Real life problems Systematic errors

— What are systematic uncertainties
— How to deal with systematic uncertainties

Wouter Verkerke, UCSB

Real-life problems

- Exact calibration of p-values/Z-values possible for counting experiment with background that is *exactly* known.

- Real-life problems are often more difficult
  - Result is (ML) fit to distribution in >=1 dimensions, with model shape uncertainties
  - Background estimate has uncertainty

- *Correct* calculation of significance no longer guaranteed!
  - Q: Is that a problem?
  - A: Yes. If your (naïve) calculation says Z=5, but it is really Z=3, there is a substantial chance your discovery is fake
  - If ATLAS and CMS use different methods one experiment may claim discovery of e.g. Higgs with only half the data of the other because of differences in significance calculation

- **Systematic errors**
  - What are systematic errors
  - How to incorporate them in measurement
  - Nuisance parameters in limits and intervals

Systematic errors vs statistical errors

- Definitions

Statistical error = any error in measurement due to
                    statistical fluctuations in data

Systematic errors = **all other errors**

Systematic uncertainty ≡ Systematic error

- But Systematic **error** ≠ Systematic **mistake**!

  – Suppose we know our measurement needs to be
    corrected by a factor of 1.05 ± 0.03

  – Not correcting the data by factor 1.05 introduces
    a systematic mistake

  – Right thing to do: correct data by factor 1.05
    and take uncertainty on factor (0.03) as a systematic error

Classification of systematic errors

- Classification by Pekko Sinerva (PhyStat 2003)

- Good, Bad and Ugly

Source of systematic errors – 'Good' and 'Bad' errors

- **'Good'** errors arise from clear causes and can be evaluated
  - Clear cause of error
  - Clear procedure to identify and quantify error
  - Example: Calibration constants,
    efficiency corrections from simulation

- **'Bad'** errors arise from clear causes, but can *not* be evaluated
  - Still clear cause
  - But no unambiguous procedure to quantify uncertainty
  - Example: theory error:
    - Given 2 or more choices of theory model you get 2 or more different answers.
    - What is the error?

Sources of systematic errors – 'Ugly' errors

- 'Ugly' errors arise from sources that have been overlooked
  - Cause unknown → error unquantifiable

- 'Ugly' errors are usually found through failed sanity checks
  - Example: measurement of CP violation on a sample of events that is known to have no CP-violation: You find $A_{CP}=0.10 \pm 0.01$
  - Clearly something is wrong – What to do?

  - 1) **Check your analysis**
  - 2) Check your analysis again
  - 3) Phone a friend
  - 4) Ask the audience
    …
  - 99) **Incorporate as systematic error as last and desperate resort!**

Wouter Verkerke, UCSB

What about successful sanity checks?

- **Do not incorporate successful checks in your systematic uncertainty**

  - Infinite number of successful sanity checks would otherwise lead to infinitely large systematic uncertainty. Clearly not right!

- Define beforehand if a procedure is a sanity check or an evaluation of an uncertainty

  - If outcome of procedure can legitimately be different from zero, it is a systematic uncertainty evaluation

  - If outcome of procedure can only significantly different from zero due to mistake or unknown cause, it is a sanity check

Common scenarios in evaluating systematic errors

- Two values – corresponding to use of two (theory) models A,B
  - What is a good estimate for your systematic uncertainty?

- I) If A and B are *extreme scenarios*, and the truth must always be between A and B
  - Example: fully transverse and fully longitudinal polarization
  - Error is root of variance with uniform distribution with width A-B

$$\sigma = \frac{|A-B|}{\sqrt{12}}$$

$$V(x) = \langle x \rangle^2 - \langle x^2 \rangle = \left(\frac{1}{2}\right)^2 - \int_0^1 x^2 dx = \frac{1}{4} - \frac{1}{3} = \frac{1}{12}$$

  - Popular method because sqrt(12) is quite small, but only justified if A,B are truly extremes!

- II) If A and B are typical scenarios
  - Example: JETSET versus HERWIG (different Physics simulation packages)
  - Error is difference divided by sqrt(2)

Factor $\sqrt{\frac{N}{N-1}}$
to get unbiased estimate of σ_{parent}

$$\sigma = \frac{|A-B|}{2} \cdot \sqrt{2} = \frac{|A-B|}{\sqrt{2}}$$

Common scenarios in evaluating systematic errors

- Two variations of the analysis procedure on the *same* data

  - Example: fit with two different binnings giving $A \pm \sigma_A$ and $B \pm \sigma_B$

  - Clearly, results A,B are correlated so $\dfrac{|A-B|}{\sqrt{\sigma_A^2 + \sigma_B^2}}$ is not a good measure of smallness of error

- Generally difficult to calculate, but can estimate uppper,lower bound on systematic uncertainty

$$\sqrt{\sigma_A^2 - \sigma_0^2} - \sqrt{\sigma_B^2 - \sigma_0^2} \leq \sigma_{A-B} \leq \sqrt{\sigma_A^2 - \sigma_0^2} + \sqrt{\sigma_B^2 - \sigma_0^2}$$

  - Where $\sigma_A > \sigma_B$ and σ_0 is the Minimum Variance Bound. $\sigma_0^2(\hat{a}) = \left\langle (\hat{a} - \langle \hat{a} \rangle)^2 \right\rangle$

  - If the better technique (B) saturates the MVB the range reduces to

$$\sigma_{A-B}^2 = \sigma_A^2 - \sigma_B^2$$

  - If MVB is not saturated (e.g. you have low statistics) you will need to use a toy Monte Carlo technique to evaluate σ_{A-B} Wouter Verkerke, UCSB

Common scenarios in evaluating systematic errors

- Common scenario in HEP analysis:
  you need to assign systematic uncertainty to shape of
  distribution obtained from simulated events

  - Measuring data in range [30,90]. Correct for acceptance effect by
    dividing by efficiency measured from simulated data.

  - Ndata[30,90] = 750

  - eff_{MC} = 7300/10000 = 0.73

  Ntotal = 750 / 0.73 = 1027

  - What is uncertainty on Ntotal from MC simulation uncertainty?
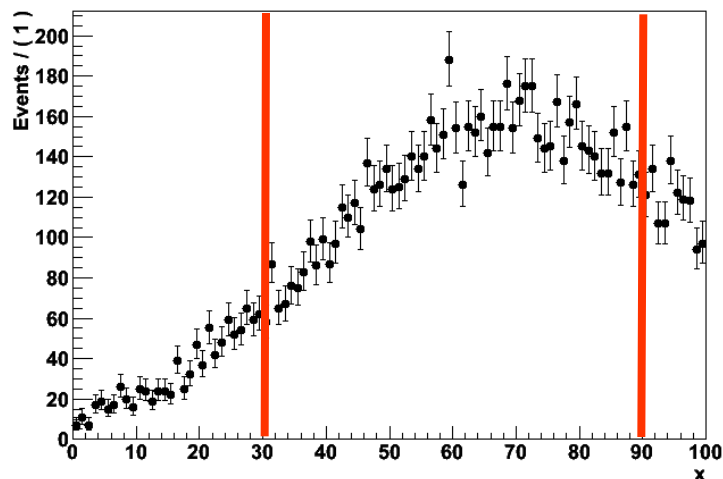
'simulation'   'data'
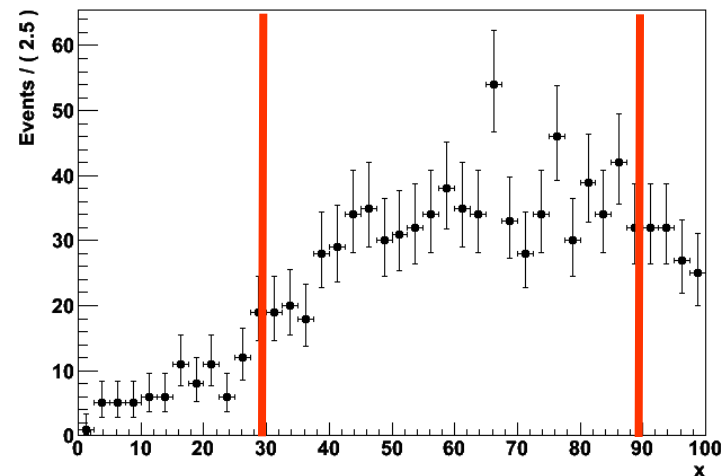
Common scenarios in evaluating systematic errors

- Popular but wrong technique: 'Cut variation'

  - Procedure: vary each of your cuts by a little bit. For each change,

  1) Measure new yield on data

  2) Correct with new MC efficiency.

  3) Difference between efficiency corrected results is systematic uncertainty.

  - Example, for a nominal cut in x at 'p' you find N(data)=105, with a MC efficiency ε_{MC}=0.835 so that N(corrected)=125.8

	N(data)	ε(MC)	N(corrected)
p+Δp	110	0.865	127.2
p-Δp	100	0.803	124.5

$$\sigma_{sys}^{p} = (127.2 - 124.5)/2 = 1.4$$

$$x = 125.8 \pm 1.4$$

- Bad thing to do →
  error can be dominated by low stats in variation range

Common scenarios in evaluating systematic errors

- **Warning I:** Cut variation does not give an precise measure of the systematic uncertainty due data/MC disagreement!

  – Your systematic error is dominated by a potentially large statistical error from the small number of events in data between your two cut alternatives

    - This holds independent of your MC statistics

  – You could see a large statistical fluctuation
    → **error overestimated**

  – You could see no change due to a statistical fluctuation
    → **error underestimated**

Wouter Verkerke, UCSB
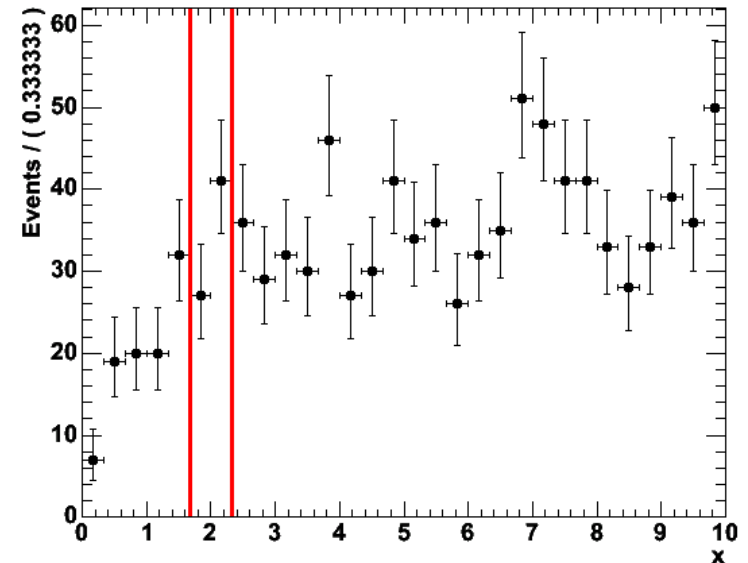
Common scenarios in evaluating systematic errors

- Warning II: Cut variation doesn't catch all types of data/MC discrepancies that may affect your analysis
    - Error may be fundamentally underestimated
    - Example of discrepancy missed by cut variation:
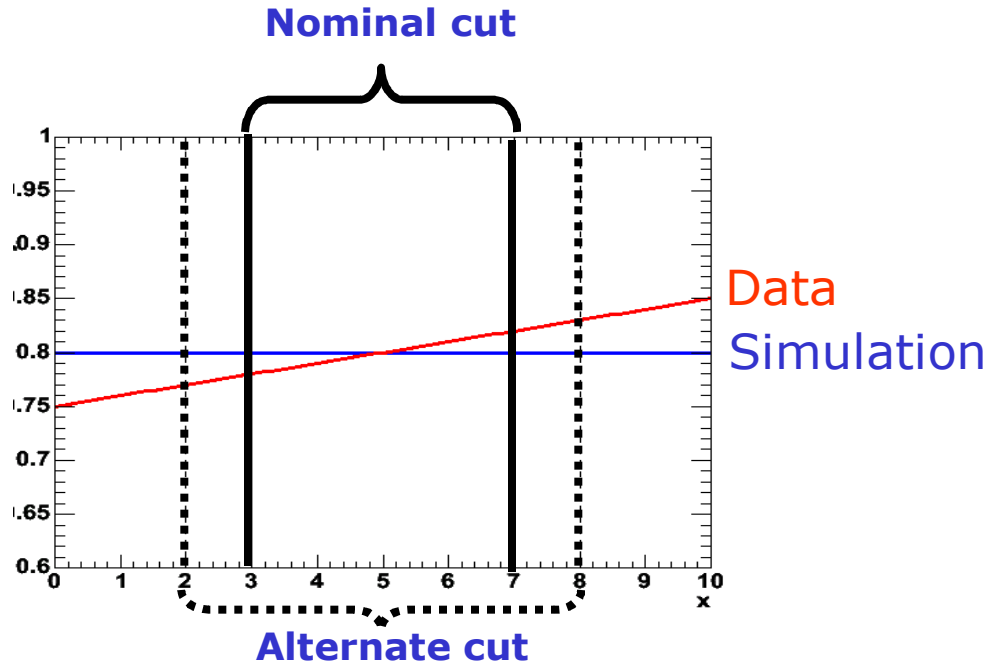
Data and Simulation give same efficiency for nominal and alternate cut, sp

Zero systematic is evaluated (in limit N→∞)

Even though data and MC are clearly different

Cut variation is a good sanity check,
but not necessarily a good estimator for systematic uncertainty

Systematic errors and correlations

- Pay attention to correlation between systematic errors

$$\sigma_{xy}^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + 2\left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right)\rho\sigma_x\sigma_y$$

- If error uncorrelated, $\rho=0$
  - Add in quadrature

- If error 100% correlated, then $\rho=1$.
  - E.g. tracking efficiency uncertainty per track for 6 tracks,

    $\sigma_{3trk} = \sigma_{trk}+\sigma_{trk}+\sigma_{trk} = 3\cdot\sigma_{trk}$ (*not* $\sqrt{3} \cdot\sigma_{trk}$)

- If errors 100% anti-correlated, then $\rho=-1$
  - This can really happen!
  - Example BF($D^{*0} \rightarrow D^0\pi^0$) =67% and BF($D^{*0}\rightarrow D^0\gamma$) =33%

Combining statistical and systematic uncertainty

- Systematic error and statistical error are independent

  - In 'simple measurements' can be added in quadrature to obtain combined error

  - Nevertheless always quote (also) separately.

  - Also valid procedure if systematic error is not Gaussian: Variances can be added regardless of their shape

  - Combined error usually approximately Gaussian anyway (C.L.T)

- You can also incorporate systematic uncertainties directly into a ML of χ^2 fit

  - Advantage: automatic and correct error propagation (including correlations)

  - For ML fit simply multiply your pdf in the parameters with a pdf that describes you systematic uncertainty

Adding uncertainties to a likelihood

- Example: Gaussian signal plus flat background
  - Mean of Gaussian, fraction of signal are parameters

- Case I : Width of Gauss is known exactly

$$F(x; \mu, f, \sigma) = f \cdot G(x, \mu, \sigma) + (1 - f) \cdot U(x)$$

$$L(x; \mu, f, \sigma) = -\sum_D F(x_i, \mu, f, \sigma)$$

- Case II: Width of Gaussian is known with 10% uncertainty
  - NB: '10% uncertainty' is ambiguous, choose Gaussian error blow

$$F'(x; \mu, f, \sigma_0, \delta\sigma) = F(x, f, \mu, \sigma) \cdot G(\sigma, \sigma_0, \delta\sigma)$$

$$L(\mu, f, \sigma_0, \delta\sigma) = -\sum_D F(x_i, \mu, f, \sigma) - \log(G(\sigma \mid \sigma_0, \delta\sigma))$$

Adding uncertainties to a likelihood

- Example 1 – Width known exactly
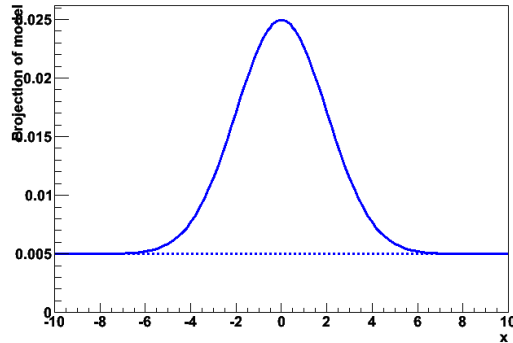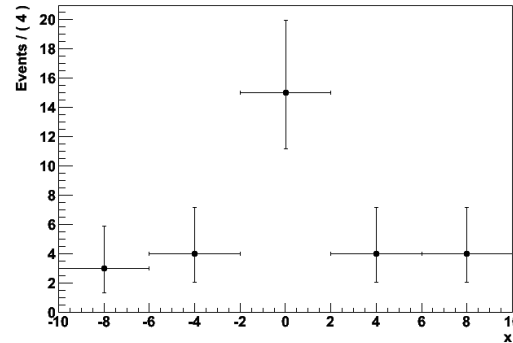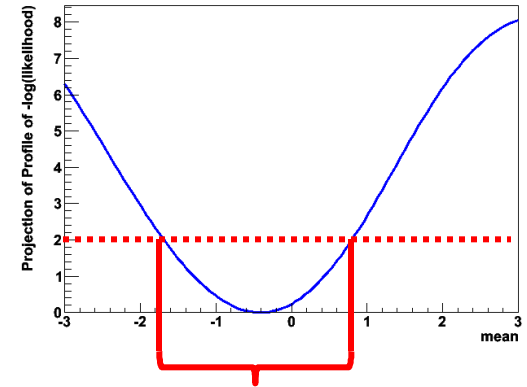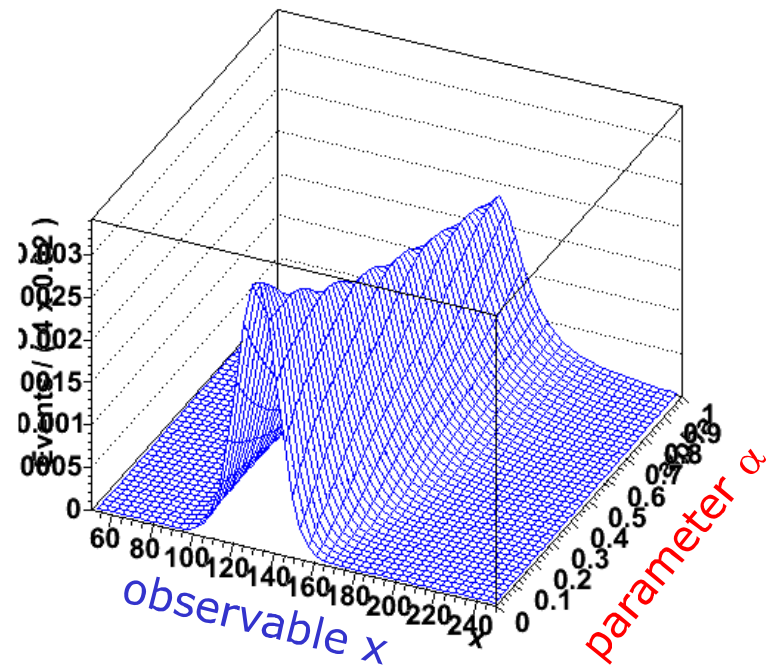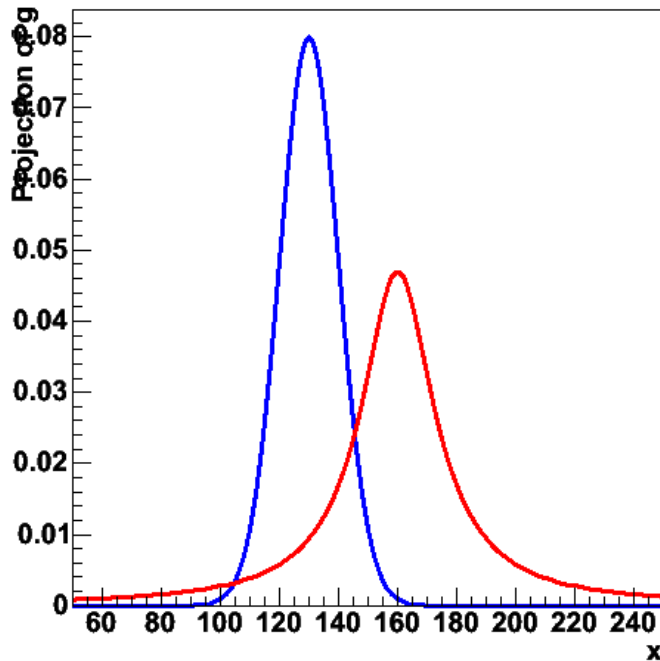

- Example 2 – Gaussian uncertainty on width
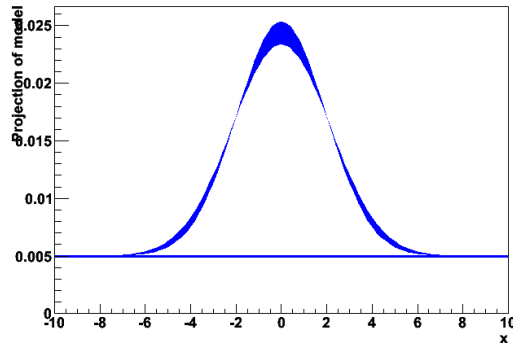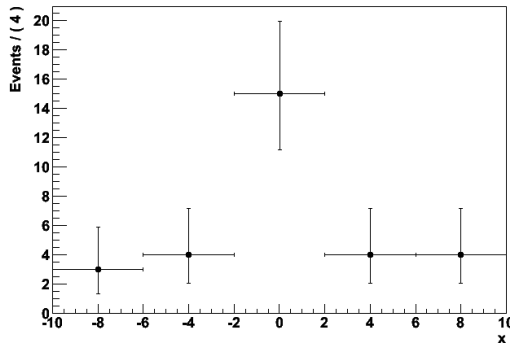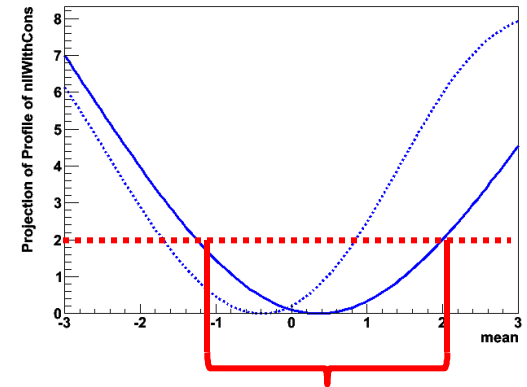
Adding uncertainties to likelihood

- Can in principle also systematic uncertainties of type 'HERWIG simulation' vs 'PYTHIA simulation' directly in likelihood using 'morphing transformation'

  - Several algorithms available in e.g. RooFit

  $F_A(x;p)$ $F_B(x;q)$ $F_{AB}(x;\alpha,p,q)$

  - Construct $-\log(L)$ with F_{AB} and let α float in fit

  - Optionally, add (Gaussian) constraint term on parameter α

Parameters of interest vs. nuisance parameters

- Typical (likelihood) fit yields many parameters

$$F(x; \mu, f, \sigma) = f \cdot G(x, \mu, \sigma) + (1 - f) \cdot U(x)$$

  – Not all are of 'physics interest'

- Terminology to distinguish parameters
  – Parameter(s) of interest → Physics we intend to publish (N_{Higgs})
  – Nuisance parameters → The others (detector resolution etc…)

- Goal: incorporate uncertainty from nuisance parameters into parameters of interest.

- Each of the three main classes of constructing intervals (Bayesian, likelihood ratio, Neyman confidence intervals) has a way to incorporate the uncertainty on the nuisance parameters in the parameters of interest. *But this remains a subject of frontier statistics research.*

- Next: introduce methods and their problems

Wouter Verkerke, NIKHEF

Dealing with nuisance parameters in Bayesian intervals

- **Elimination of nuisance parameters in Bayesian interval**
  - Construct a multi-D prior pdf P(parameters)
    for the space spanned by all parameters;
  - Multiply by P(data|parameters) for the data obtained;
 - Integrate over the full subspace of all nuisance parameters;

$$ p(s \mid x) = \int \left(L(s, \vec{b}) \, p(s, \vec{b}) \right) d\vec{b} $$

  - You are left with the posterior pdf for the parameter of interest. The math is now reduced to the case of no nuisance parameters.

- **Issues**
  - The multi-D prior pdf is a problem for both subjective and non-subjective priors.
  - In HEP there is almost no use of the favored non-subjective priors (reference priors of Bernardo and Berger), so we do not know how well they work for our problems.
  - In case of many nuisance parameters, the high-D numeric integral can be a technical problem (use of Markov Chain Monte Carlo)
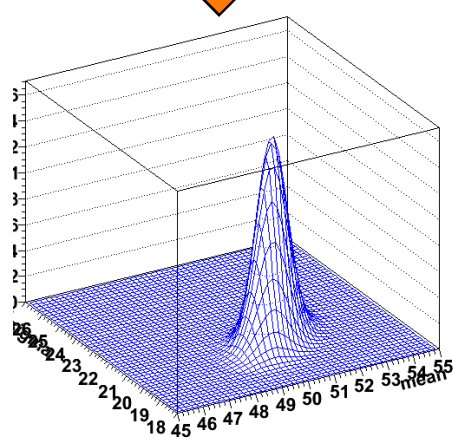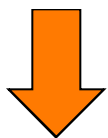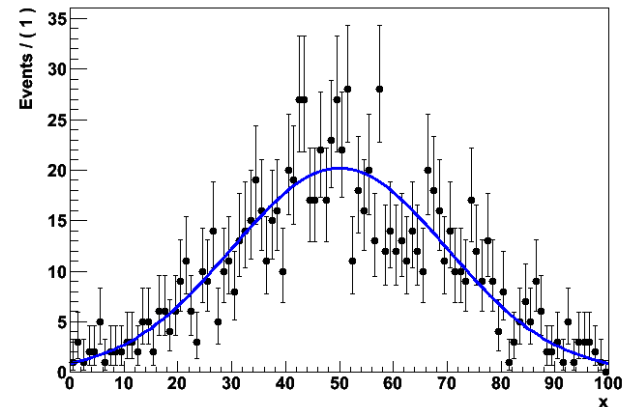
Illustration of nuisance parameters in Bayesian intervals

- Example: data with Gaussian model (mean,sigma)

-logLR(mean,sigma)
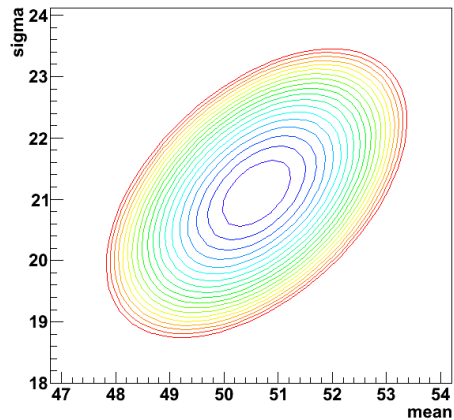
MLE fit fit data
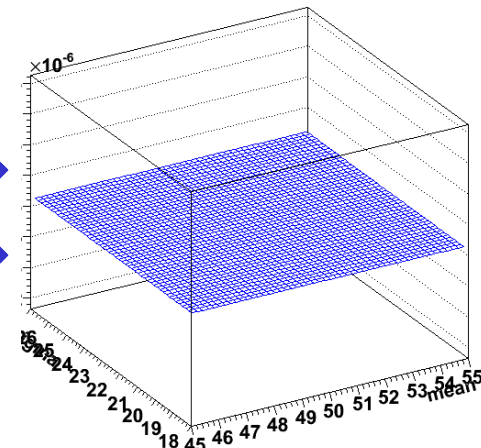
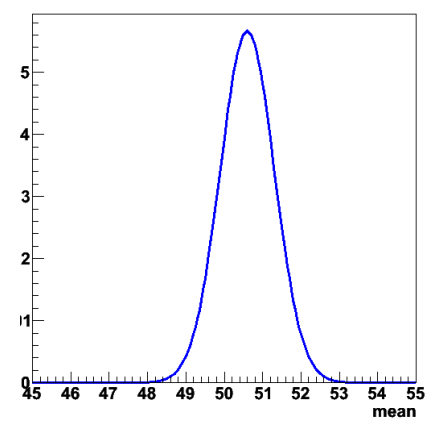

∫ LR(mean,sigma) × prior(mean,sigma) = posterior(mean)

Dealing with nuisance parameters in Frequentist intervals

- *(Full) Neyman construction*:

    - The goal is that the parameter of interest should be covered at the stated confidence **for every value of the nuisance parameter**

    - if there is any value of the nuisance parameter which makes the data consistent with the parameter of interest, that parameter point should be considered:

    - eg. don't claim discovery if any background scenario is compatible with data

- Issues

    - Significant over coverage common problem

    - Wilks theorem may not apply due to e.g. 'look elsewhere effects' in nuisance parameters → must rely on toy MC approach, can get very cumbersome
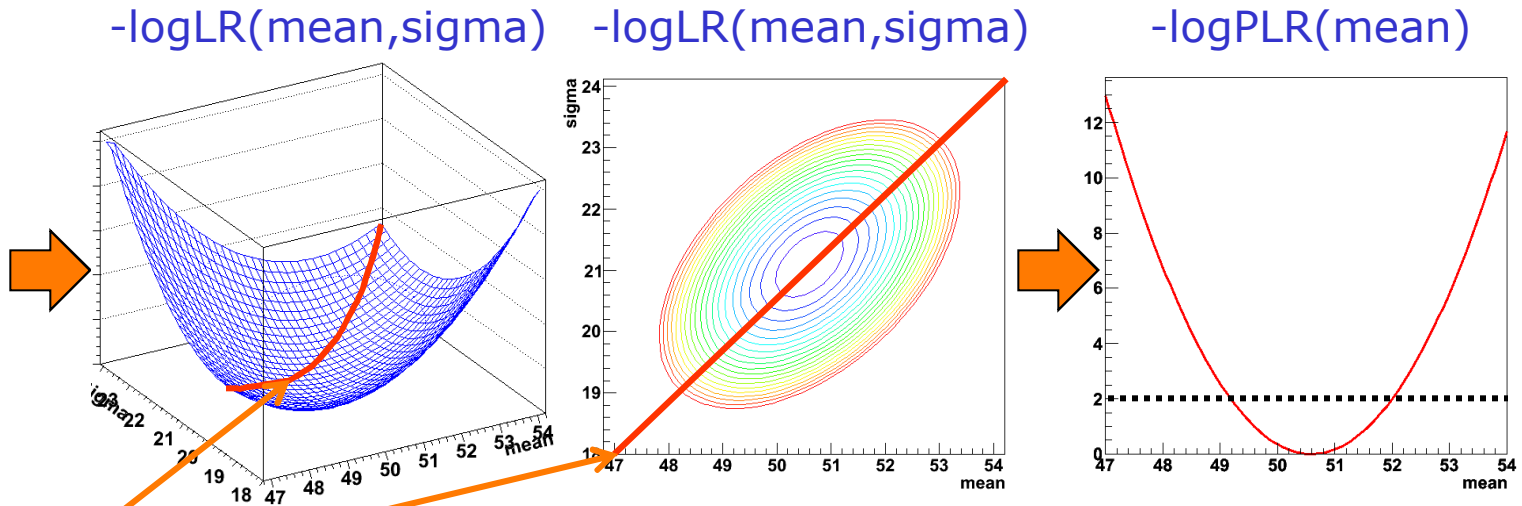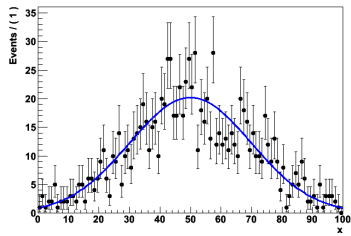

full construction
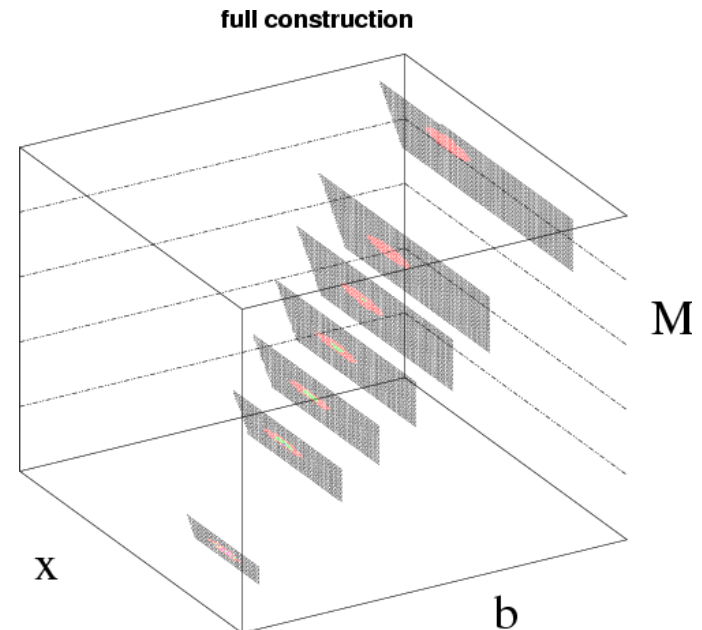
Wilks theorem and nuisance parameters

- Wilks's theorem holds if the **true distribution is in the family** of functions being considered

    - eg. we have sufficiently flexible models of signal & background to incorporate all systematic effects

    - but we don't believe we simulate everything perfectly

    - ..and when we parametrize our models usually we have further approximated our simulation.

- E.g. if a model has a floating mass, it is clear that there is a degradation in significance due to the look-elsewhere effect (if you look into a wide enough mass range, your always find 'some peak' in the background)

    - Formally, the conditions required for Wilks's theorem do not hold because floating mass parameter makes no sense in a background-only model.

Dealing with nuisance parameters in Likelihood ratio intervals

- ## Nuisance parameters in LR interval

  - For each value of the parameter of interest, search the full subspace of nuisance parameters for the point at which the likelihood is maximized.

  - Associate that value of the likelihood with that value of the parameter of interest → 'Profile likelihood'
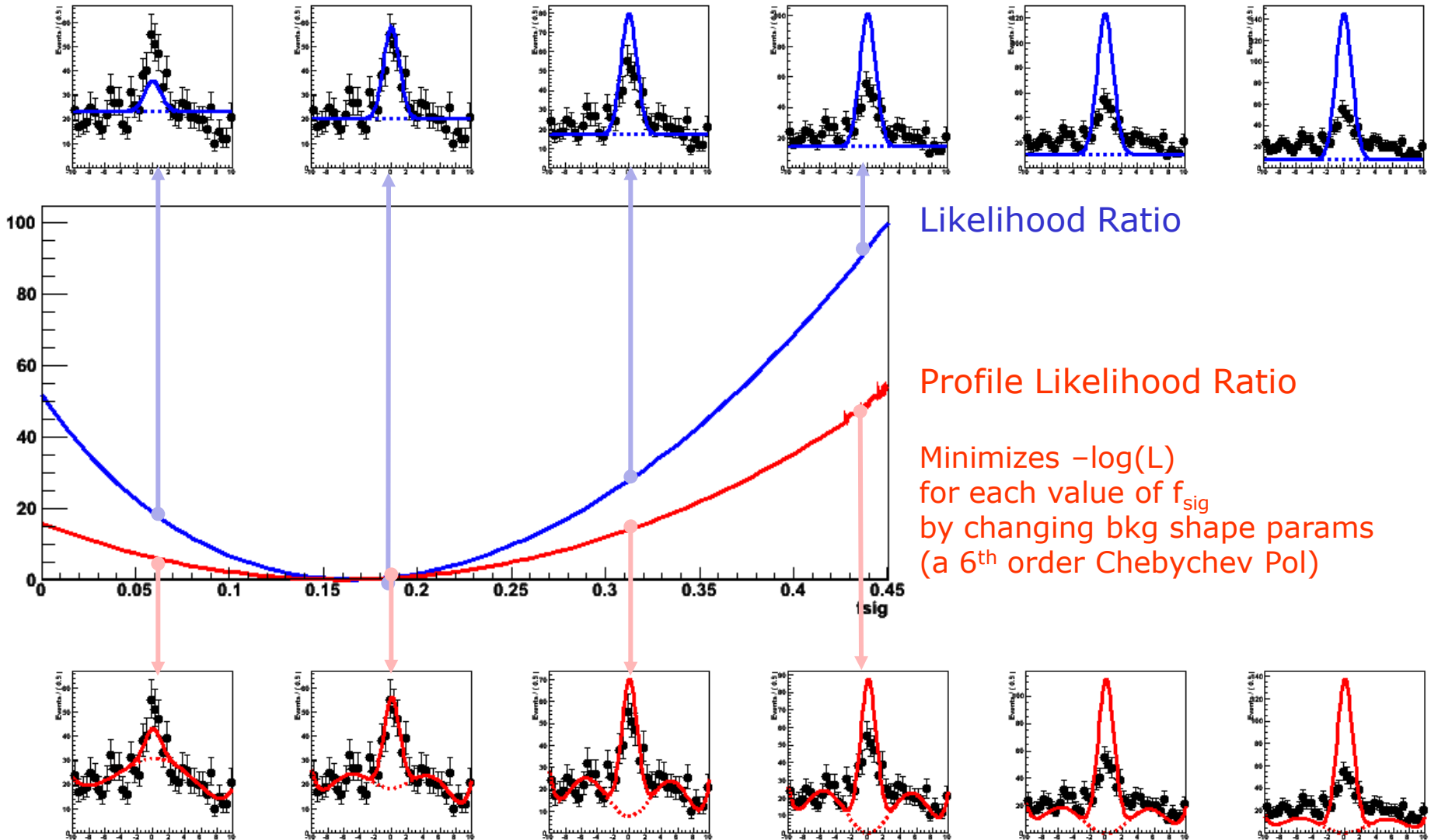

MLE fit fit data

-logLR(mean,sigma)   -logLR(mean,sigma)   -logPLR(mean)

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\sigma}}(\mu))}{L(\hat{\mu}, \hat{\sigma})}$$

best L(μ) for any value of s

best L(μ,σ)

Dealing with nuisance parameters in Likelihood ratio intervals

Likelihood Ratio

Profile Likelihood Ratio

Minimizes −log(L)
for each value of f_{sig}
by changing bkg shape params
(a 6th order Chebychev Pol)

Link between MINOS errors and profile likelihood

Parameter of interest

Nuisance parameter
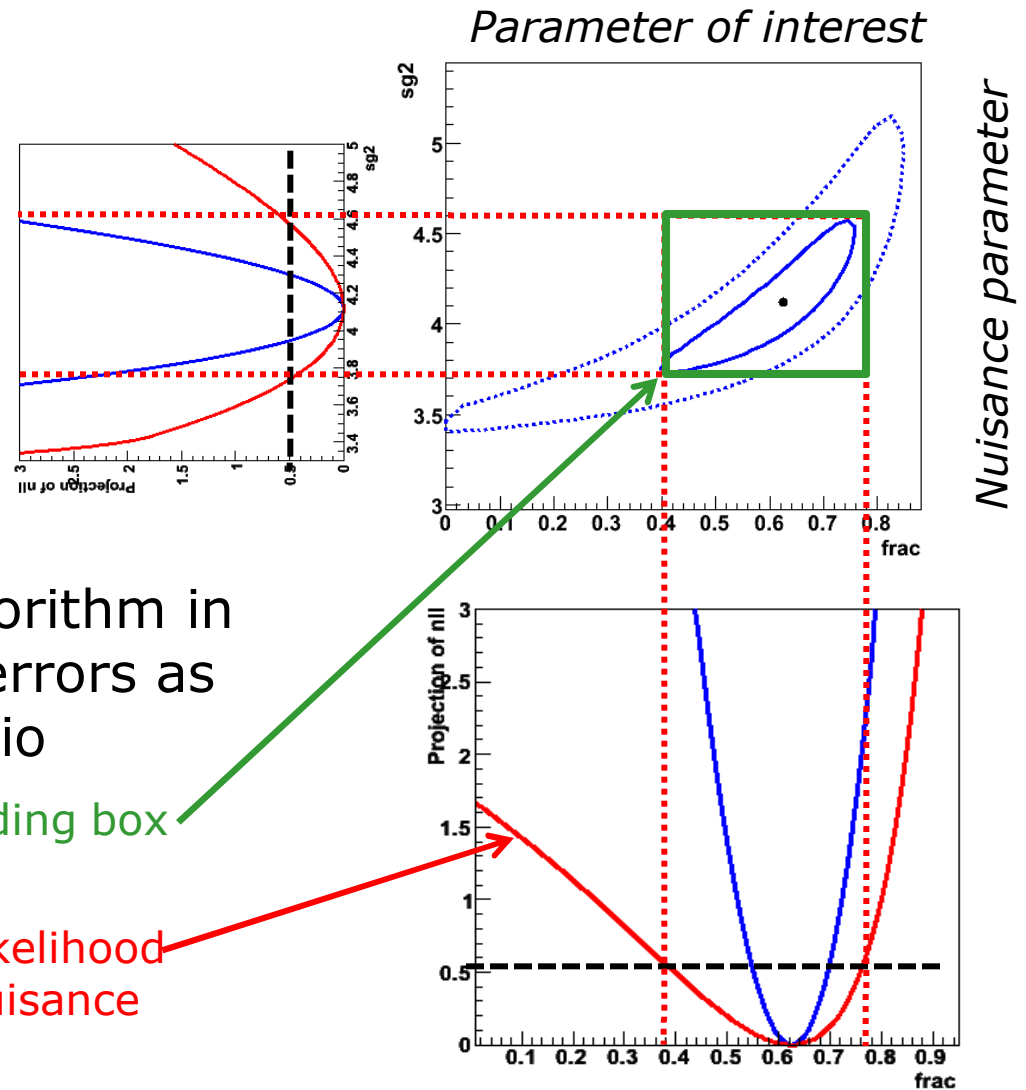
- Note that MINOS algorithm in MINUIT gives same errors as Profile Likelihood Ratio

  - MINOS errors is bounding box around $\lambda(s)$ contour

  - Profile Likelihood = Likelihood minimized w.r.t. all nuisance parameters

Dealing with nuisance parameters in Likelihood ratio intervals

- ## Issues with Profile Likelihood

  - By using best-fit value of the nuisance parameters corresponding to each value of the parameter of interest, this has a reputation of underestimating the true uncertainties.

  - In Poisson problems, this is partially compensated by effect due to discreteness of n, and profile likelihood (MINUIT MINOS) gives good performance in many problems.

- ## NB: Computationally Profile Likelihood is quite manageable, even with a large number of nuisance parameters

  - Minimize likelihood w.r.t. 20 parameters quite doable

  - Especially compared to numeric integration over 20 parameters, or constructing confidence belt in 20 dimensions…

Wouter Verkerke, NIKHEF

Recent comparisons results from PhyStat 2007

A Prototype Problem

BROOKHAVEN
NATIONAL LABORATORY

What is significance Z of an observation $x = 178$ events in a signal like region, if my expected background $b = 100$ with a 10% uncertainty?

- ‣ if you use the ATLAS TDR formula $Z_{5'} = 5.5$
- ‣ if you use Cousins-Highland $Z_N = 5.0$

The question seems simple enough, but it is not actually well-posed

- ‣ what do I mean by 10% background uncertainty?

Typically, we consider an auxiliary measurement y used to estimate background (Type I systematic)

- ‣ eg: a sideband counting experiment where background in sideband is a factor τ bigger than in signal region

$$L_P(x, y | \mu, b) = Pois(x | \mu + b) \cdot Pois(y | \tau b).$$
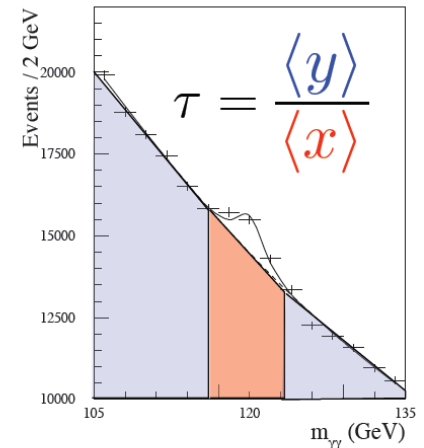
Example Sideband Measurement

BROOKHAVEN
NATIONAL LABORATORY

Sideband measurement used to extrapolate / interpolate the background rate in signal-like region

For now ignore uncertainty in extrapolation.

$$\tau = \frac{\langle y \rangle}{\langle x \rangle}$$

$$L_P(x, y | \mu, b) = Pois(x | \mu + b) \cdot Pois(y | \tau b).$$

Comparison of Methods for Prototype Problem

BROOKHAVEN
NATIONAL LABORATORY

In my contribution to PhyStat2005, I considered this problem and compared the coverage for several methods

‣ See Linnemann's PhyStat03 paper

Major results:

‣ Cousins–Highland result (Z_N) badly under–covers (only 4.2σ)!

  • rate of Type I error is 110 times higher than stated!

  • much less luminosity required
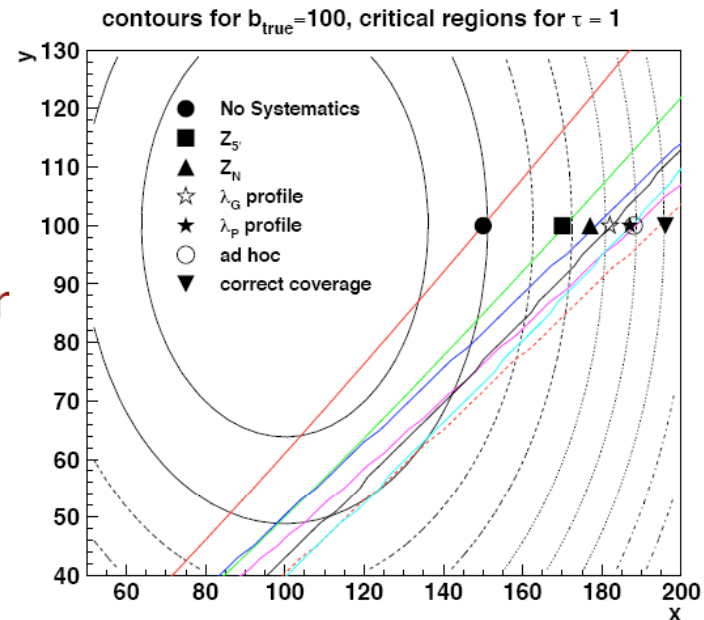
‣ Profile Likelihood Ratio (MINUIT/ MINOS) works great out to 5σ!

contours for b_{true}=100, critical regions for τ = 1

● No Systematics
■ $Z_{5'}$
▲ Z_N
☆ λ_G profile
★ λ_P profile
○ ad hoc
▼ correct coverage

Figure 7. A comparison of the various methods critical boundary $x_{crit}(y)$ (see text). The concentric ovals represent contours of L_G from Eq. 15.

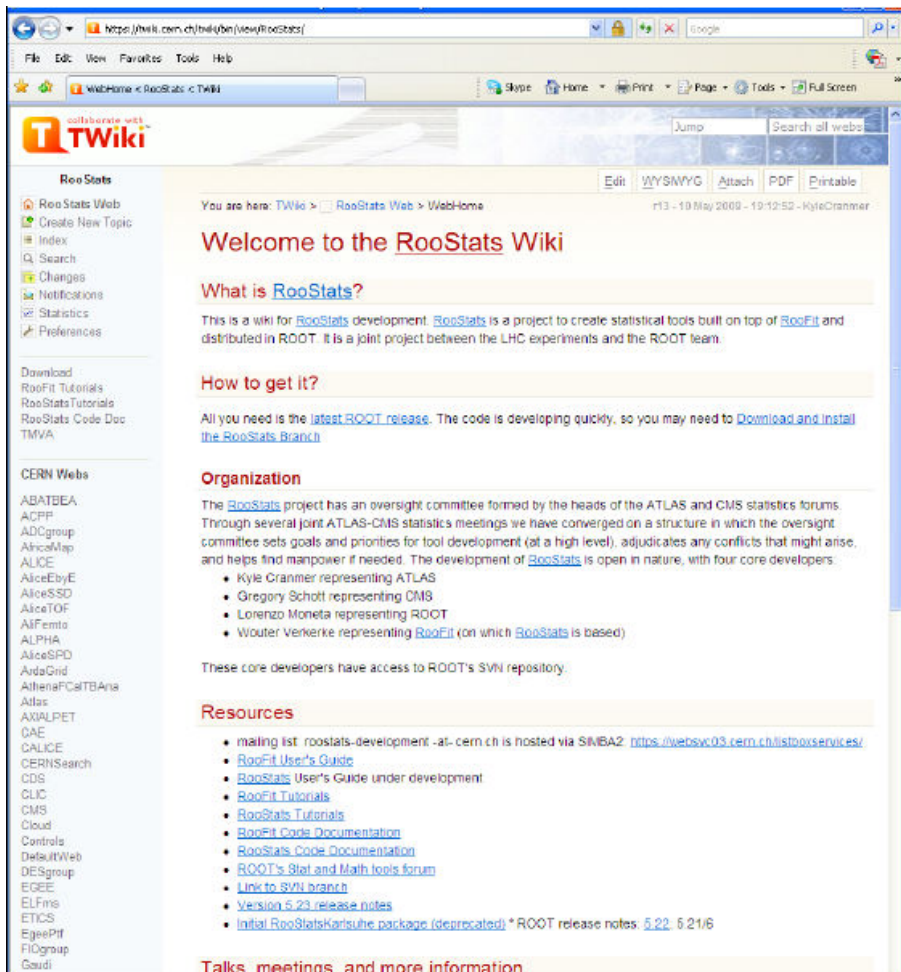| Method	L_G ($Z\sigma$)	L_P ($Z\sigma$)	$x_{crit}(y = 100)$
No Syst	3.0	3.1	150
$Z_{5'}$	4.1	4.1	171
Z_N (Sec. 4.1)	4.2	4.2	178
ad hoc	4.6	4.7	188
$Z_\Gamma = Z_{Bi}$	4.9	5.0	185
profile λ_P	5.0	5.0	185
profile λ_G	4.7	4.7	∼182

Hybrid Techniques: Introduction to Pragmatism

- Given the difficulties with all three classes of interval estimation, especially when incorporating nuisance parameters, it is common in HEP to relax foundational rigor and:

  - Treat nuisance parameters in a Bayesian way while treating the parameter of interest in a frequentist way, or

  - Treat nuisance parameters by profile likelihood while treating parameter of interest another way, or

  - Use the Bayesian framework (even without the priors recommended by statisticians), but evaluate the frequentist performance. In effect (as in profile likelihood) one gets approximate coverage while respecting the L.P.

Goal for the LHC a Few Years Ago

- Have in place tools to allow computation of results using a variety of recipes, for problems up to intermediate complexity:
  - Bayesian with analysis of sensitivity to prior
  - Frequentist construction with approximate treatment of nuisance parameters
  - Profile likelihood ratio (Minuit MINOS)
  - Other "favorites" such as LEP's CLS(which is an HEP invention)

- The community can then demand that a result shown with one's preferred method also be shown with the other methods, *and sampling properties studied.*

- When the methods all agree, we are in asymptopic nirvana.

- When the methods disagree, we learn something!
  - The results are answers to different questions.
  - Bayesian methods can have poor frequentist properties
  - Frequentist methods can badly violate likelihood principle

ATLAS/CMS/ROOT Project: RooStats built on RooFit

- *Core developers:*

- K. Cranmer (ATLAS)

- Gregory Schott (CMS)

- Wouter Verkerke (RooFit)

- Lorenzo Moneta (ROOT)

- Open project, all welcome to contribute.

- Included in ROOT production releases since v5.22, more soon to come

- Example macros in `$ROOTSYS/tutorials/roostats`

- RooFit *extensively documented, RooStats manual catching up, code doc in ROOT.*

RooStats Project – Example

- Create a model - Example

$$Poisson(x\,|\,s\cdot r_s + b\cdot r_b)\cdot Gauss(r_s,1,0.05)\cdot Gauss(r_b,1,0.1)$$

Create workspace with above model (using factory)

```
RooWorkspace* w = new RooWorkspace("w");
w->factory("Poisson::P(obs[150,0,300],
                    sum::n(s[50,0,120]*ratioSigEff[1.,0,2.],
                        b[100,0,300]*ratioBkgEff[1.,0.,2.]))");
w->factory("PROD::PC(P, Gaussian::sigCon(ratioSigEff,1,0.05),
                        Gaussian::bkgCon(ratioBkgEff,1,0.1))");
```

Contents of workspace from above operation

```
RooWorkspace(w) w contents

variables
---------
(b,obs,ratioBkgEff,ratioSigEff,s)

p.d.f.s
-------
RooProdPdf::PC[ P * sigCon * bkgCon ] = 0.0325554
  RooPoisson::P[ x=obs mean=n ] = 0.0325554
    RooAddition::n[ s * ratioSigEff + b * ratioBkgEff ] = 150
  RooGaussian::sigCon[ x=ratioSigEff mean=1 sigma=0.05 ] = 1
  RooGaussian::bkgCon[ x=ratioBkgEff mean=1 sigma=0.1 ] = 1
```

e, NIKHEF

RooStats Project – Example

- Confidence intervals calculated with model
  - Profile likelihood
    ```
    ProfileLikelihoodCalculator plc;
    plc.SetPdf(w::PC);
    plc.SetData(data); // contains [obs=160]
    plc.SetParameters(w::s);
    plc.SetTestSize(.1);
    ConfInterval* lrint = plc.GetInterval(); // that was easy.
    ```

  - Feldman Cousins
    ```
    FeldmanCousins fc;
    fc.SetPdf(w::PC);
    fc.SetData(data); fc.SetParameters(w::s);
    fc.UseAdaptiveSampling(true);
    fc.FluctuateNumDataEntries(false);
    fc.SetNBins(100); // number of points to test per parameter
    fc.SetTestSize(.1);
    ConfInterval* fcint = fc.GetInterval(); // that was easy.
    ```

  - Bayesian (MCMC)
    ```
    UniformProposal up;
    MCMCCalculator mc;
    mc.SetPdf(w::PC);
    mc.SetData(data);  mc.SetParameters(s);
    mc.SetProposalFunction(up);
    mc.SetNumIters(100000); // steps in the chain
    mc.SetTestSize(.1); // 90% CL
    mc.SetNumBins(50); // used in posterior histogram
    mc.SetNumBurnInSteps(40);
    ConfInterval* mcmcint = mc.GetInterval();
    ```

RooStats Project – Example

- Retrieving and visualizing output

  ```
  double fcul = fcint->UpperLimit(w::s);
  double fcll = fcint->LowerLimit(w::s);
  ```


Profile Likelihood Ratio and Posterior for S
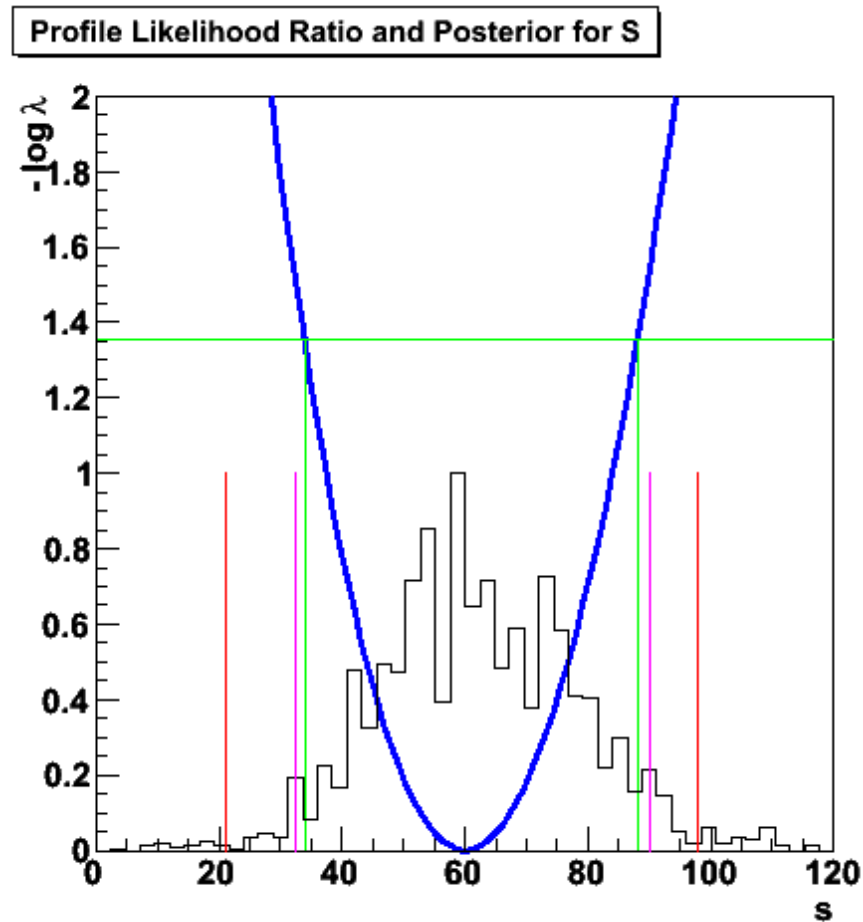
RooStats Project – Example

- ## Some notes on example

  - Complete working example (with output visualization) shipped with ROOT distribution (*$ROOTSYS/tutorials/roofit/rs101_limitexample.C*)

  - **Interval calculators make no assumptions on internal structure of model**. Can feed model of arbitrary complexity to same calculator (computational limitations still apply!)
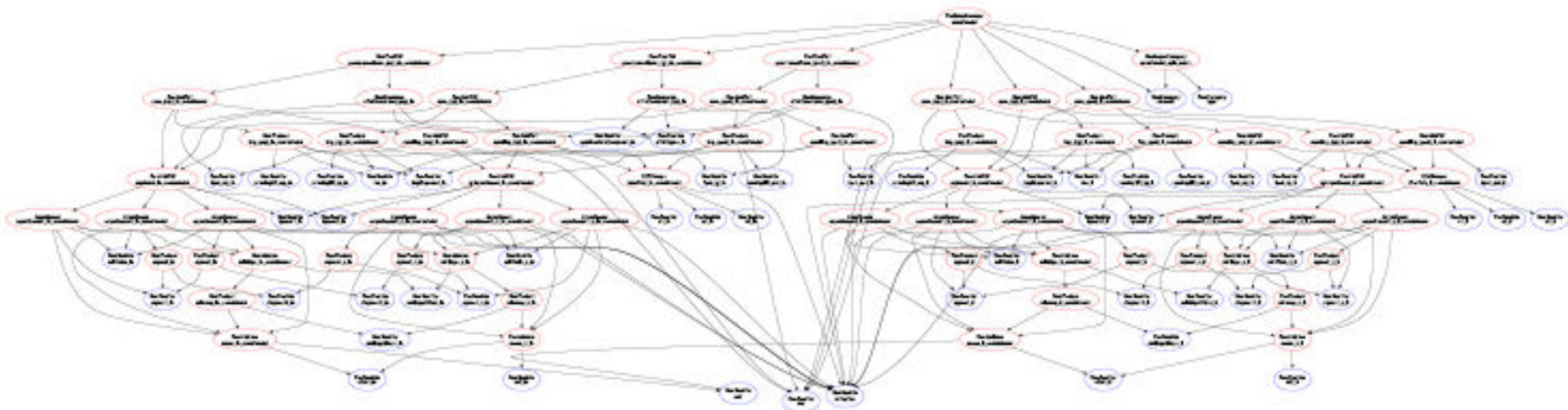
'Digital' publishing of results

- A likelihood may be considered the ultimate publication of a measurement

- Interesting to be able to digitally publish actual likelihood rather than

  - Parabolic version (i.e. you publish your measurement and an error)

  - Some parameterized form. Cumbersome in >1 dimension. No standard protocol for exchanging this time of information

- This is trivially possible with RooFit/RooStats

  - Many potential applications, e.g. combining of Higgs channels,
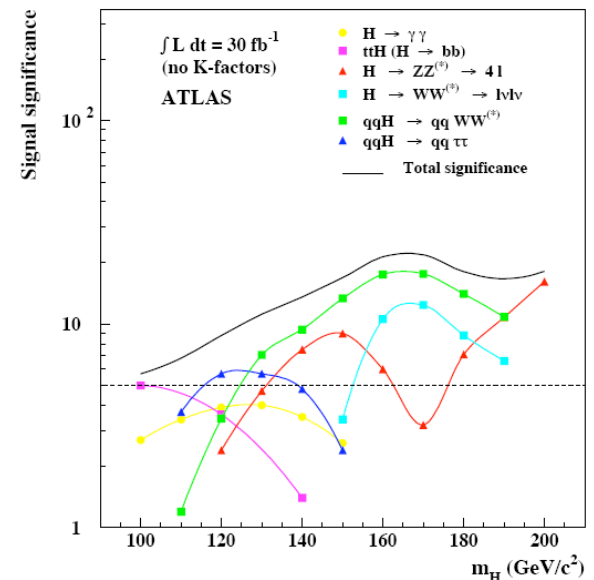
Using persisted p.d.f.s.

- Using both model & p.d.f from file


```
TFile f("myresults.root") ;
RooWorkspace* w = f.Get("w") ;
```

Make plot of data and p.d.f
```
RooPlot* xframe = w::x.frame() ;
w::d.plotOn(xframe) ;
w::g.plotOn(xframe) ;
```
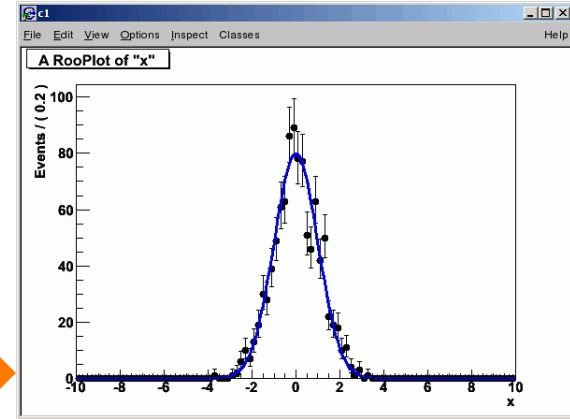
Construct likelihood & profile LH
```
RooAbsReal* nll = w::g.createNLL(w::d)
RooAbsReal* pll = nll->createProfile(w::mean) ;
```
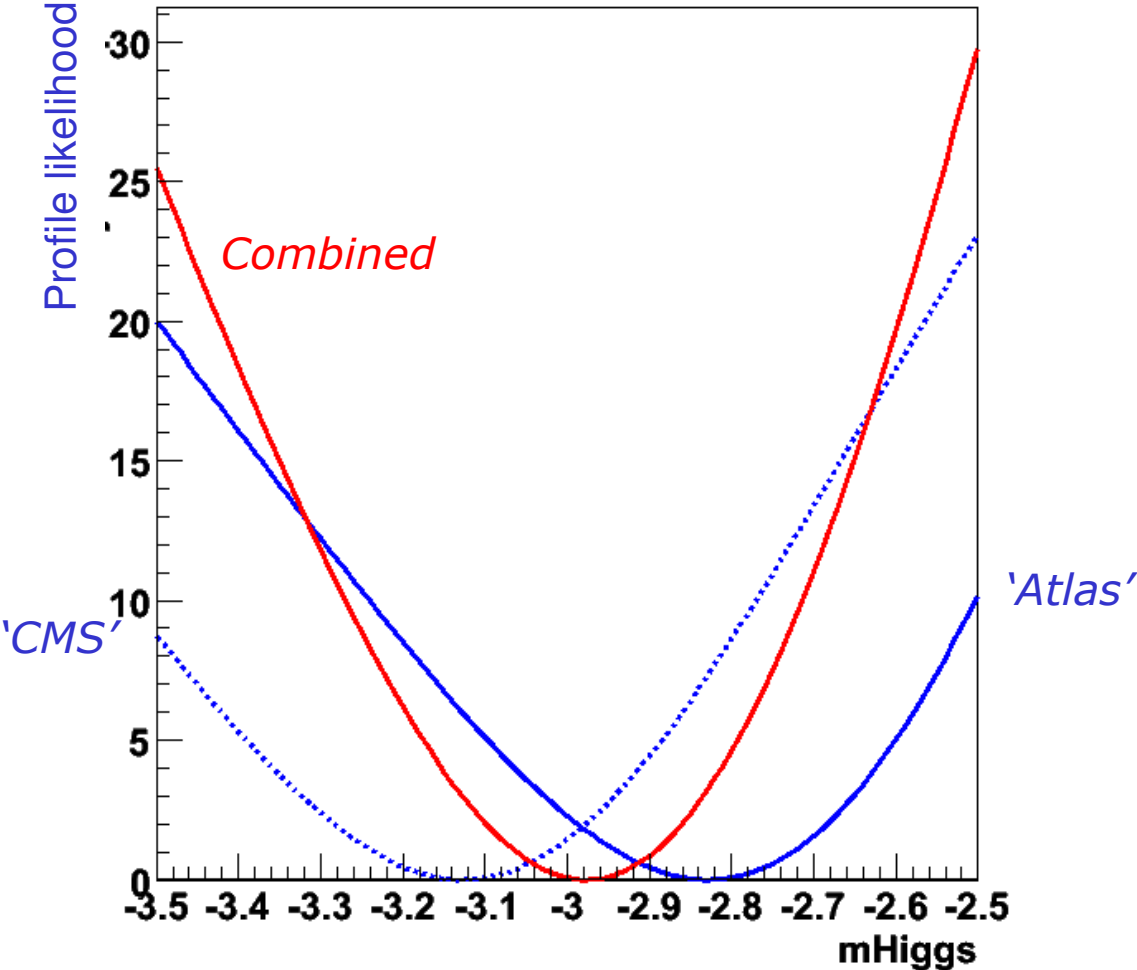
Draw profile LH
```
RooPlot* mframe = w::m.frame(-1,1) ;
pll->plotOn(mframe) ;
mframe->Draw()
```
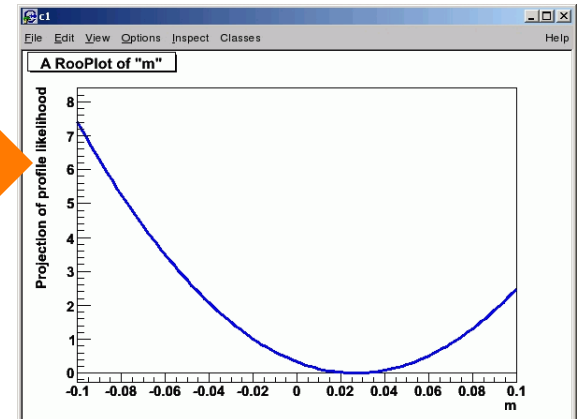

– **Note that above code is independent of actual p.d.f in file → e.g. full Higgs combination would work with identical code**

A toy combination example

Work in progress on realistic Higgs limit combination

Combining the inputs

Using the same code as last time, with a few extra lines for the new channels, we arrive at the combined dataset & model

‣ here the only common parameter is mu, the master signal strength
  · could easily make Higgs mass be the same for all three channels
‣ the combined model has 27 nuisance parameters
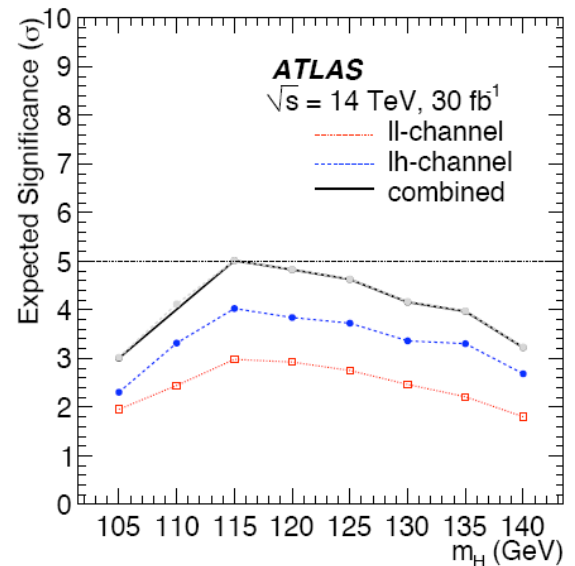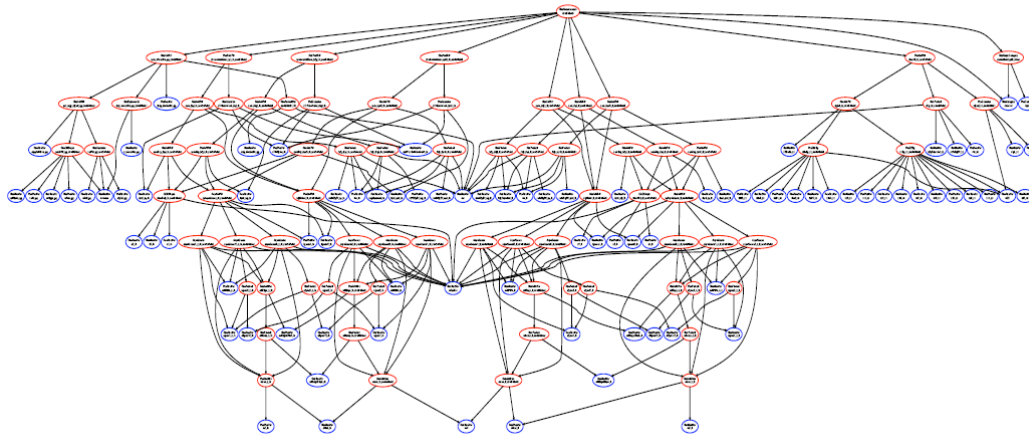
The end – Recommended reading

- Easy
  - R. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley, 1989
  - L. Lyons, *Statistics for Nuclear and Particle Physics*, Cambridge University Press
  - Philip R. Bevington and D.Keith Robinson, *Data Reduction and Error Analysis for the Physical Sciences*
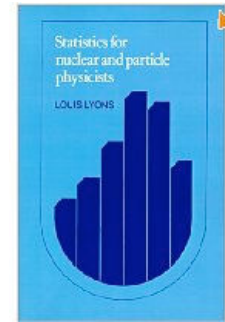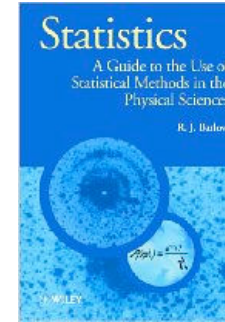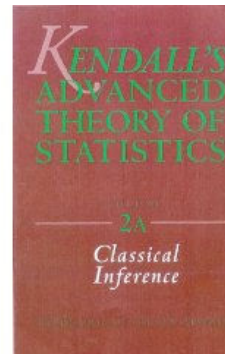
- Intermediate
  - Glen Cowan, Statistical Data Analysis (Solid foundation for HEP)
  - Frederick James, *Statistical Methods in Experimental Physics, World Scientific,* 2006. (This is the second edition of the influential 1971 book by Eadie et al., has more advanced theory, many examples)
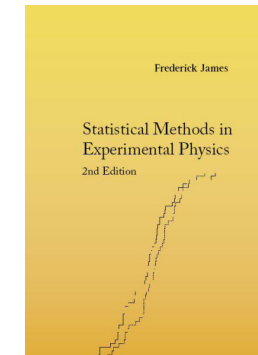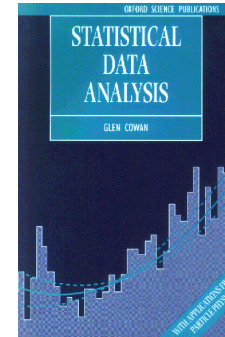
- Advanced
  - A. Stuart, K. Ord, S. Arnold, Kendall's Advanced Theory of Statistics, Vol. 2A, 6th edition, 1999; and earlier editions of this "Kendall and Stuart" series. (Authoritative on classical frequentist statistics; anyone contemplating a NIM paper on statistics should look in here first!)

- PhyStat conference series:
  - Beginning with Confidence Limits Workshops in 2000, links at http://phystat-lhc.web.cern.ch/phystat-lhc/ and http://www.physics.ox.ac.uk/phystat05/

Wouter Verkerke, NIKHEF