A Lecture on some (tiny) Aspects of Statistics

(as used in HEP

and especially LHC)

Philip Bechtle
Physikalisches Institut, Universität Bonn

March 13th 2014

universitätbonn

1. Introduction

2. Interpretation of Statistics

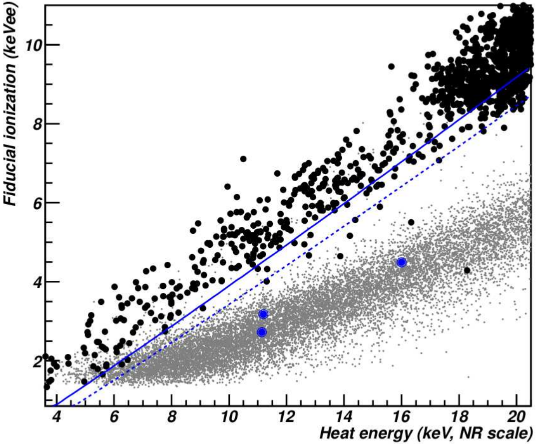
3. Limits and Measurements at LHC
   - Introductory Example
   - The Profile Likelihood Technique at the LHC
   - Examples from the Results

With inspiration (and some material) from very nice talks from

Roger Barlow, Alex Read, Kyle Cranmer

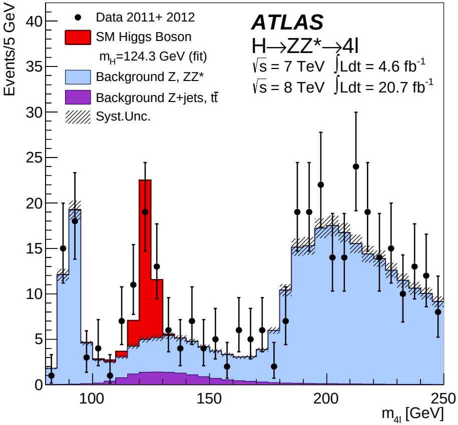
The Task

- Statistics can be used for very many purposes
- I guess here we are most concerned about
  - Finding or excluding a signal
  - Determining uncertainties



The Task

- Statistics can be used for very many purposes
- I guess here we are most concerned about
  - Finding or excluding a signal
  - Determining uncertainties



The Definition of the Probability

- For most of the talk: Define Probability P of X as

$$P(X) = N(X)/N \quad \text{for} \quad N \to \infty$$

  Examples: coins, dice, cards

- For continuous x extend to Probability Density

$$P(x \,\text{to}\, x + \mathrm{d}x) = p(x)\mathrm{d}x$$

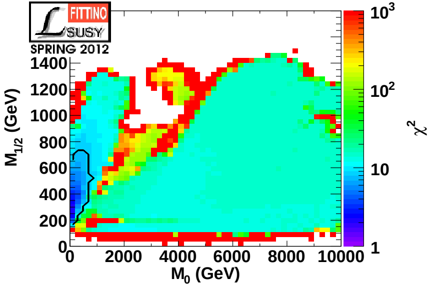
  $p(x)$ is the *probability density function (pdf)*

- Examples:
  - Measuring continuous quantities ($p(x)$ often Gaussian, Poisson, ...)
  - Counting rates
  - Physical Quantities: Parton momentum fractions (proton pdfs) ...
- Alternative: Define Probability $P(X)$ as "degree of belief that X is true"

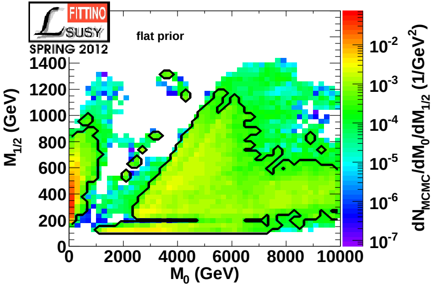
The Likelihood

- Probability distribution of random variable x often depends on some parameter a
- Joint function $p(x, a)$:
  - Considered as $p(x)|a$ this is the pdf.
  - Normalised: $\int p(x)\mathrm{d}x = 1$
  - Considered as $p(a)|x$ this is the Likelihood $L(a)$ (or $\mathcal{L}(a)$)
  - Not "likelihood of a" but "likelihood that a would give x"
  - Not normalised. Indeed, must never be integrated.
- This is going to be one of the central concepts/quantities for the rest of the talk
- If we want to know a parameter a, we are looking for the point where the likelihood that a would predict the data x is maximized
- If we want to test a Hypothesis H_0 against another one (H_1), we want to compare their likelihoods
- If we want to know what a cannot be, we want to know where $\mathcal{L}(a)|x$ is small

Examples for Frequentist and Bayesian Interpretations



Frequentist Profile Likelihood



Bayesian, Flat prior

- Quantify the agreement between each model point and the data:

$$\chi^2 = \sum_{i=1}^{n_{Obs}} \frac{(M_i - O_i(\vec{P}))^2}{\sigma_i^2} + Constraints$$

- Advanced MCMC scans with automatically adapting proposal density width

Frequentist Reasoning

It's pretty simple, I think:

- Probability of an event is the relative frequency of its occurrence

Frequentist Reasoning

It's pretty simple, I think:

- Probability of an event is the relative frequency of its occurrence
- Need something which (at least in a simulation) can in principle repeated indefinitely, otherwise there exists no probability

Frequentist Reasoning

It's pretty simple, I think:

- Probability of an event is the relative frequency of its occurrence
- Need something which (at least in a simulation) can in principle repeated indefinitely, otherwise there exists no probability
- Since the universe can't be repeated (we don't know how to simulate its genesis before the big bang, therefore the parameters of the Universe are not random variables): there exists no probability density in theory/parameter space

Frequentist Reasoning

It's pretty simple, I think:

- Probability of an event is the relative frequency of its occurrence
- Need something which (at least in a simulation) can in principle repeated indefinitely, otherwise there exists no probability
- Since the universe can't be repeated (we don't know how to simulate its genesis before the big bang, therefore the parameters of the Universe are not random variables): there exists no probability density in theory/parameter space
- Therefore, the only statements we can make are:
  If theory H is true (which we will *never* know), then the probability of the observed outcome D of our experiment $P(D|H)$ is…

Frequentist Reasoning: Examples

- Can't say
  "It will probably rain tomorrow."
  There is only one tomorrow. P is either 1 or 0
- Have to say
  "The statement 'It will rain tomorrow.' is probably true."
  Can then even quantify (meteorology).

Frequentist Reasoning: Examples for interpeting physics results

- Can't say
  "m_t has a 68% probability of lying between 171 and 175 GeV"
- Have to say "The statement 'm_t lies between 171 and 175 GeV' has a 68% probability of being true"
- Be aware:
  - In this context, a certain value of m_t has no probability. It is either true or false.
  - But the interval $[171, 175]$ depends on the data and does fluctuate. If you repeat the experiment, you will get different intervals each time, and 68% of them should cover the invariant true value.
- if you always say a value lies within its error bars, you will be right 68% of the time
- Say "m_t lies between 171 and 175 GeV" with 68% Confidence. Or 169 to 177 with 95% confidence.
- That is the Confidence Level CL

Bayesian Reasoning

- “I’m a frequentist”, thus probably I cannot do justice to Bayesian resoning, even though I try

Bayesian Reasoning

- “I’m a frequentist”, thus probably I cannot do justice to Bayesian resoning, even though I try
- Mathematically, Bayes theorem is unquestioned and simple:

$$P(H|D) = \frac{P(D|H)}{P(D)} P(H)$$

$$P(D) = \sum_{i=1}^{i<n} P(D|H_i) P(H_i)$$

with

  - $P(H|D)$: “Posterior”, belief in H given D
  - $P(D|H)$: “Likelihood”, probability of D given H
  - $P(H)$: “Prior” belief in H, given nothing
  - $P(D)$: “Evidence”: believe in D, given all possible hypotheses H

My Thoughts on Bayesian Reasoning

- The math is certainly fine.

My Thoughts on Bayesian Reasoning

- The math is certainly fine.
- There is no problem if
  - You can "measure" $P(H)$ (e.g. systematics measured elsewhere). Probabaly this is not always easy, and there might be guesswork involved, but at least one can get an objective hint.
  - You know $P(H)$ from your model (but wait… where in physics do we know the model? Then we would not need to do physics…)

My Thoughts on Bayesian Reasoning

- The math is certainly fine.
- There is no problem if
  - You can "measure" $P(H)$ (e.g. systematics measured elsewhere). Probabaly this is not always easy, and there might be guesswork involved, but at least one can get an objective hint.
  - You know $P(H)$ from your model (but wait... where in physics do we know the model? Then we would not need to do physics...)
- I have big problems with the physical and philosophical meaning of what is done here if applied to fundamental parameters

Bayesian Reasoning – oh, why?

- Given this – in my very personal view – philosophical mess: Why is it used by many people?

Bayesian Reasoning – oh, why?

- Given this – in my very personal view – philosophical mess: Why is it used by many people?

Stolen from Ben Farmer and Martin White

- $P(\text{ball under cup i}) = P(i) = 1/1000$
  Assume we looked under 999 cups. No ball found (D)!

$$P(1000|D) = \frac{P(D|1000)}{\sum_{i=1}^{i\leq 1000} P(D,i)P(i)} P(1000)$$

$$P(1000|D) = \frac{1}{999 \times 0 + 1 \times 1/1000} \times \frac{1}{1000} = 1$$

Bayesian Reasoning – oh, why?

- Given this – in my very personal view – philosophical mess: Why is it used by many people?

Stolen from Ben Farmer and Martin White

- $P(\text{ball under cup i}) = P(i) = 1/1000$
  Assume we looked under 999 cups. No ball found (D)!

$$P(1000|D) = \frac{P(D|1000)}{\sum_{i=1}^{i\leq 1000} P(D,i)P(i)} P(1000)$$

$$P(1000|D) = \frac{1}{999 \times 0 + 1 \times 1/1000} \times \frac{1}{1000} = 1$$

- In an "experiment" where the "theory" consists of a fixed number of known individually testable basic theorems this is fine.

Bayesian Reasoning – oh, why?

- But on second thought, isn't that result from the previous slide a bit strange?

Bayesian Reasoning – oh, why?

- But on second thought, isn't that result from the previous slide a bit strange?
- We have a hypothesis "There is exactly one ball under one of the 1000 cups".
- We look under 999 cups and don't find a ball
- We do the bayesian math and realize that the fact that we did not find the ball in 99.9 % of the possible experiments increased or belief that we will find it in the next trial

Bayesian Reasoning – oh, why?

- But on second thought, isn't that result from the previous slide a bit strange?
- We have a hypothesis "There is exactly one ball under one of the 1000 cups".
- We look under 999 cups and don't find a ball
- We do the bayesian math and realize that the fact that we did not find the ball in 99.9 % of the possible experiments increased or belief that we will find it in the next trial
- Of course this is mathematically correct
- But in reality we never are sure about our hypothesis. So we cannot use the prior, and thus we cannot say that not fining the ball can increase our degree of belief that we will find it next time.

Bayesian Reasoning – why not?

- Can we apply this to the belief in the CMSSM parameter point x?

$$P(CMSSM\,x|D) = \frac{P(D|CMSSM\,x)}{\sum_{i=1}^{\infty} P(D|CMSSM\,i)P(CMSSM\,i)} P(CMSSM\,x)$$

Bayesian Reasoning – why not?

- Can we apply this to the belief in the CMSSM parameter point x?

$$P(CMSSM\,x|D) = \frac{P(D|CMSSM\,x)}{\sum_{i=1}^{\infty} P(D|CMSSM\,i)P(CMSSM\,i)} P(CMSSM\,x)$$

no, we clearly can't; We can't execute the sum, and we don't know anything about $P(CMSSM\,x)$, it does not exist for philosophical reasons! Our theories are *never* finite sums of basic theorems

Bayesian Reasoning – why not?

- Can we apply this to the belief in the CMSSM parameter point x?

$$P(CMSSM\,x|D) = \frac{P(D|CMSSM\,x)}{\sum_{i=1}^{\infty} P(D|CMSSM\,i)P(CMSSM\,i)} P(CMSSM\,x)$$

  no, we clearly can't; We can't execute the sum, and we don't know anything about $P(CMSSM\,x)$, it does not exist for philosophical reasons! Our theories are *never* finite sums of basic theorems

- So can we apply it to answer the question: How unlikely did the CMSSM become relative to the SM, given that we found no SUSY?

$$\frac{P(CMSSM|D)}{P(SM|D)} = \frac{P(D|CMSSM)}{P(D|SM)} \frac{P(CMSSM)}{P(SM)}$$

  It looks like we at least could say something about how the present data D modifies our prior belief in the CMSSM and the SM

Bayesian Reasoning – why not?

- Can we apply this to the belief in the CMSSM parameter point x?

$$P(CMSSM\,x|D) = \frac{P(D|CMSSM\,x)}{\sum_{i=1}^{\infty} P(D|CMSSM\,i)P(CMSSM\,i)} P(CMSSM\,x)$$

  no, we clearly can't; We can't execute the sum, and we don't know anything about $P(CMSSM\,x)$, it does not exist for philosophical reasons! Our theories are *never* finite sums of basic theorems

- So can we apply it to answer the question: How unlikely did the CMSSM become relative to the SM, given that we found no SUSY?

$$\frac{P(CMSSM|D)}{P(SM|D)} = \frac{P(D|CMSSM)}{P(D|SM)} \frac{P(CMSSM)}{P(SM)}$$

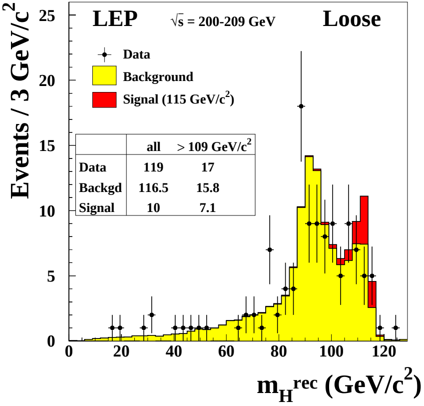
  It looks like we at least could say something about how the present data D modifies our prior belief in the CMSSM and the SM

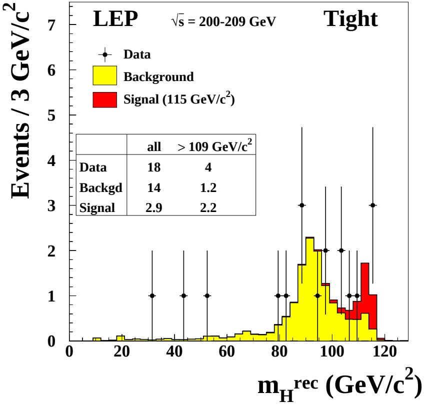
- Well, actually no:

$$P(D|CMSSM) = \prod_i \int_{\theta_i} P(D|CMSSM\,\theta_i)P(CMSSM\,\theta_i)\mathrm{d}\theta_i$$

Do we see a Higgs mass peak? Use LEP for simplcity

- Are there many of these candidates?

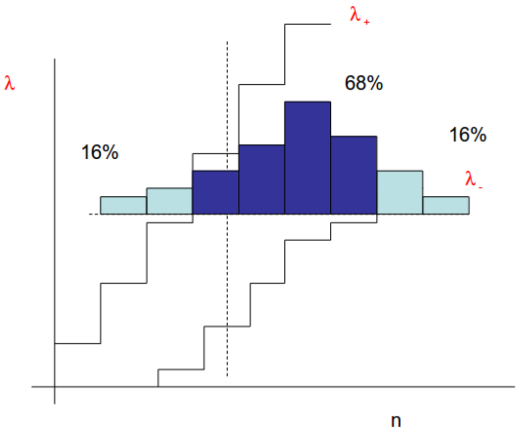




- How significant is the small excess? Need advanced statistical analysis

The Neyman Interval

- Let's neglect systematics for the time being . . .
- Use Poisson-Distribution $p(n; \lambda) = e^{-\lambda}\lambda^n/n!$
- For any true λ the probability that $(n|\lambda)$ is within the belt is 68% (or more) by construction
- For any n, $[\lambda_-, \lambda_+]$ covers the true λ at 68% confidence
- Only integrated over n, not over λ!

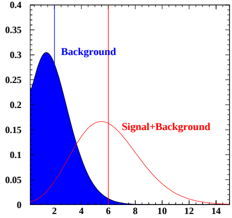


Technique technically works for every CL, and single or double sided

Getting the most out of the availale events?

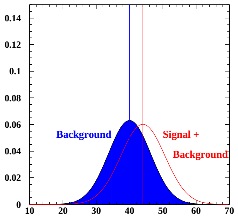
- If hypothesis exists with $d \approx s+b$ on a significant level: Higgs found
- If not: Calculate, how improbable d is under a certain hypothesis s: $\rightarrow$ exclusion
- First example: Add all s, b, d of all channels (Counting Experiment)
- If $s \neq 0$ only in one channel: this degrades sensitivity

Poisson-distributions for s=4,b=2



b=2, s=4

Poisson-distributions for s=4,b=40



b=40, s=4

Not the most sensitive method . . .

Avoiding a big problem?

- Observe $d = 5$ events. Expected background b of 0.9 events
  Data d = signal s + background b
- Say with 68% confidence: $[2.84, 8.38]$ covers $s + b$
- So say with 68% confidence: $[1.94, 7.48]$ covers s

Avoiding a big problem?

- Observe $d = 5$ events. Expected background b of 0.9 events
  Data $d =$ signal $s +$ background b
- Say with 68% confidence: $[2.84, 8.38]$ covers $s + b$
- So say with 68% confidence: $[1.94, 7.48]$ covers s
- Suppose expected background 10.9 events?
  "We say at 68% confidence that $[-8.06, -2.52]$ covers s with 68% CL"

Avoiding a big problem?

- Observe $d = 5$ events. Expected background b of 0.9 events
  Data $d =$ signal $s +$ background b
- Say with 68% confidence: $[2.84, 8.38]$ covers $s + b$
- So say with 68% confidence: $[1.94, 7.48]$ covers s
- Suppose expected background 10.9 events?
  "We say at 68% confidence that $[-8.06, -2.52]$ covers s with 68% CL"
- This is technically correct. We are allowed to be wrong 32% of the times

Avoiding a big problem?

- Observe $d = 5$ events. Expected background b of 0.9 events
  Data d = signal s + background b
- Say with 68% confidence: $[2.84, 8.38]$ covers $s + b$
- So say with 68% confidence: $[1.94, 7.48]$ covers s
- Suppose expected background 10.9 events?
  "We say at 68% confidence that $[-8.06, -2.52]$ covers s with 68% CL"
- This is technically correct. We are allowed to be wrong 32% of the times
  While it is mathematically correct, it makes no sense physically

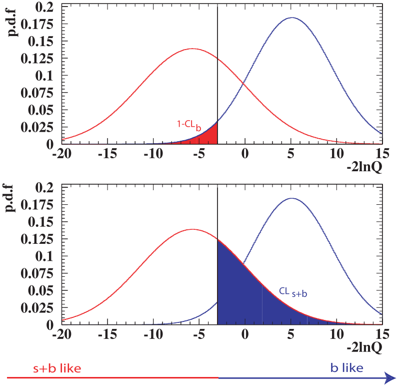
Avoiding a big problem?

- Observe $d = 5$ events. Expected background b of 0.9 events
  Data $d =$ signal $s +$ background b
- Say with 68% confidence: $[2.84, 8.38]$ covers $s + b$
- So say with 68% confidence: $[1.94, 7.48]$ covers s
- Suppose expected background 10.9 events?
  "We say at 68% confidence that $[-8.06, -2.52]$ covers s with 68% CL"
- This is technically correct. We are allowed to be wrong 32% of the times
  While it is mathematically correct, it makes no sense physically
- We know that the background happens to have a downward fluctuation. How can we incorporate that knowledge?

We assume *here* that the background is calculated correctly
Deal with systematics later using nuisance parameters

A simple choice of a better test statistics Q

- For optimal sensitivity, do just not add the total channel contents
  but use the information of full (mass) distributions
- Define the test statistics Q as a likelihood ratio
  $Q = \prod_i P_{d_i}(s_i + b_i)/P_{d_i}(b_i)$
- Define $1 - \mathrm{CL}_b$: Probability of a b-experiment to give a less background like result than the observed one
- Define CL_{s+b}: Probability of a s+b-experiment to give a more background like result than the observed one



Conservative limit:
$\mathrm{CL}_s = \mathrm{CL}_{s+b}/\mathrm{CL}_b$

The Likelihood Ratio: Neyman-Pearson-Lemma

- We are performing a hypothesis test between two hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$
- the likelihood-ratio test which rejects H_0 in favour of H_1 when the test statistics

$$Q(d) = \frac{L(d|\theta_0)}{L(d|\theta_1)} \leq \eta$$

  with

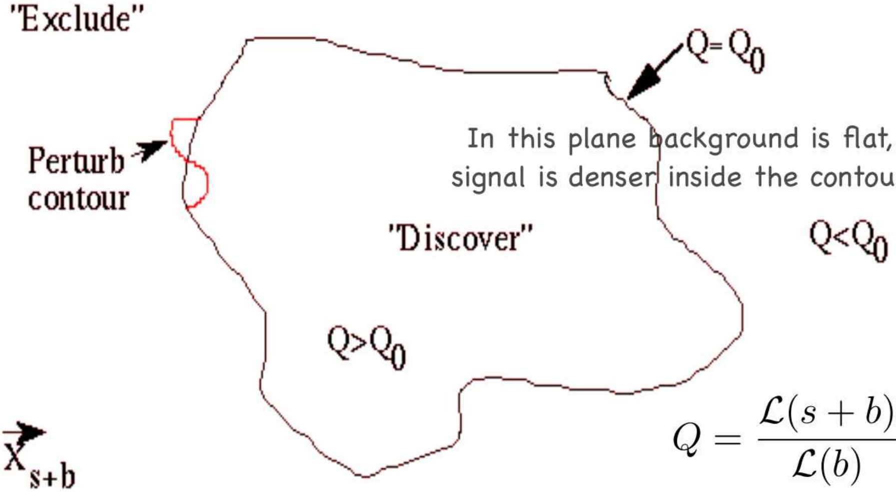
$$P(Q(d) \leq \eta \mid H_0) = \alpha$$

  is the most powerful test of size α
- What does that mean? And what are H_0 and H_1?

The Likelihood Ratio: Neyman-Pearson-Lemma

- We are performing a hypothesis test between two hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$
- the likelihood-ratio test which rejects H_0 in favour of H_1 when the test statistics
$$Q(d) = \frac{L(d|\theta_0)}{L(d|\theta_1)} \leq \eta$$
with
$$P(Q(d) \leq \eta \mid H_0) = \alpha$$
is the most powerful test of size α
- What does that mean? And what are H_0 and H_1?
- We want α ("Type I" error) very small
- We want the power
$$P(\text{reject } H_0 | H_0 \, \text{is false}) = \beta$$
to be as large as possible. $1 - \beta$ is the "Type II" error.

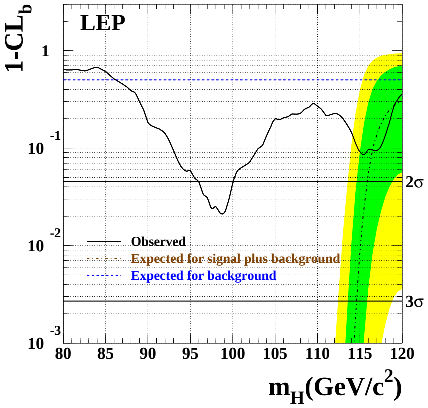
The Likelihood Ratio: Neyman-Pearson-Lemma



In this plane background is flat, signal is denser inside the contour

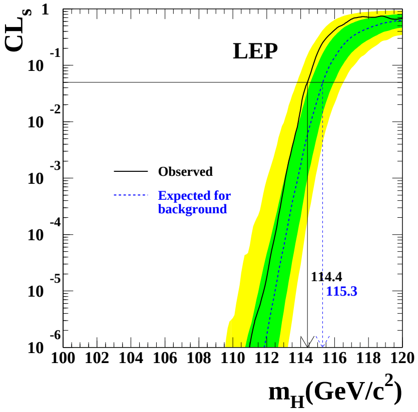
$$Q = \frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}$$

Is there a Significant Excess?



- $(1 - CL_b)$ is a measure of the 'background-likeness' of an experiment. If $(1 - CL_b)$ is e.g. 5 %, then the probability of this outcome to be caused by a fluctuation of the background is 5 %
- No excess above 3σ
- Be aware of the 'look-elsewhere' effect!

No Significant Excess: What's the Limit?



- CL_s is a measure of how signal-like the outcome of an experiment is. If CL_s is small, it is very unlikely that there is a signal. Hence, a 95 % CL corresponds to $CL_s = 0.05$
- Final word from LEP on the SM Higgs:

$$m_h > 114.4\,\mathrm{GeV}$$

Developments since LEP: Profile Likelihood

- Already at LEP: The important thing is to split the the statistics in bins with high s_i/b_i and low s_i/b_i
- New: Introduce signal strength scaling parameter μ
- Assume you measure d_i and try to explain it with $\mu s_i + b_i$ as assumed expectation values
- In addition, measure m_k background bins and try to explain with $u_k(\vec{\theta})$ as expectation value

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \quad \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

- Significance test is based on profile likelihood test statistics:

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

maximizes L for specified μ

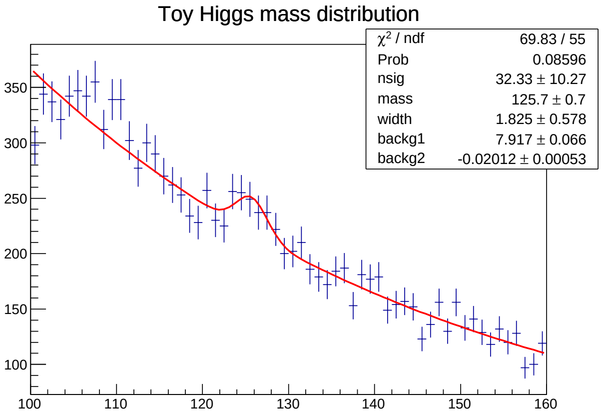
maximize L

See how this is similar to a fit?

The Profile Likelihood Technique in a fit

- In a fit to measurements $\vec{x}$, you vary the parameters $\vec{a}$ and either maximize the Likelihood $\ln \mathcal{L}(\vec{x};\vec{a})$ (or minimize the χ^2)
- In special cases:

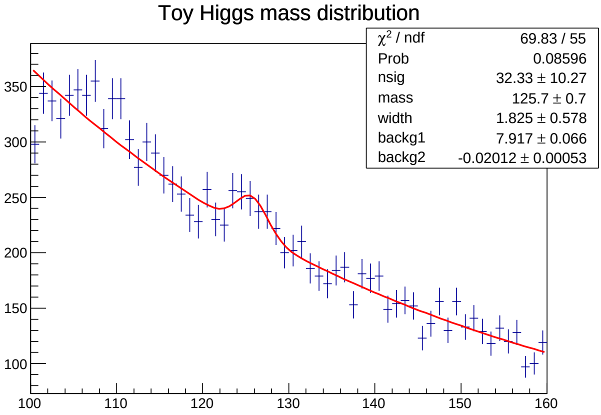
$$-2\ln \mathcal{L} = \chi^2 = (\vec{x} - \vec{\bar{x}}(\vec{a}))^T C^{-1} (\vec{x} - \vec{\bar{x}}(\vec{a}))$$



The Profile Likelihood Technique in a fit

- In a fit to measurements $\vec{x}$, you vary the parameters $\vec{a}$ and either maximize the Likelihood $\ln \mathcal{L}(\vec{x};\vec{a})$ (or minimize the χ^2)
- In special cases: (and no correlations)

$$-2\ln \mathcal{L} = \chi^2 = \sum_i \frac{(x_i - \bar{x}_i(\vec{a}))^2}{\sigma_i^2}$$

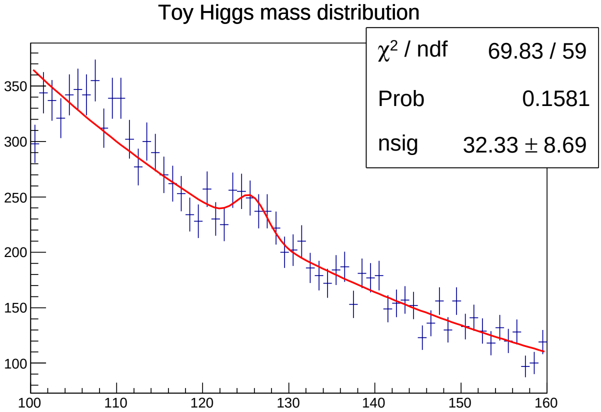


The Profile Likelihood Technique in a fit

- In the above fit, the uncertainty on the number of signal events seems to be larger than the poisson uncertainty $\sqrt{N}$. Why?

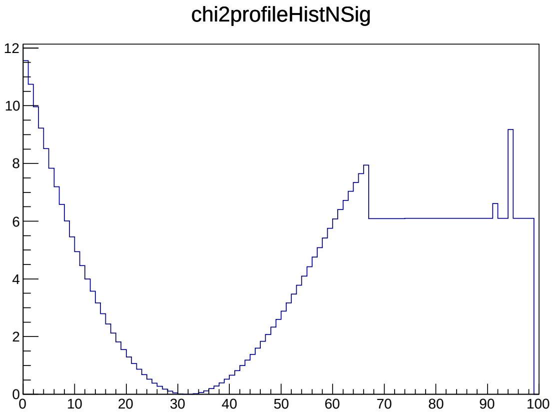
The Profile Likelihood Technique in a fit

- In the above fit, the uncertainty on the number of signal events seems to be larger than the poisson uncertainty $\sqrt{N}$. Why?
- Obviously that is because there is an uncertainty on the background model. Let's fix everything apart from NSig:



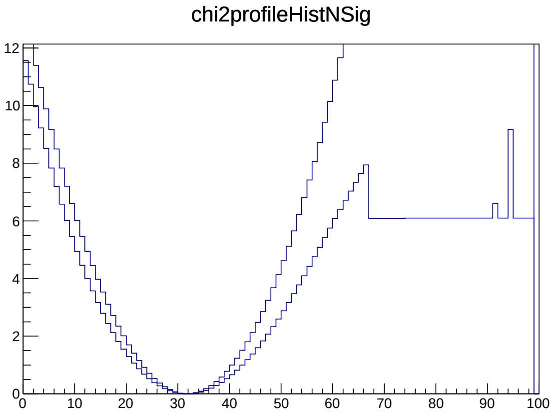
The Profile Likelihood Technique in a fit

- So what does "profiling" mean?
- Study how the χ^2 (or more precisely $-2\ln\mathcal{L}$) behaves if one parameter of interest is varied and if all other nuisance parameters are varied such that they give the lowest possible $-2\ln\mathcal{L}$ for each given parameter of interest



The Profile Likelihood Technique in a fit

- So what does "profiling" mean?
- Study how the χ^2 (or more precisely $-2\ln\mathcal{L}$) behaves if one parameter of interest is varied and if all other nuisance parameters are varied such that they give the lowest possible $-2\ln\mathcal{L}$ for each given parameter of interest



The Profile Likelihood Technique in a fit

- The test statistics chosen at LHC for the exclusion of a given signal hypothesis with strength μ is

$$\lambda(\mu) = \frac{\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})}$$

The Profile Likelihood Technique in a fit

- The test statistics chosen at LHC for the exclusion of a given signal hypothesis with strength μ is

$$\lambda(\mu) = \frac{\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})}$$

- Let's rewrite that:

$$-2 \ln \lambda(\mu) = -2 \ln \mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}}) + 2 \ln \mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})$$

- that looks mightily familiar to the fit. There, we plotted

$$-2\Delta \ln \mathcal{L} \approx \Delta\chi^2 = \chi^2(\mu) - \chi^2_{min}$$

The Profile Likelihood Technique in a fit

- The test statistics chosen at LHC for the exclusion of a given signal hypothesis with strength μ is

$$\lambda(\mu) = \frac{\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})}$$

- Let's rewrite that:

$$-2\ln\lambda(\mu) = -2\ln\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}}) + 2\ln\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})$$

- that looks mightily familiar to the fit. There, we plotted

$$-2\Delta\ln\mathcal{L} \approx \Delta\chi^2 = \chi^2(\mu) - \chi^2_{min}$$

- The choice of $\lambda(\mu)$ is optimal (Neyman-Pearson) for distinguishing the hypothesis μ from what is observed ($\hat{\mu}$). I.e. it is optimal for excluding ranges of μ.

The Profile Likelihood Technique in a fit

- The test statistics chosen at LHC for the exclusion of a given signal hypothesis with strength μ is

$$\lambda(\mu) = \frac{\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})}$$

- Let's rewrite that:

$$-2\ln\lambda(\mu) = -2\ln\mathcal{L}(d; \mu, \hat{\hat{\vec{\theta}}}) + 2\ln\mathcal{L}(d; \hat{\mu}, \hat{\vec{\theta}})$$

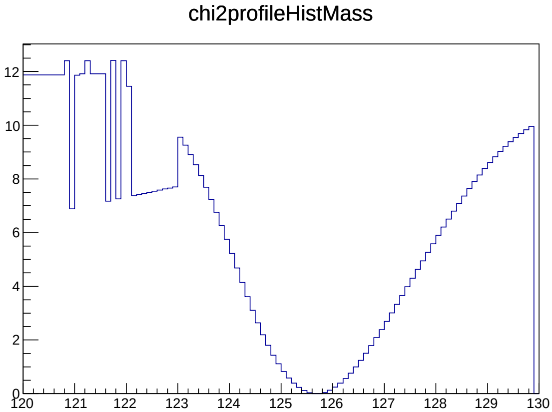
- that looks mightily familiar to the fit. There, we plotted

$$-2\Delta\ln\mathcal{L} \approx \Delta\chi^2 = \chi^2(\mu) - \chi^2_{min}$$

- The choice of $\lambda(\mu)$ is optimal (Neyman-Pearson) for distinguishing the hypothesis μ from what is observed ($\hat{\mu}$). I.e. it is optimal for excluding ranges of μ.
- Example: If we exclude $\mu = 0$: Exclude that there is no Higgs
- If we exclude $\mu = 1$: Exclude that there is a SM Higgs

The Profile Likelihood Technique in a fit

- We can do this with every parameter . . . here it's the mass:

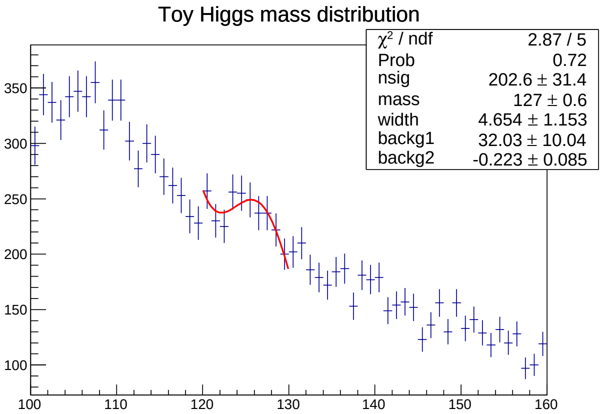


The Profile Likelihood Technique in a fit

- So fitting the nuisance parameters is a great thing because we automatically include our systematics (i.e. the uncertainty of the background description) into the limit or fit result.

The Profile Likelihood Technique in a fit

- So fitting the nuisance parameters is a great thing because we automatically include our systematics (i.e. the uncertainty of the background description) into the limit or fit result.
- In addition, it can be (depends on the experimental situation) an elegant way of determining the background in the first place:



The Profile Likelihood Technique in a fit

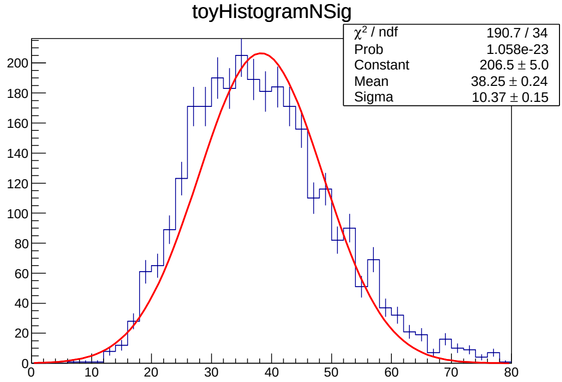
- So how do we know the uncertainty of our measurements?

The Profile Likelihood Technique in a fit

- So how do we know the uncertainty of our measurements?
- Either we just read it off at $\Delta\chi^2 = 1$ (or $\Delta \ln \mathcal{L} = 1/2$)
  If we know that the errors are gaussian, and the relation between all parameters and all observables is linear

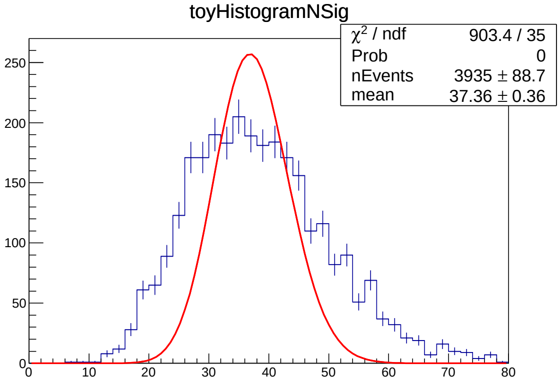
The Profile Likelihood Technique in a fit

- So how do we know the uncertainty of our measurements?
- Either we just read it off at $\Delta\chi^2 = 1$ (or $\Delta \ln \mathcal{L} = 1/2$)
  If we know that the errors are gaussian, and the relation between all parameters and all observables is linear
- Or we throw toys



The Profile Likelihood Technique in a fit

- So how do we know the uncertainty of our measurements?
- Either we just read it off at $\Delta\chi^2 = 1$ (or $\Delta \ln \mathcal{L} = 1/2$)
  If we know that the errors are gaussian, and the relation between all parameters and all observables is linear
- Or we throw toys



Developments since LEP

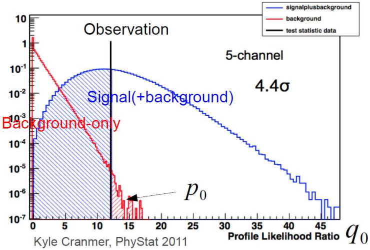
	Test statistic	Test statistic	Nuisance parameters	Pseudo-experiments
LEP	$-2\ln\frac{L(\mu,\tilde{\theta})}{L(0,\tilde{\theta})}$	Simple LR	Fixed by MC	Nuisance parameters randomized about MC
Tevatron	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}$	Ratio of profiled likelihoods	Extracted from priors	Nuisance parameters randomized from priors
LHC	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$	Profile likelihood ratio	Profiled (fit to data)	New nuisance parameters fitted for each pseudo-exp.

Limits at the LHC: Setting the CL

- Try to reject the background hypothesis based on q_0, independent of s_i

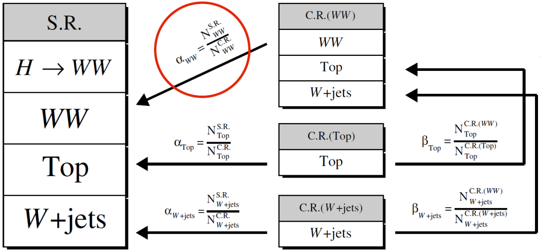
$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

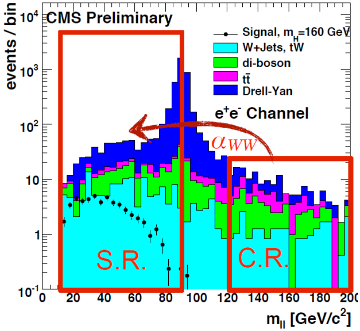
- E.g. could get the following: if p_0 small, reject SM! Found new physics! But it doesn't tell us whether we found the SM Higgs. We might have found something else!
- To get a hint whether a new observation could be the SM Higgs, $\hat{\mu}$ must be compatible with 1



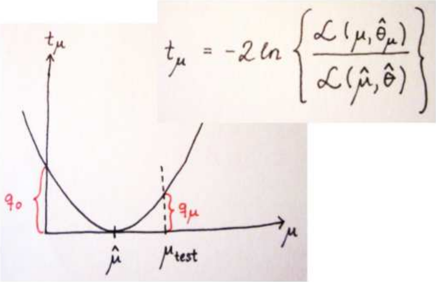
Limits at the LHC: How to control θ

- The big thing since LEP: Ged rid of partly bayesian techniques by fitting the systematic uncertainties to the data during limit setting at each toy MC



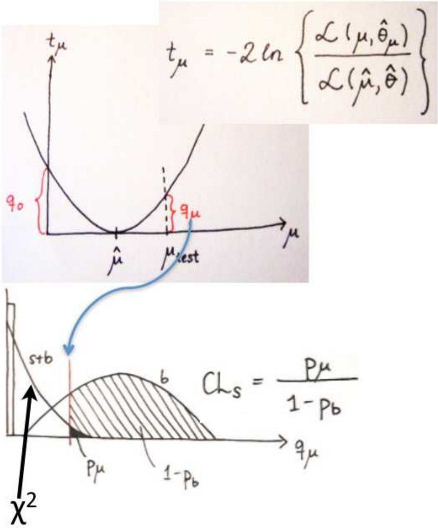


CL_s based Exclusions

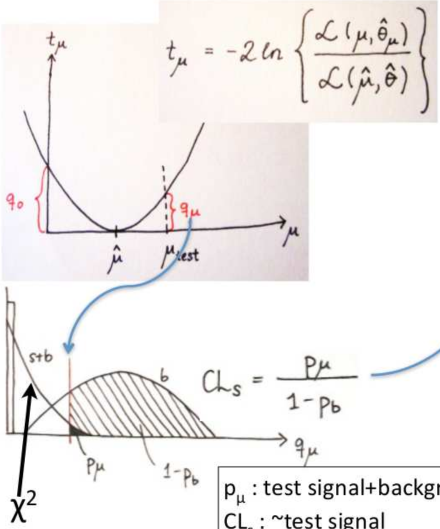


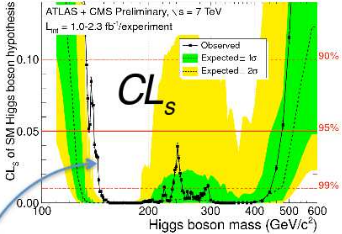
χ^2

CL_s based Exclusions



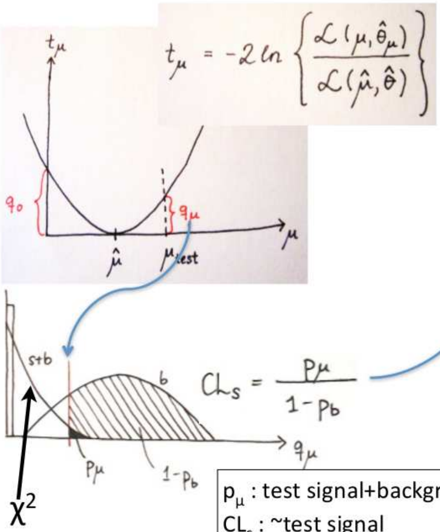
CL_s based Exclusions

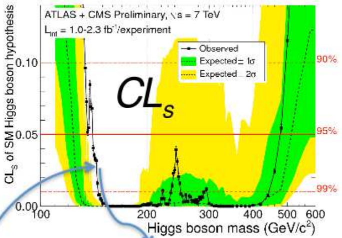




p_μ : test signal+background
CL_s : ~test signal

CL_s based Exclusions

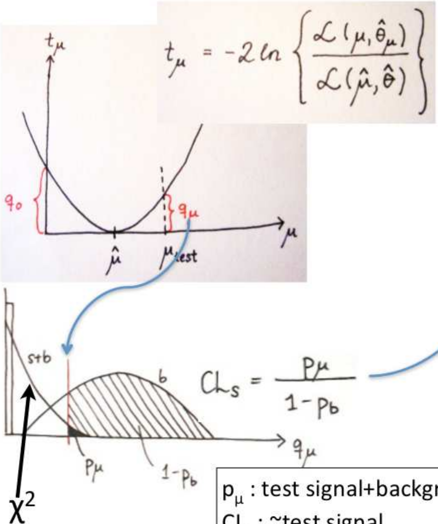


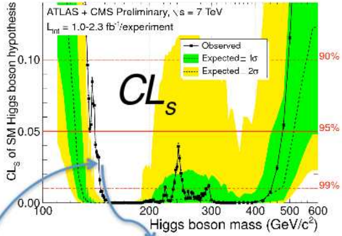


$$\mu_{95}^{up} = \mu(CL_s = 0.05)$$

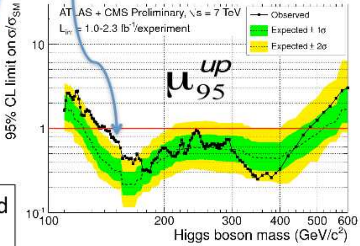
p_μ : test signal+background
CL_s : ~test signal

CL_s based Exclusions





$$\mu_{95}^{up} = \mu(CL_s = 0.05)$$

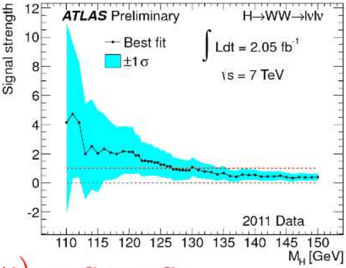


p_μ : test signal+background
CL_s : ~test signal

Measurements of Signal Rates

LHCHCG Combination Procedures

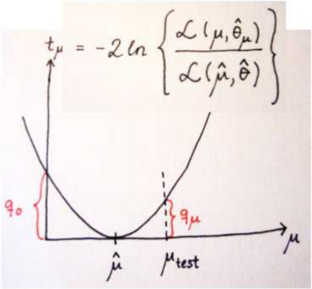
$$= -2 \ln \left\{ \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right\}$$

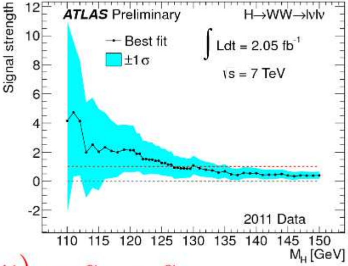


P.S. $q_{LEP}(\mu) = q_\mu - q_0$

Measurements of Signal Rates

LHCHCG Combination Procedures

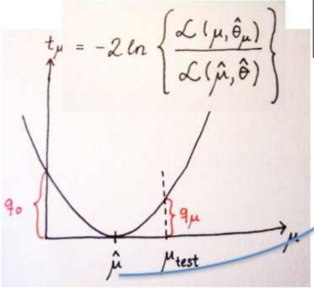




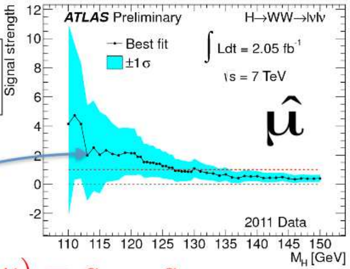
P.S. $q_{LEP}(\mu) = q_\mu - q_0$

Measurements of Signal Rates

LHCHCG Combination Procedures



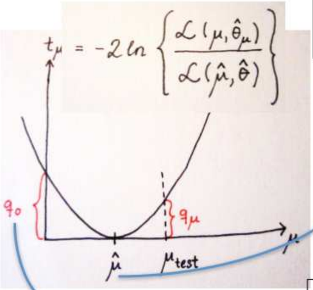
$\hat{\mu}$ to estimate signal strength



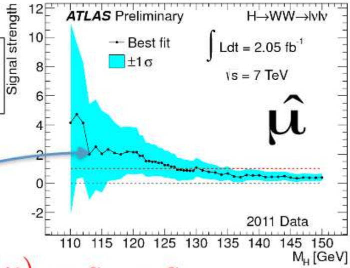
P.S. $q_{LEP}(\mu) = q_\mu - q_0$

Measurements of Signal Rates

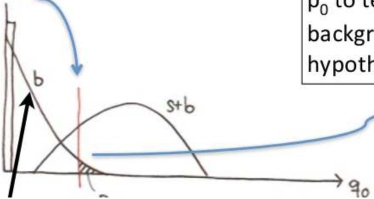
LHCHCG Combination Procedures



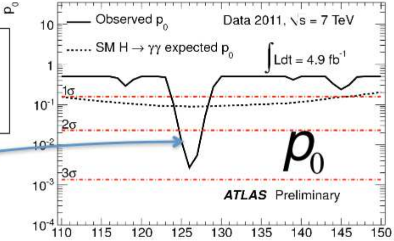
$\hat{\mu}$ to estimate signal strength



P.S. $q_{LEP}(\mu) = q_\mu - q_0$

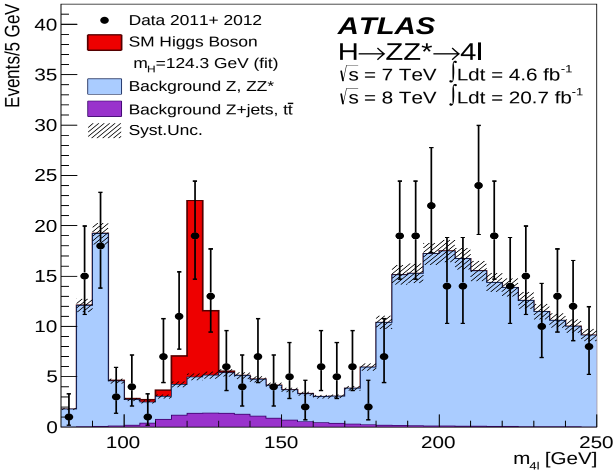


p_0 to test background hypothesis



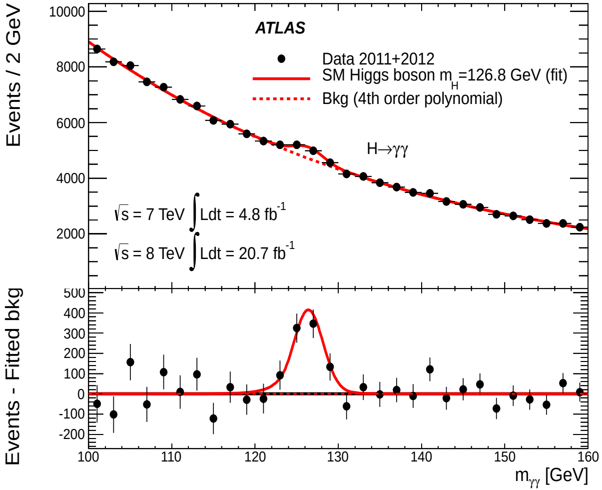
Examples: Present Results from ATLAS

- The observables: E.g. binned mass distributions



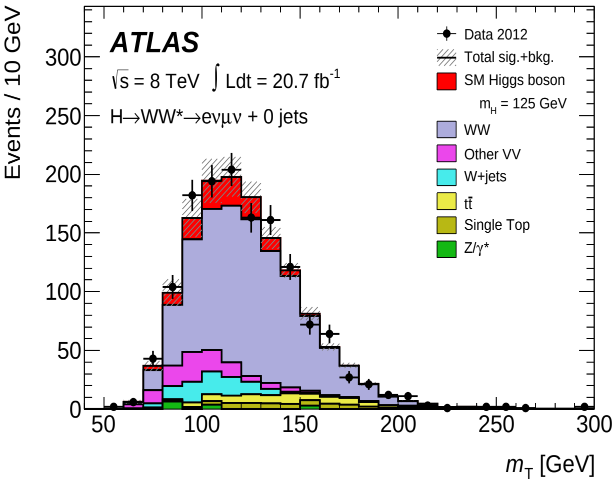
Examples: Present Results from ATLAS

- The observables: E.g. binned mass distributions



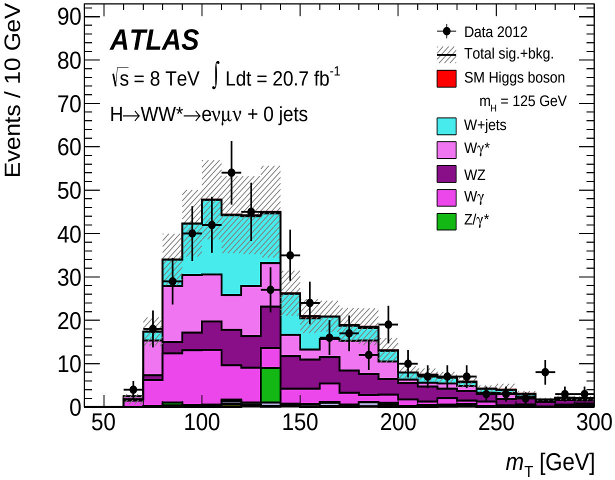
Examples: Present Results from ATLAS

- With signal regions and control regions, all used in the same fit



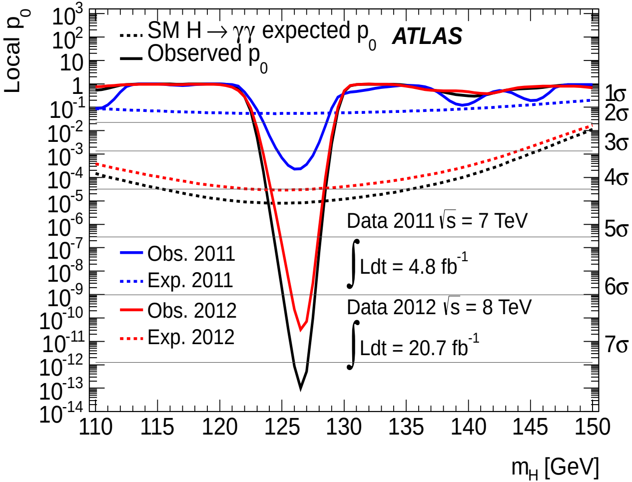
Examples: Present Results from ATLAS

- With signal regions and control regions, all used in the same fit



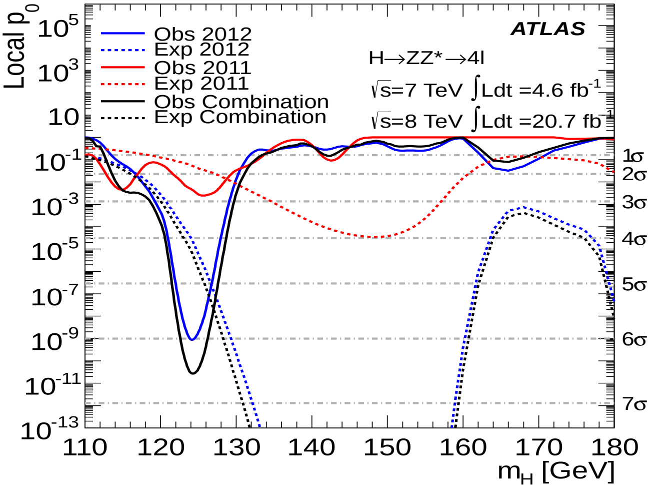
Examples: Present Results from ATLAS

- Make sure that we are sure we measured a signal



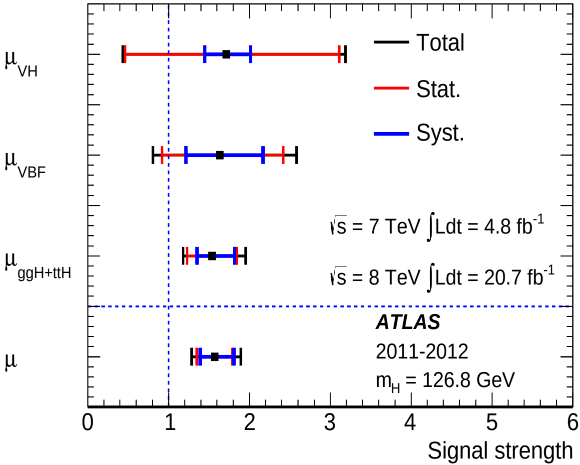
Examples: Present Results from ATLAS

- Make sure that we are sure we measured a signal



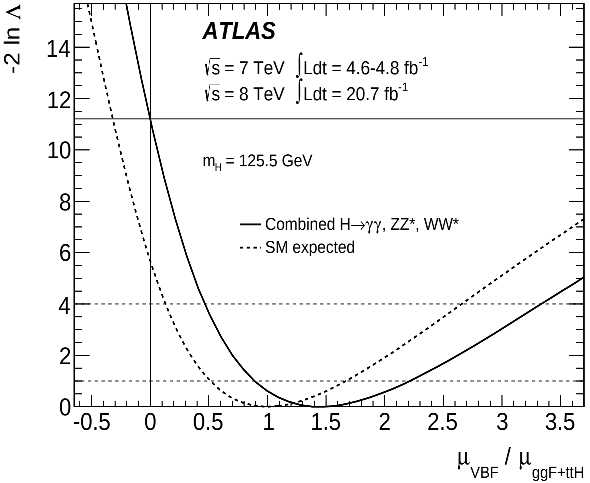
Examples: Present Results from ATLAS

- And measure properties of the signal



Examples: Present Results from ATLAS

- And measure properties of the signal



Summary

- I hope that's kind of what you were interested in
- Too much to summarize on one slide anyway . . .
- There are so many things I could not cover in the available time (I guess), like
    - So much on variances, expectation values, pdfs, . . .
    - The Look Elsewhere Effect
    - Doing justice to careful applications of Bayesian statistics
    - . . .

Backup Slides