Treatment of Errors in Analysis of Particle Physics Experiments

Roger Barlow
Manchester University

Hadron Collider Physics Summer School, Fermilab

21st August 2010

Outline

- Why errors are Gaussian
- Combination of Errors
- Errors and Likelihood
- Systematic Errors
  - Working with them
    - Combination of errors
    - Signal+background
  - Finding them
- Consistency checks
- Asymmetric Errors

Why are Gaussians Normal?

The Central Limit Theorem:

If a random number is formed by convoluting N random numbers with means μ_i and variances σ_i^2 then

- The means add: $\mu=\Sigma\mu_i$
- The variances add: $\sigma^2 =\Sigma\sigma_i^2$
- The distribution tends to a Gaussian for large N – whatever the distribution of the separate contributions

If an error is due to many causes, it will be Gaussian

Don't be scared of the non Gaussian

Adding variances works even for non-Gaussian errors.

Even if one source is non Gaussian, it will be mixed up with others.

The only point that doesn't work is the $\pm\sigma$ = 68% etc.

But you can avoid invoking that till the end (be careful in the tail.)

Combination of Errors

Challenge:

Average 2 measurements

$+0.0094 \pm 0.0112 \pm 0.0214$

$-0.00736 \pm 0.00266 \pm 0.00305$

and get

$-0.00957 \pm 0.00251 \pm 0.00146$

Combination of Errors

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$$

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + 2\rho \left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right)\sigma_x \sigma_y + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$$

$$\mathbf{V}' = \tilde{\mathbf{G}} \mathbf{V} \mathbf{G}$$

$$V_{ij} = \rho_{ij} \sigma_i \sigma_j \qquad G_{ij} = \left(\frac{\partial f_j}{\partial x_i}\right)$$

- Does not assume Gaussian behaviour
- Does assume linear dependence

Best Combination

For any w one can form an estimate

$$x = w\,x_1 + (1-w)\,x_2$$

With error

$$\sigma^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2\,w\,(1-w)\,\rho\,\sigma_1\,\sigma_2$$

To minimise this (differentiate) use

$$w = \frac{\sigma_2^2 - \rho\,\sigma_1\,\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\,\rho\,\sigma_1\,\sigma_2}$$

More algebra...

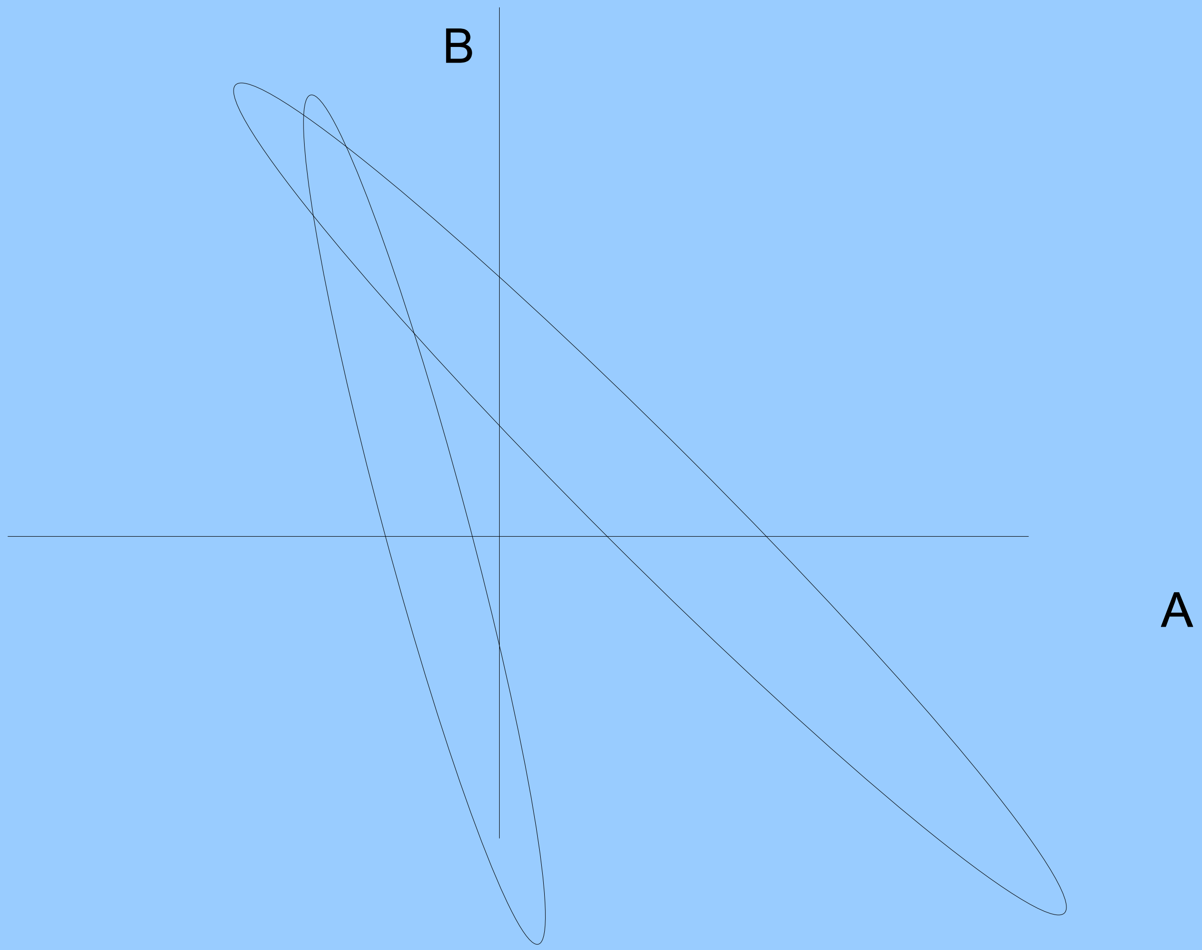
$$x=\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}x_1+\frac{\sigma_1^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}x_2$$

Suppose $\sigma_2=10\sigma_1$, $\rho=0.5$

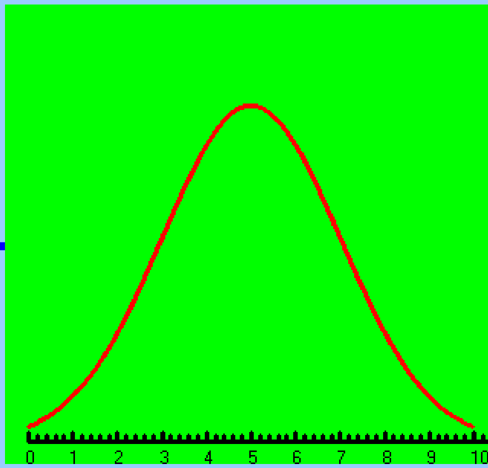
$$x=\frac{95}{91}x_1+\frac{-4}{91}x_2$$

And the error falls to $(72/91)\ \sigma_1$

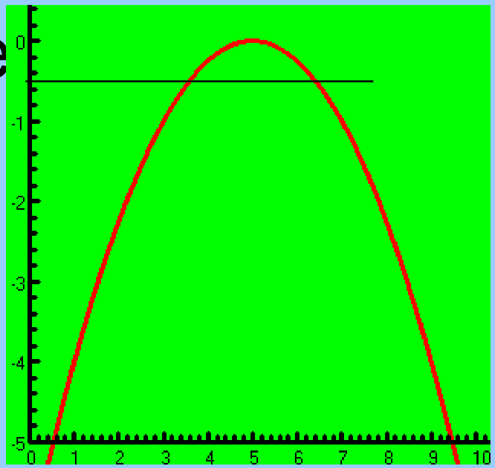
Graphically



Maximum Likelihood



Estimate a model paramete 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

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

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

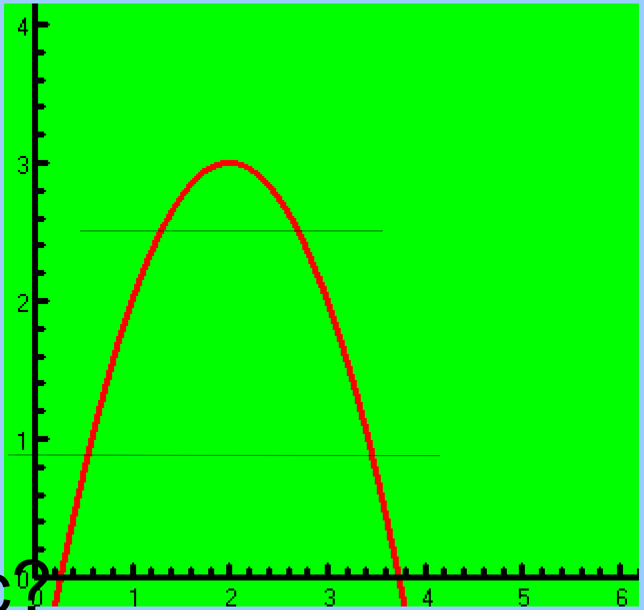
Neat way to find Confidence Intervals

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

$\Delta \ln L = -2$ for 95.4% CL (2σ)

Or whatever you choose

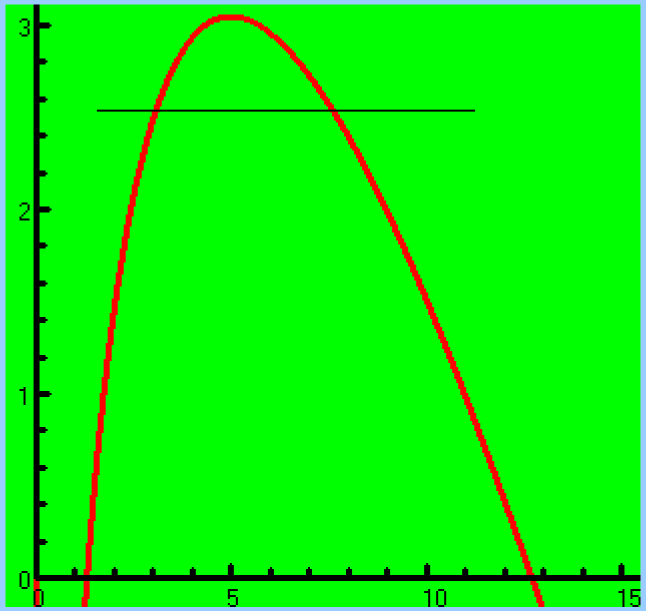
2-sided or 1-sided



But what if isn't Gaussian/parabolic?
You can transform to some μ' for which it is.
Carry out above procedure. Translate back to μ
These limits are what you get from the $\Delta \ln L = -\frac{1}{2}$ procedure anyway – so can omit intermediate step

A Poisson measurement

1. You detect 5 events. Best value 5. But what about the errors?



Find points where log likelihood falls by ½.

Gives upper error of 2.58, lower error of 1.92

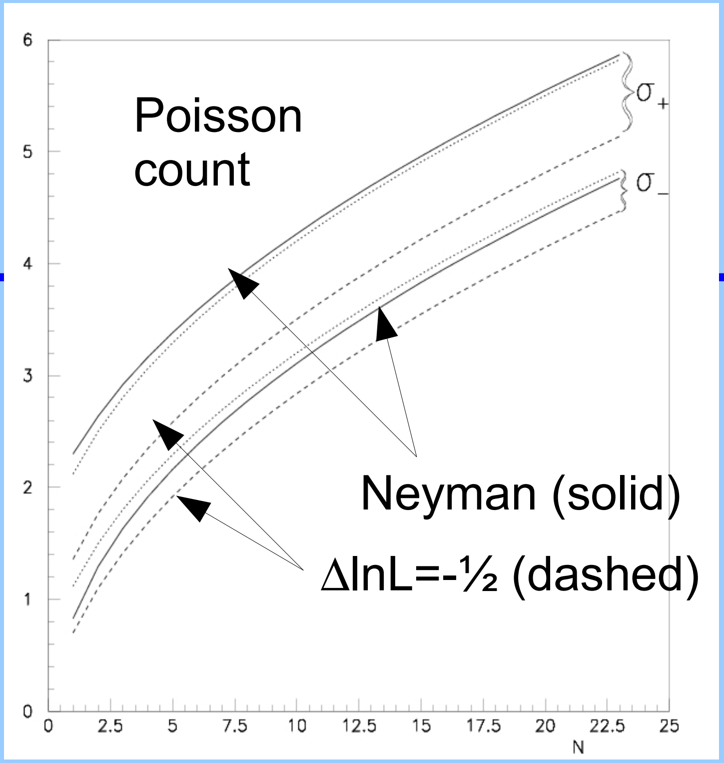
Is it valid?

- 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 $\Delta \ln L = -\frac{1}{2}$

- 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

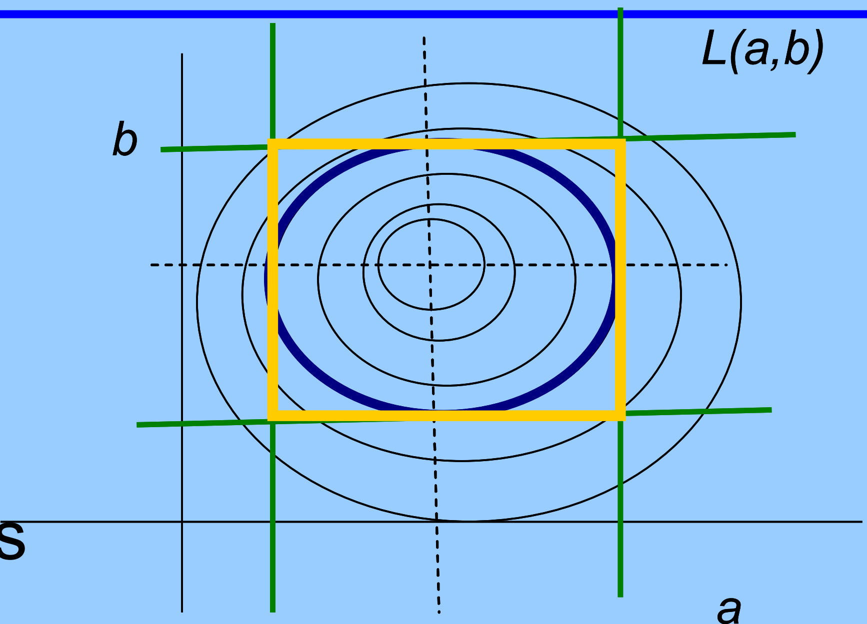


- Will people stop using it? No
- But be careful when doing comparisons

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

More dimensions

- Suppose 2 uncorrelated parameters, a and b
- For any fixed b, $\Delta \ln L=-½$ 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, $\Delta \ln L=-½$ gives 39% CL region
- for 68% need $\Delta \ln L=-1.15$

More dimensions, other limits

- Wilks' theorem

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

- 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

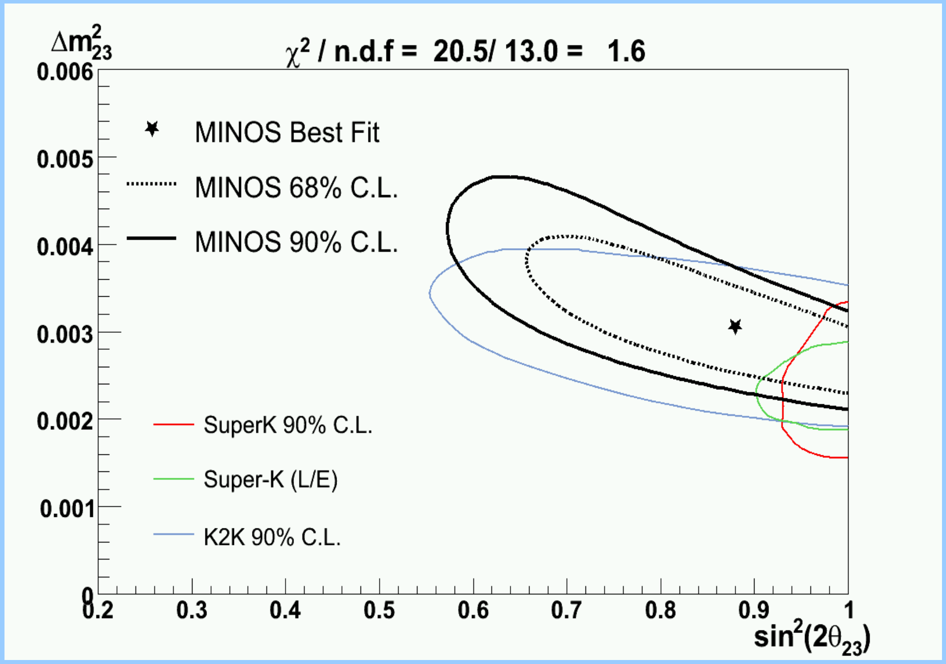
Small N non-Gaussian measurements

No longer ellipses/ellipsoids

Use ΔlnL to define confidence regions, mapping out contours

Probably not totally accurate, but universal

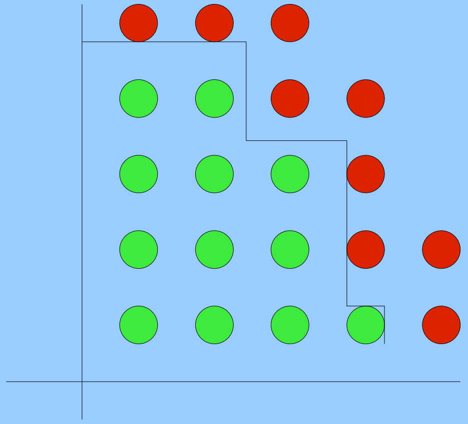
See Tim Gershon's talk for even more amazing contours



To do it properly...

- Have dataset, Max L at $M = \hat{M}$
- 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, find $\hat{M}$ and corresponding T_{MC}
- 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

Systematic Errors

Some persistent myths

Systematic Errors should be added linearly

[Young]

A reproducible inaccuracy introduced by faulty equipment, calibration, or technique [Bevington].

These errors are difficult to detect and cannot be analyzed statistically [Taylor]

Systematic Errors: What are they?

"Systematic effects" is a general category which includes effects such as background, selection bias, scanning efficiency, energy resolution, angle resolution, variation of counter efficiency with beam position and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a "systematic error".

J Orear

Systematic Errors have nothing to do with mistakes..

Systematic Errors: Examples

Measuring tracks in a magnetic field

$P_T = 0.3\ B\ R$

Error on R gives statistical error on momentum

Error on B gives systematic error on momentum

Measuring photons in a calorimeter

$E = \alpha D + \beta$

Error on D gives statistical error on energy

Errors on α, β give systematic error on energy

Measuring a cross section by counting events

$\sigma = N/(\eta L)$

Error on N gives statistical error on cross section

Errors on η,L give systematic error on cross section

Systematic Errors: features

(1) Repetition does not improve them. (Not repeating this measurement, anyway)

(2) Different measurements are affected in the same way. Hence correlations in combination-of-errors

(3) They do not reveal themselves through bad χ^2

Systematics: Note 1- repetition

E.g. More hits on a track improves the measurement of curvature, but not the measurement of the magnetic field.

Many systematic effects are measured by ancillary experiments. More data in these can help.

Sometimes the 'ancillary experiment' is a part of the main experiment... whether an uncertainty is 'systematic' or 'random' may be debatable.

Systematics: Note 2- Correlation

The matrices do it all:

E.g. Two photons measured:

$$E_1 = \alpha D_1 + \beta$$
$$E_2 = \alpha D_2 + \beta$$

$$\boldsymbol{V} = \begin{pmatrix} D_1 & 1 & \alpha & 0 \\ D_2 & 1 & 0 & \alpha \end{pmatrix} \begin{pmatrix} \sigma_\alpha^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 \\ 0 & 0 & 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} D_1 & D_2 \\ 1 & 1 \\ \alpha & 0 \\ 0 & \alpha \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha^2 D_1^2 + \sigma_\beta^2 + \alpha^2 \sigma_1^2 & \sigma_\alpha^2 D_1 D_2 + \sigma_\beta^2 \\ \sigma_\alpha^2 D_1 D_2 + \sigma_\beta^2 & \sigma_\alpha^2 D_2^2 + \sigma_\beta^2 + \alpha^2 \sigma_2^2 \end{pmatrix}$$

Similar approach when combining experimental results that share uncertainties

Systematics: Note 3- Concealment

If you have underestimated (or ignored) a systematic uncertainty then this does not show up in a bad χ^2 or whatever.

This is a challenge. But there are other consistency checks. And there is hard work and common sense, asking colleagues, etc...

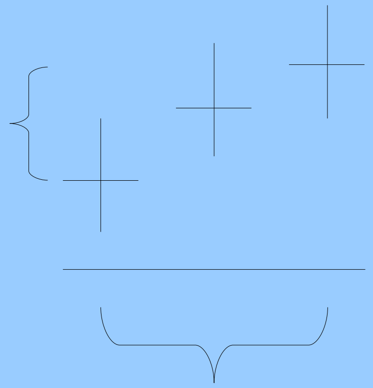
Serious Systematics

If there isn't some nice equation to differentiate...

Example: Complicated analysis (cuts, isolation, combinations...) depending on energies of jets. Jet energy scale not precisely established.

Procedure: repeat analysis

at $+\sigma$ and $-\sigma$



Notes on Serious Systematics

1. Don't sweat the small stuff. Concentrate effort and computer time on the biggest one (or two) effects.

2. You don't have to use $\pm\sigma$. You can take larger/smaller steps to evaluate the gradient. And more values (but see 1)

3. Do not fold in any (statistical) error on the check results.

4. Do not take an asymmetric error too seriously

Systematics may be Bayesian

- A systematic error is an uncertainty in an effect
- Sometimes this is well understood and determined experimentally – e.g. Energy calibration
- Often (more often?) they are estimates - “Theory Errors”
- These are intrinsically Bayesian. Can/must be treated as such

Theory Errors

Typical example – calculated background

Ask (politely) more than one theorist

Be prepared to translate from tolerances to standard deviations.

Common uncertainty – hadronisation model. Run analysis with two different programs. There are cases for

Averaging and taking half the difference as σ

Choosing one and taking the difference as σ

Averaging and taking the difference over $\sqrt{12}$ as σ- but only if these are two extremes

Systematic Errors =Nuisance Parameters

Suppose the result of an experiment depends on a parameter of interest M and a 'nuisance parameter' N

$$P'(M,N|R) \propto L(R|M,N)\ P_M(M)\ P_N(N)$$

We are interested in

$$P'(M|R)=\int P'(M,N|R)\ dN \propto P_M(M) \int L(R|M,N)\ P_N(N)\ dN$$

This is called Marginalisation. Frequentists cannot do it as it involves integrating the Likelihood. For Bayesians it's obvious. (Depends on prior $P_N(N)$)

Poisson revisited

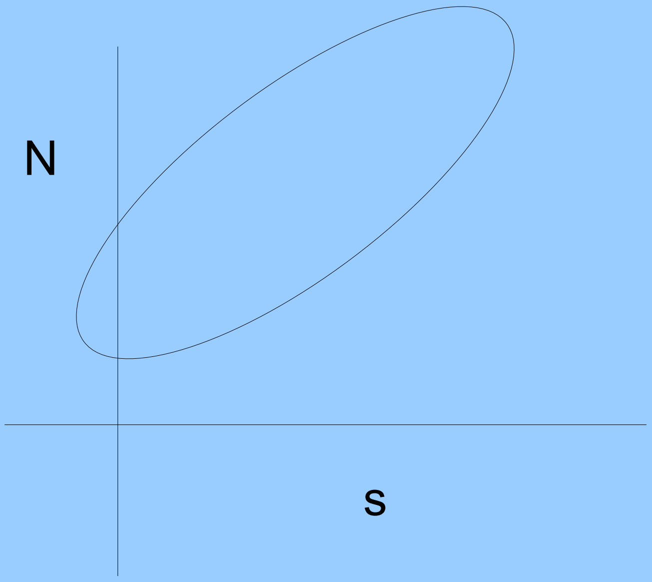
Signal strength $\mu = As + b$

A is 'sensitivity', b is background. Uncertainties on both values are systematic errors. Group them together as N (for nuisance)

4 approaches known. (Plus? This is under development. See Banff workshop proceedings...)

Approach 1

Quote joint CL contours for N and s



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

Approach 2

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

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

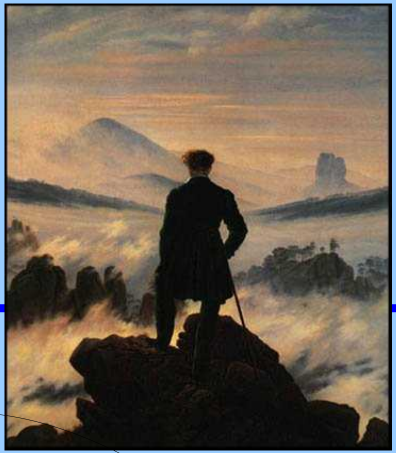
Approach 3: Cousins & Highland

- Integrate out N to get L(s,r)
- This can be done analytically or numerically
- Study L(s,r) and use ΔlnL=-½ or equivalent

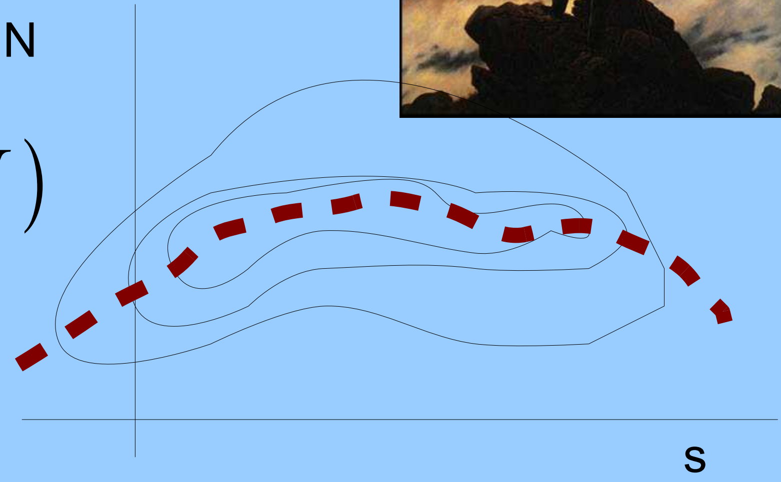
This is a frequentist/Bayesian hybrid. Acceptable (?) if the effects are small

Prior dependence for A – 10% uncertainty in A gives limits ~1% different for different priors

Approach 4



- Profile Likelihood
- Use $\hat{L}(r,s)=L(r,s,\hat{N})$
- 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



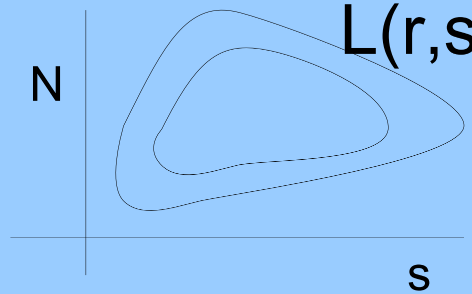
Intuitively sensible

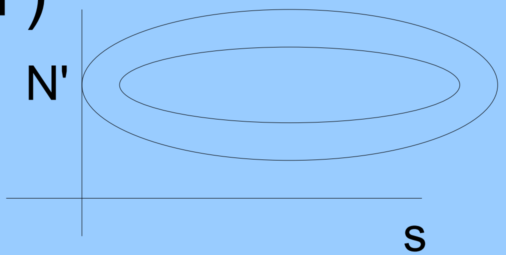
Studies show it has reasonable properties

Justification (?) for using profile likelihood technique

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

$$L(r,s,N)=L(r,s)\ L'(r,N')$$





There are cases where this obviously works (and where it obviously doesn't) but OK for simple peaks.

Assuming it does, the shape of L(r,s) can be found by fixing N' (at any value).

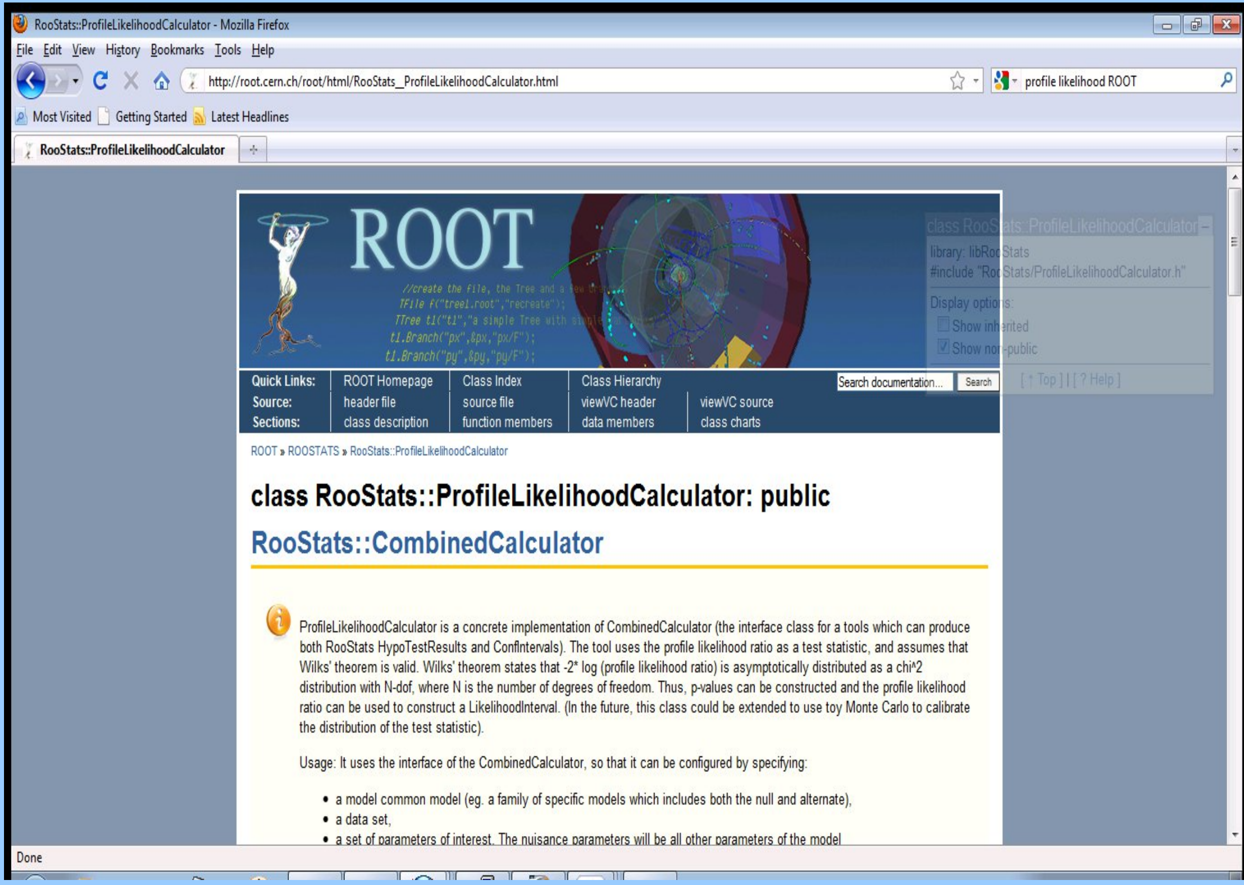
Can fix N' by taking the peak for given s, as L'(r,N') is independent of s and peak is always at the same N'

Profile Likelihood

Provided by Minuit

Available in Roostats

Use it!



Searching for Systematics

"As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know."

- Donald Rumsfeld

Finding them

Think of everything -

Ask around – widely

Run consistency checks

Consistency checks

Are mass peaks in the right place?

Is the result consistent for different runs?

Is the result consistent with different cuts?

Is an analysis of the Monte Carlo consistent with what you put in?

Etc etc

All well known and noncontroversial...

What is consistent?

'Consistent within errors' is less impressive than it looks, as the two analyses share the same data.

Taking the difference in quadrature between the two results as a benchmark has a lot to recommend it.

What next?

Run a consistency check.

Decide if the result is consistent.

If it is – tick the box and move on. Do NOT add it to the systematic error!

If it is not – find out why! Do NOT add it to the systematic error!

- well, only as a last resort.

Point of confusion

Add these

Uncertainty	Resulting change
Jet energy scale 10%	1.2
Muon acceptance 2%	2.3
Lumi 5%	3.4
...	

Do not add these

Check	Change	Verdict
Loose muons	+0.5	OK
2010 data only	+0.3	OK
Change E_{cut} + 1 GeV	-1.1	OK

Asymmetric Errors

Arise from 2 sources:

Statistical errors from asymmetric confidence intervals, typically non-parabolic In likelihood plots

Systematic errors from serious systematic checks where the upward and downward shifts are different

Standard Method

Combine the positive errors in quadrature

Combine the negative errors in quadrature

$$a^{+3}_{-5} + b^{+4}_{-12} = (a+b)^{+5}_{-13}$$

This is clearly wrong.

$$a^{+1}_{-2} + b^{+1}_{-2} + c^{+1}_{-2} + d^{+1}_{-2} = (a+b+c+d)^{+2}_{-4}$$

Standard Method

Combine the positive errors in quadrature

Combine the negative errors in quadrature

$$a^{+3}_{-5} + b^{+4}_{-12} = (a+b)^{+5}_{-13}$$

This is clearly wrong.

$$a^{+1}_{-2} + b^{+1}_{-2} + c^{+1}_{-2} + d^{+1}_{-2} = (a+b+c+d)^{+2}_{-4}$$

- It violates the Central Limit Theorem

Alternative

Given a value and asymmetric errors, use these 3 points on the log likelihood plot to draw a parabola-like curve.

Good results from

$$\ln L(a) = -\frac{1}{2}\frac{(x-\hat{x})^2}{V}$$

with $V = V_0 + V'(x-\hat{x})$

For details see arXiv:physics/0406120

Recommendation: Avoid asymmetric Errors

Replace $12.3^{+3.3}_{-2.9}$ by 12.3 ± 3.1

Conclusion

Do not inflate your Systematic Errors.
“Conservative” is not a valid reason.

Statistics is a Science, not a folk tradition.
Respect it, and it will serve you well.