Discovery, exclusion, nuisance parameters

— Formulating discovery, exclusion
— Dealing with syst. uncertainties as nuisance parameters

Wouter Verkerke, UCSB

Introduction

- In the previous module extensive focus on fundamental statistics and the meaning of confidence interval and limits, i.e. what does a statement "We exclude SuperSymmetry at 95% C.L." precisely mean
- Today we explore other issues that are of crucial importance to be able to make such statement on real-life theories
  - Models can have uncertainties (in the form of unconstrained or weakly constrained parameters) → How do we incorporate that in our statement
  - Making statements about what we *expect* from both H0(SUSY+SM) and H1(SM) in addition to what we observe (especially relevant for theories with uncertainties we'll generally expect a range of possibilities)
  - Incorporating p-value of the alternate hypothesis in the conclusion. Example: If your data says P(SM+SUSY)=0.001 we 'exclude SUSY' but if p(SM)=0.003 on that same data, should we believe our conclusion?

Theories with uncertainties - Nuisance parameters

- Have so far considered problems with one model parameter

- Hypothetical case for “SuperSymmetry” discovery
  - Simulation for SM – Predicts 3 events (Poisson, μ exactly known)
  - Simulation for SUSY – Predicts 6 events → 9 events in total
  - Observed event count in data: 8 events
- How do you conclude (or not) that you’ve discovered supersymmetry?
  - You expect 9 events (with SUSY), you see 8, looks promising

- Discussed three types of solution to above problem.
- What do we do if background is **not** exactly known?
  - E.g. $\mu = 3.0 \pm 1.0$ (NB: this statement does not unique fix $P(\mu)$)

Nuisance parameters

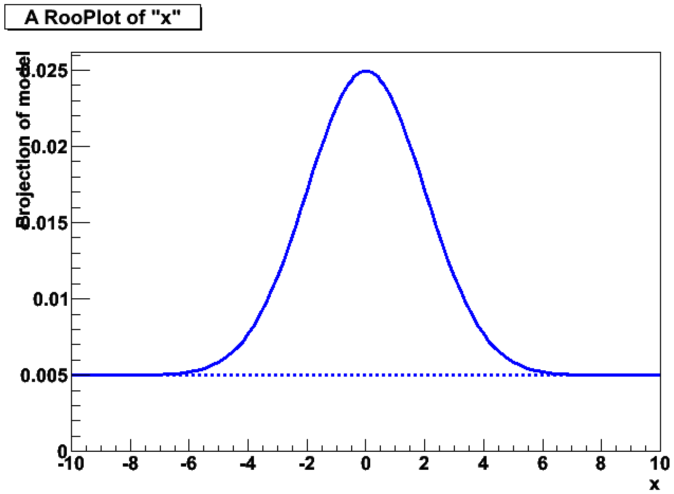
- In real life, background rate, shape of background model are usually not *exactly* known
  - Need procedure to incorporate uncertainty on these ‘nuisance parameters’ into account when setting limits etc.
- For preceding problems (with precisely defined null hypotheses) procedures exist to calculate intervals and significances could be exactly
- When dealing with nuisance parameters, this generally not possible anymore
- Q: Is that a problem?
  - A: Yes. If your (approximate) 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

Treatment of nuisance parameters

- 1 – Definition of nuisance parameters
  - A nuisance parameter is any parameter of the model that is not a parameter-of-interest (for physics).
    - Example: for Higgs discovery N(higgs) is of interest, everything else is nuisance
- 2 – Introduction of nuisance parameters in Likelihood
  - Sometimes nuisance parameter arise naturally in the likelihood.
  - Systematic uncertainties always introduce nuisance parameters, but explicit parameterization not always obvious (e.g. how to parameterize effect of Pythia-vs-Herwig?)
- 3 – Treatment of nuisance parameters in inference
  - Each of the three main classes of constructing intervals (Bayesian, likelihood ratio, Neyman confidence intervals) has a different way to incorporate the uncertainty on the nuisance parameters in the parameters of interest.

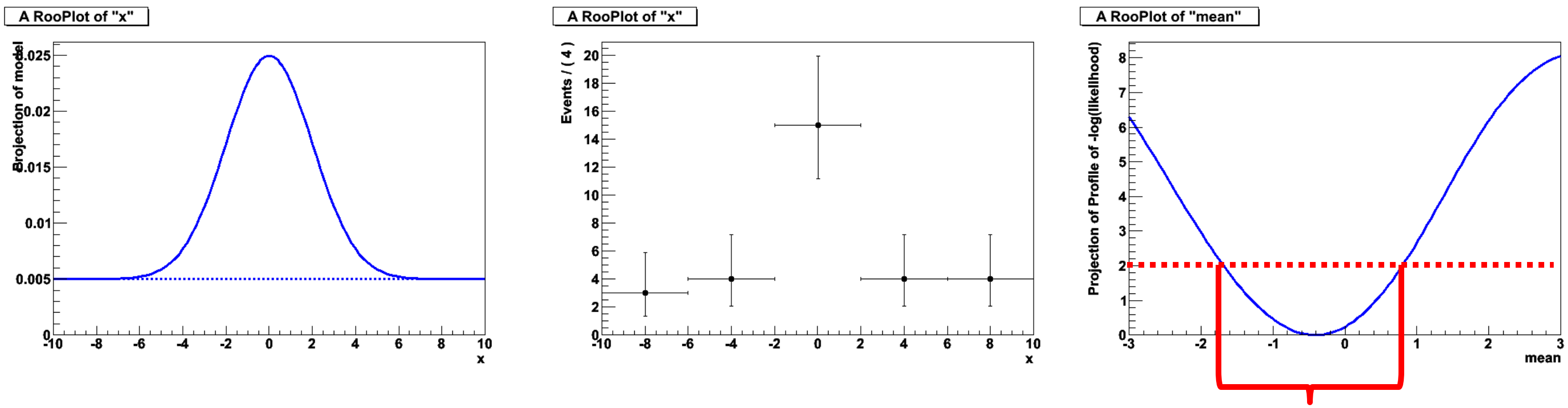
Likelihood fit – Definition of nuisance parameters

- In ML fits, any floating fit parameter that is not the parameter of interest is a *nuisance* parameter
- Model = Nsig*Gauss(x,m,s)+Nbkg*Uniform(x)
  - N_{sig} is parameter of interest
  - m,s,N_{bkg} are nuisance parameters, if not exactly known/fixed
  - Uncertainty on nuisance parameters will increase uncertainty on parameter of interest Nsig
  - In this example, the nuisance parameters can be constrained from the data along with the parameter of interest (given a sufficiently large dataset) → Even without any prior knowledge on m,s,N_{bkg} one is still capable of making a statement on Nsig

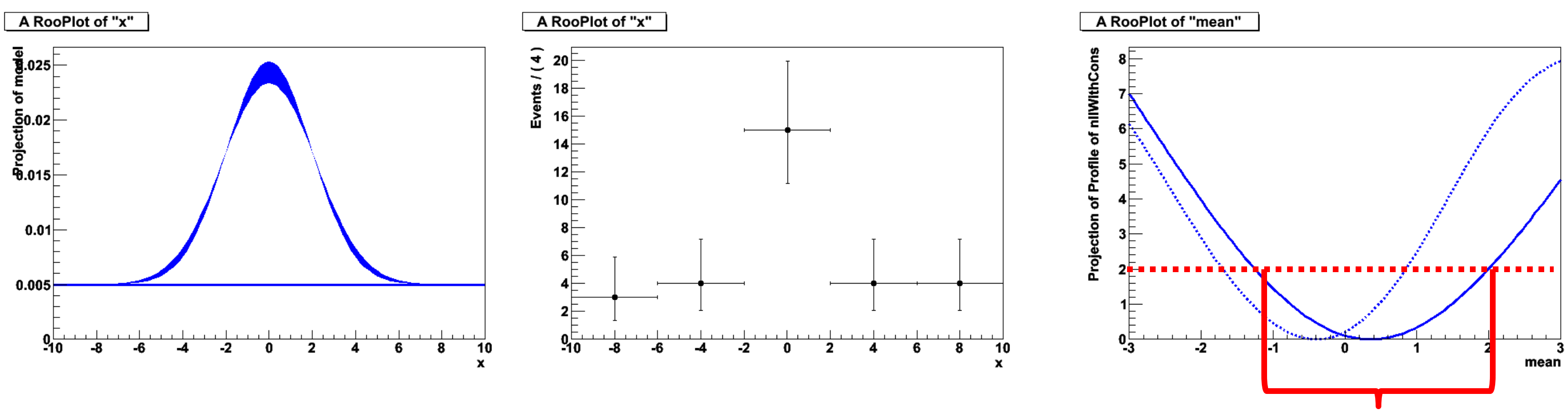


Adding uncertainties to a likelihood

- Example 1 – Width known exactly



- Example 2 – Gaussian uncertainty on width



Counting with sideband – Nuisance parameters

- In other examples, a the nuisance parameter may *not* be constrainable from the data
  - An 'auxiliary measurement/constraint' term must be introduced to define the magnitude of the uncertainty of the NP so that a statement can be made on the POI.
- Example:
  counting experiment with sideband: $\text{Poisson}(N_{sig}|s+b)$
  - Must have some external information on b to be able to do measurement
- Example of external constraint on b:
  - We have a control region where we measure background only.

    Model: $\text{Poisson}(N_{sig}|s+b) \cdot \text{Poisson}(N_{ctl}|\tau \cdot b)$
  - Measurement now consistent of two numbers: N_{sig}, N_{ctl}
  - NB: Mathematically and conceptual identical to concept of 'simultaneous measurements' discussed earlier

Constraining nuisance parameters

- Also in cases where the nuisance parameters can be measured from the data, it is possible that external information exists that is more constraining the inference from the data sample
  - Can incorporate this in the same form as 'auxiliary measurement'
- Ex: Model = Nsig*Gauss(x,m,s)+Nbkg*Uniform(x)
- → $-\log L = (-\sum_{data} \mathrm{Model}(x_i,N_{sig},N_{bkg},m,s)) - \log C(m,\ldots)$
- Typical shape of C is Gaussian, or Poisson
- Note notational convention difference in C(m) for Frequentist and Bayesian formalism
  - Freq: C = Gaussian(y,m,s) 'auxiliary measurement in observable y'
  - Bayes: C = Gaussian(m,m_0,s) 'prior on m'
  - Note that for a Gaussian shape both are mathematically equivalent as Gauss(x,m,s) = Gauss(m,x,s), but this is not necessarily true for other shapes

Shape of auxiliary measurement likelihood

- Shape of auxiliary measurements requires some careful thought – especially when evaluating high Z limits
- Option A: Rescaled Poisson: Poisson($N|\tau\cdot b$)
  - Most suitable if uncertainty on B is dominated by statistical uncertainty from a sideband or control region
  - (It is the exact solution for a counting measurement in a sideband)
- Option B: Gaussian: Gauss(b,b_0,σ_b) or Gauss(b_0,b,σ_b)
  - Usually chosen if source information is known in form $b_0 \pm \sigma_b$
  - Also often chosen if true shape is unknown (e.g. 'theory uncertainty')
  - Central Limit Theorem → Sum of many uncertainties is asymptotically Gaussian
  - But beware of relatively large Gaussian uncertainties
    → These can result in optimistically biased significance calculations

Shape of auxiliary measurement likelihood

- Option B: Gaussian: Gauss(b,b_0,σ_b) [continued]
  - Illustration of danger of large Gaussian uncertainties

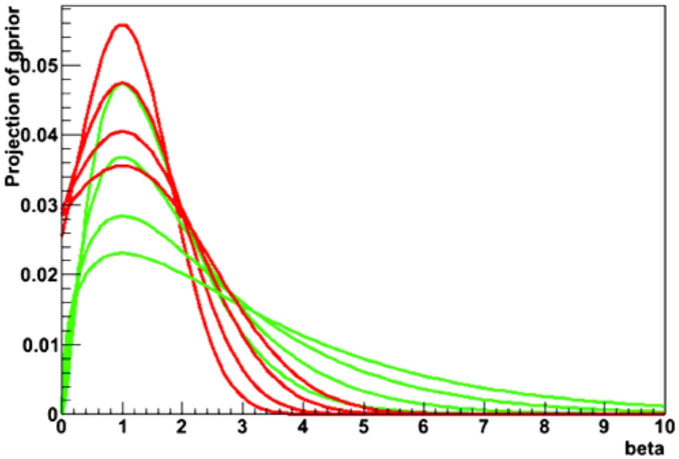
    Model = Poisson($N_{sig}|s+b$) · Gaussian(b,b_0,σ_b)
    with $\mathbf{b_0 = 3, \sigma_b = 1}$ (33%)

    If we look at 5σ fluctuations we in principle allow the Gaussian term to move 5σ off its center
    → Allow downward fluctuation to b=-2 !

    In reality b must be greater than zero
    → **Significance of result will be optimistically biased**

- Option C: Gamma(b,b_0,σ_b)

$$f(x;k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\,\Gamma(k)} \text{ for } x \geq 0 \text{ and } k,\theta > 0.$$



  - Longer positive tail than Gaussian
  - Better behavior at 0 than Gaussian
  - Asymptotically Gaussian
  - Good ‘alternate model’ for systematics

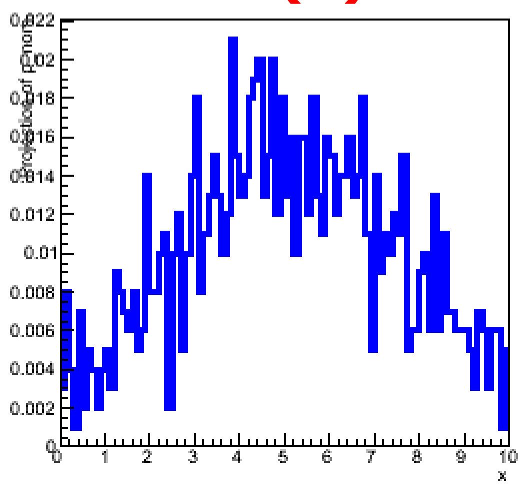
Turning *all* uncertainties into nuisance parameters

- For 'simple' measurements systematic uncertainties are traditionally included 'a posteriori' with error propagation
  - E.g. measure cross-section using Herwig, then again using Pythia and use the difference in these cross-section values for the fragmentation uncertainty on the cross-section (and add all such systematic uncertainties in quadrature)
- For fundamental techniques (e.g. Bayesian, Likelihood ratio) techniques these sources must be incorporated in the likelihood
  - No accurate 'a posteriori' prescription exists to include these
  - Inclusion ensures consistent treatment of these systematics as nuisance parameters in inference analysis (limit or confidence interval)
- But certain types of systematics are difficult to parameterize...

Parameterized template pdfs

- In many LHC analyses (e.g. Higgs searches, top cross-section), shapes are defined by histograms rather than analytical functions. While these shapes are fixed (no parameters), nuisance parameters can be introduced *a posteriori* through 'morphing techniques'
- Original (fixed) shape for e.g. the signal in an observable x → Histogram defines shape of probability density function
  - But this distribution, obtained from simulation has many systematic uncertainties associated with it (originating from the simulation → How do you turn these into nuisance parameters

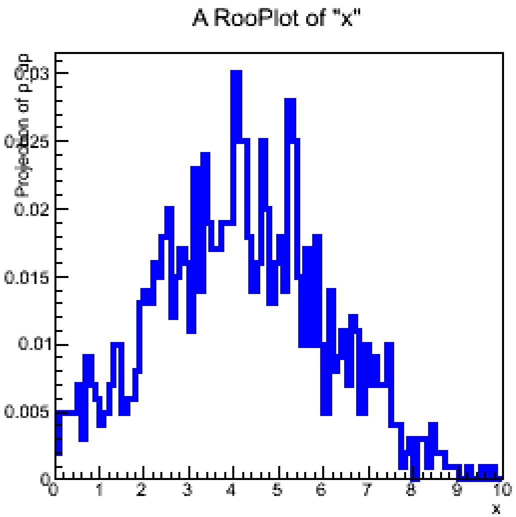
F(x)



Parameterized template pdfs

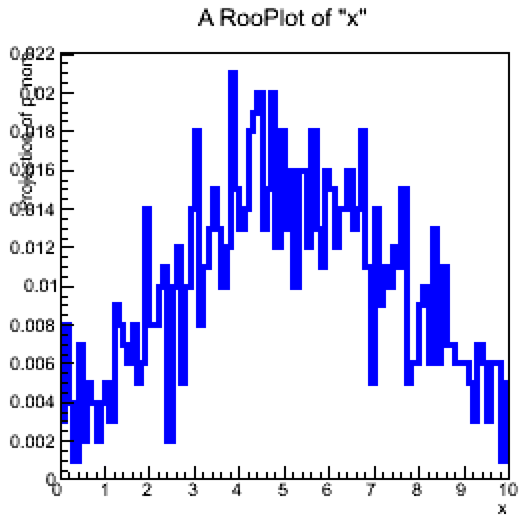
- First quantify effect by regenerating histogram from the source (e.g. full simulation) with a source of systematic uncertainty set to a shifted value
- Example: Jet Energy scale up / down by X%

distribution for JES at -1σ



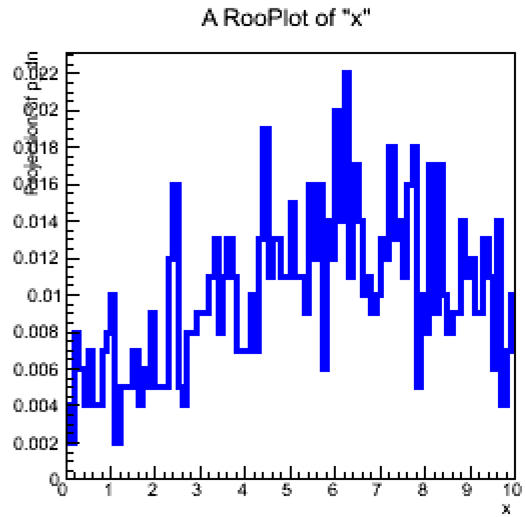
$F_-(x)$

nominal distribution



$F_0(x)$

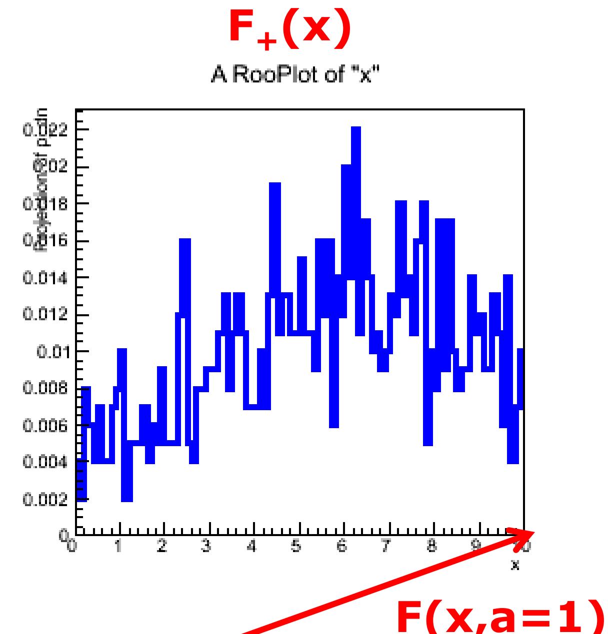
distribution for JES at $+1\sigma$

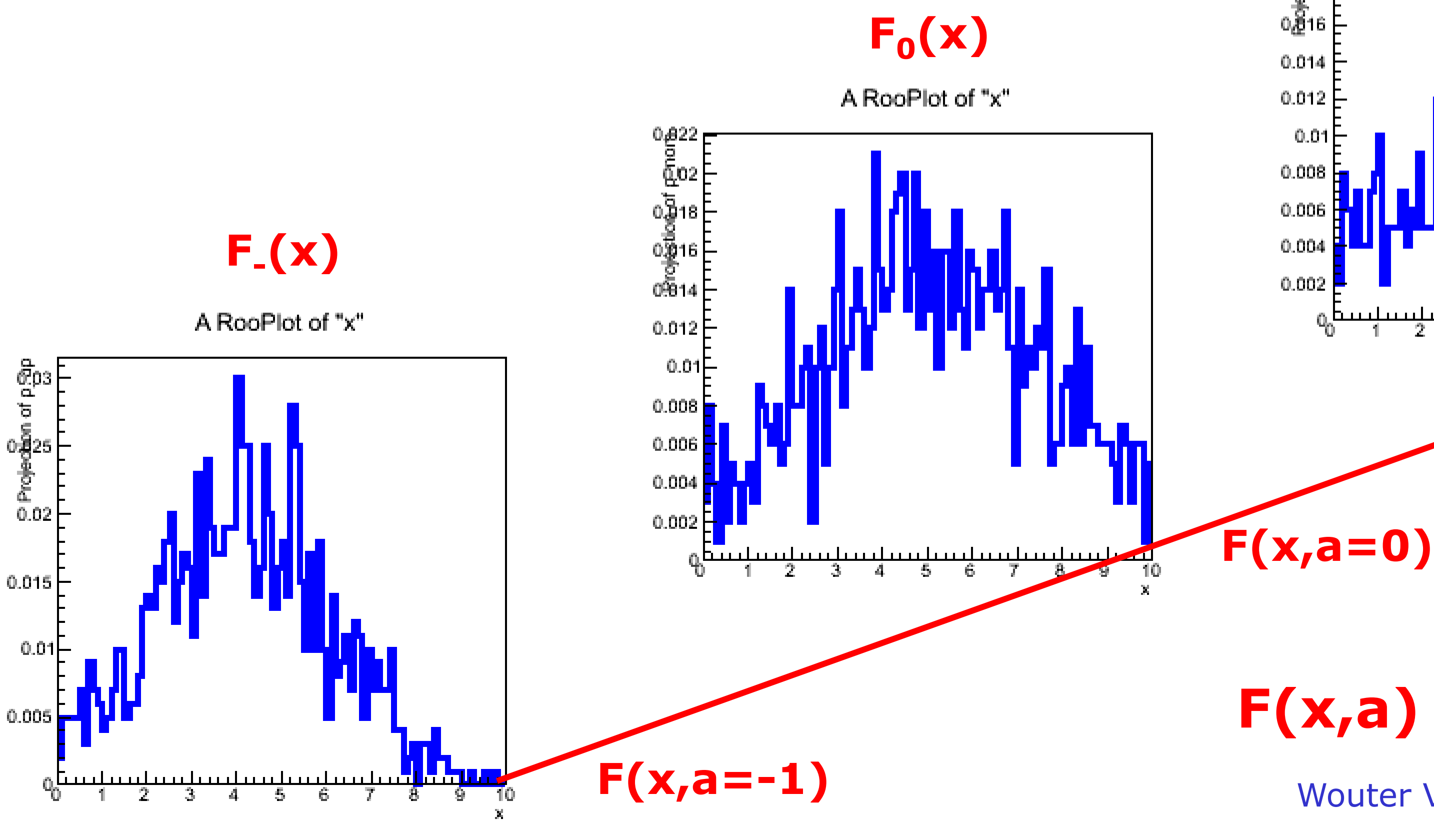


$F_+(x)$

Parameterized template pdfs

- Final step is to make a pdf that interpolates (bin-by-bin) between the histograms introducing a newly introduced nuisance parameter
  - Constrain size of systematic by introducing 'constraint' on a in likelihood





Parameterized template pdfs

- Final step is to multiply likelihood with constraint that defines magnitude of uncertainty as 'one sigma'

$$-\log L = (-\Sigma_{data} F(x_i,a)) - \log \mathrm{Gauss}(a,0,1)$$

- Can repeat this procedure for any number of systematic uncertainties

$$-\log L = (-\Sigma_{data} F(x_i,a,b,\ldots)) - \log G(a,0,1)\, G(b,0,1)\ldots$$

- Note that data may also constrain magnitude of NP
  - In such cases the uncertainty on a,b will be less than 1

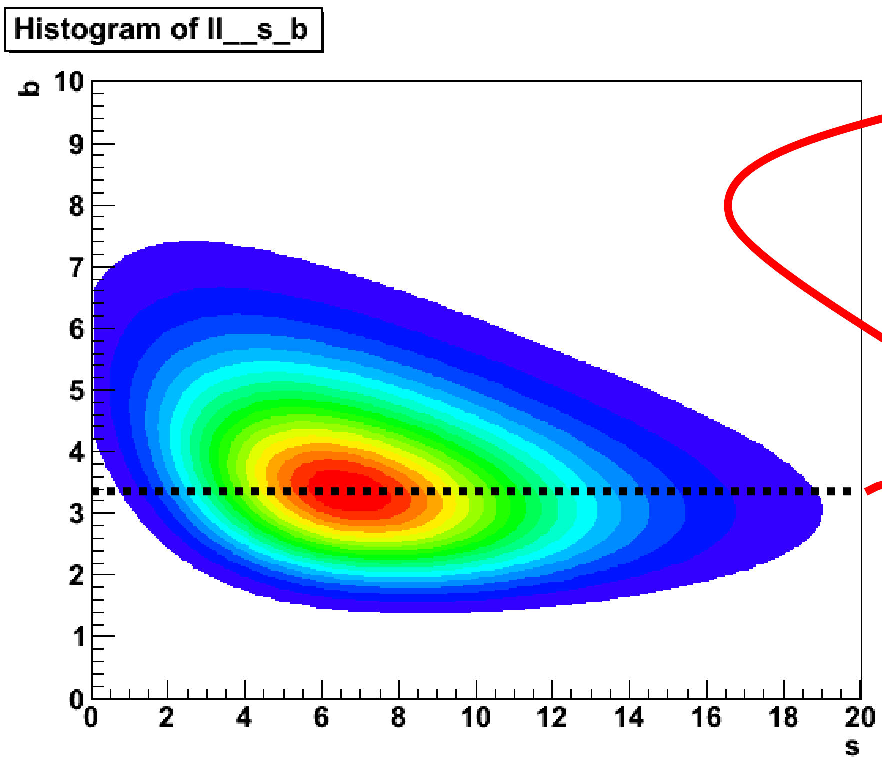
Treatment of nuisance parameters

- Effort so far has been to incorporate systematic uncertainties as explicit nuisance parameters in model
  - In analytical pdfs, the free parameters of these models are the nuisance parameters
  - In template-based pdfs, parameters can be introduced by morphing/interpolation techniques
  - In either case the magnitude of the uncertainty represented by the NP can be constraint with an 'auxiliary measurement' type of constraint in the likelihood
    "$L(s,b) = \text{Poisson}(N_{sig}|s+b) \cdot \textbf{Gaussian}(b,b_0,\sigma_b)$"
- The next step is to include the effect of all these nuisance parameters on the statistical inference on the parameter-of-interest
- Will first discuss procedure in each of the three 'fundamental' approaches

Counting with sideband – Nuisance parameters

- Model: Poisson(N_{sig}|s+b)Poisson(N_{ctl}|τ·b), τ=3 (exact)

- Visualization of Likelihood
  - Nsig=10, Nctl=10

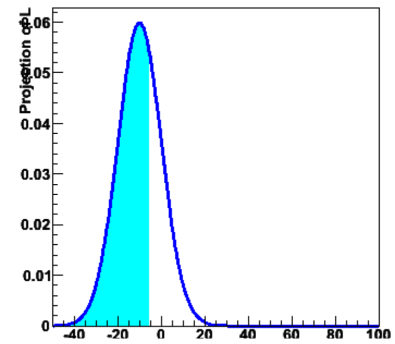


We know how to set interval on s ***given*** a fixed b

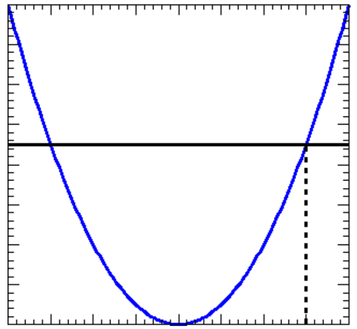
Now need to incorporate uncertainty on b…

Reminder: intervals defined ways

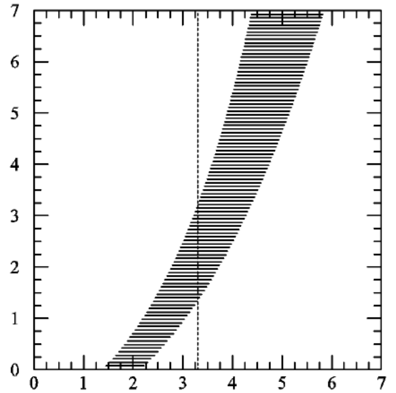
- Bayesian interval
  at 90% credibility:
  find μ_u such that posterior probability $p(\mu>\mu_u) = 0.1$.



- Likelihood ratio method for approximate 90% C.L. U.L.:
  find μ_u such that $L(\mu_u) / L(3)$ has prescribed value.
  - Asymptotically identical to Frequentist interval (Wilks theorem)
  - **Equivalent to MINOS errors**



- Frequentist one-sided 90% C.L. upper limit: find μ_u such that $P(n \leq 3 \mid \mu_u) = 0.1$.



Dealing with nuisance parameters in Bayesian intervals

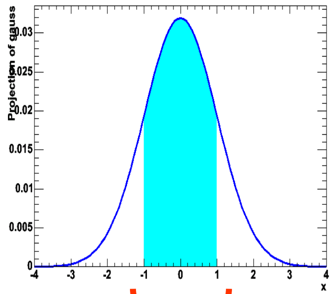
- Reminder: definition of Bayesian intervals

$$p(\mu|x_0) \propto L(x_0|\mu)\ p(\mu),$$

where:

- $p(\mu|x_0)$ = posterior pdf for μ, given the results of this experiment
- $L(x_0|\mu)$ = Likelihood function of μ from the experiment
- $p(\mu)$ = prior pdf for μ,

$p(\mu|x_0)$



Area that integrates X% of posterior

- If you have nuisance parameters **a**, equation becomes
- $p(\mu,\mathbf{a}|x_0) \propto L(x_0|\mu,\mathbf{a})\ p(\mu)\ p(\mathbf{a})$

Dealing with nuisance parameters in Bayesian intervals

- Elimination of nuisance parameters in Bayesian interval: Integrate over the full subspace of all nuisance parameters;

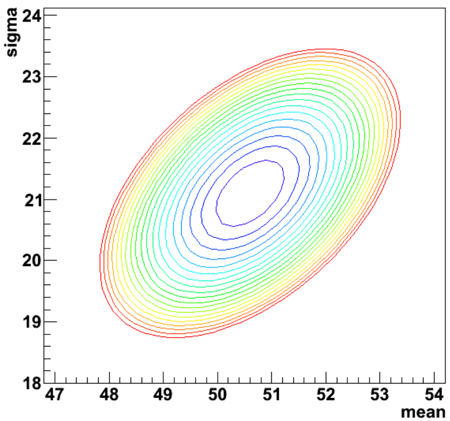
$$p(s \mid x) = \int \left(L(s, \vec{b})\, p(s, \vec{b}) \right) d\vec{b}$$

- You are left with the posterior pdf for the parameter of interest.

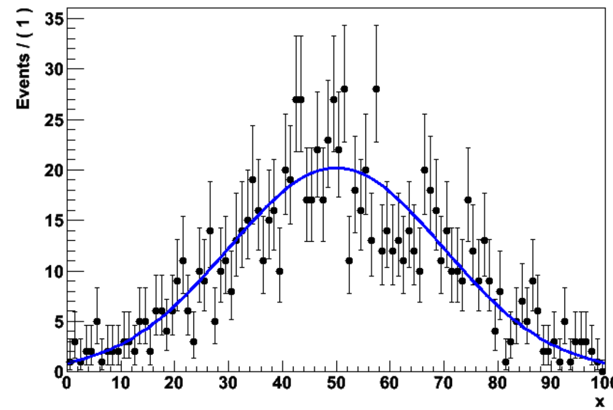
Illustration of nuisance parameters in Bayesian intervals

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

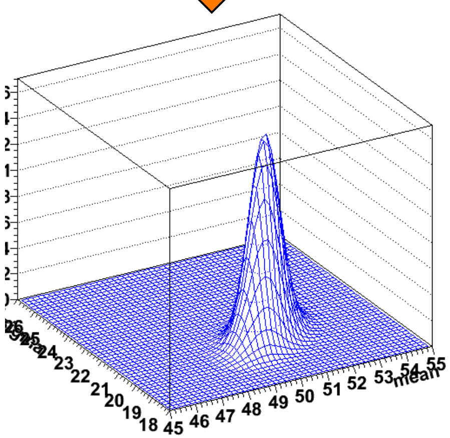
-logLR(mean,sigma)



MLE fit fit data

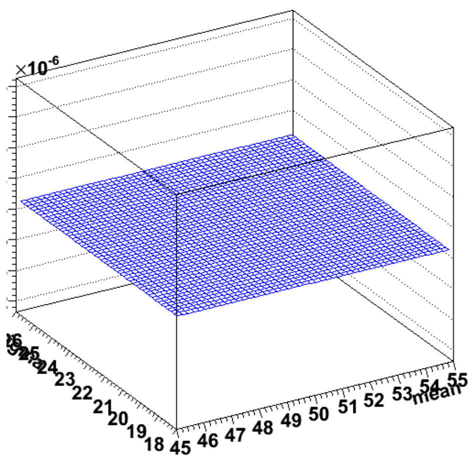


∫



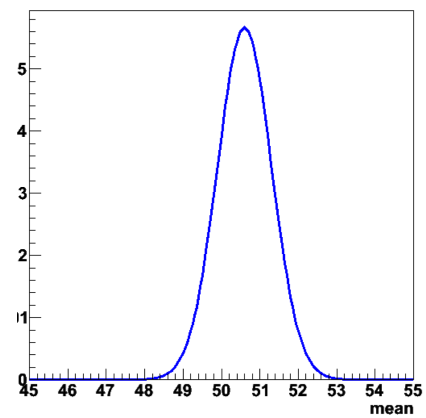
LR(mean,sigma)

×



prior(mean,sigma)

=



posterior(mean)

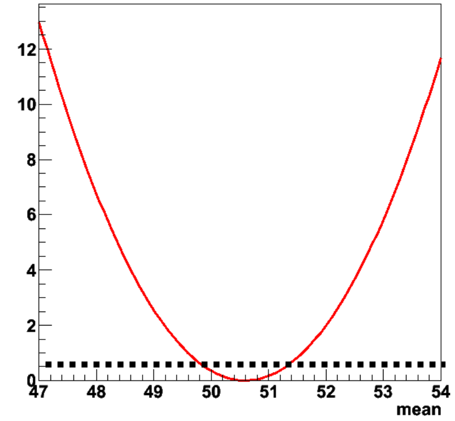
Dealing with nuisance parameters in Bayesian intervals

- 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-dimensional numeric integral can be a technical problem (use of Markov Chain Monte Carlo can help)

Dealing with nuisance parameters in Likelihood ratio intervals

- Likelihood ratio intervals with one parameter

$$LR(\mu) = \frac{L(\mu)}{L(\hat{\mu})}$$



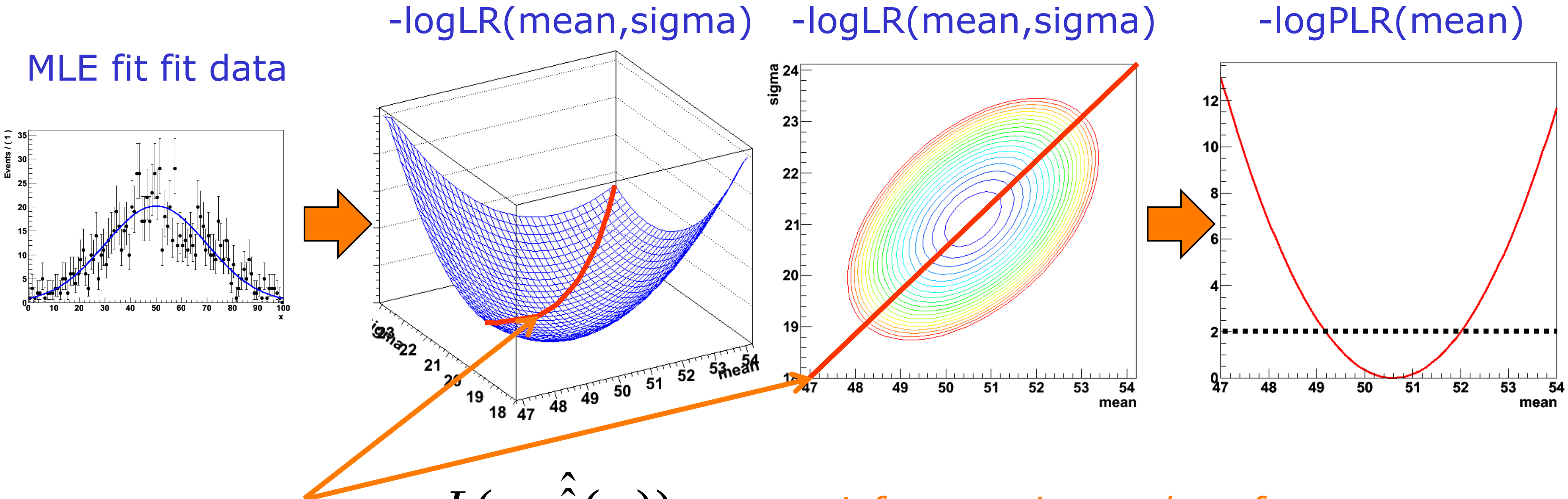
- With nuisance parameters – modify definition of the likelihood ratio

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

$\hat{\mu}$ is MLE estimate of μ
$\hat{\sigma}$ is MLE estimate of σ
$\hat{\hat{\sigma}}(\mu)$ is MLE estimate of s conditional on given value of μ

Dealing with nuisance parameters in Likelihood ratio intervals

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

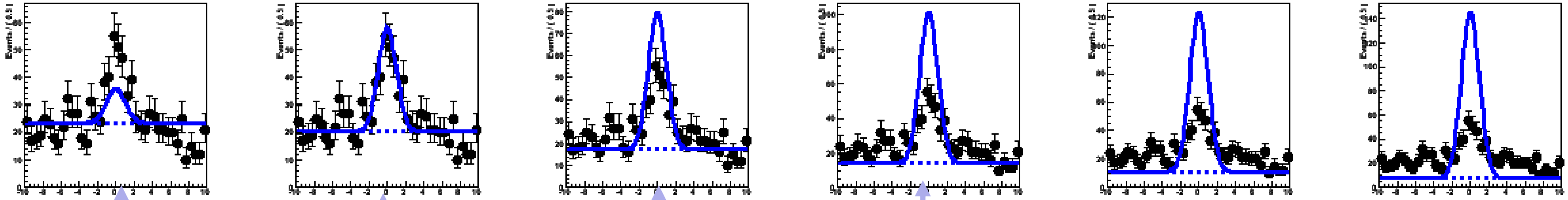


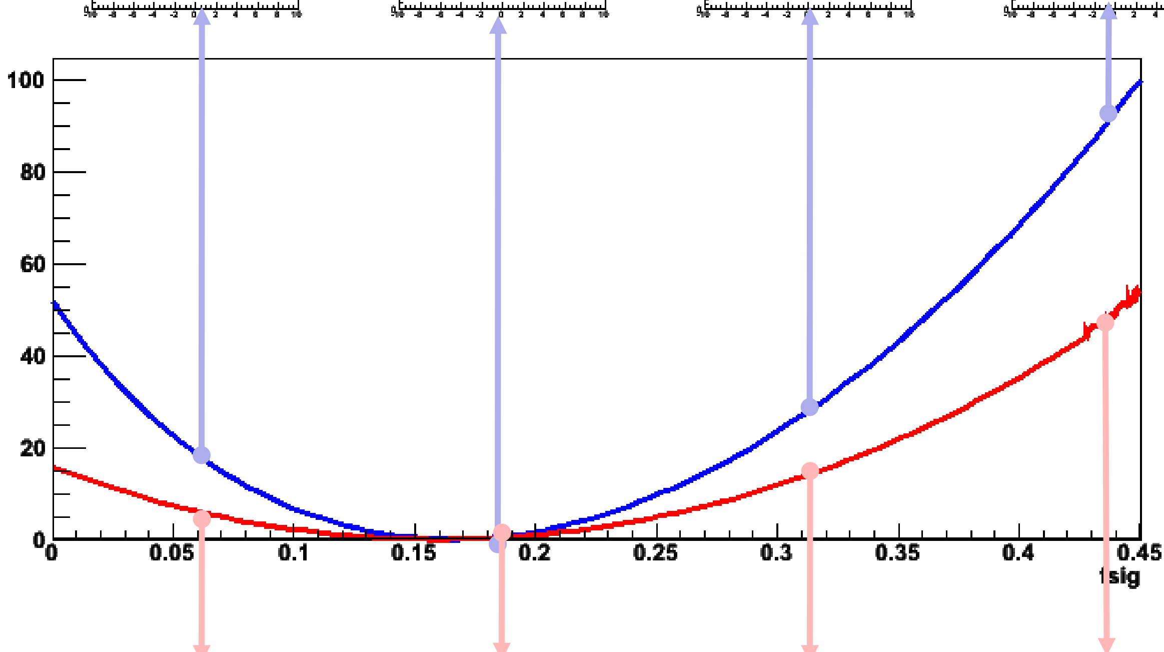
$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\sigma}}(\mu))}{L(\hat{\mu}, \hat{\sigma})}$$

max L for any given value of μ

max L

Dealing with nuisance parameters in Likelihood ratio intervals

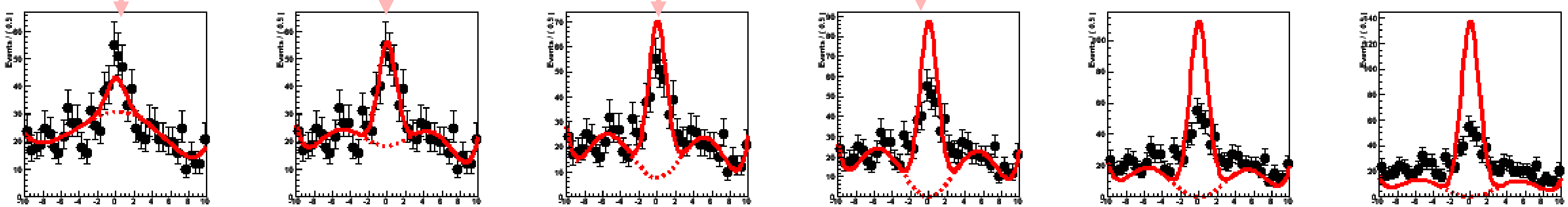




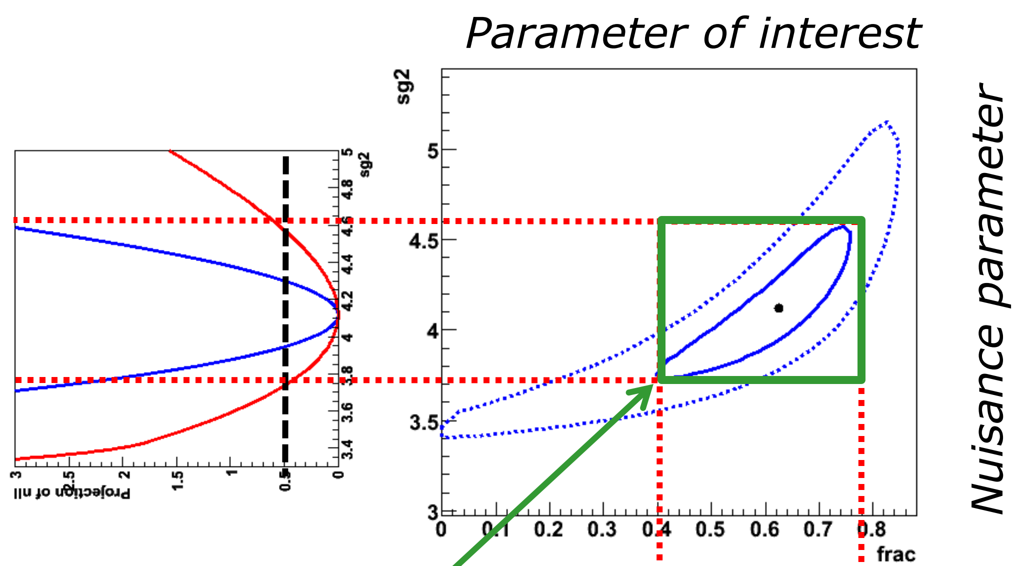
Likelihood Ratio

Profile Likelihood Ratio

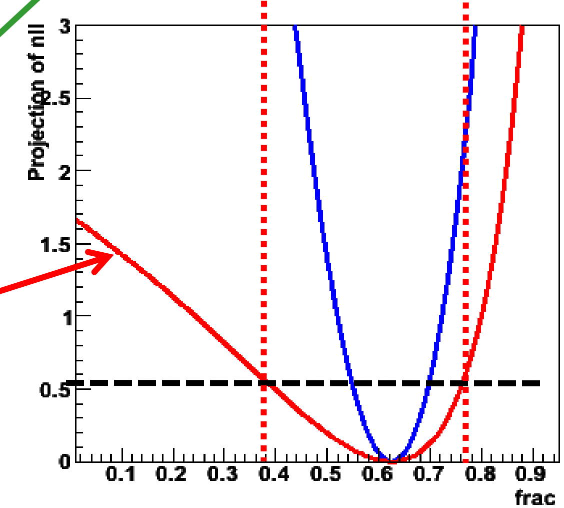
Minimizes –log(L)
for each value of f_{sig}
by changing bkg shape params
(a 6th order Chebychev Pol)



Link between MINOS errors and profile likelihood



- 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

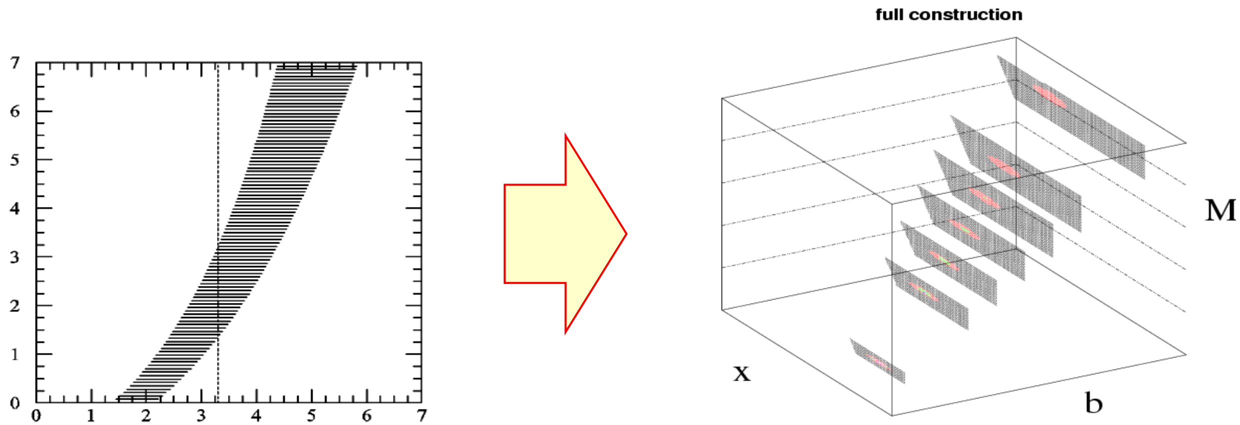


Dealing with nuisance parameters in Likelihood ratio intervals

- Profile likelihood ratio intervals asymptotically equivalent to frequentist intervals
  - Just like plain LR intervals asymptotically equivalent to frequentist intervals
- Issues with Profile Likelihood
  - 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…
  - But beware of finding the wrong minimum, General problem with algorithmic minimization
    - But in profile likelihoods many minimizations are performed with incrementally different starting points → How to choose starting point?

Dealing with nuisance parameters in Frequentist intervals

- Incorporating nuisance parameters in the Neyman construction of the confidence belt makes the belt multi-dimensional



  - 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
  - But: technically very challenging and significant problems with over-coverage → Practical approach: absorb nuisance parameters in profile likelihood then make confidence belt on PLR and POI

How much do answers differ between methods?

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

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

These slide discuss the earlier shown problem:

Poisson(N_{sig}|s+b) · Poisson(N_{ctl}|τ·b)

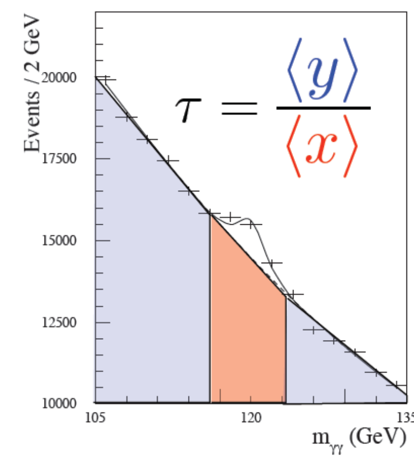
NB: This is one of the very few problems with nuisance parameters with can be *exactly* calculation

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.



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

Recent comparisons results from PhyStat 2007

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!

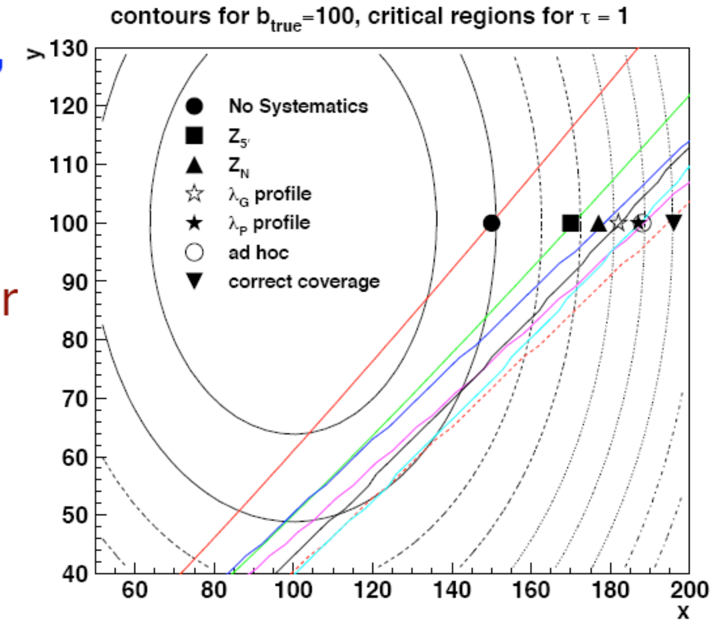


Figure 7. A comparison of the various methods critical boundary $x_{crit}(y)$ (see text). The concentric ovals represent contours of L_G from Eq. 15.

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	$\sim$182

Exact solution

Summary on incorporating nuisance parameters

- 1 – Definition of nuisance parameters
  - A nuisance parameter is any parameter of the model that is not a parameter-of-interest (for physics).
    - Example: for Higgs discovery N(higgs) is of interest, everything else is nuisance
- 2 – Introduction of nuisance parameters in Likelihood
  - Sometimes nuisance parameter arise naturally in the likelihood.
  - In other cases they can be introduces with techniques like template morphing to parameterize underlying systematic uncertaintis
  - For proper treatment of systematic uncertainties in fundamental methods all uncertainties *must* be described as nuisance parameters (i.e. no ad-hoc solutions allows as can be done for 'simple' measurements)
- 3 – Treatment of nuisance parameters in inference
  - 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. Answers can differ sizable – even for simple problems. *This remains a subject of frontier statistics research.*

Discovery, exclusion – expectation and observation

- Now that we have discussed techniques on how formulate model uncertainties in likelihood and how to incorporate these in the statements on the p-value of models (or correspondingly on intervals)
- How to we use those models to make sensible physics conclusions, e.g.
  - Did we discover the Higgs?
  - Can we exclude Supersymmetry?

- First need to decided on formulating answer as discovery or exclusion
  - Discovery: Make statement based on p-value of background only hypothesis (i.e. data is inconsistent with SM). Low p-value means data is inconsistent with SM only, we've discovered 'something' in the data
  - Exclusion: Make statement based on p-value of background + new physics hypothesis (p_{s+b}). Low p-value means new physics is excluded at some C.L.
- While to first order discover and exclusion amount to swapping H_0 and H_1 we treat these scenarios in practice somewhat asymmetrically

Discovery, exclusion – expectation and observation

- Discovery (this scenario has not been exercised much yet at the LHC unfortunately)
  - Canonically aim for '5 sigma' → p-value (SM resulting in excess seed on data or better) of $1.2 \cdot 10^{-7}$
  - Can (and should) in principle also make a statement on p-value of 'SM+new physics' scenario to quantify to which extent the observed result is consistent with the alternate hypothesis, but this is quite complicated in practice → models have usually several free parameters, so what precisely you want to compare to.
  - So, can publish exclusion of SM first, and follow up later what it means precisely
- Exclusion (this scenario *has* been exercised much already!)
  - Canonically aim for 95% → p-value of (SM+NewPhysics) of 5%
  - Explicitly look at p-value of SM-only of the data also. Did you *expect* to be able to exclude this point?

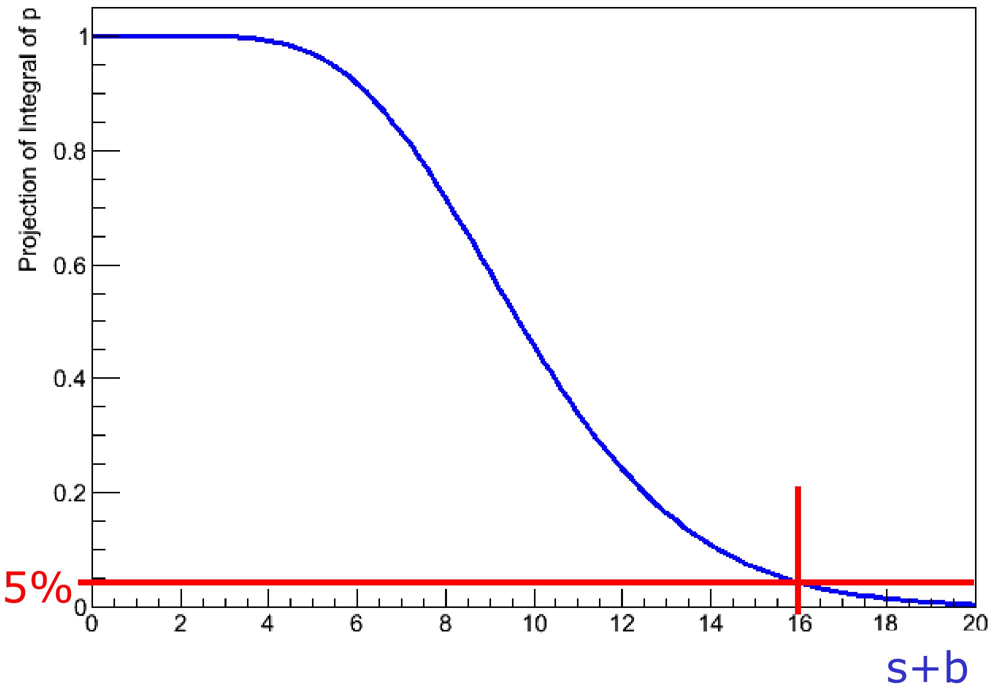
Exclusion statements – in a bit more detail

- Simple example again –
  counting experiment, no nuisance parameters
  - N_{exp}(SM+NewPhys) = 20
  - N_{exp}(SM only) = 10
  - N_{obs} = 10
- p-value (SM+NewPhys) = $$p = \int_0^{10} Poisson(n;\mu = 20)dn = 0.005$$
- p-value (SM) = $$p = \int_0^{10} Poisson(n;\mu = 10)dn = 0.58$$
- New physics scenario predicting 10 additional events is clearly excluded at high C.L. (p=0.005).

Exclusion statements – in a bit more detail

- Usually reframe this statement: rather than excluding a fixed N_{sig} at some C.L, fix C.L. at 95% and find Nsig that is just excluded at that C.L.

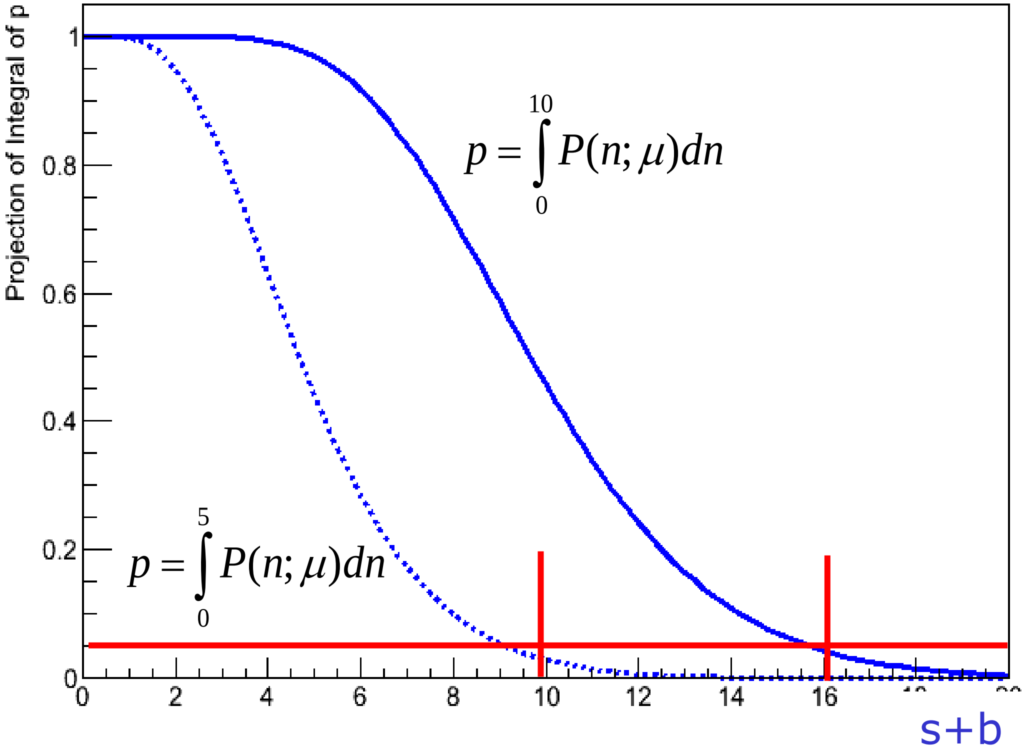
$$p = \int_0^{10} Poisson(n;\mu)dn$$



- Plot p-value for N_{obs}=10 vs mu=s+b
  - Can do this analytically for a Poisson counting exp w/o nuisance params
  - In this example: can exclude N(NewPhys) of ~6 events at 95% C.L.

Exclusion statements – low fluctuations

- What if we observe only 5 events (instead of 10?)
  - N_{exp}(SM+NewPhys) = 20, N_{exp}(SM only) = 10
  - **N_{obs} = 5**
- We can exclude N_{sig}=0 at >95% C.L! Hmm....
  - But we don't *expect* to have sensitivity to do that.
  - Also p(SM)≈6%



Exclusion statements – CLs

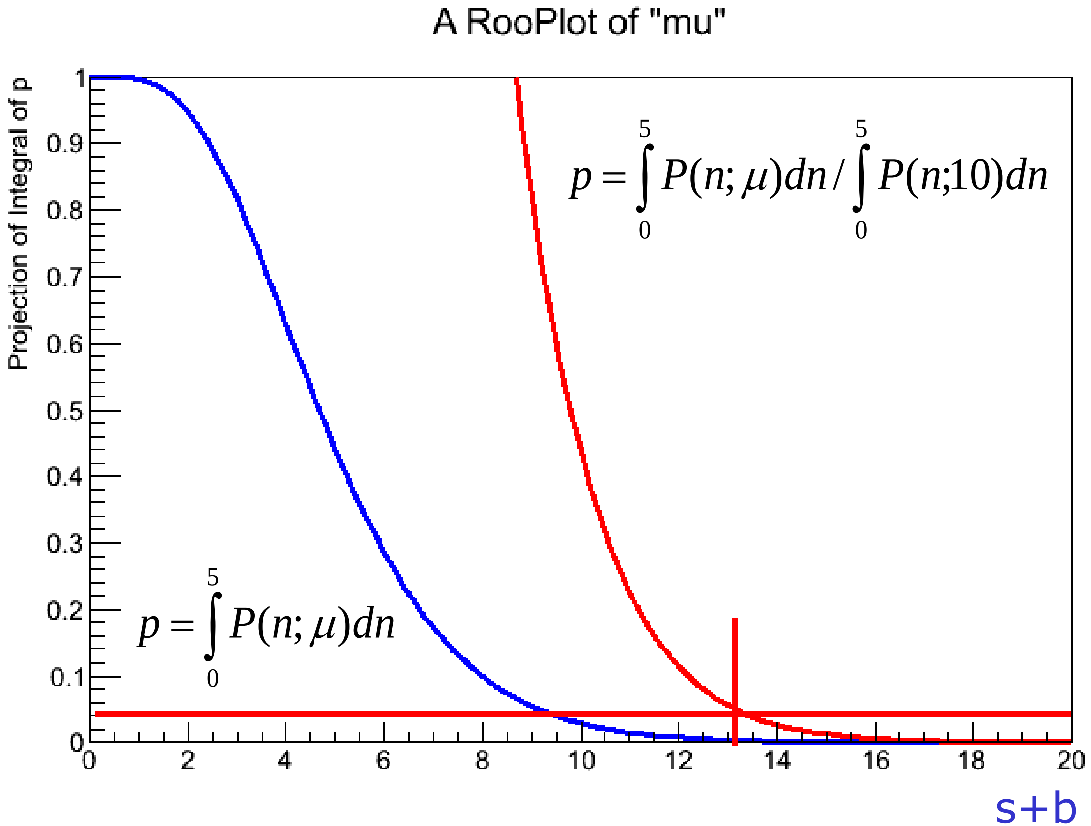
- What should you do with such cases?
  - I.e. low stat fluctuation in observed data gives limit on signal that (far) exceeds expected sensitivity
- HEP invented procedure do deal with such situations is a technique called ‘CL_S’
  - Instead of taking the p-value of S+B case
    $p(S+B) = 7.2 \cdot 10^{-5}$ (Nobs=5, Nexp=20)

    take the ratio of p-values of p(S+B)/p(B)

    $p(S+B) = 7.2 \cdot 10^{-5}$ (Nobs=5, Nexp=20)
    $p(B) = 6.7 \cdot 10^{-2}$ (Nobs=5, Nexp=10)
    ‘CLs’ $= 1.1 \cdot 10^{-3}$ (Nobs=5)

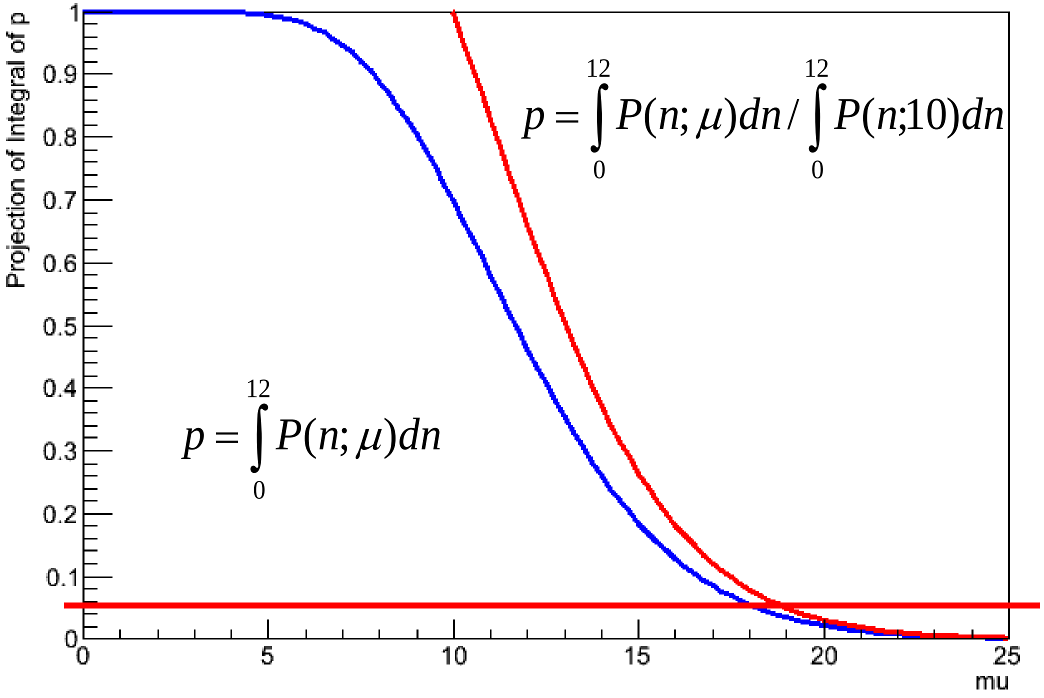
Exclusion statements – CLs

- Effectively CLs prevents you from making exclusions in areas where you do not *expect* to have sensitivity
  - Example case with Nobs is 5, signal limit is now ~3.2 events at 95% C.L. (from excluding Nsig=0 at >95%C.L.)



Exclusion statements – CLs

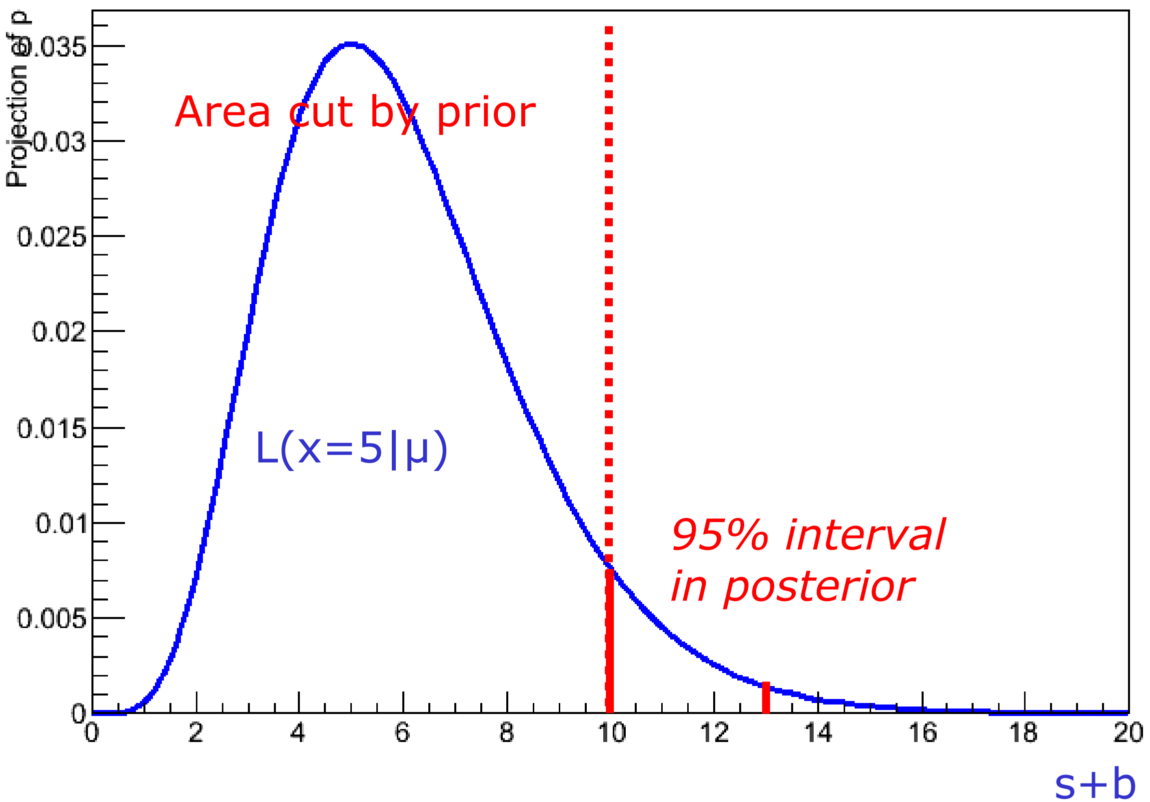
- If you are in a region where there Nobs > Nexp(bkg) difference between CLs and p_{s+b} is small



- CLs tends to overcover a bit when N_{obs} ~<= N_{exp}(bkg)

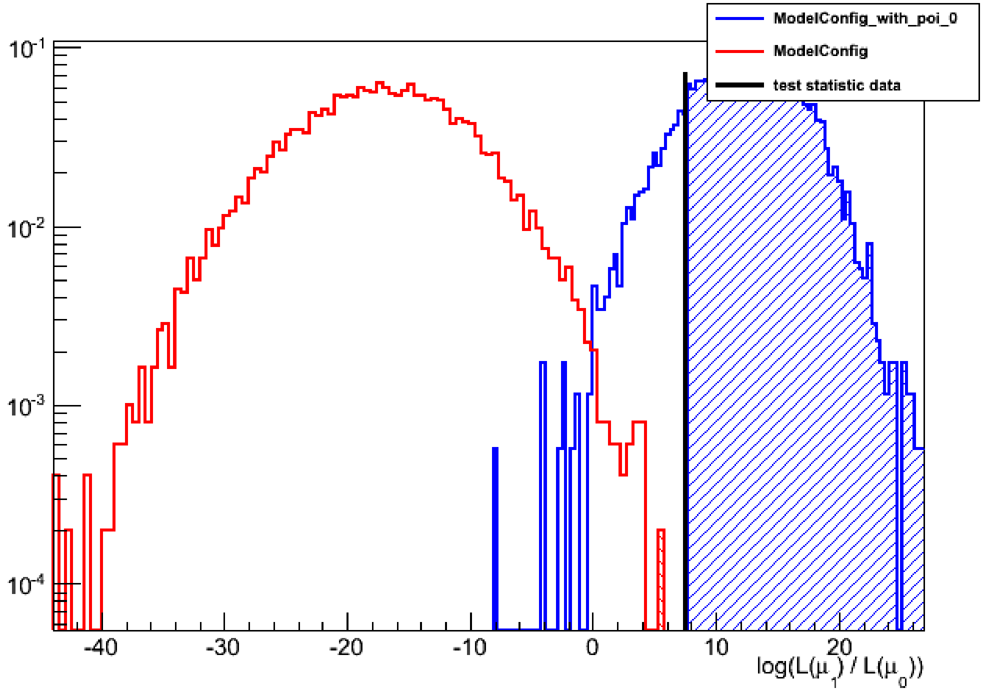
Poisson case: Bayes with flat prior vs CLs

- Note that for the *case of the simple Poisson counting experiment*, the CL_S limit is identical the Bayesian limit assuming a flat prior for s>0 and zero prior for s<0



CLs on realistic analysis

- On more complex analysis, distribution of p values for B and S+B cannot be calculated easily as was done here for Poisson number counting case
- Solution: Obtain distributions from pseudo-experiments and then follow procedure as usual
  - But can take a lot of computing time...

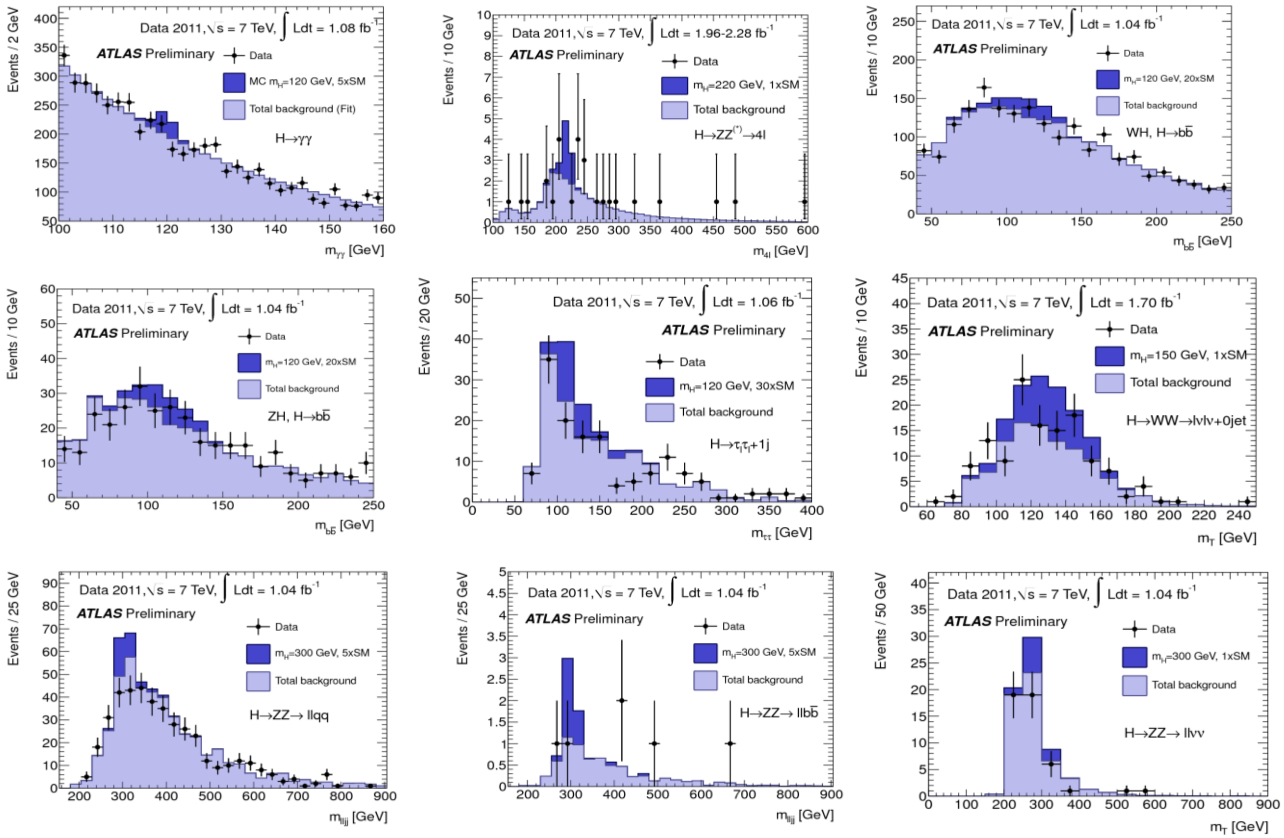


Putting it all together – The Higgs search as example

- The LHC search for the Higgs boson uses almost all of the techniques that we have discussed
  - The search for the Higgs is complicated: It's mass is unknown, it can decay in many possible ways, and preferred decay modes depend on the mass
- Strategy:
  - Search for Higgs boson in many decay channels.
  - Set exclusion limit as function of (unknown) Higgs mass, i.e. assume m_H=X, determine Higgs production cross-section that can be excluded assuming that mass, and then reiterate procedure for $m_H=X+\Delta X$.
  - Goal for now: try to rule out range of Higgs mass. (If Higgs boson exist, exclusion will remain impossible in some region of Higgs mass.) Goal for later: switch from exclusion to discovery mode
  - Increase sensitivity by combining results from all channels in limit calculation
  - Increase sensitivity by combining results from ATLAS and CMS (should be public for the first time *this* week)

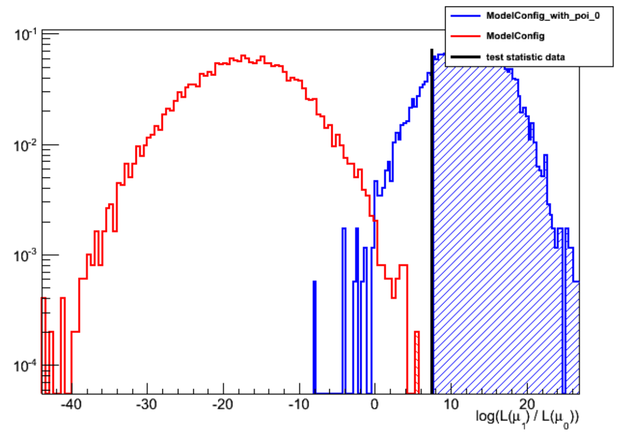
Putting it all together – The Higgs search as example

- Step 1 – Event selection
  - Defined separately for each Higgs decay channel (gg,WW,ZZ,ττ)
  - Quantify SM and Higgs expected distributions for each channel (the latter for a large range of mHiggs hypotheses)
  - **NB: This is a substantial effort for each channel**



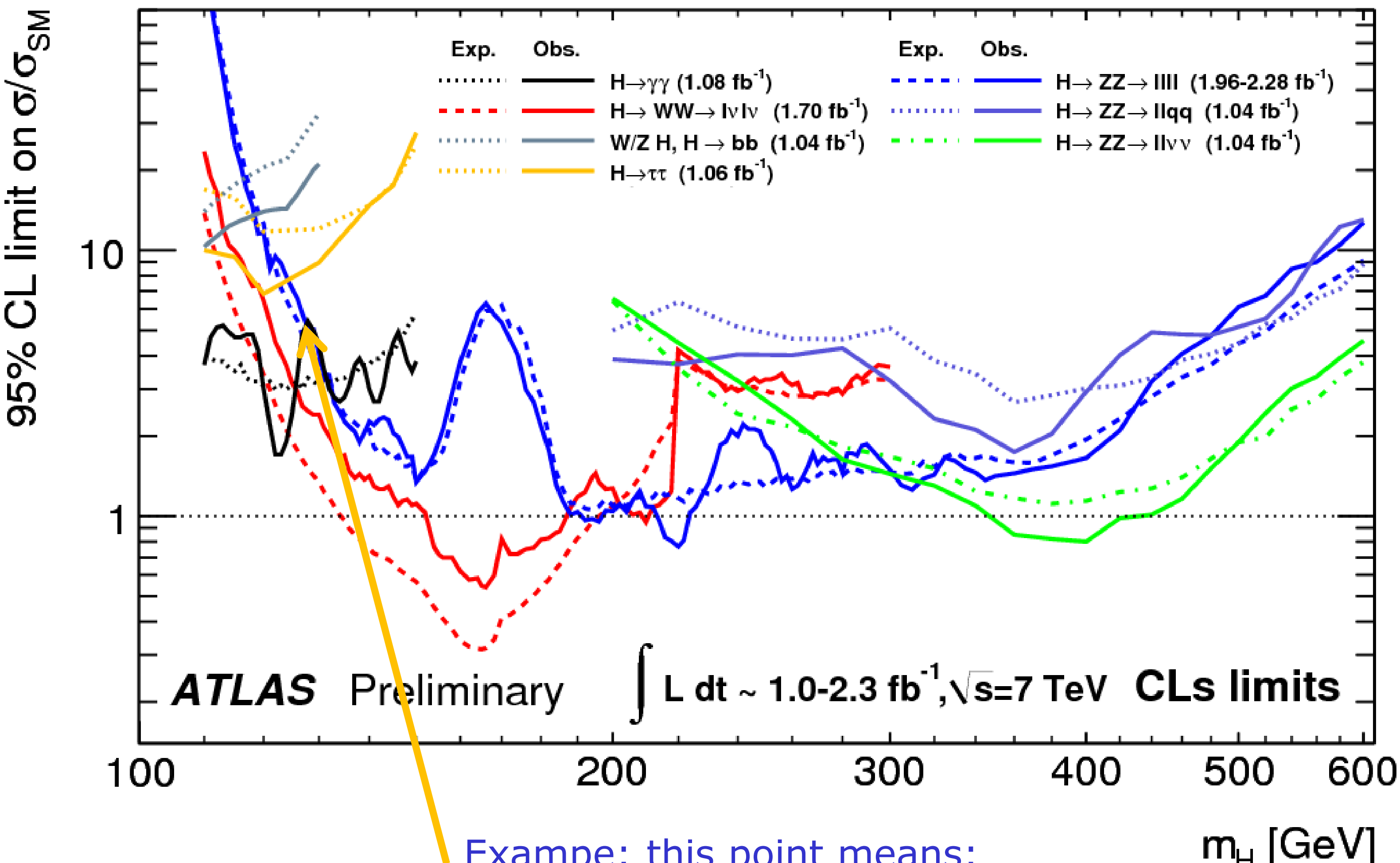
Putting it all together – The Higgs search as example

- Step 2 – Statistical analysis of each channel
- Use 'profile likelihood ratio' as test statistic, as input to CLs construction. For each channel and m(Higgs):
  - Generate distribution of expected values for test statistic in background only hypothesis and signal+background hypothesis



  - Calculated observed value of p(s+b), p(b) → observed CLs limit on Higgs→XX at m(Higgs)=YY GeV
  - Calculate expected CLs limit for each point too (central value of expectation and 1-sigma and 2-sigma intervals)

Putting it all together – The Higgs search as example

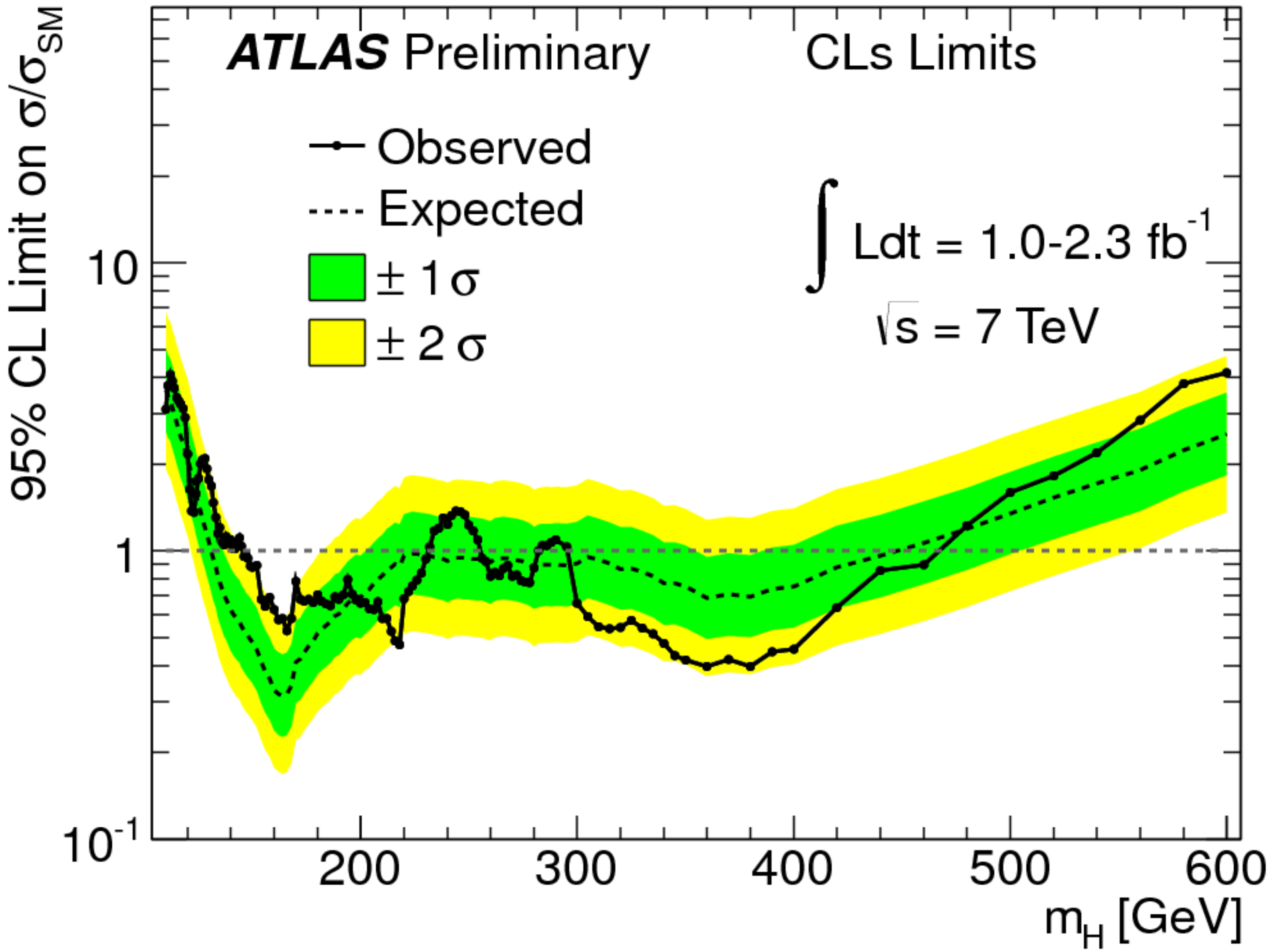


Exampe: this point means:
H→gg analysis can exclude Higss production at
~5x the SM predicted rate at m_H=130 GeV

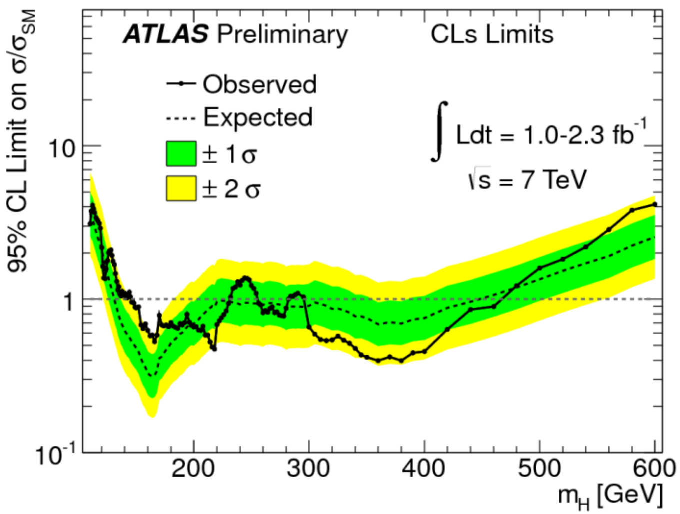
Putting it all together – The Higgs search as example

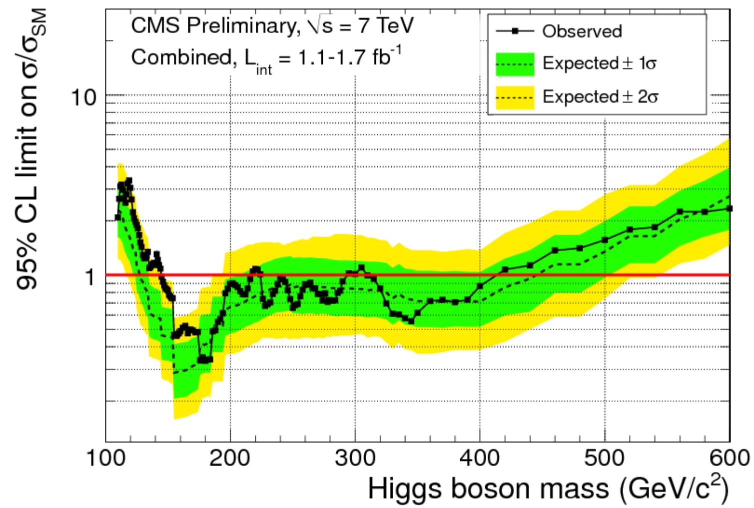
- Step 3 – Combination effort
- Construct joint likelihood of all channels (and/or experiments)
  - Take Lcomb = $L_{gg}(m_H,....)\cdot L_{WW}(m_H,...)\cdot L_{ZZ}(m_H,..)$
  - Profile over all nuisance parameters (can be >100 parameters), (amounts to minimizing the likihood with >100 parameters!)
  - Perform CLs construction (generate pseudo-data, minimize likelihood, iterate a few 100K times)
- Need to this carefully about correlations between nuisance parameters from various channels
  - i.e. Jet Energy Scale uncertainty will be common in all channels, backgrounds rates are usually not etc etc
- Repeat exercise for many values of m(Higgs)

Putting it all together – The Higgs search as example



Coming (very) soon – LHC combination

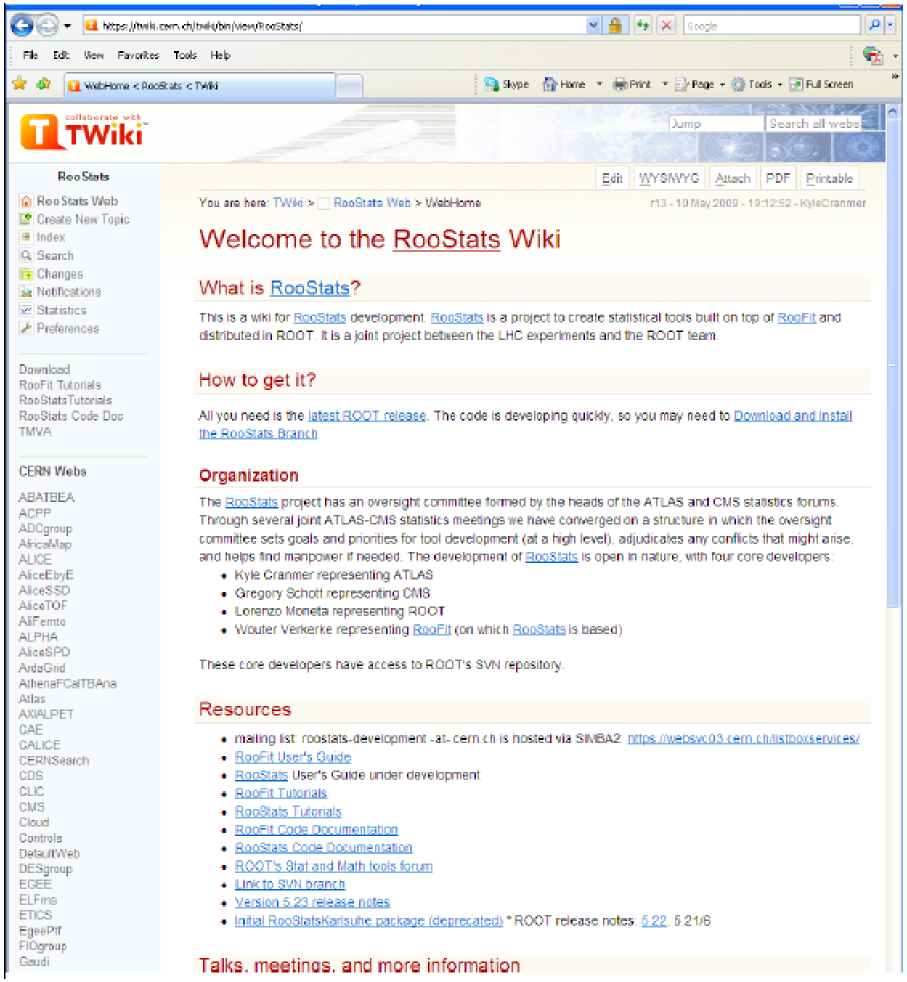




Software for combinations and limit setting

- A lot of software has been developed in the past years to simplify the technical implementation of likelihoods, combining of likelihoods and to perform limit calculations
- Tools for modeling of pdfs already existed in some form (RooFit) → has been extended to make this models persistable → Can write actual likelihood to file
- Combination effort consists then of picking up these files, constructing the joint likelihood and minimizing this
  - Technically easy (but still a lot of effort and thinking required)
- Write a new series of 'standard' tools that perform Bayesian, Frequentist and Likelihood-based limit calculations on such models
  - Goal: each calculator can handle *any* model (very ambitious!)

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.*

B. Cousins: 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 asymptotic regime.
- 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

RooStats Project – Example

- Create a model - Example

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

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
```

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

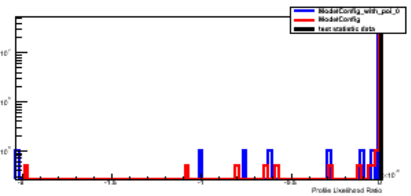
```
FeldmanCousins fc;
fc.SetPdf(w::PC);
fc.SetData(data); fc.SetParameters(w::s);
fc.UseAdaptiveSampling(true);
fc.FluctuateNumDataEntries(false);
fc.SetNBins(100); // number of points to test per parameter
fc.SetTestSize(.1);
ConfInterval* fcint = fc.GetInterval(); // that was easy.
```

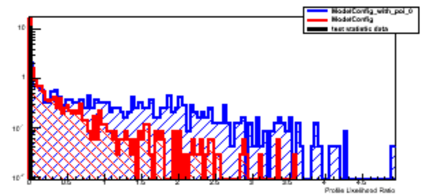
  - Bayesian (MCMC)

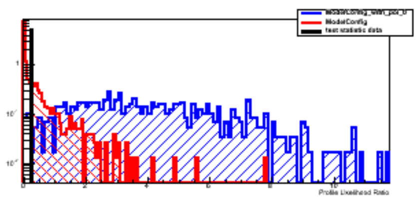
```
UniformProposal up;
MCMCCalculator mc;
mc.SetPdf(w::PC);
mc.SetData(data);  mc.SetParameters(s);
mc.SetProposalFunction(up);
mc.SetNumIters(100000); // steps in the chain
mc.SetTestSize(.1); // 90% CL
mc.SetNumBins(50); // used in posterior histogram
mc.SetNumBurnInSteps(40);
ConfInterval* mcmcint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

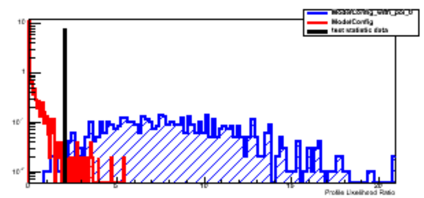
RooStats Project – Example

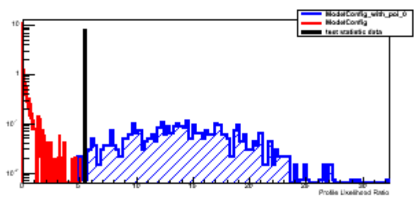
- Retrieving and visualizing output

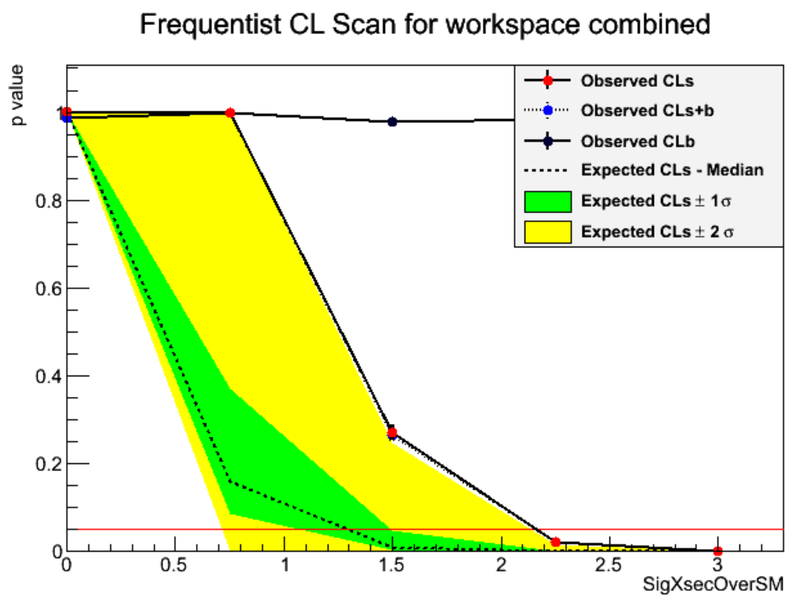






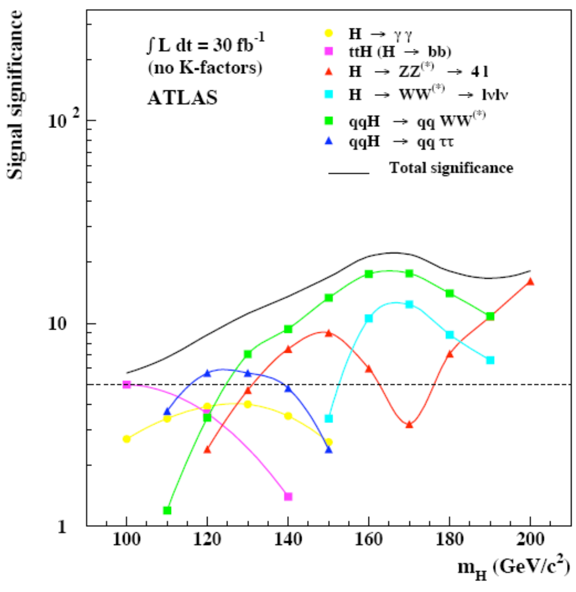






‘Digital’ publishing of results

- 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 applications, e.g. now used in combining of Higgs channels, sharing of models between ATLAS and CMS for combination effort



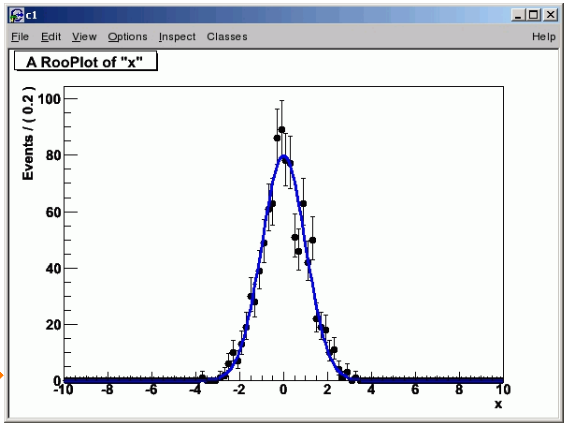
Using persisted p.d.f.s.

- 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) ;
```

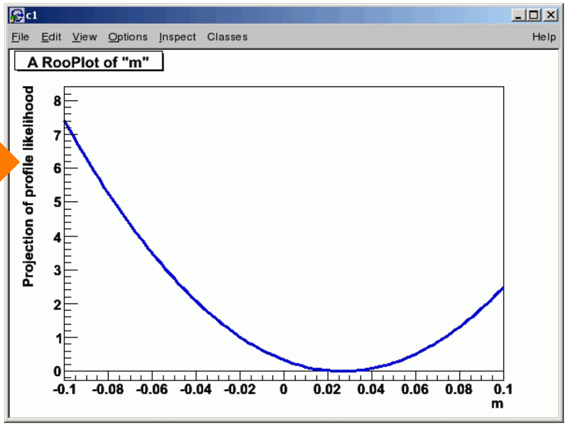


Construct likelihood & profile LH

```
RooAbsReal* nll = w::g.createNLL(w::d)
RooAbsReal* pll = nll->createProfile(w::mean) ;
```

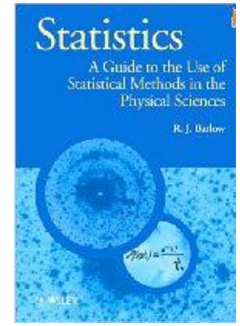
Draw profile LH

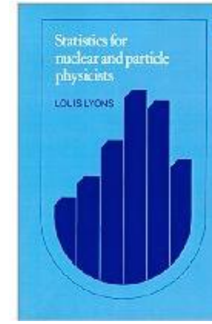
```
RooPlot* mframe = w::m.frame(-1,1) ;
pll->plotOn(mframe) ;
mframe->Draw()
```



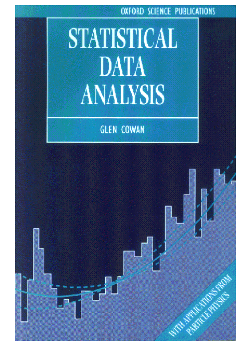
– **Note that above code is independent of actual p.d.f in file → e.g. full Higgs combination would work with identical code**

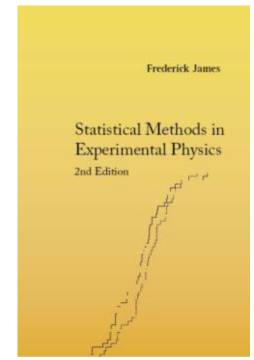
The end – Recommended reading



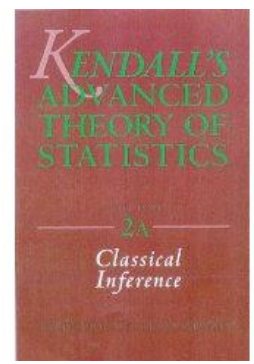


- Easy
  - R. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley, 1989
  - L. Lyons, *Statistics for Nuclear and Particle Physics*, Cambridge University Press
  - Philip R. Bevington and D.Keith Robinson, *Data Reduction and Error Analysis for the Physical 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)

- 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/