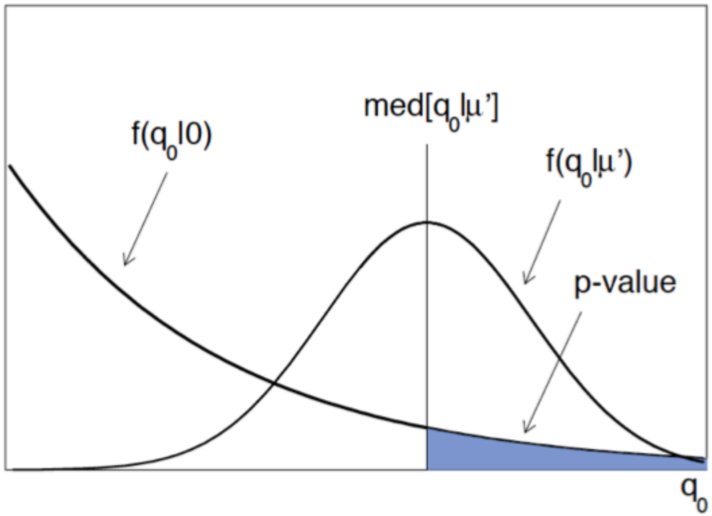
Advanced Statistics Course – Part IV

W. Verkerke (NIKHEF)

Expected sensitivity & asymptotic distributions

Expected significance

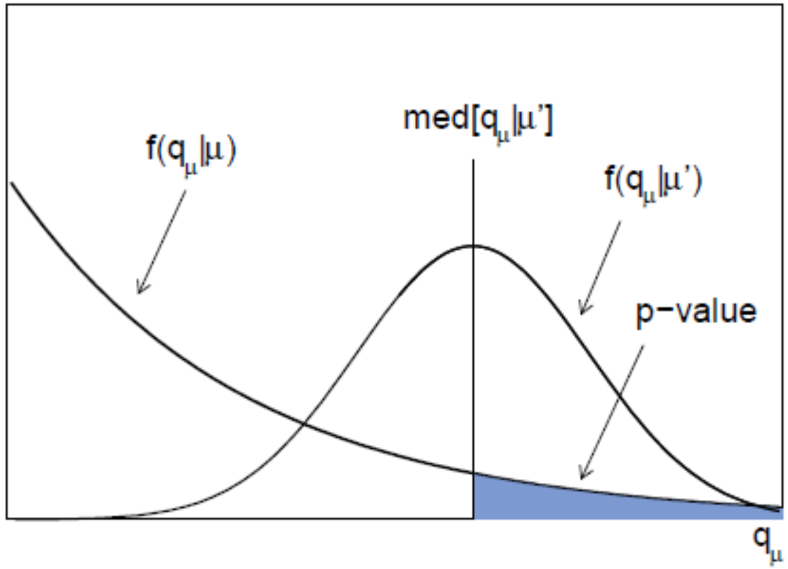
- When planning an experiment, we want to quantify how sensitive we are to a potential discovery, e.g. by giving the median significance, assuming some non-zero signal strength μ’



- For p-value, need $f(q_0|0)$ → find q_0 for which p-values is e.g. 0.05 ($=q_0^{disc}$)
- For sensitivity need $f(q_0|\mu')$ → find value of μ’ for which median q_0 value is q_0^{disc}

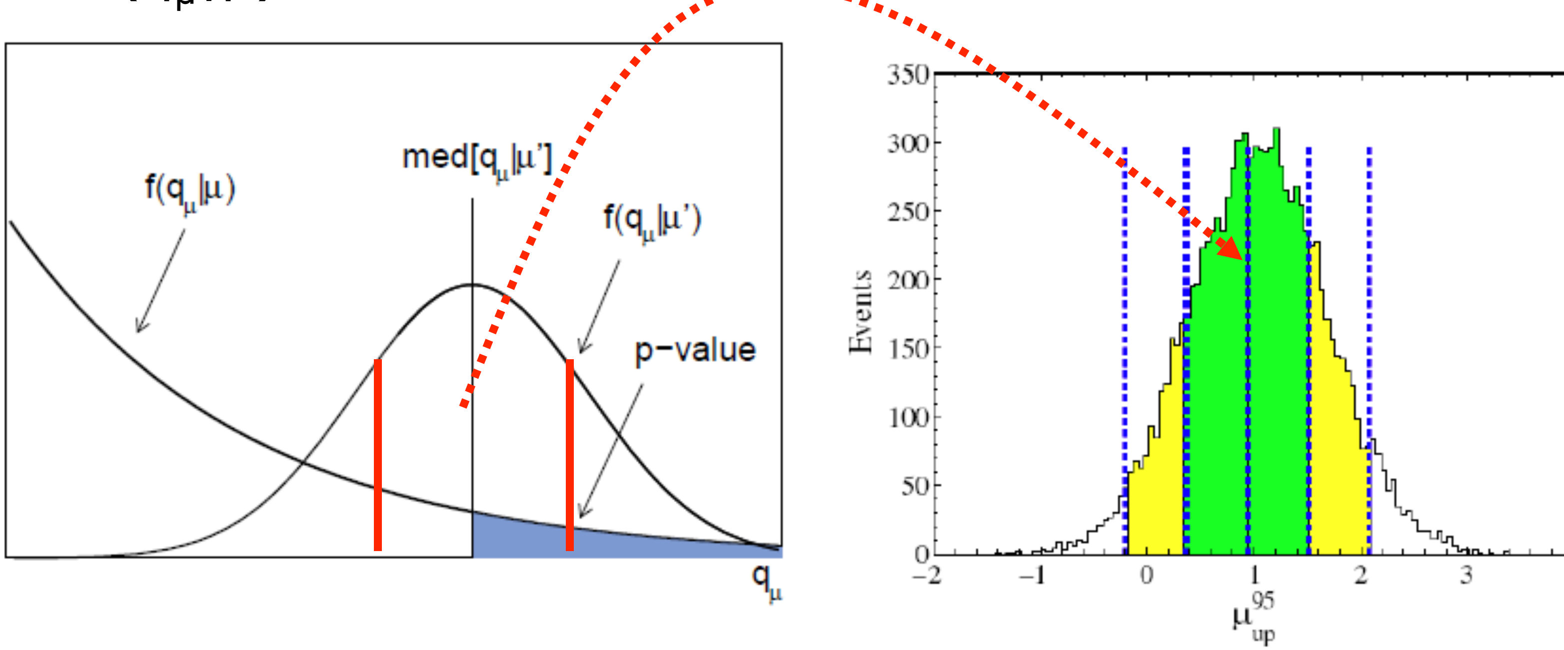
Expected (upper) limits

- Similarly, can construct expected (upper) limits.
- Allows to compare observed limits to what we would expected (based on the sensitivity of the experiment)
  - Find value of μ_{up} for which background-only data would result in (e.g.) 95% exclusion of μ_{up}.
  - For given μ, use $f(q_\mu|0)$ to find median value of q_μ for background-only data.
  - Find μ_{up} for which $med[q_{\mu\text{-}up}|0]$ results in $p_{\mu\text{-}up}$ of (e.g.) 0.05



Uncertainty bands on expected limits

- Upper limits are subject to the effect of statistical fluctuations in the data. Customary to quantify this effect with error bands on the expected limit
- Example for N·σ error bands on upper limit → Recalculate limits using med[q_μ|μ′]±σ(q_μ|μ′) to define value of q_μ that should result in 5% p-value for f(q_μ|μ)



Asymptotic distributions

- Using toy experiments to calculate limits can be very computationally very expensive.
- Fortunately analytical 'asymptotic' forms exists for all needed test statistic distributions that are valid in the limit $N \rightarrow \infty$ for $f(q_\mu|\mu')$
- For t_μ, the profile likelihood ratio test statistic, defined as

$$t_\mu = -2\ln\lambda(\mu), \quad \lambda(\mu) = \frac{L(data \mid \mu, \hat{\hat{\theta}}(\mu))}{L(data \mid \hat{\mu}, \hat{\theta})}$$

the asymptotic form for t_μ for an assumed strength μ' is

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\frac{\mu-\mu'}{\sigma}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right)\right]$$

(this is known as a "non-central χ^2 distribution")

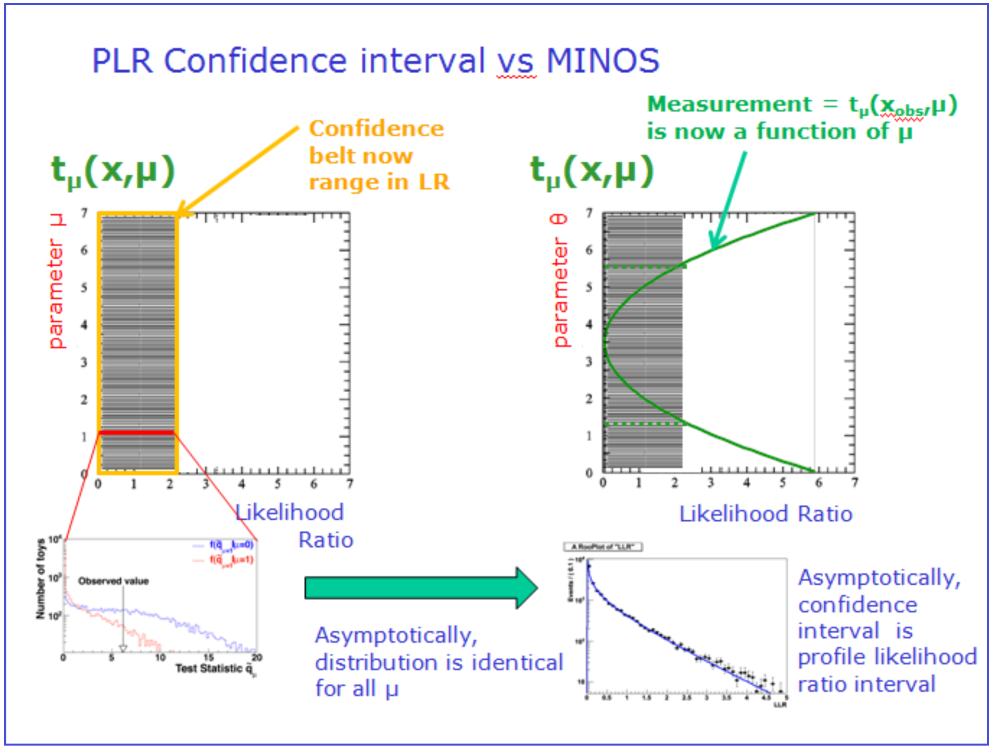
Asymptotic distributions

- For $f(t_\mu|\mu)$ this simplifies to

$$f(t_\mu|\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t_\mu}} e^{-t_\mu/2} \; .$$

  a simply χ^2 distribution for one degree of freedom

- This we'd already seen this before when discussing the relation between MINOS intervals and confidence intervals based on t_μ



Asymptotic forms of q_0 and q_μ

- Similarly, asymptotic forms exist for q_0 and q_μ
- For the discovery test statistic q_0: full form

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

- Simplified form for μ=μ′=0

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} .$$

δ function at zero
(modeling cases with μ-hat<0)

Chi-squared distribution
for 1 degree of freedom

- Finally, one can also show that the discovery significance is asymptotically simply $\sqrt{q_{0,obs}}$

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} .$$

Asymptotic forms of q_0 and q_μ

- For the exclusion test statistic q_μ: full form

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu}-\frac{(\mu-\mu')}{\sigma}\right)^2\right]$$

- Simplified solution for μ=μ'

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$

δ function at zero
(modeling cases with μ-hat<0)

Chi-squared distribution
for 1 degree of freedom

Using asymptotic forms for μ≠μ’

- Similar to case for t_μ, asymptotic distributions for q_0,q_μ for μ≠μ’ depend on σ, which cannot be calculated analytically as it depends on the nuisance parameters

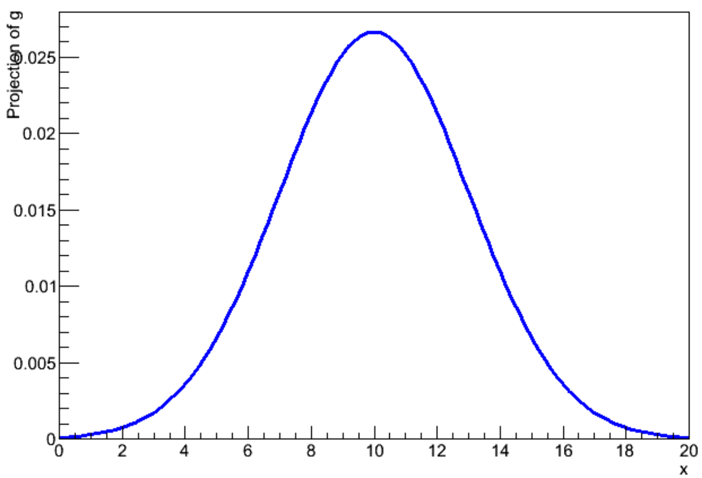
$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)^2\right]$$

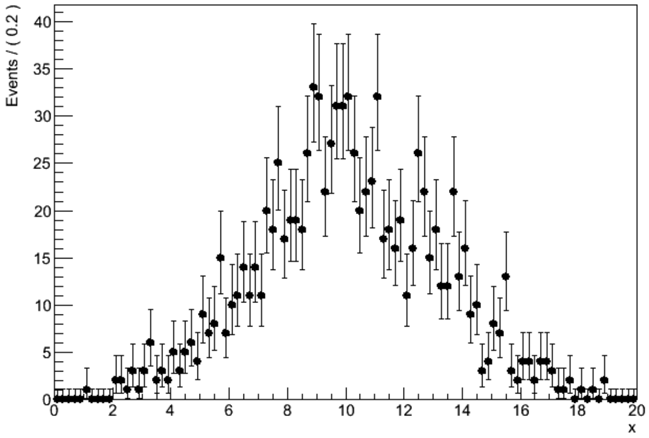
- But for these test statistics we can calculate this in a computationally inexpensive way using ‘Asimov Datasets’

The Asimov dataset illustrated

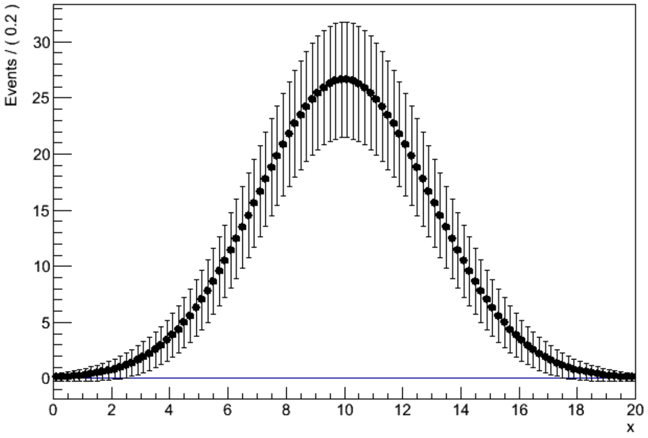
f(x)



Regular data set of 1000 events



Asimov dataset
(a weighted dataset
with a sum of weights of 1000)



NB: Asimov datasets can be trivially generated from any model in RooFit by adding the Asimov() argument to RooAbsPdf::generate()

How to Asimov dataset helps

- With a test statistic evaluated on an Asimov dataset one can calculate directly (in the asymptotic limit)
- The variance σ^2 needed for the full asymptotic expressions

Generally

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}} ,$$

For discovery: $f(q_0|\mu')$

$$\sigma_A^2 = \frac{\mu'^2}{q_{0,A}} .$$

For limits: $f(q_\mu|0)$

$$\sigma_A^2 = \frac{\mu^2}{q_{\mu,A}} ,$$

Value of test statistics for Asimov dataset

- Note that with Asimov datasets, one can also trivially obtain the median p-values needed for the median expected limit and observation significance

$$\text{med}[Z_0|\mu'] = \sqrt{q_{0,A}} ,$$

$$\text{med}[Z_\mu|0] = \sqrt{q_{\mu,A}} .$$

(Corresponding p-values trivially calculable from Z-values)

Asymptotic independence on nuisance parameters

- While full asymptotic forms depends nuisance parameters via σ, cases for μ≠μ' do **not**

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} .$$

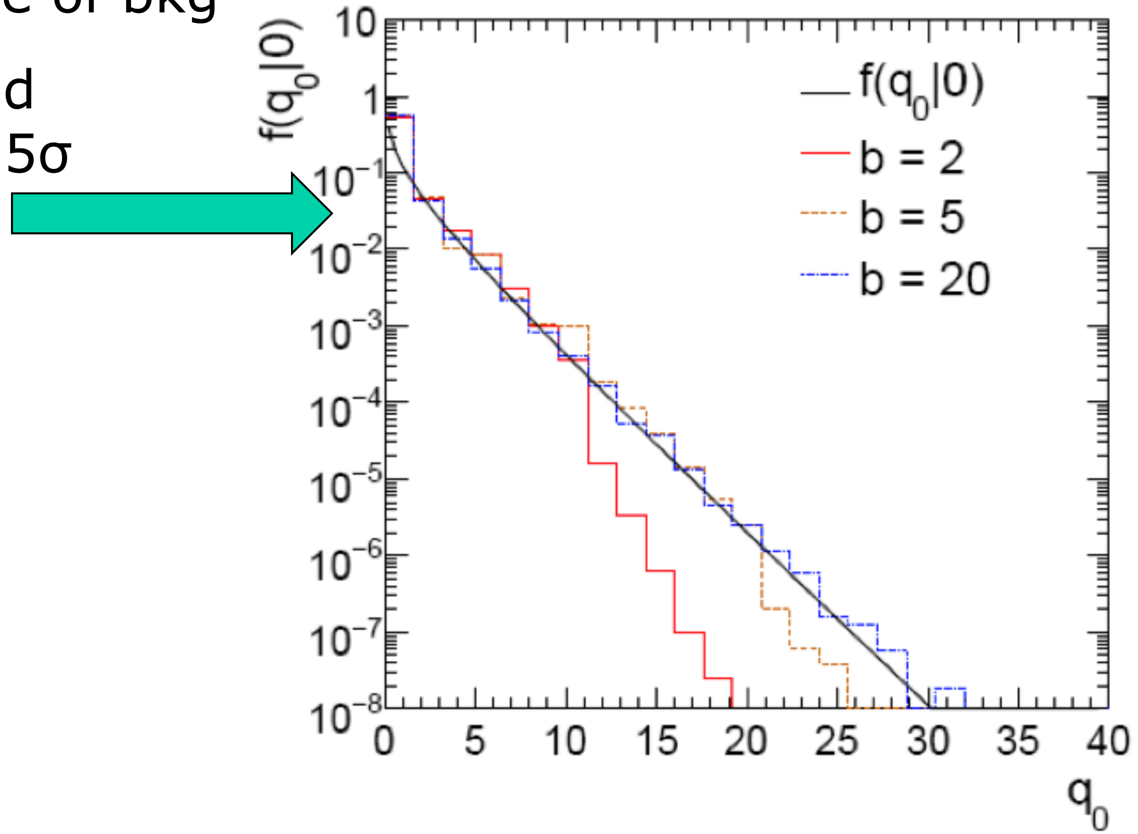
- This means that the distributions that are used to calculation observation p-values and limits are asymptotically independent of the value of the nuisance parameters.
- This is good → If distributions are truly independent of nuisance parameters, exact coverage is restored.
- Note that this does not hold for all test statistics, e.g. for $\tilde{t}_\mu$ this is not the case (NB: $\tilde{t}_0$ corresponds to Feldman Cousins)

How quickly is asymptotic behavior achieved?

- Investigate using on/off problem

  Poisson(n|μs+b)Poisson(m|τ·b)   with τ=1

- Discovery: p-value of bkg
- Asymptotic is good approximation to 5σ already at b=20!

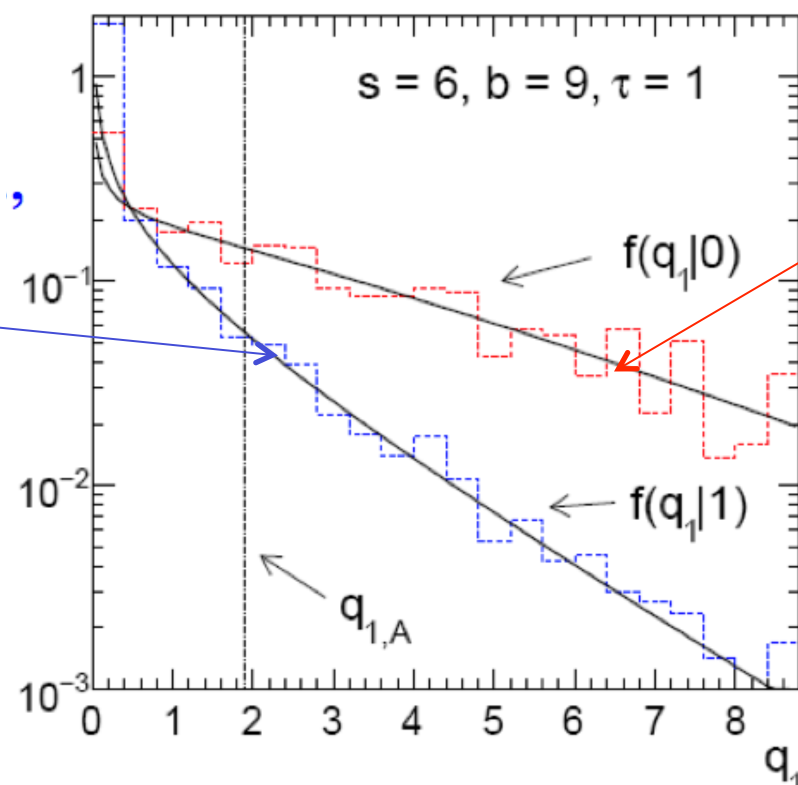


How quickly is asymptotic behavior achieved?

- Exclusion limits: look at distributions for q_1 test statistics for assumed hypothesis $\mu=0$ and $\mu=1$
  - Already good behavior for s=6, b=9

Poisson(n|μs+b)Poisson(m|τ·b) with τ=1

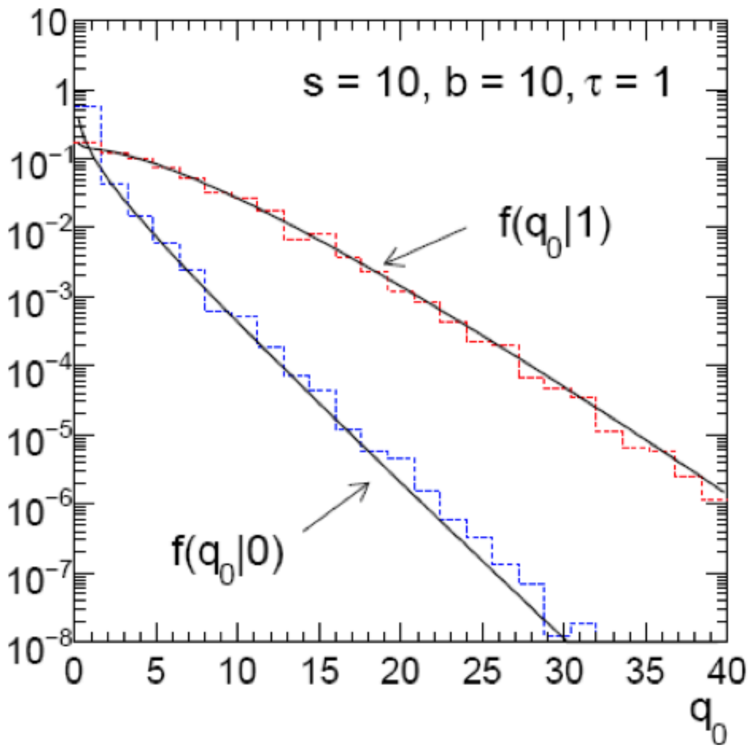
Distribution defines p-value of μ=1 hypothesis

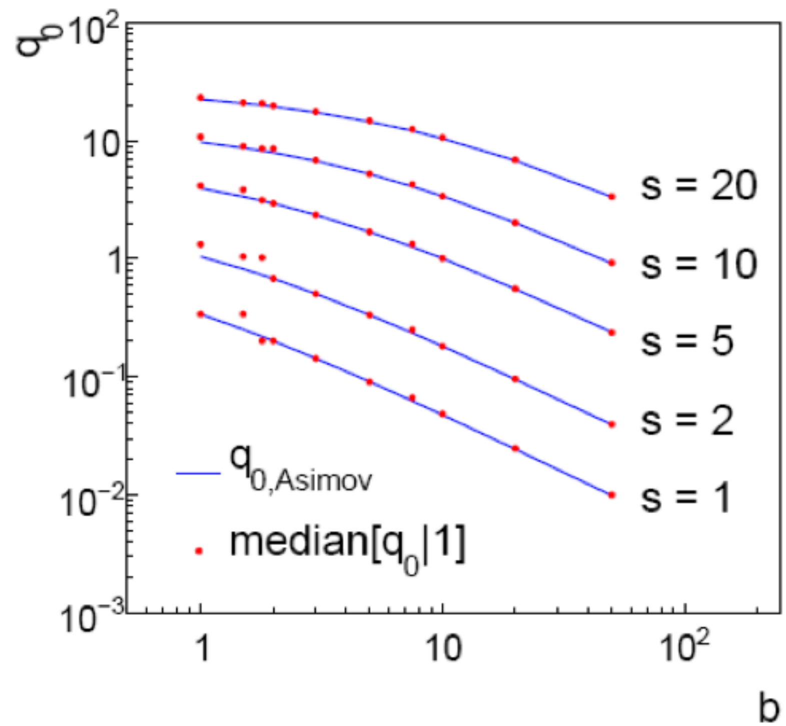


Median of distribution defines expected exclusion sensitivity

How quickly is asymptotic behavior achieved?

- Same, for s=10, b=10, med[q_0|1] for s=1..20 and b=1...50





- For all asymptotic formulae and lots of details

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

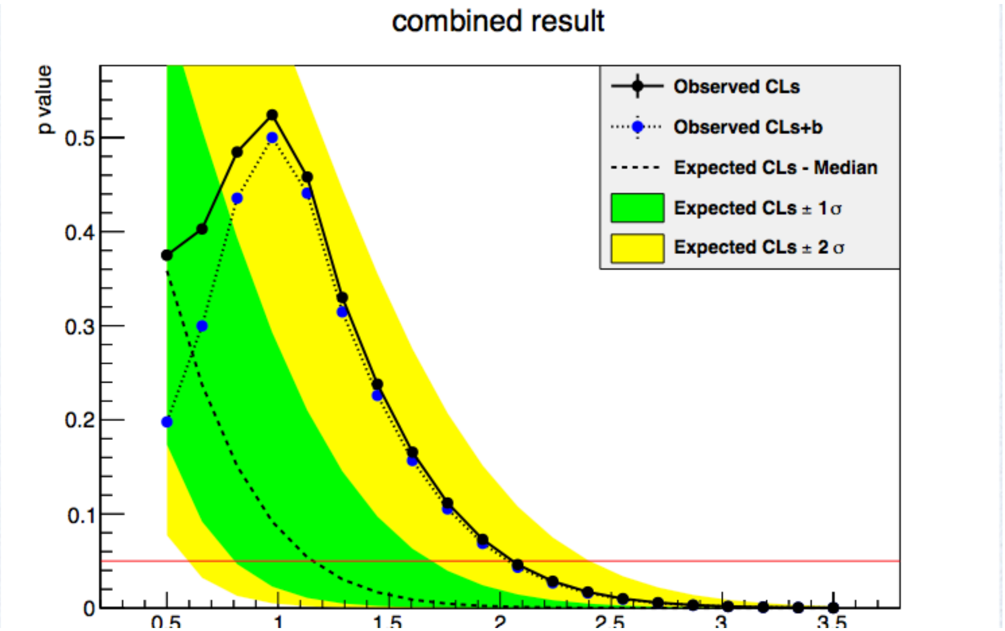
Using asymptotical formulae in RooStats

- All asymptotic formulae are implemented in RooStats in the class AsymptoticCalculator
  - Specify instead of FrequentistCalculator when you want to use asymptotic forms.

```
// create first HypoTest calculator (N.B null is s+b model)
AsymptoticCalculator ac(*data, *bModel, *sbModel);

HypoTestInverter calc(*ac);
// run inverter same as using other calculators
........
```

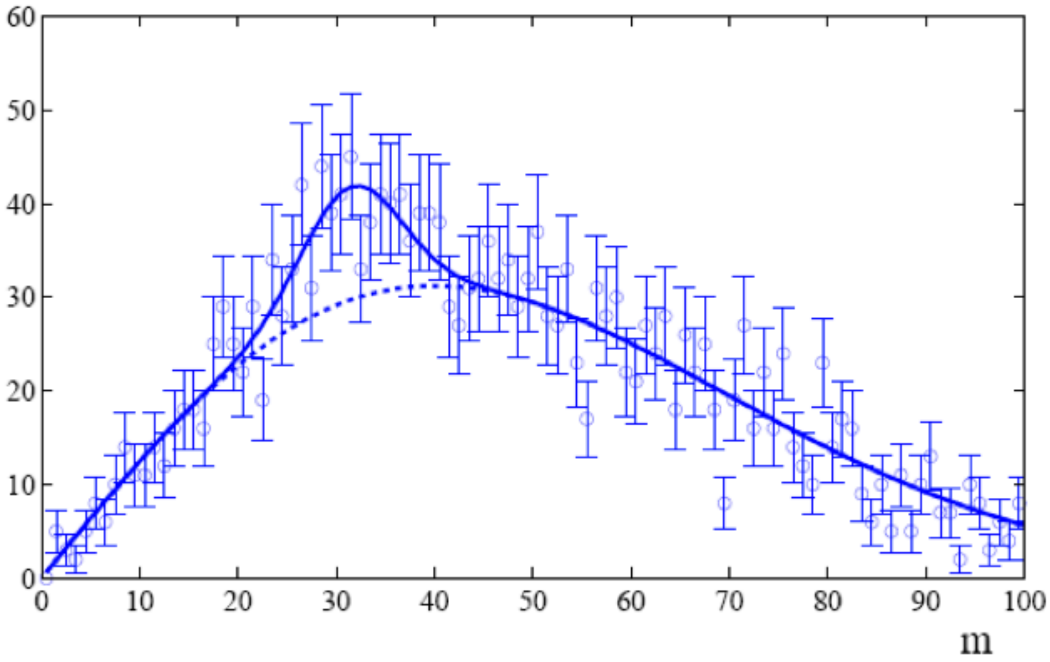
- Also available in StandardHypoTestInvDemo.C: Specify calculatorType=3 (instead of 0)



The look-elsewhere effect

The 'Look Elsewhere Effect'

- Suppose a model for a mass distribution allows for a peak at mass m with with amplitude μ
- The data shows a bump at mass m_0.
- How consistent is this with the no-bump hypothesis?



Calculating the p-value for a fixed mass

- If the mass m_0 of the peak is known a priori, problem reduces to a likelihood ratio test statistic with one parameter of interest (μ)

$$t_{fix} = \frac{L(0, m_0)}{L(\hat{\mu}, m_0)}$$

- P-value for background-only hypothesis calculated as

$$p_{0,fix} = \int_{t_{fix,obs}}^{\infty} f(t_{fix} \mid 0) dt_{fix}$$

  specifies probability to observe t_{fix} or larger
  at the specified value of m_0

- Expect that asymptotic form of $f(t_{fix}|0)$ can be used with a sufficiently large data sample.

Calculating the p-value for an unspecified mass

- If we don't where to expect a peak in the distribution, we want the probability to find a peak at least as significant anywhere in the distribution:

$$t_{float} = \frac{L(0)}{L(\hat{\mu}, \hat{m})}$$

Note that m is not a parameter of the background-only model

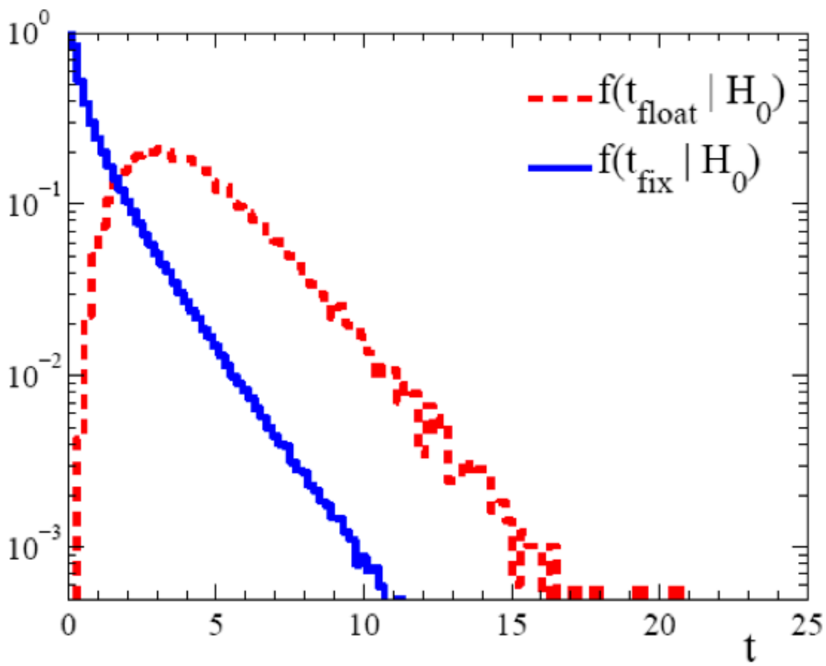
- Calculate p-value as usual

$$p_{0,float} = \int_{t_{float,obs}}^{\infty} f(t_{float} \mid 0) dt_{float}$$

- What about the asymptotic distribution for $f(t_{float}|0)$?

Asymptotic distributions for t_{fix} and t_{float}

- For sufficiently large data sample t_{fix} has chi-squared distribution for 1 degree of freedom
- For t_{float}, one naively expected a chi-squared distribution with 2 degrees of freedom (μ,m)



But **observed** distribution is very different for t_{float}

Wilks Theorem does not hold since parameter m is not defined in μ=0 model

- Obtaining $f(t_{float}|0)$ from toy experiments can be very expensive and difficult → Any other ways?

The look-elsewhere effect and the trials factor

- The fact that p_{fix} and p_{float} are different is called the 'look elsewhere effect'
- Probability to obtained observed result under background hypothesis increases if you look in a broad range instead of in a specific place
- The ratio p_{float}/p_{fix} is also called the 'trials factor'
- Generally, if we know the trials factor for a given experiment, we only need to calculate p_{fix} to obtained p_{float}

Approximating the trials factor

- Gross and Vitells (arXiv 1005.1891) show that

$$p_{float} \approx p_{fix} + \langle N(c) \rangle$$

where N(c) is the number of ‘upcrossings’ of

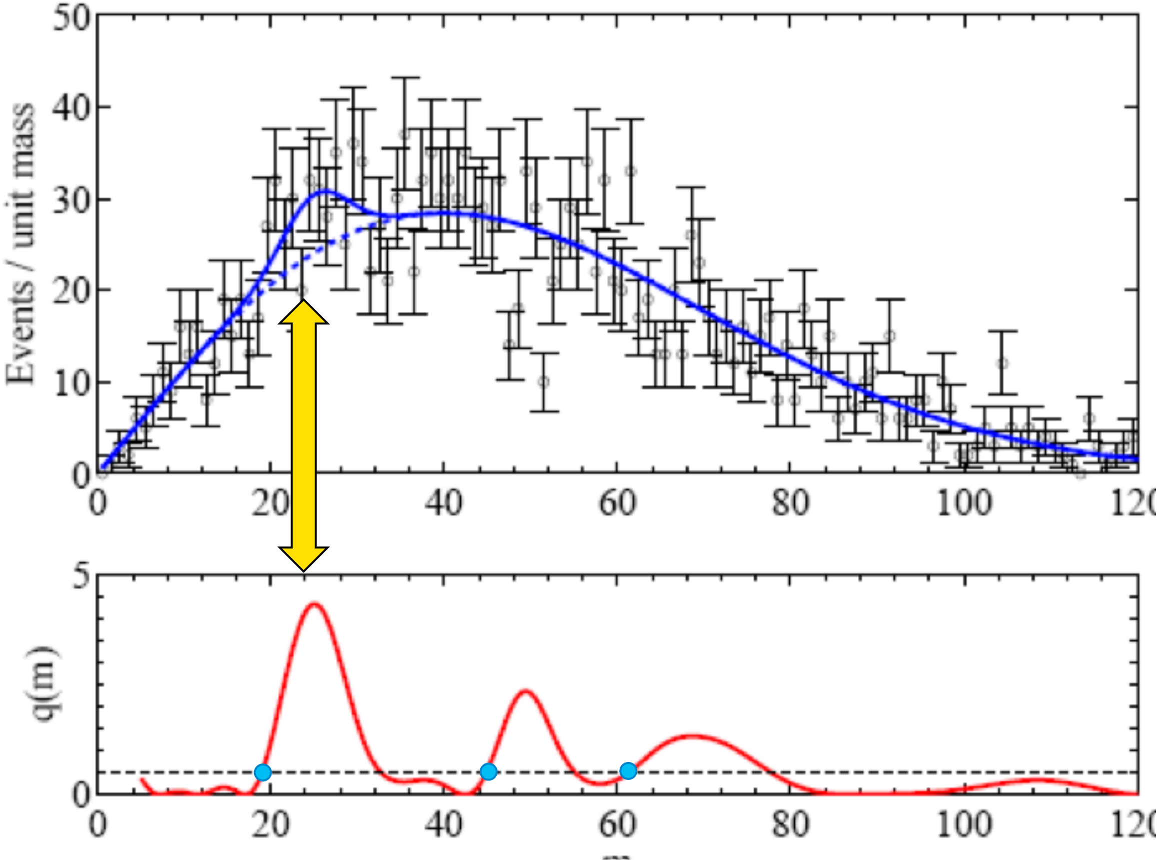
$$\lambda(\mu, m) = -2\log\frac{L(\mu, m)}{L(\hat{\mu}, m)}$$

in the examined range of m, where $c=Z_{fix}^2$ is a threshold set by the significance Z_{fix} corresponding to p_{fix}

- With this approximate relation, can do fixed mass analysis and apply correction factor to get approximate p-value for floating mass scenario.

Counting upcrossings

- Illustration of upcrossings over threshold c of q(m) vs m



Here 3 upcrossings over threshold c=0.5

Counting up-crossings at high Z_{fix}

- Typically, we are interested at Z_{float} when Z_{fix} high, e.g. 5σ.
- Would need <N(c=25)> for the correction factor, but this is very hard to obtain → Very few upcrossings at c=25 in background-only samples
- Can use another approximation

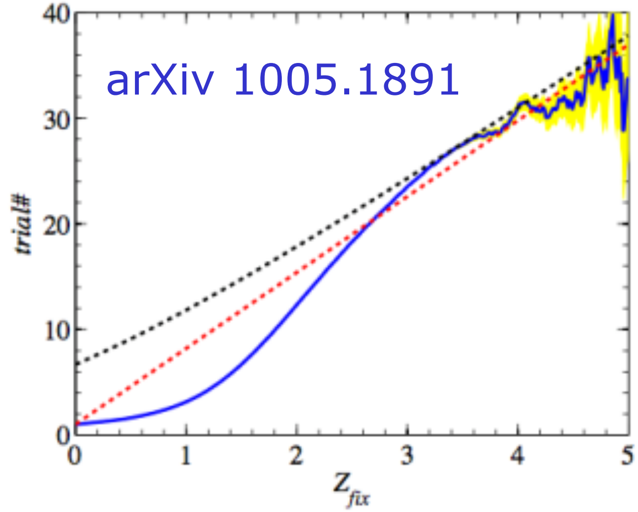
$$\langle N(c) \rangle \approx \langle N(c_0) \rangle e^{-(c-c_0)/2}$$

- Can estimate e.g. N(c=0.5) from background simulation (without need for high statistics), and approximate N (c=25)
- Example: $Z_{fix} = 5\sigma$ → pfix = 2.9 10^{-7}
  <N(0.5)>=8 → N(25) = 8·$e^{-12.25}$ =3.8 10^{-5}
  $p_{float} \approx (p_{fix}=2.9\ 10^{-7}) + 1.4\ 10^{-5} = 3.9\ 10^{-5}$
  $Z_{float} \approx 3.9\sigma$ (trials factor = 134)

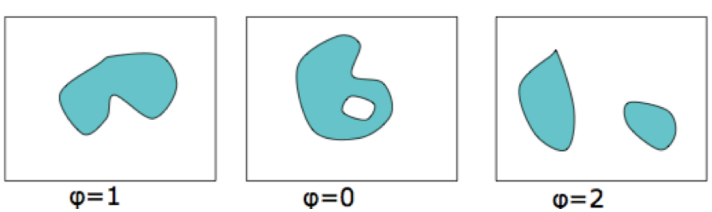
Convergence of approximate trials factor

- Approximation of trials factor due to look-elsewhere-effect works best at high Z_{fix}

$$trial\# \approx 1 + \sqrt{\frac{\pi}{2}} \langle N(c) \rangle Z_{fix}$$



- Note: similar procedure can also be defined for N floating parameters. Replace <#upcrossings> by expectation value of Euler characteristic
  φ = #(disconnected components) - #(holes)



Summary of look-elsewhere-effect

- Look-Elsewhere-Effect occurs when testing a single model (e.g. SM) with multiple observations (different experimental selections, e.g. invariant mass regions)
- There is no LEE when considering exclusion limits - we test specific models and say wether each is excluded
- Approximate LEE should be sufficient for most applications, and both p_{float} and p_{fix} should always be reported

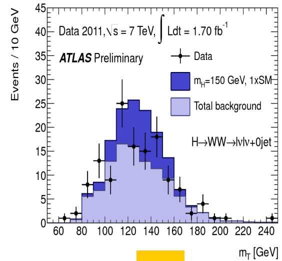
(Higgs) Combinations

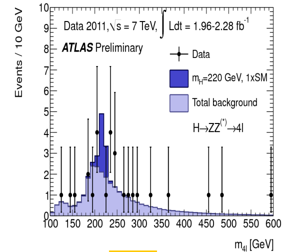
Combining multiple measurement for a limit

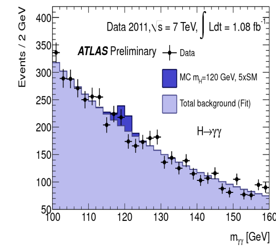
- For many types of (new) physics being investigated, there exist multiple experimental signatures sensitive to the same (new) physics parameters
- Example: Search for Standard Model Higgs boson is conducted in O(10) experimental signatures
  - $H \to ZZ \to llll$
  - $H \to WW \to lqlq$
  - $H \to \gamma\gamma$
  - $H \to bb$
- Each search result is a likelihood $L(\mu,\boldsymbol{\theta})$ that can be used to construct a confidence interval on $\mu=\sigma/\sigma_{SM.}$
- How do you combine the information from all searches into a single (most powerful) confidence interval on μ?

Combining Higgs channels (and experiments)

- Procedure: define joint likelihood







$$L(\mu,\theta_{comb}) = L_{H\rightarrow WW}(\mu,\theta_{WW}) \cdot L_{H\rightarrow ZZ}(\mu,\theta_{ZZ}) \cdot L_{H\rightarrow \gamma\gamma}(\mu,\theta_{\gamma\gamma}) \cdot \ldots$$

$$L(\mu,\theta_{LHC}) = L_{ATLAS}(\mu,\theta_{ATLAS}) \cdot L_{CMS}(\mu,\theta_{CMS}) \cdot \ldots$$

- Correlations between θ_{WW},$\theta_{\gamma\gamma}$ etc and between θ_{ATLAS},θ_{CMS} requires careful consideration!
- The construction profile likelihood ratio test statistic from joint likelihood and proceed as usual

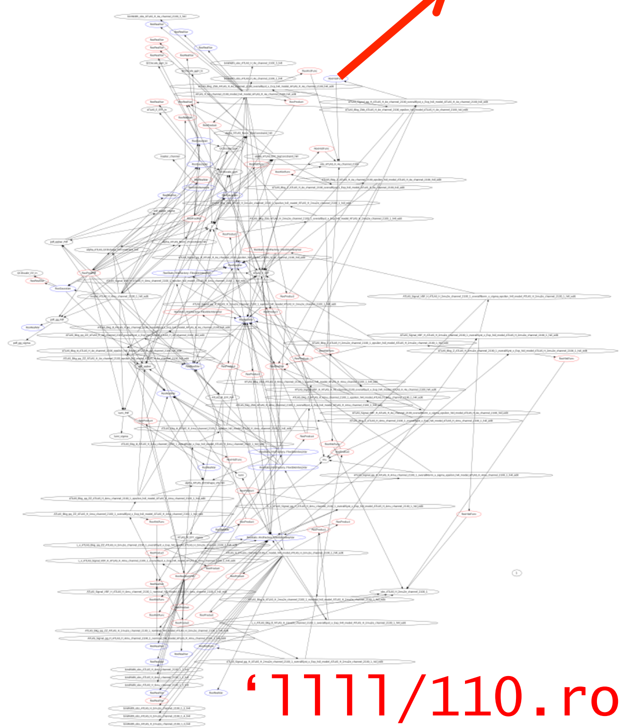
$$\tilde{q}_{\mu} = -2\ln\frac{L(data\,|\,\mu,\hat{\hat{\theta}}_{\mu})}{L(data\,|\,\hat{\mu},\hat{\theta})}$$

Wouter Verkerke, NIKHEF

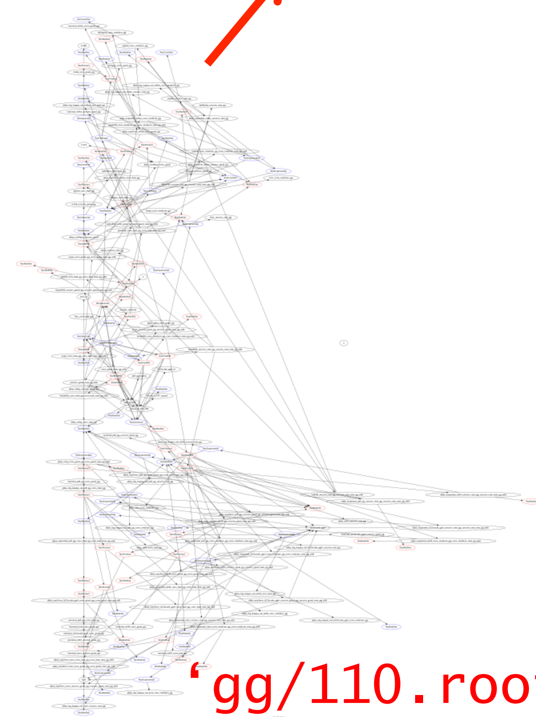
Combinations in RooFit/RooStats

- The workspace concept in RooFit greatly simplifies the practical aspects of such combined model building
- Each (analysis channel/experiment) builds its own probability model for each channel and exports it in a RooFit workspace
- Combined model is built by 'just' combining workspaces

$$L(\mu,\theta_{comb}) = L_{H\to WW}(\mu,\theta_{WW})\cdot L_{H\to ZZ}(\mu,\theta_{ZZ})\cdot L_{H\to\gamma\gamma}(\mu,\theta_{\gamma\gamma})\cdot\ldots$$



'llll/110.root'



'gg/110.root'

Joint distributions in RooFit

- At the likelihood level a joint likelihood is constructed by simply multiplying the component likelihoods

$$L(\mu,\theta_{comb}) = L_{H\rightarrow WW}(\mu,\theta_{WW})\cdot L_{H\rightarrow ZZ}(\mu,\theta_{ZZ})\cdot L_{H\rightarrow\gamma\gamma}(\mu,\theta_{\gamma\gamma})\cdot\ldots$$

- But for frequentist you also need the joint probability model (probability density function) to be able to sample toy experiments. How do you represent a joint measurement as the probability model
- A 'simultaneous pdf' of $f_A(x)$ and $f_B(y)$ can be defined as

  f(x,y,i) = $f_A(x)$ if i=A
  $f_B(y)$ if (i=B)

- From observation sets **x** and **y** joint dataset is constructed D(x,y,i) that contains all values of x and y of the original datasets and a label i that is A or B, indicating the origin of each event

Example of constructing a joint model

- An example constructing a joint model of class RooSimultaneous using the factory operator SIMUL

Create example models for channels A,B

Create index observable

Join models and mapping each to a state of the index observable

```
// Pdfs for channels 'A' and 'B'
w.factory("Gaussian::pdfA(x[-10,10],mean[-10,10],sigma[3])") ;
w.factory("Uniform::pdfB(x)") ;

// Create discrete observable to label channels
w.factory("index[A,B]") ;

// Create joint pdf
w.factory("SIMUL::joint(index,A=pdfA,B=pdfB)") ;
```

Example of constructing a joint dataset

- A joint dataset is made with similar ease

Channel datasets

Index observable
(as already defined in the workspace)

```
RooDataSet *dataA, *dataB ;
RooDataSet dataAB("dataAB","dataAB",Index(w::index),
                  Import("A",*dataA),Import("B",*dataB)) ;
```

Joint dataset

Associate each dataset with
an index observable label

- Can choose between copying channel data [Import()] and linking to channel data [Link()]
  - When linking the joint dataset is just a 'virtual layer' that redirects to the original datasets

Tools exist to import models from other workspaces

- For practical combinations, channel models will be provided in workspaces in separate files
- Convenient tools exist to import such models from file-based workspaces into your own

```
w.importFromFile("atlas_higgs_ZZ_125.root:wspace:hzz_top_model",…) ;
```

filename

workspace object name

model object name

- Specifying the 'top node' of a model will automatically import all components

Tools exist to import models from other workspaces

- Option exist to rename variable, pdf component names, and dataset names to avoid unintended clashes
- NB: When constructing joint models in RooFit:

  same parameter name = same parameter

  i.e. when a parameter 'mu' already exists in the workspace and new model is imported that has its own copy of mu, it will be linked to the existing copy, so that all models in the same workspace share the same mu
- In contrast: when pdf (components) and functions have the same name, this is refused by default, as it is likely a mistake (there is switch to override if this is intentional).
- Tools exist for comprehensive renaming upon import to avoid unintended clashes

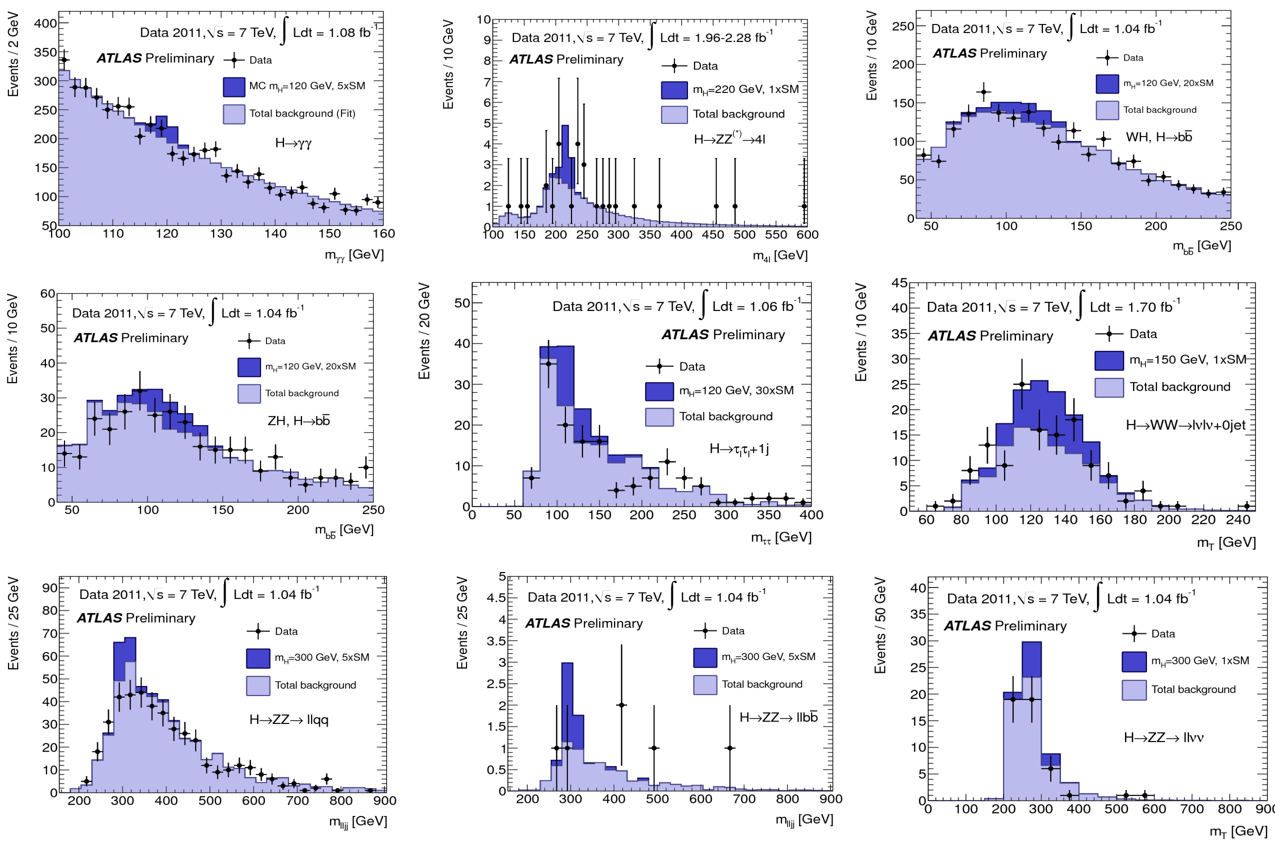
```
w.import(atlasHiggsZZ,
         RenameAllVariablesExcept("mHiggs","aHZZ"),
         RenameVariable("HiggsXS","mu") ;
```

Things to worry about in combinations

- Just focused on technical aspects of joint model building.
- Construction of joint models present vast array of difficult questions to answer in terms of common nuisance parameters between channels and experiments
  - Is the jet-energy-scale between H→ZZ and H→WW correlated?
  - Is the jet-energy-scale between ATLAS and CMS correlated?
  - Is the W+jets cross-section (background) correlated between H→WW (lqlq) and H→ZZ(llqq)
- Remember that frequentist use of likelihood treats all contributions on similar footing:
  main measurement *and* subsidiary measurements.
- E.g is it OK if main measurement of H→ZZ constrains nuisance parameter for JES that is also used in H→WW more than the subsidiary measurement that was introduced for H→WW?

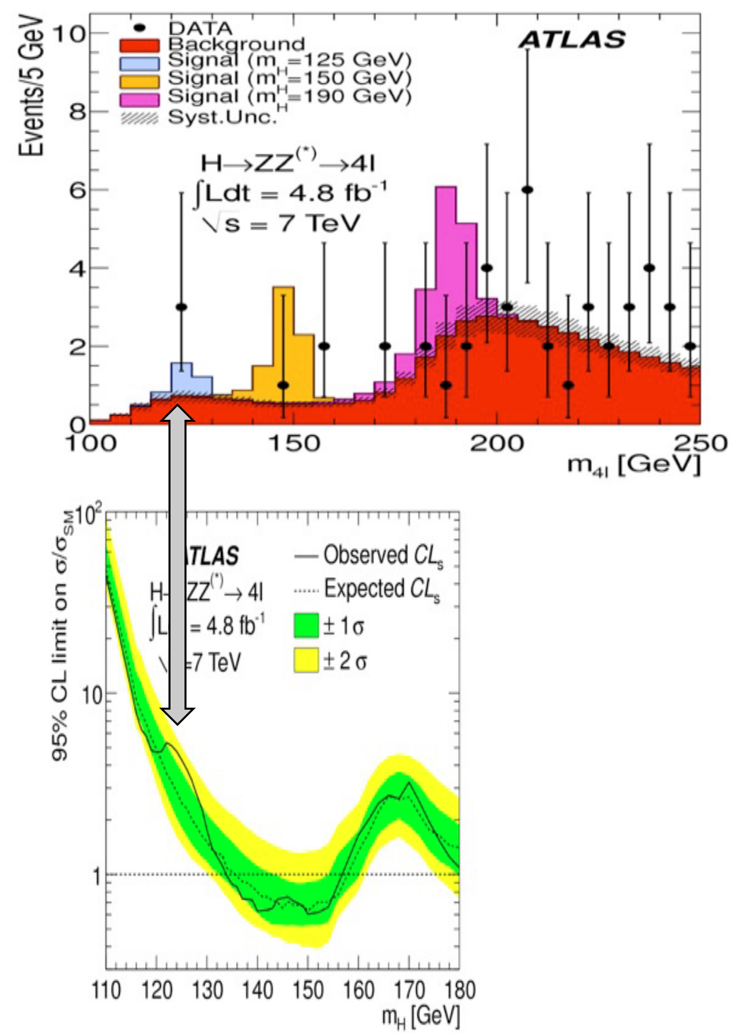
Putting it all together – The Higgs search as example

- 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)



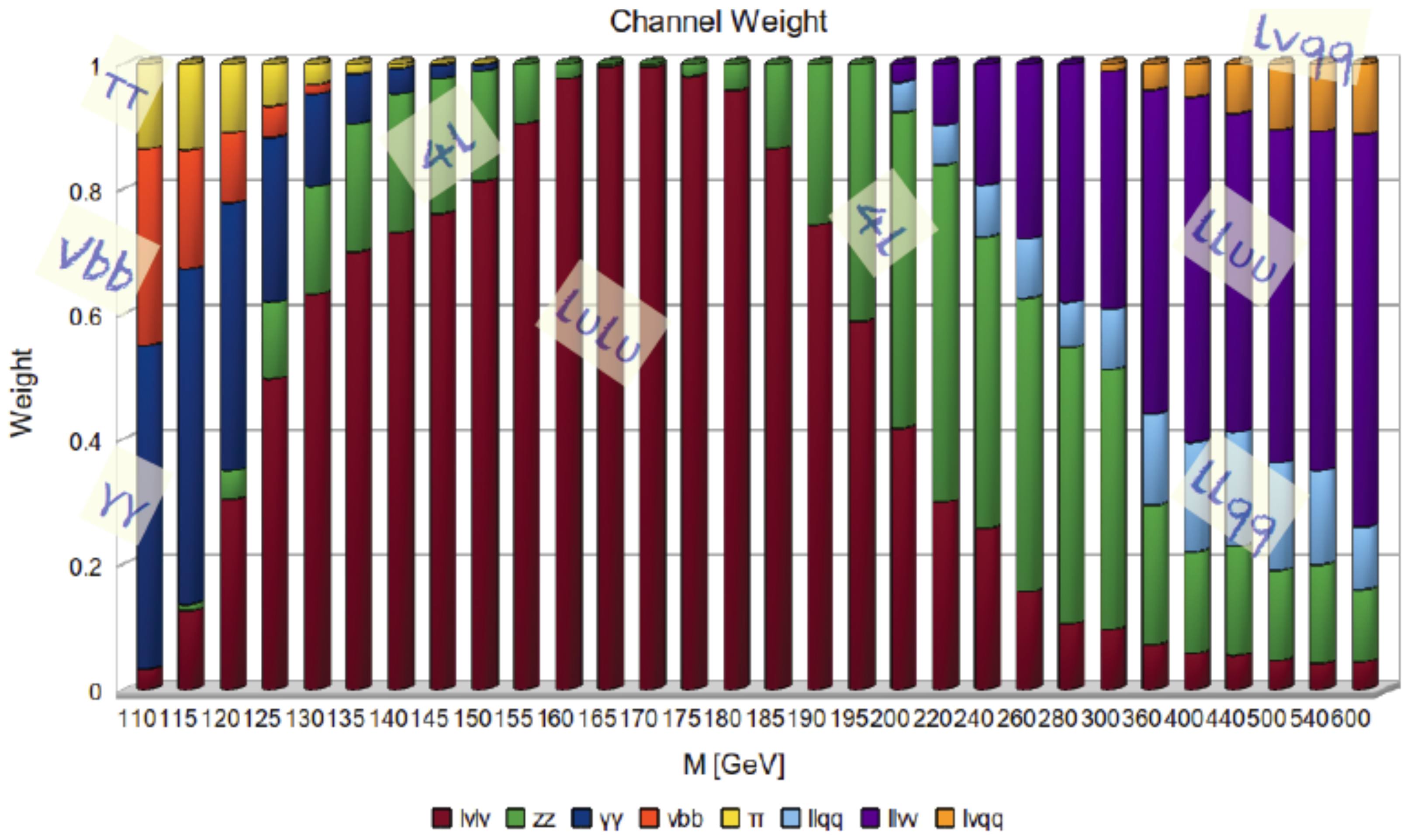
Each channel is sensitive to a different range of m_H

- Can execute limit-setting procedure for each channel separately
- Limit on $\mu = \sigma/\sigma_{SM}$ always calculated at ***fixed*** m_H
  - Trials factor will be discussed later
- For each m_H calculate
  - Observed limit (using CL_S)
  - Expected limits and 1,2σ bands
  - p-value of background hypothesis (will show later)



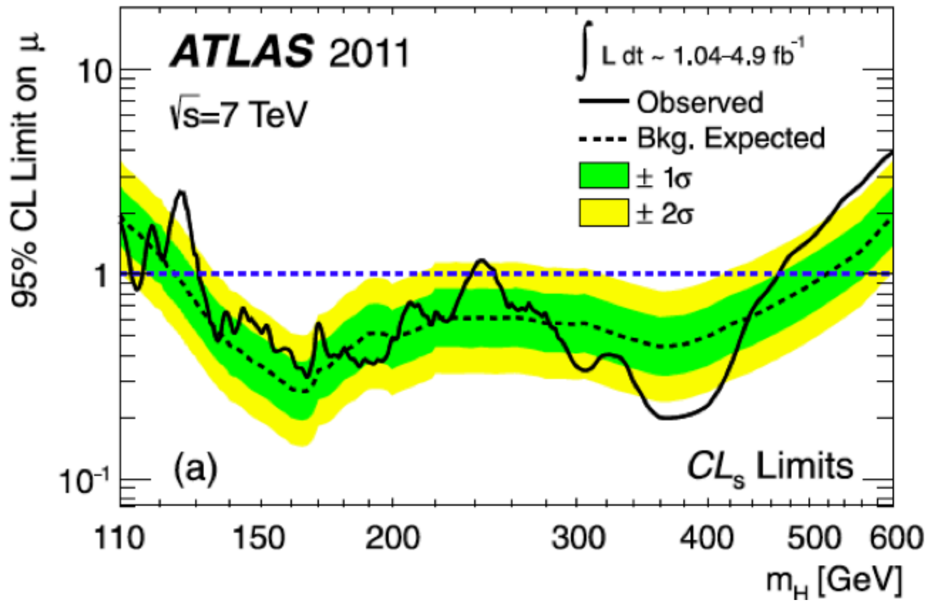
Then combine all channels

- Different channels contribute in different ranges of m_H...



How to read the upper limit plot

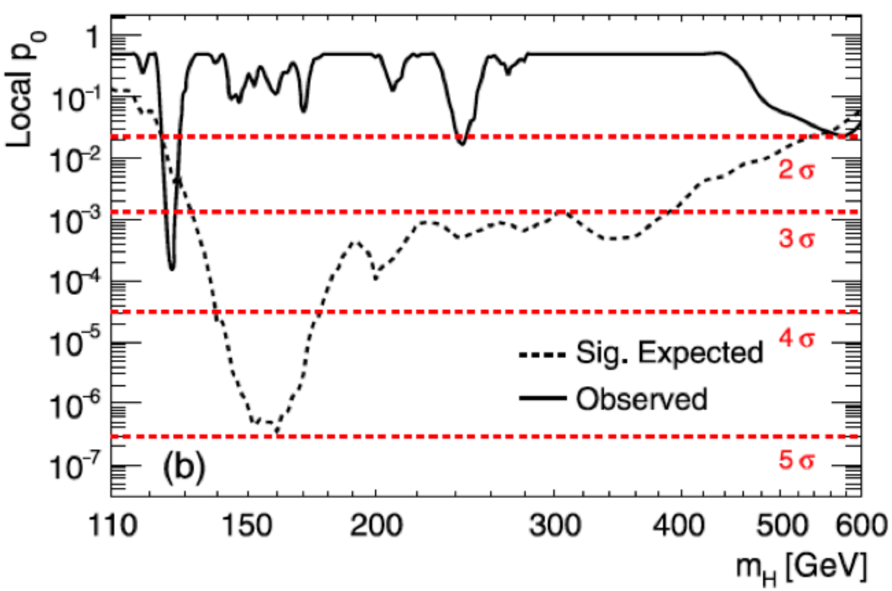
- For each value of m_H, find the CL_S upper limit on μ
- For each m_H, determine distribution of upper limits μ_{up} one would obtain under the hypothesis μ=0
  - The dashed curve is the median of μ_{up} and the green(yellow) and give the ±1σ(2σ) regions of this distribution
- Range(s) of m_H for which $\mu_{up} < 1$ is region where we *expect* to be able to exclude SM Higgs boson (at μ=1)
- Range(s) of m_H where observed limit is <1 is what we actually exclude



ATLAS, Phys. Lett. B 710 (2012) 49-66

How to read the p0 plot

- The "local" p0 means the p-value for the background-only hypothesis obtained from the test of μ=0 at each individual m_H, without any correction for the LEE
- The expected curve gives the median p0 under the assumption of the SM Higgs (μ=1) at high m_H
  - Low expected p_0 means high observation sensitivity, (***not*** higher probability of the Higgs having this mass)



ATLAS, Phys. Lett. B 710 (2012) 49-66

How to read the "blue band" plot

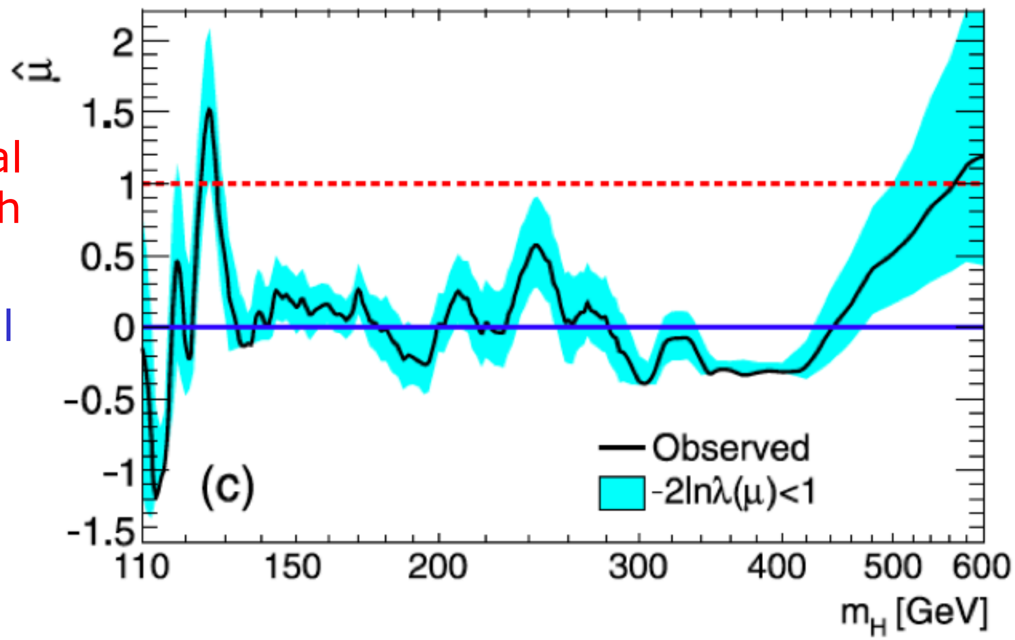
- On the plot of μ-hat (fit to σ/σ_{SM}) versus m_H, the blue band is defined by

$$\ln L(\mu) > \ln L(\hat{\mu}) - \frac{1}{2}$$

"Minos error" (≈ 68% C.L.)

SM signal strength

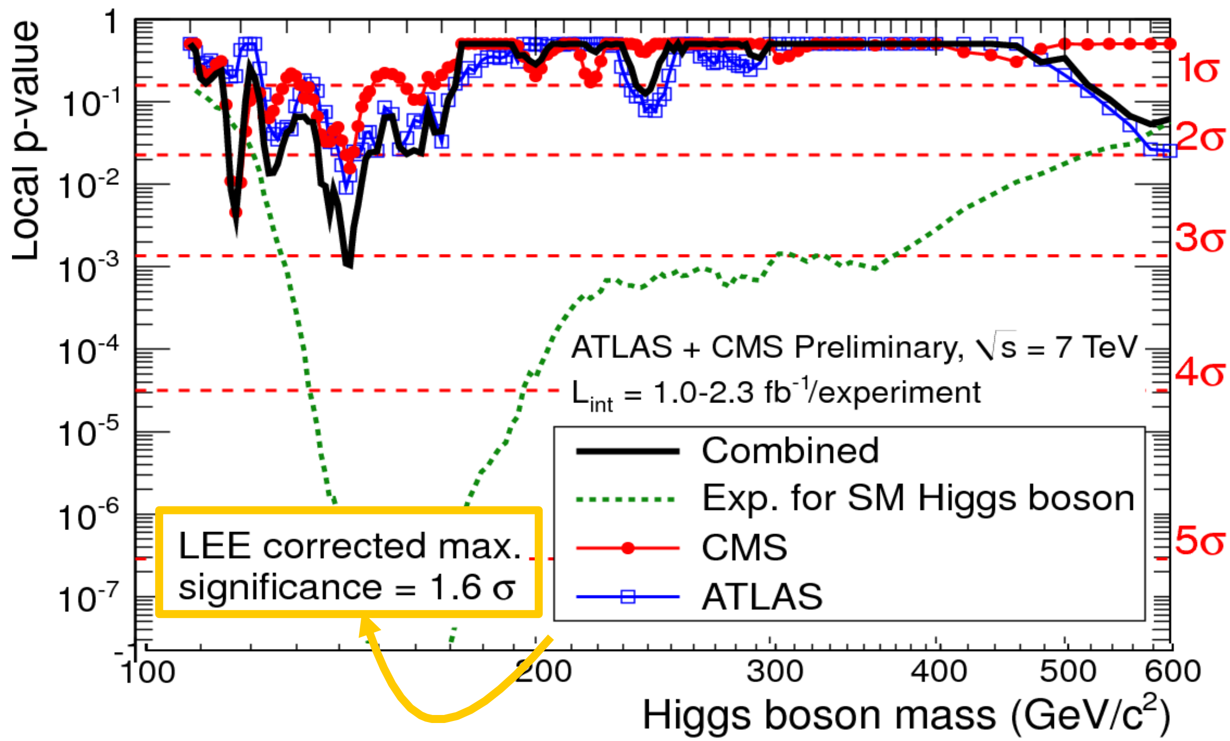
No signal



ATLAS, Phys. Lett. B 710 (2012) 49-66

The Look Elsewhere Effect

- Indications of the global p-value (correction for the look-elsewhere effect), if given is calculated from the "upcrossing counting method"
  - Keep in mind that this approximation works best for high Z_{local}
  - Example below for preliminary ATLAS+CMS Higgs combination performed in Fall 2011



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*

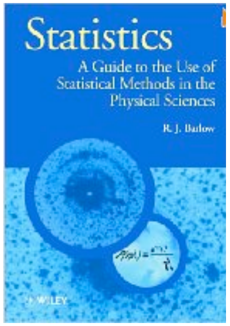
- Intermediate
  - Glen Cowan, Statistical Data Analysis (Solid foundation for HEP)
  - Frederick James, *Statistical Methods in Experimental Physics, World Scientific,* 2006. (This is the second edition of the influential 1971 book by Eadie et al., has more advanced theory, many examples)

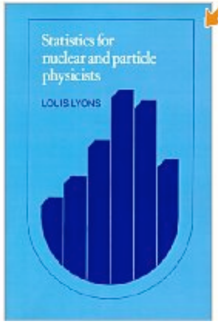
- Advanced
  - A. Stuart, K. Ord, S. Arnold, Kendall's Advanced Theory of Statistics, Vol. 2A, 6th edition, 1999; and earlier editions of this "Kendall and Stuart" series. (Authoritative on classical frequentist statistics)

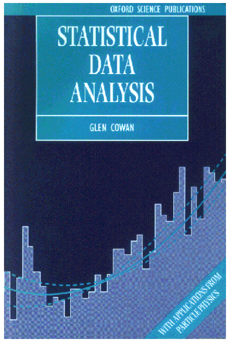
- Recent papers (covered in these lectures)
  - Asymptotic Distributions

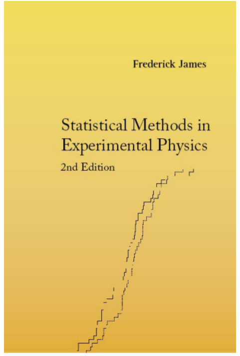
    Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554
  - Look-Elsewhere-Effect

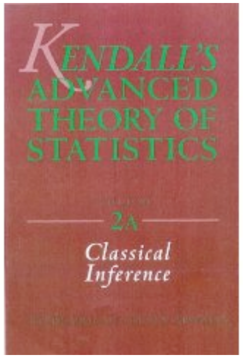
    arXiv 1005.1891











Exercise A

- Repeat the limit setting exercises of module 3 with the Asymptotic Calculator instead of the Frequentist calculator (use calculatorType=3) and compare the results.