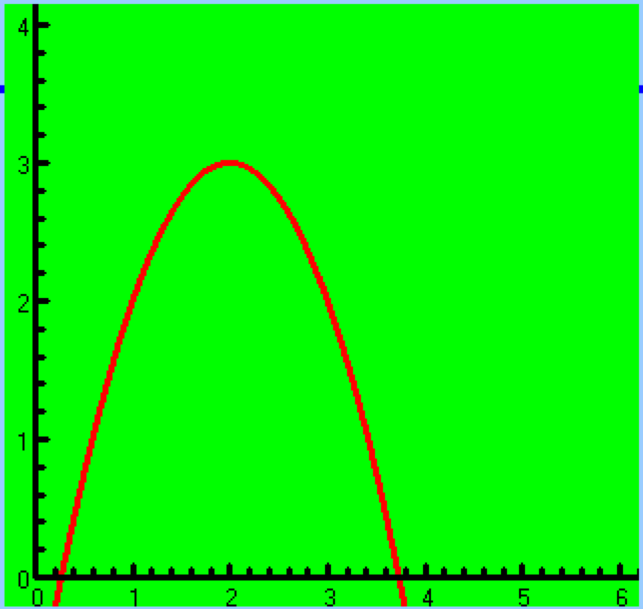
Roger Barlow

SLUO Lectures on Statistics

August 2006

Contents

1 The Likelihood Ratio



Estimate a model parameter M by maximising the likelihood

In the large N limit

i) This is unbiassed

ii) The error is given by

$$\frac{1}{\sigma^2} = -\left\langle \frac{d^2 \ln L}{dM^2} \right\rangle$$

iii) ln L is a parabola

$$L = L_{max} - \frac{1}{2} C \left(M - \hat{M} \right)^2$$

iv) We can approximate

$$C \equiv \frac{-d^2 \ln L}{dM^2} \Big|_{M=\hat{M}} = -\left\langle \frac{d^2 \ln L}{dM^2} \right\rangle$$

v) Read off σ from ΔlnL=-½

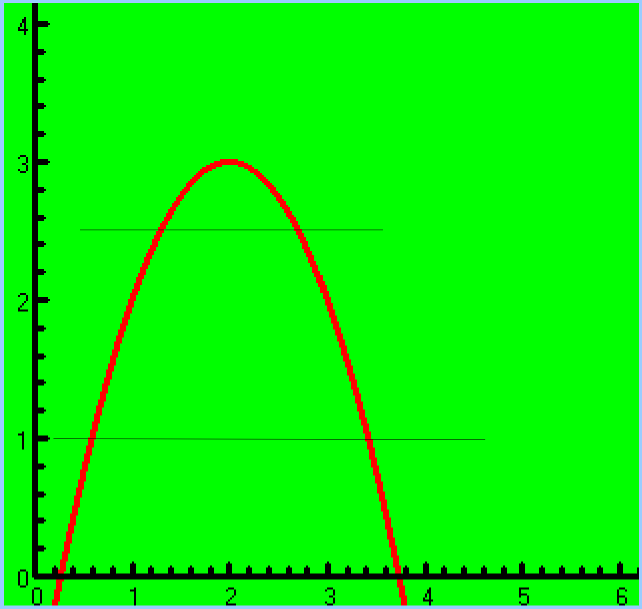
Neat way to find Confidence Intervals

Take $\Delta \ln L = -\frac{1}{2}$ for 68% CL (1σ)

$\Delta \ln L = -2$ for 95.4% CL (2σ)

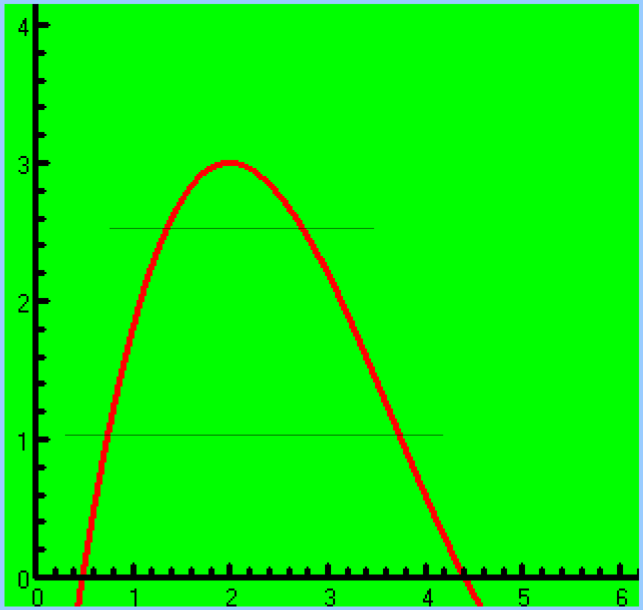
Or whatever you choose

2-sided or 1-sided



For finite N

- None of the above are true
- Never mind! We could transform from M $\rightarrow$M' where it was parabolic, find the limits, and transform back
- Would give $\Delta \ln L=-\frac{1}{2}$ for 68% CL etc as before
- Hence asymmetric errors



Everybody does this

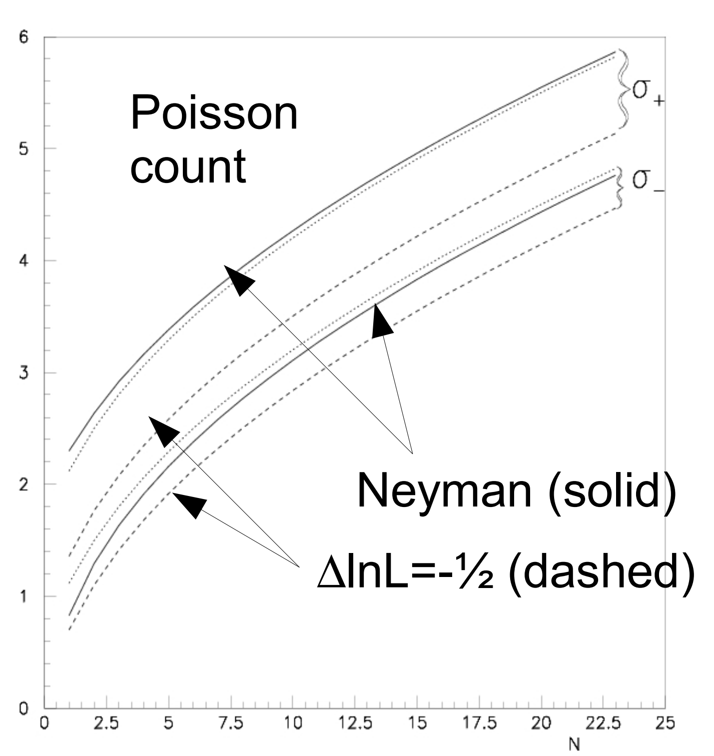
Is it valid?

- Try and see with toy model (lifetime measurement) where we can do the Neyman construction
- For various numbers of measurements, N, normalised to unit lifetime
- There are some quite severe differences!

N	Exact		$\Delta \ln L = -\frac{1}{2}$	
	σ_-	σ_+	σ_-	σ_+
1	0.457	4.787	0.576	2.314
2	0.394	1.824	0.469	1.228
3	0.353	1.194	0.410	0.894
4	0.324	0.918	0.370	0.725
5	0.302	0.760	0.340	0.621
6	0.284	0.657	0.318	0.550
7	0.270	0.584	0.299	0.497
8	0.257	0.529	0.284	0.456
9	0.247	0.486	0.271	0.423
10	0.237	0.451	0.260	0.396
15	0.203	0.343	0.219	0.310
20	0.182	0.285	0.194	0.261
25	0.166	0.248	0.176	0.230

Conclusions on ΔlnL=-½

- Is it valid? No
- We can make our curve a parabola, but we can't make the actual 2nd derivative equal its expectation value
- Differences in 2nd significant figure



- Will people stop using it? No
- But be careful when doing comparisons

Further details in NIM **550** 392 (2005) and PHYSTAT05

2: More dimensions

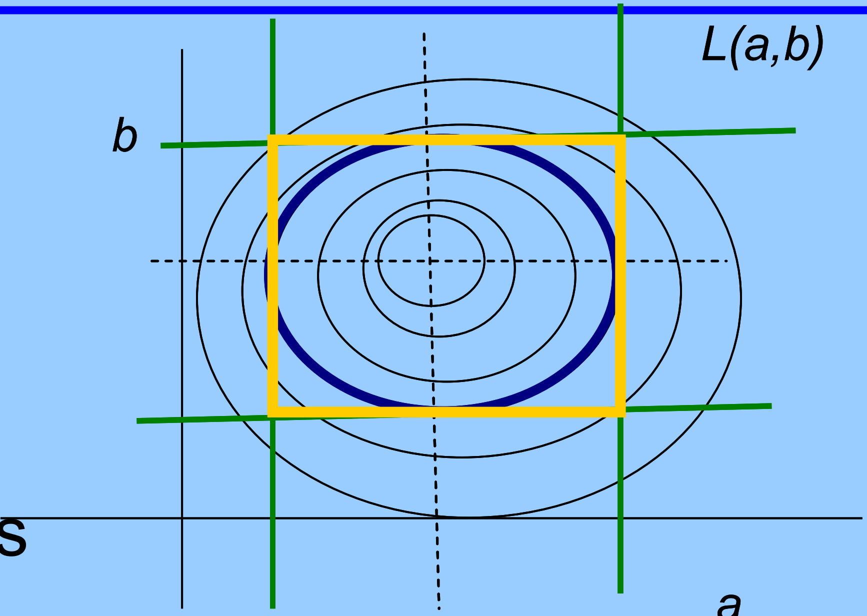
Suppose 2 uncorrelated parameters, a and b

For fixed b, $\Delta \ln L = -\frac{1}{2}$ will give 68% CL region for a

And likewise, fixing a, for b

Confidence level for square is $0.68^2 = 46\%$

Confidence level for ellipse (contour) is 39%



Jointly, $\Delta \ln L = -\frac{1}{2}$ gives 39% CL region

for 68% need $\Delta \ln L = -1.15$

More dimensions, other limits

- Useful to write

$$-2\Delta \ln L = \chi^2$$

- Careful! Given a multidimensional Gaussian, $\ln L = -\chi^2/2$. But $-2\Delta \ln L$ obeys a χ^2 distribution only in the large N limit...

Level is given by finding χ^2 such that $P(\chi^2, N) = 1 - CL$

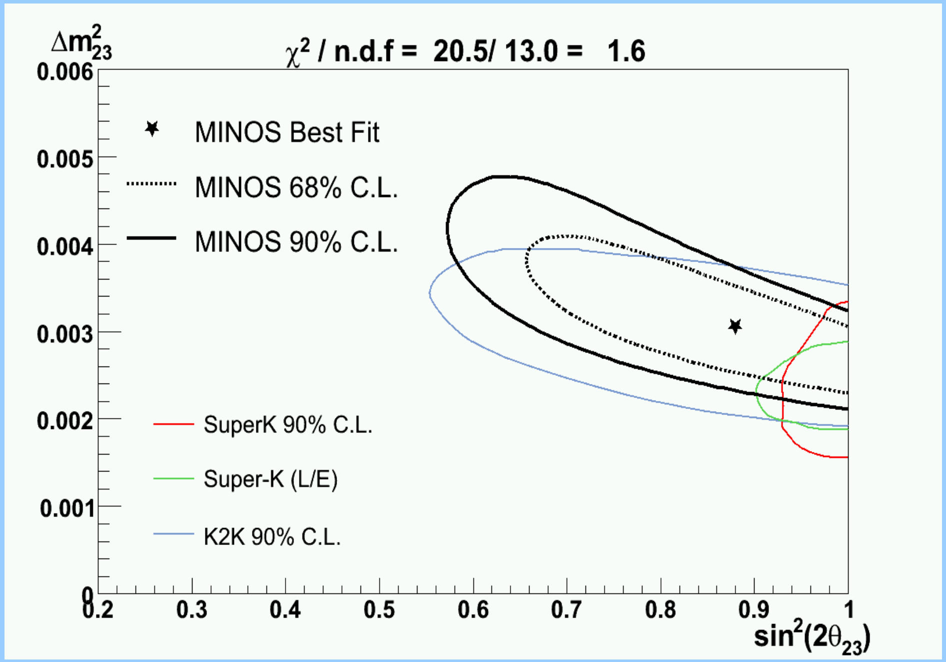
- Generalisation to correlated gaussians is straightforward
- Generalisation to more variables is straight forward. Need the larger $\Delta \ln L$

	68%	95%	99%
1	0.5	1.92	3.32
2	1.15	3.00	4.60
3	1.77	3.91	5.65

etc

Small N non-Gaussian measurements

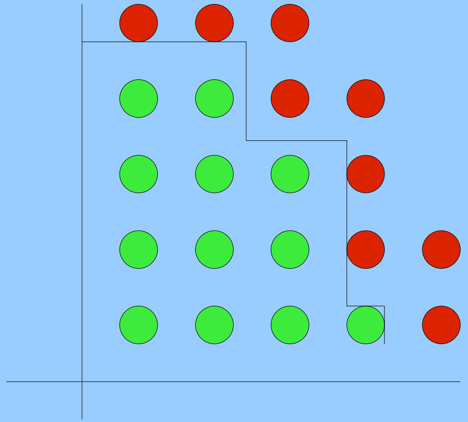
- No longer ellipses/ellipsoids
- Use $\Delta \ln L$ to define confidence regions, mapping out contours
- Probably not totally accurate, but universal



What's the alternative? Toy Monte Carlo

- Have dataset
- Take point M in parameter space. Is it in or out of the 68% (or ...) contour?
- Find $T = \ln L(R|\hat{M}) - \ln L(R|M)$

  clearly small T is 'good'
- Generate many MC sets of R, using M
- How often is $T_{MC} > T_{data}$?
- If more than 68%, M is in the contour



We are ordering the points by their value of T (the Likelihood Ratio) – almost contours but not quite

3: Nuisance parameters Systematic Errors

<u>Formalism</u>

Model parameter M

Result R

Nuisance parameter(s) N

Likelihood is

- L(M,N|R) from experiment
- L'(N) about N

These are combined

<u>Example: Poisson counting</u>

- Source strength S
- R events seen
- Background b

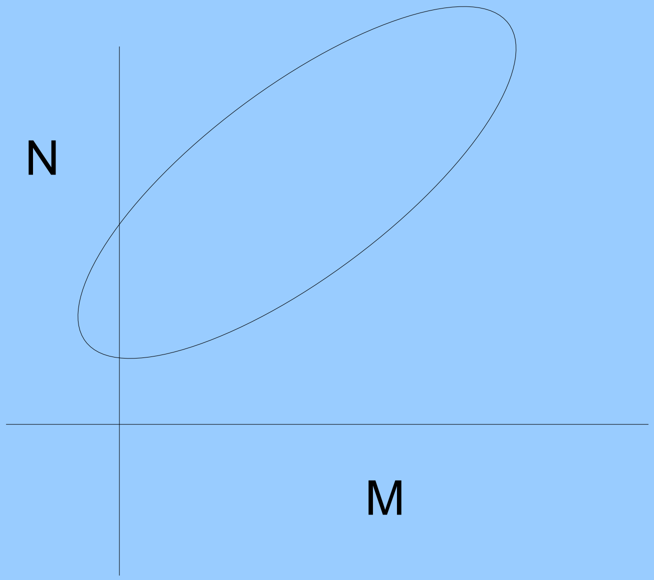
  $b=b_0\pm\sigma_b$

- Poisson(R,S+b)
- Gauss(b,b_0,σ_b)

$$e^{-(s+b)}\frac{(s+b)^R}{R!}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b-b_o)^2/2\sigma_b^2}$$

Approach 1

Quote joint CL contours for N and M



- This is a non-starter. Nobody cares about N. You're losing information about M. (N may be multidimensional)

Approach 2

- Set N to central values to get quoted result for M. Then shift N up one sigma, repeat, and get (systematic) error on M

- No theoretical justification
- Grossly over-estimates error on M
- Still in use in some backward areas

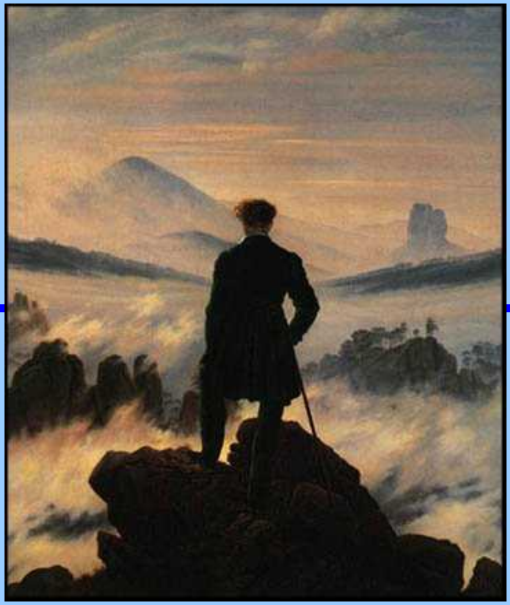
Approach 3

- Integrate out N to get L(M,R)
- This can be done analytically or numerically
- Study L(M,R) and use $\Delta \ln L = -\frac{1}{2}$ or equivalent

This is a frequentist/Bayesian hybrid. Acceptable (?) if the effects are small.

Approach 4

- Profile Likelihood
- Use $\hat{L}(R,M)=L(R,M,\hat{N})$
- Find maximum $\hat{\hat{L}}$
- See how it falls off and use ΔlnL=-½ or equivalent, maximising by adjusting N as you step through M



Intuitively sensible

Studies show it has reasonable properties

4: Justification (?) for using profile likelihood technique

Suppose {M,N} can be replaced by {M,N'} such that

$$L(R,M,N)=L(R,M)\ L'(R,N')$$

There are cases where this obviously works

There are cases where it obviously doesn't work

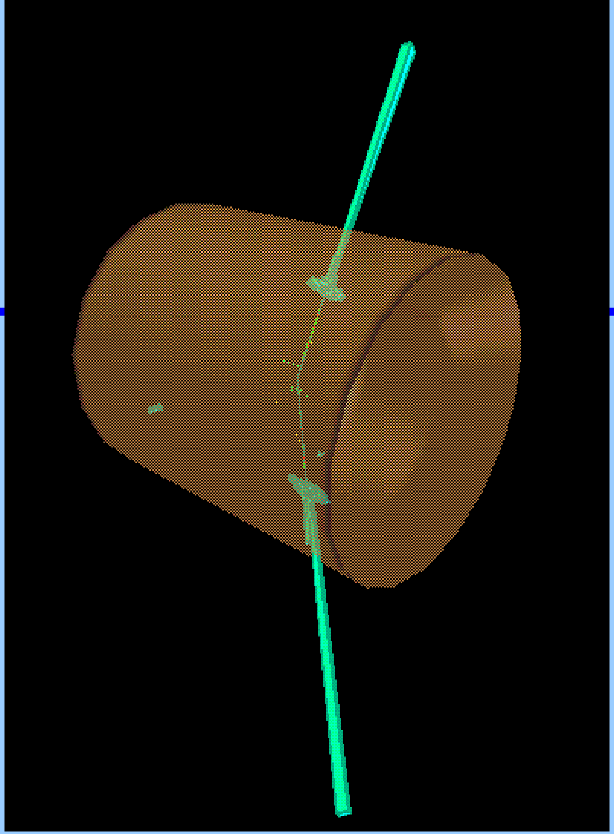
Assuming it does, the shape of L(R,M) can be found by fixing N'.

Can fix N' by taking the peak for given M, as L'(R,N') is independent of M and peak is always at the same N'

Profile Likelihood

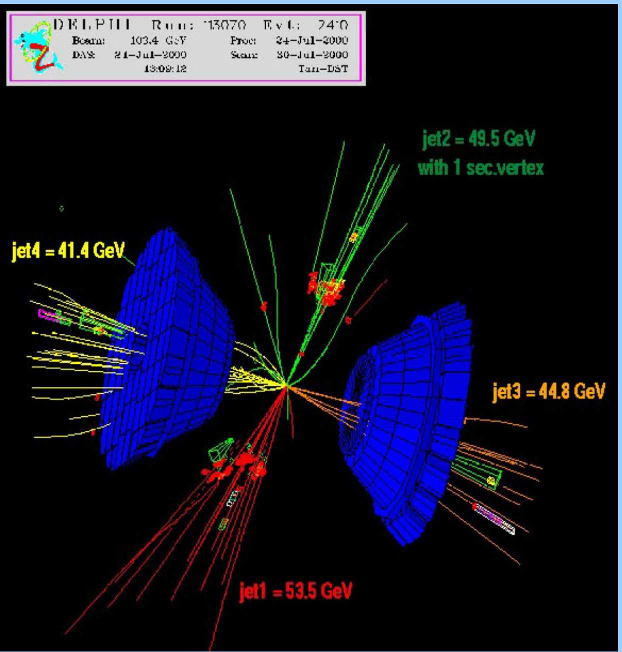
- Provided by Minuit
- Available in ROOT as TRolke
- Use it!

5: The CL_s Technique



Used for Higgs searches by the combined LEP experiments.

'Frequentist-motivated'



Different experiments selected events with varying degrees of Higgsishness

Combining information

Define test-statistic Q which increases with signal s.
Use Likelihood ratio $L(R|M_H)/L(R|\text{no } H)$

Properties known (as function of M_H) from Monte Carlo

Measure some Q_{obs}

Define $CL_b = P_b(Q \leq Q_{obs})$ $CL_{s+b} = P_{s+b}(Q \leq Q_{obs})$

$$CL_s = CL_{s+b}/CL_b$$

Treat this as a CL, even though it isn't. It therefore overcovers.

Why divide?

If you see a small number of events, you know that the background has a downward fluctuation. (In the limit of N=0, we know the background is zero)

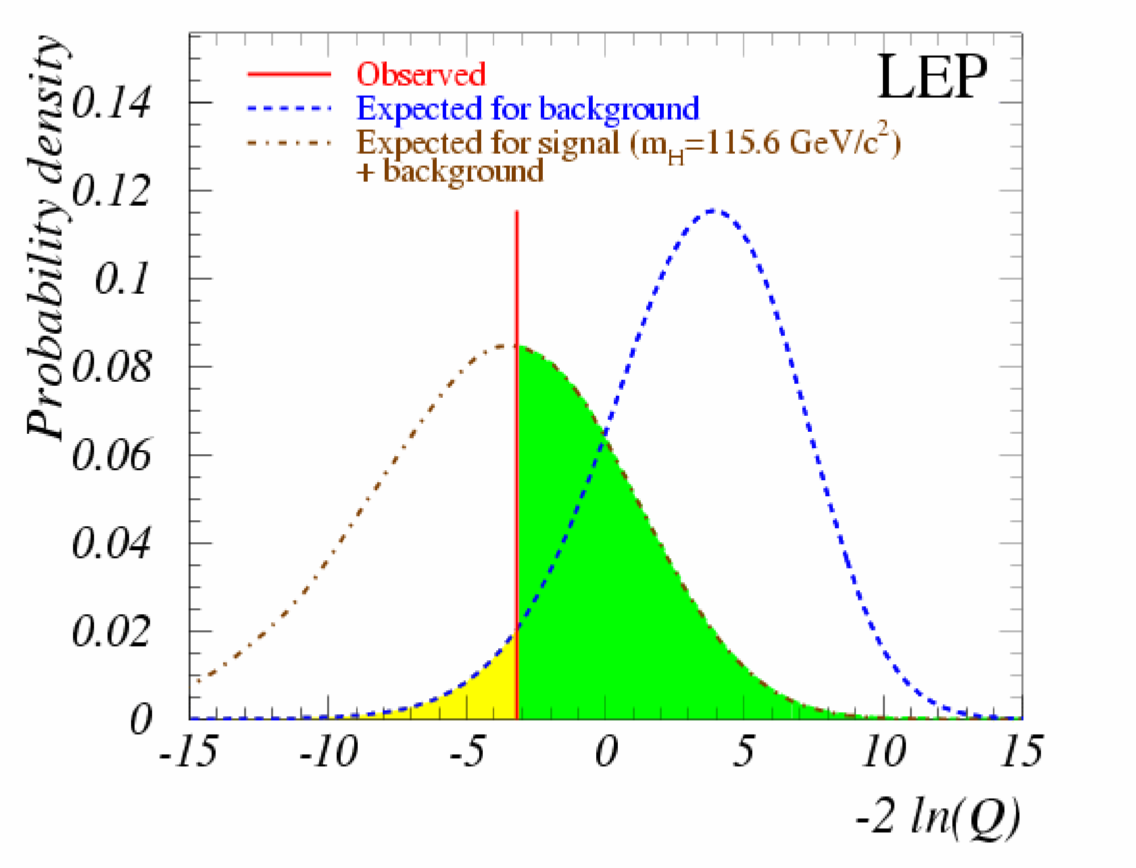
This is like the Bayesian formula

$$\alpha = \frac{\sum_0^n e^{-(s+b)} (s+b)^r / r!}{\sum_0^n e^{-b} b^r / r!}$$

What happens..

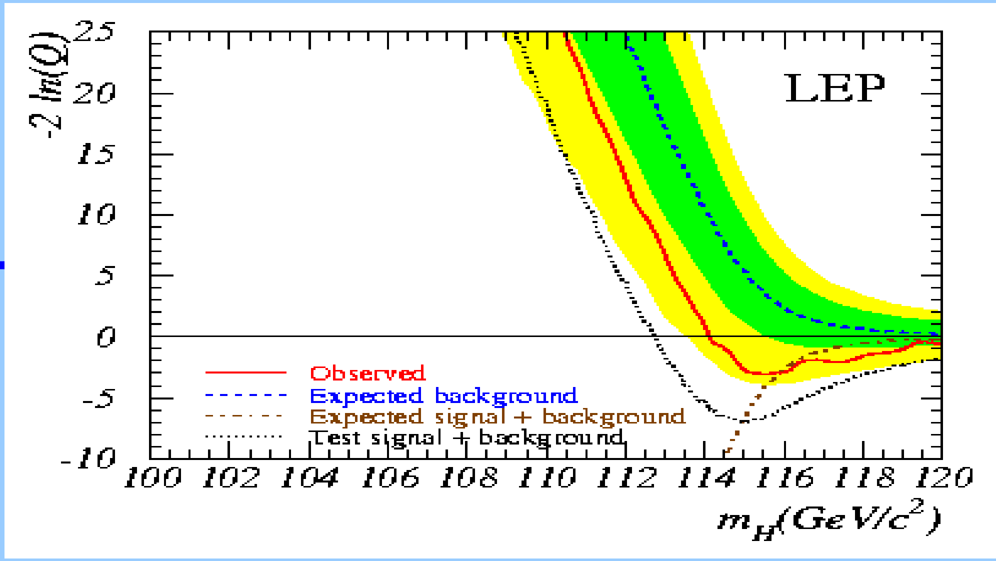
Yellow is $1\text{-}CL_b$

Green is CL_{s+b} for given m_H



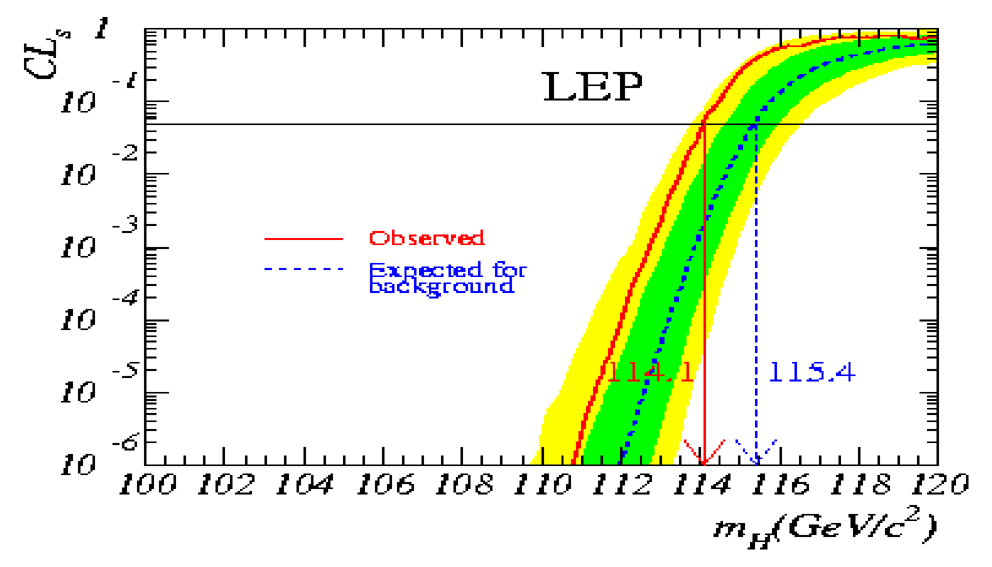
Results

(as of 2002)



Rule out M_H up to 114.1 GeV

(>114.1 GeV @ 95%)



Summary on CL_s

Used for several searches at LEP and elsewhere

Adaptive and sensible.

Frequentist but 'behaves like P(theory|Data)'

Well adapted to exclusion.

See Alex Read's talks at CERN and Durham workshops

6: The Banff Workshop outcome

Particle+statistics workshop in Banff, July 15-20 2006

- Task: given model

$$n = \text{Poisson}(\varepsilon s + b)$$

$$y = \text{Poisson}(tb)$$

$$z = \text{Poisson}(u\varepsilon)$$

t, u known. n,y,z known

Put limit(s) on s at 90% and 99% CL.

Many Models

- Bayesian
- Frequentist
- Hybrid
- ...

Part 1: 10,000 'experiments' have been generated

Participants to run their models and report results

Will be scored for coverage and shortness

Part 2: same again but with 10 separate channels per experiment (same s, different t,u,y,z and n)

Results to be announced in due course...

7: Further reading

- The Particle Data Book
- Textbooks by Glen Cowan, Louis Lyons, R.B.
- "Recommended Statistical Procedures for BaBar" BAD 318
- PHYSTAT proceedings (all Ed. Louis Lyons):
  - CERN 2000-05
  - Durham 2002 IPPP 02/39
  - SLAC 2003 SLAC-R-703
  - Oxford 2005 "Statistical problems in Particle Physics", Imperial College Press (2006)