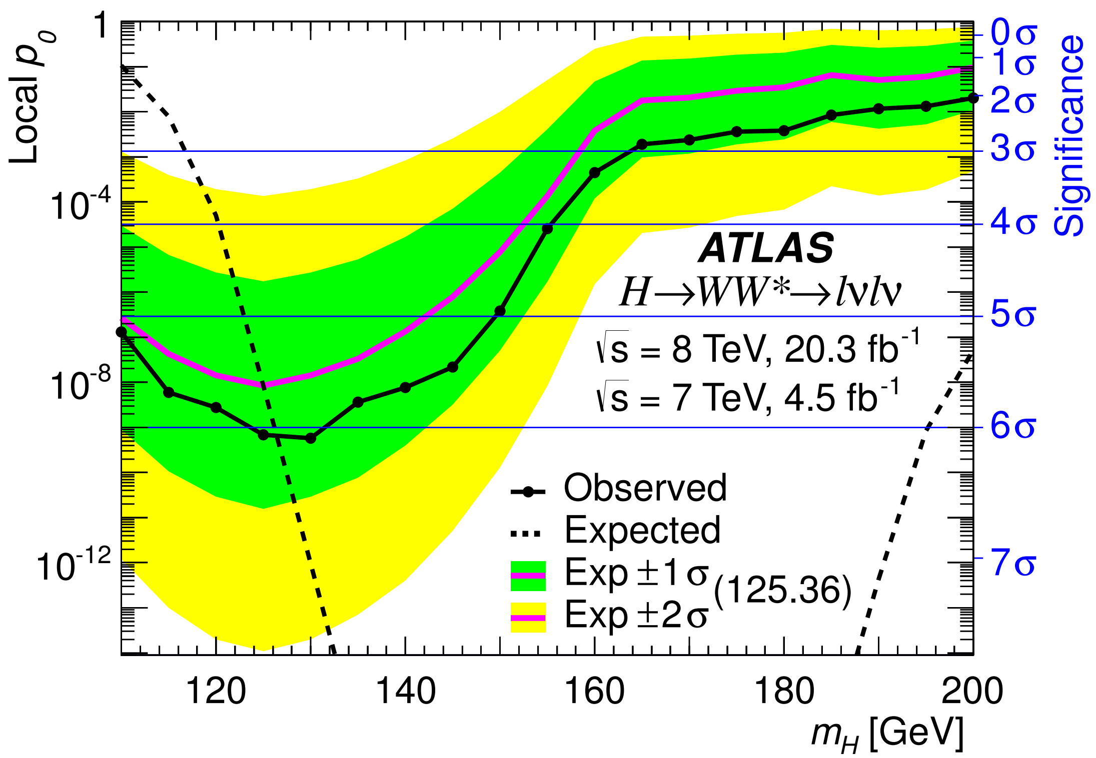
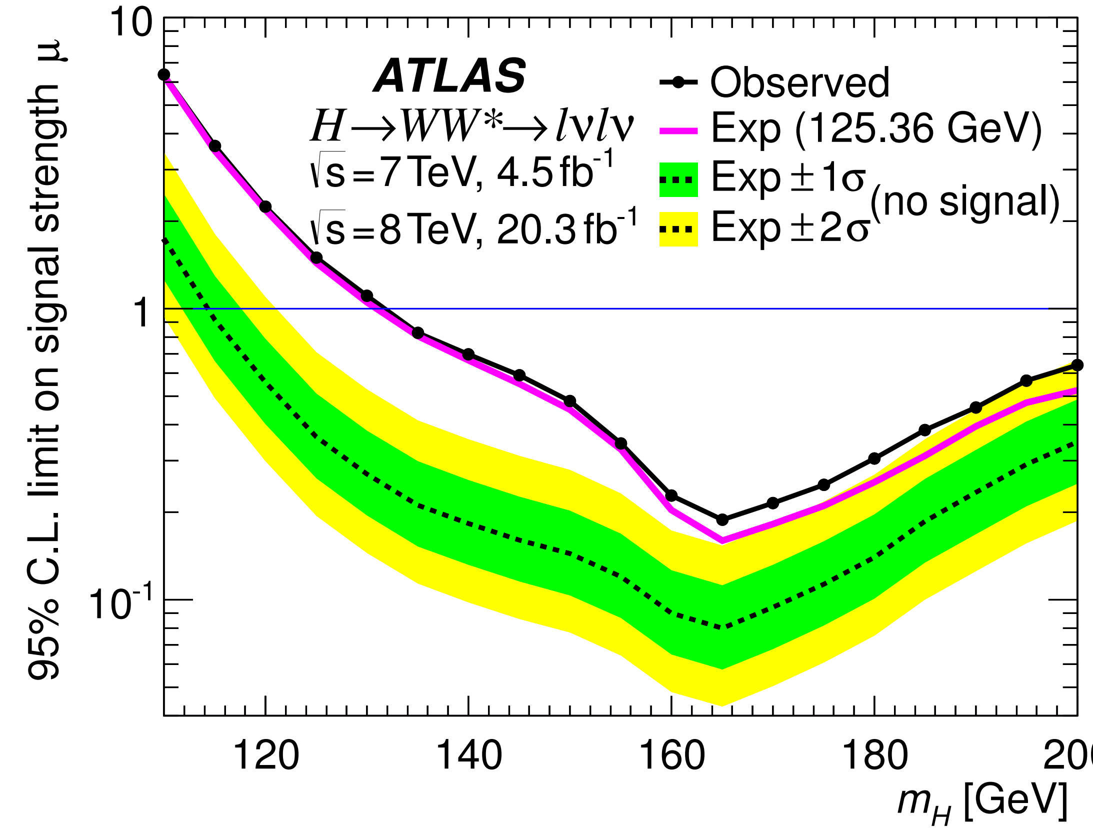
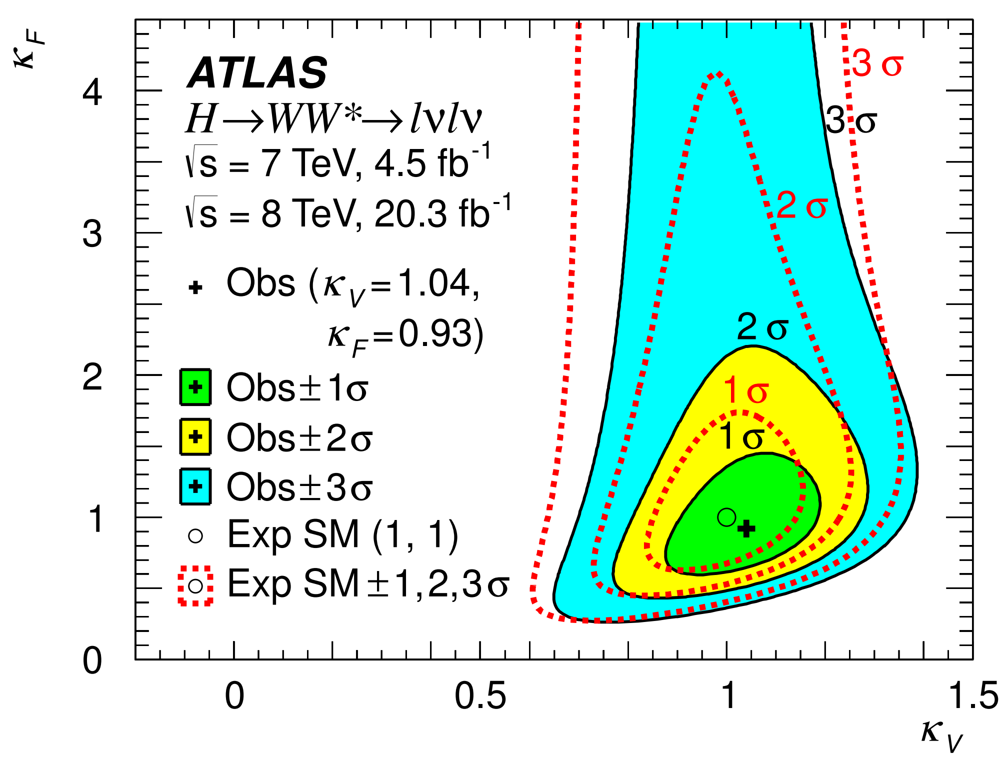
# Statistical tools in Higgs search and discovery




Slides prepared using mainly material from E. Gross and W. Vekerke talks.

# Introduction

Enormous effort to search for new particles, for example the Higgs boson, we can have signatures in many decay channels

- Results → many plots with signal, background expectations, each with (systematic) uncertainties, and data

- Q: How do you conclude from this that you've seen the Higgs (or not)?

– we want an answer of type: 'We can exclude that the Higgs (or a new particle) exists at 95% CL", or "The significance of the observed excess is $5\sigma$"

# Quantifying discovery and exclusion – Frequentist approach

- Consider the simplest case – a counting experiment

– Observable: N (the number of events)

– Model F(N | s+b): Probability to get N events given an assumed value of signal expectation (s) and background expectation (b)

Let's assume to know exactly the expected background b=5.

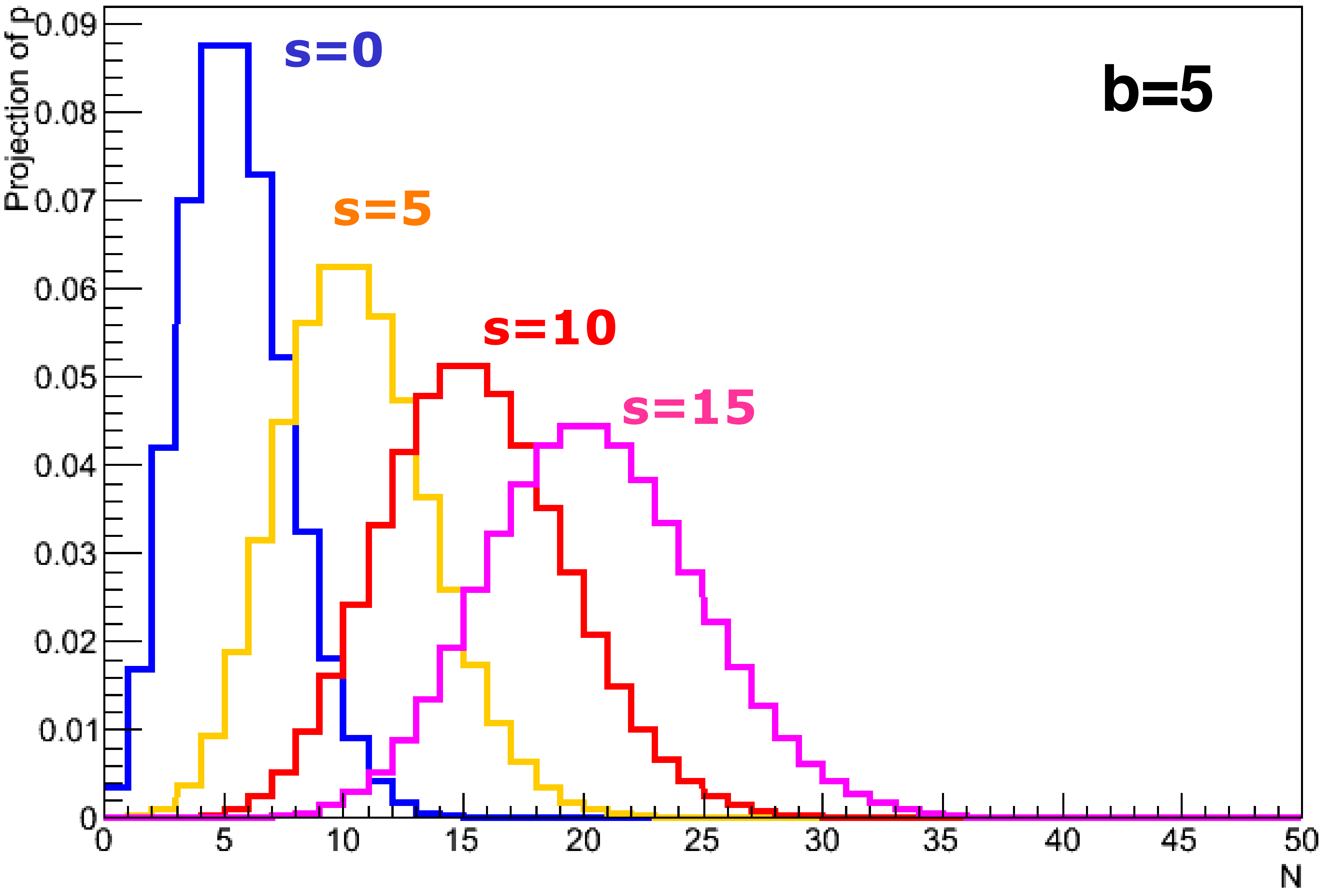
F is given by Poisson(N | s+b)

$$F(N|y) = \frac{y^N}{N!}e^{-y} \Rightarrow F(N|s+b) = \frac{(s+b)^N}{N!}e^{-(s+b)}$$



[from G.Herten]

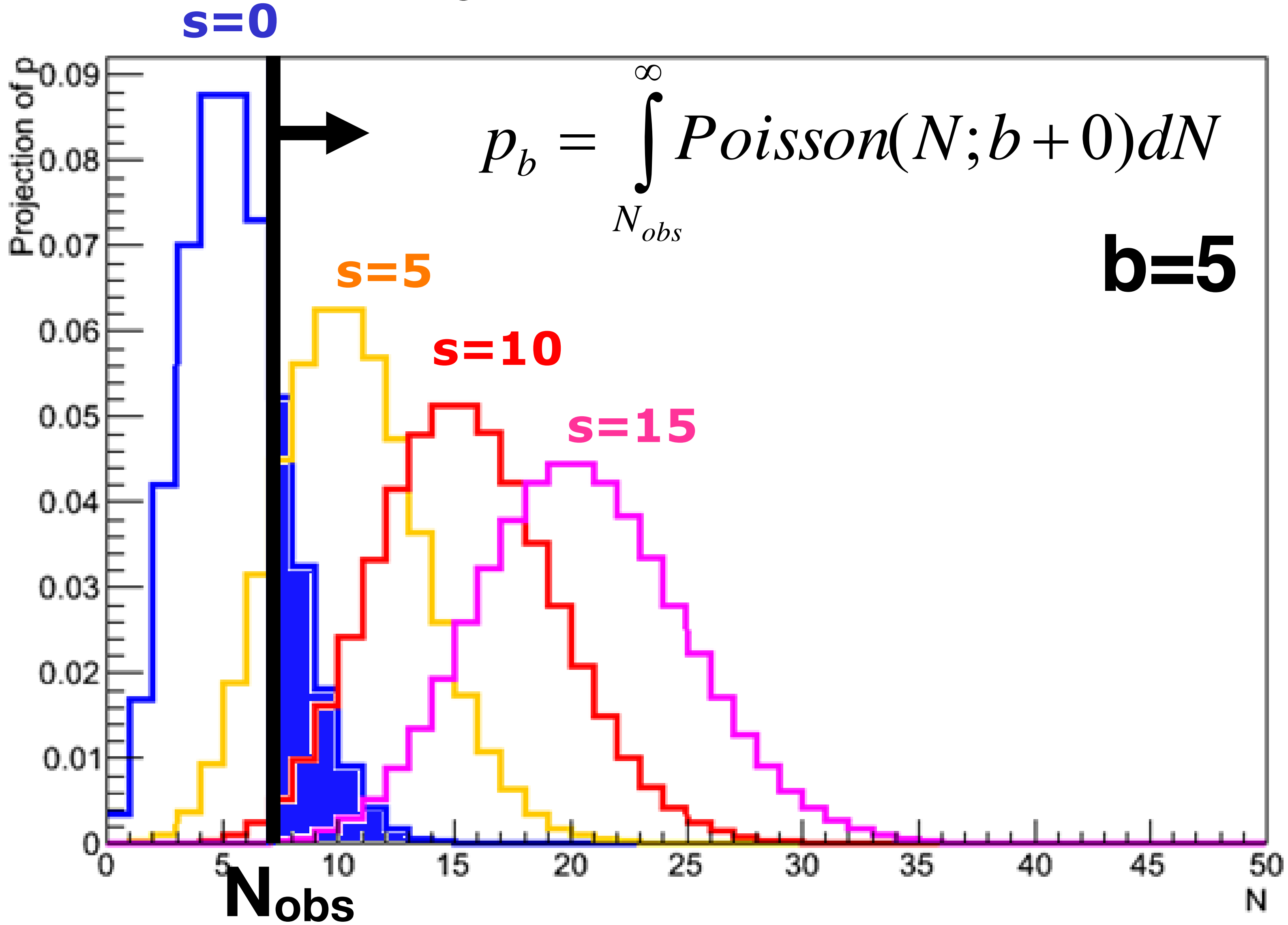
# Quantifying discovery and exclusion – Frequentist approach



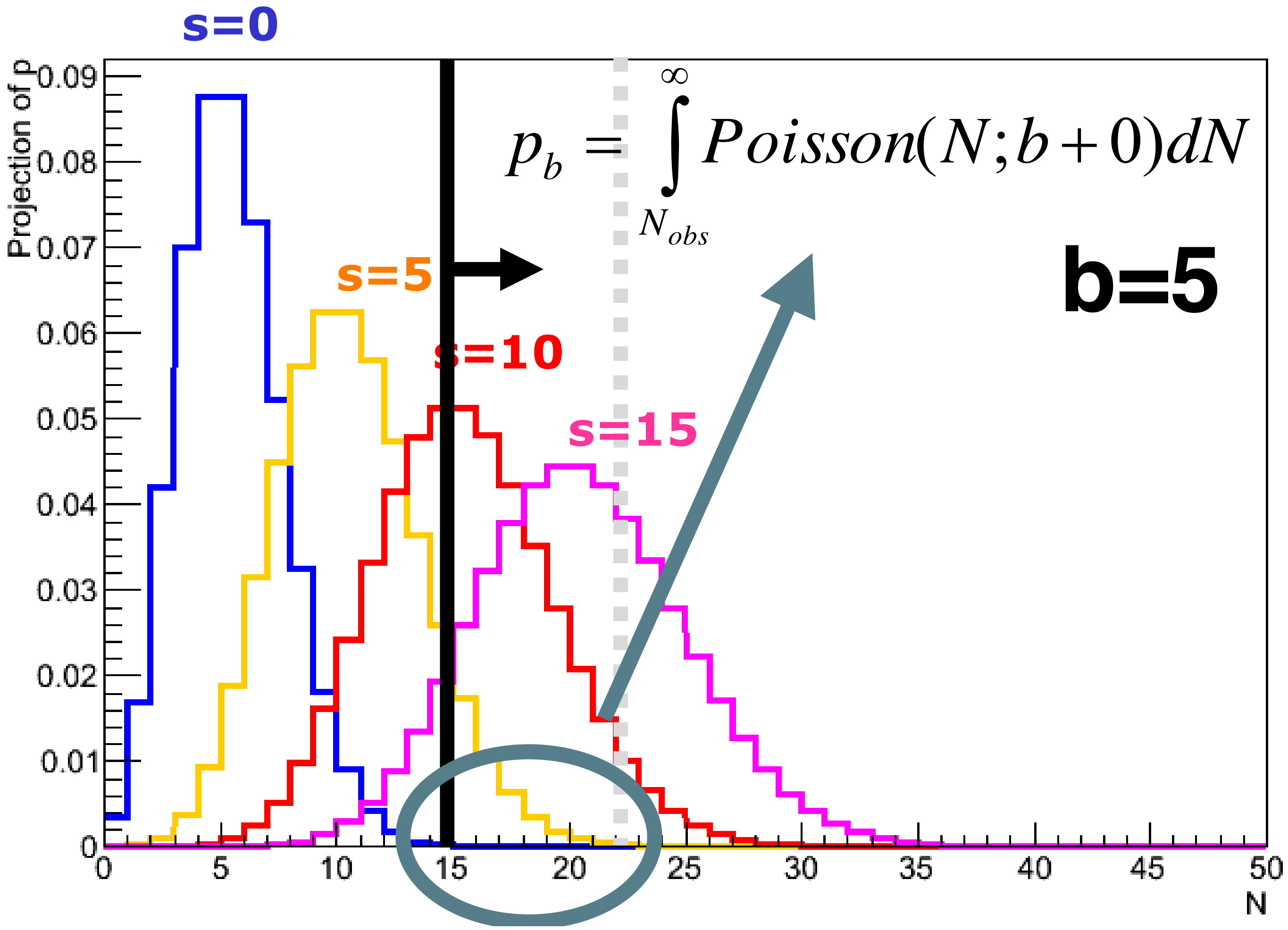
# Quantifying discovery and exclusion – Frequentist approach

- Now make a measurement $N=N_{obs}$ (example $N_{obs}=7$)
- Can now define p-value(s), e.g. for bkg hypothesis
  - Fraction of future measurements with N=Nobs (or larger) if s=0
    (probability that the background can fluctuate up to $N_{obs}$ or above)

$$p_b = \int_{N_{obs}}^{\infty} Poisson(N; b+0)dN \quad (=0.23)$$



- p-values of background hypothesis is used to quantify 'discovery' = excess of events over background expectation
- Another example: $N_{obs}$=15 for same model, what is $p_b$?




$$p_b = \int_{N_{obs}}^{\infty} Poisson(N; b+0)dN$$

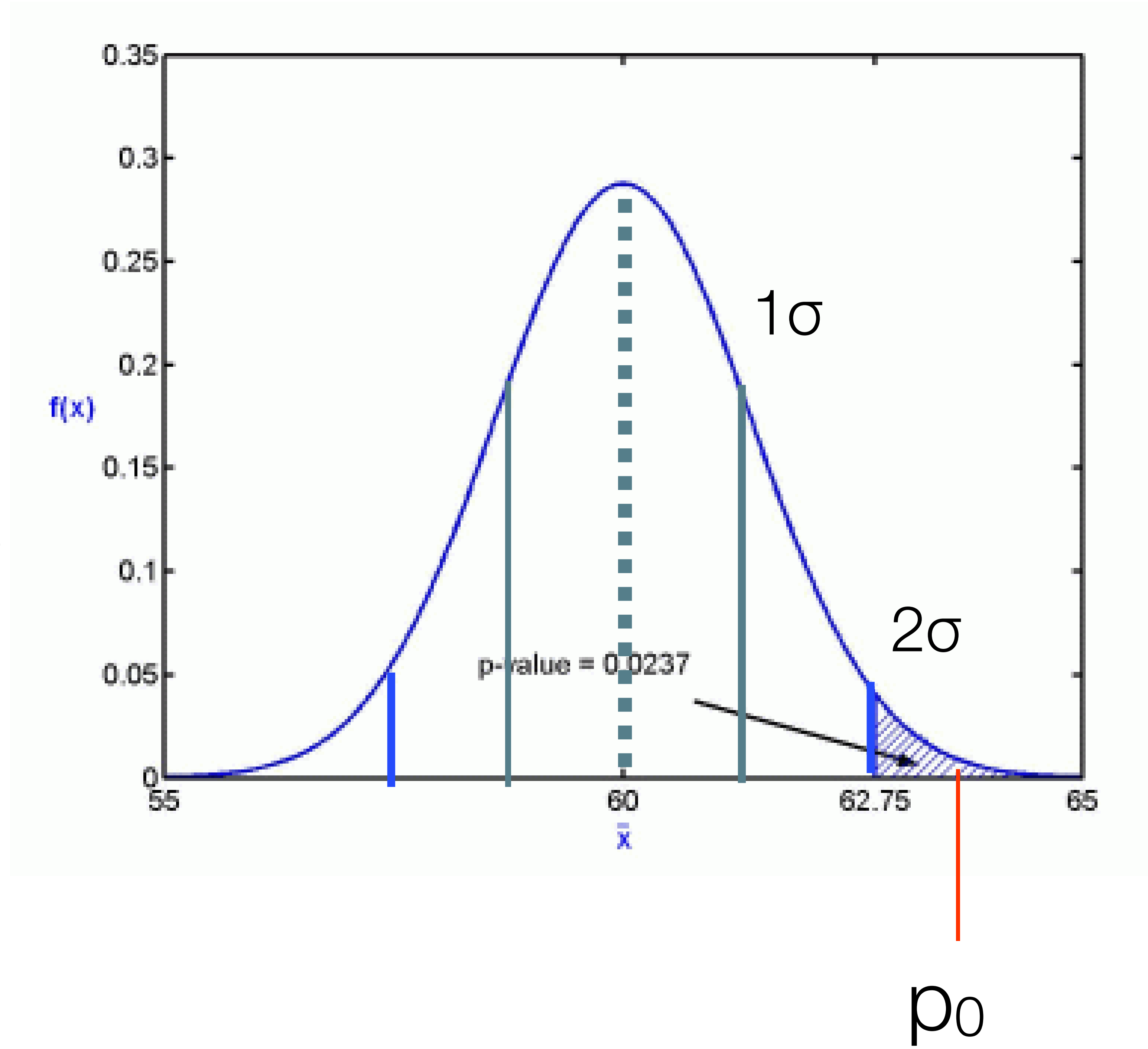
$$(= 0.00022)$$

$$p_b = p_{s+b}(s=0) = p_0$$

For large **b** the Poisson distribution becomes a gaussian distribution

# From $p_0$ to number of $\sigma$

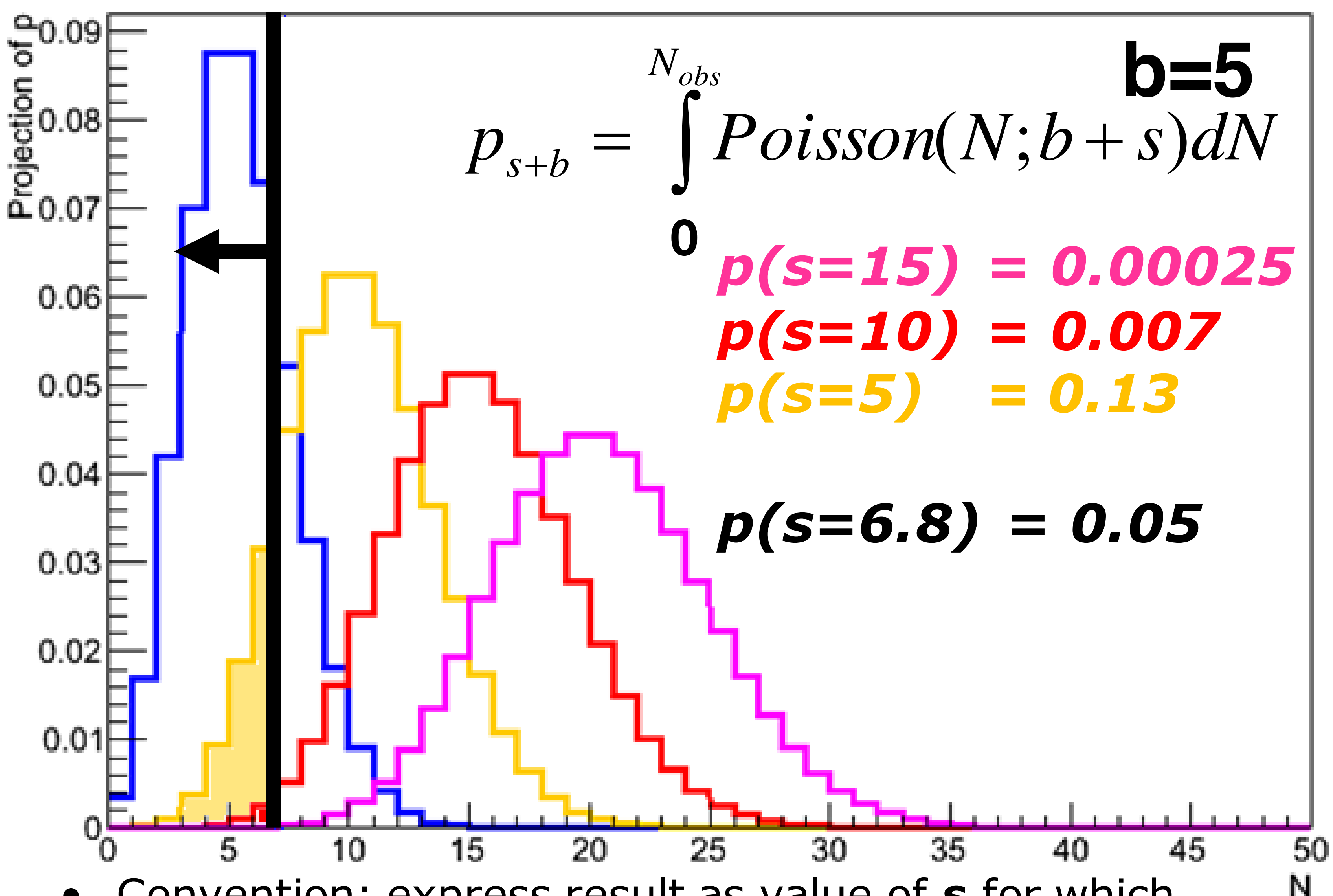
An observed excess is n**σ**
if the integral of the right tail above the region delimited by the n**σ** interval is equal to the observed $p_0$



# Quantifying exclusion – Frequentist approach

We want to exclude a signal hypothesis s.
The question is: are my data compatible with the signal+background hypothesis?
or: what is the probability that s+b under fluctuates below the observed yield $N_{obs}$?

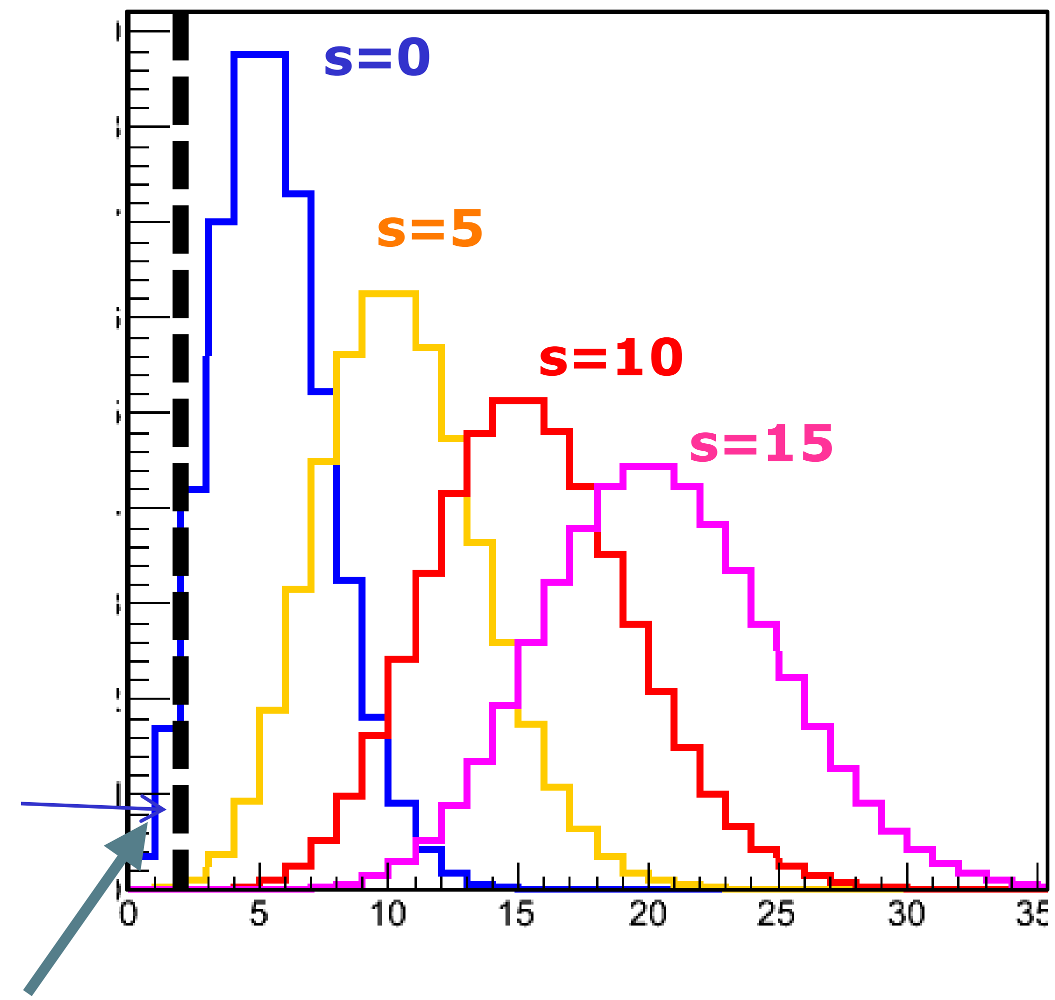
**$CL_{s+b} = p_{s+b}$**




- Convention: express result as value of **s** for which p(s+b)=5% → **"s>6.8 is excluded at 95% C.L."**

# Small signals and background under-fluctuations

- $\langle N_{obs} \rangle = s+b$ leads to the physical requirement that $N_{obs} > b$
- A very small expected s might lead to an anomaly when $N_{obs}$ fluctuates far below the expected background, b.
- At one point DELPHI alone had $CL_{s+b}=0.03$ for $m_H=116$ GeV
- However, the cross section for 116 GeV Higgs at LEP was too small and Delphi actually had no sensitivity to observe it
- The frequntist would say: Suppose there is a 116 GeV Higgs….
  In 3% of the experiments the true signal would be rejected… (one would obtain a result incompatible or more so with m=116) i.e. a 116 GeV Higgs is excluded at the 97% CL…..




The background hypothesis is not very likely, excluding background automatically excludes any signal

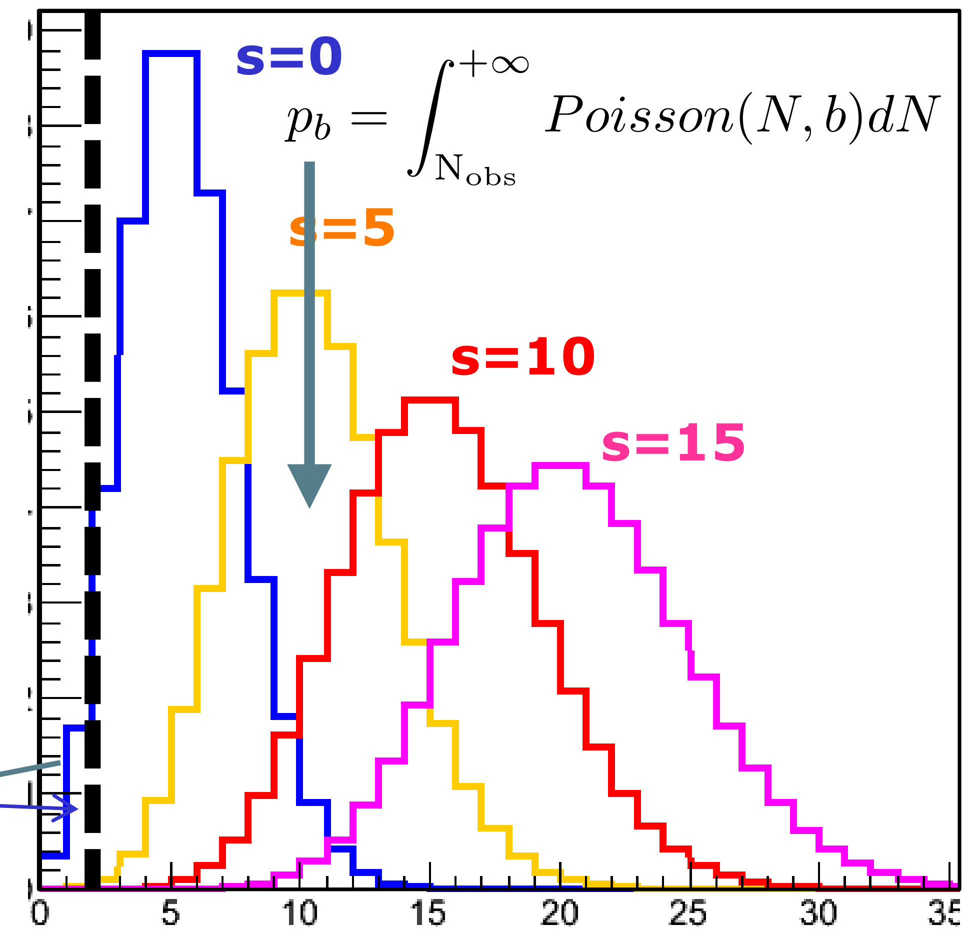
The problem of this method is that it ignores sensitivity to signal. Even if you expect s=0.000001 you would exclude any signal if your background under-fluctuates.

# Small signals and background under-fluctuations

- $\langle N_{obs} \rangle = s+b$ leads to the physical requirement that $N_{obs} > b$
- A very small expected s might lead to an anomaly when $N_{obs}$ fluctuates far below the expected background, b.
- At one point DELPHI alone had $CL_{s+b}=0.03$ for $m_H=116$ GeV
- However, the cross section for 116 GeV Higgs at LEP was too small and Delphi actually had no sensitivity to observe it
- The frequntist would say: Suppose there is a 116 GeV Higgs….
  In 3% of the experiments the true signal would be rejected… (one would obtain a result incompatible or more so with m=116) i.e. a 116 GeV Higgs is excluded at the 97% CL…..

$CL_b$

$1 - p_b$




if the background hypothesis is not very likely $1-p_b \rightarrow 0$ compensating the numerator

the $CL_s$ method was introduced

$$CL_s = \frac{p_{s+b}}{1-p_b} = \frac{CL_{s+b}}{CL_b}$$

$$p_{s+b} = \int_0^{N_{obs}} Poisson(N;b+s)dN \xrightarrow{s=0} = 1 - \int_{N_{obs}}^{+\infty} Poisson(N;b)dN = 1 - p_b$$

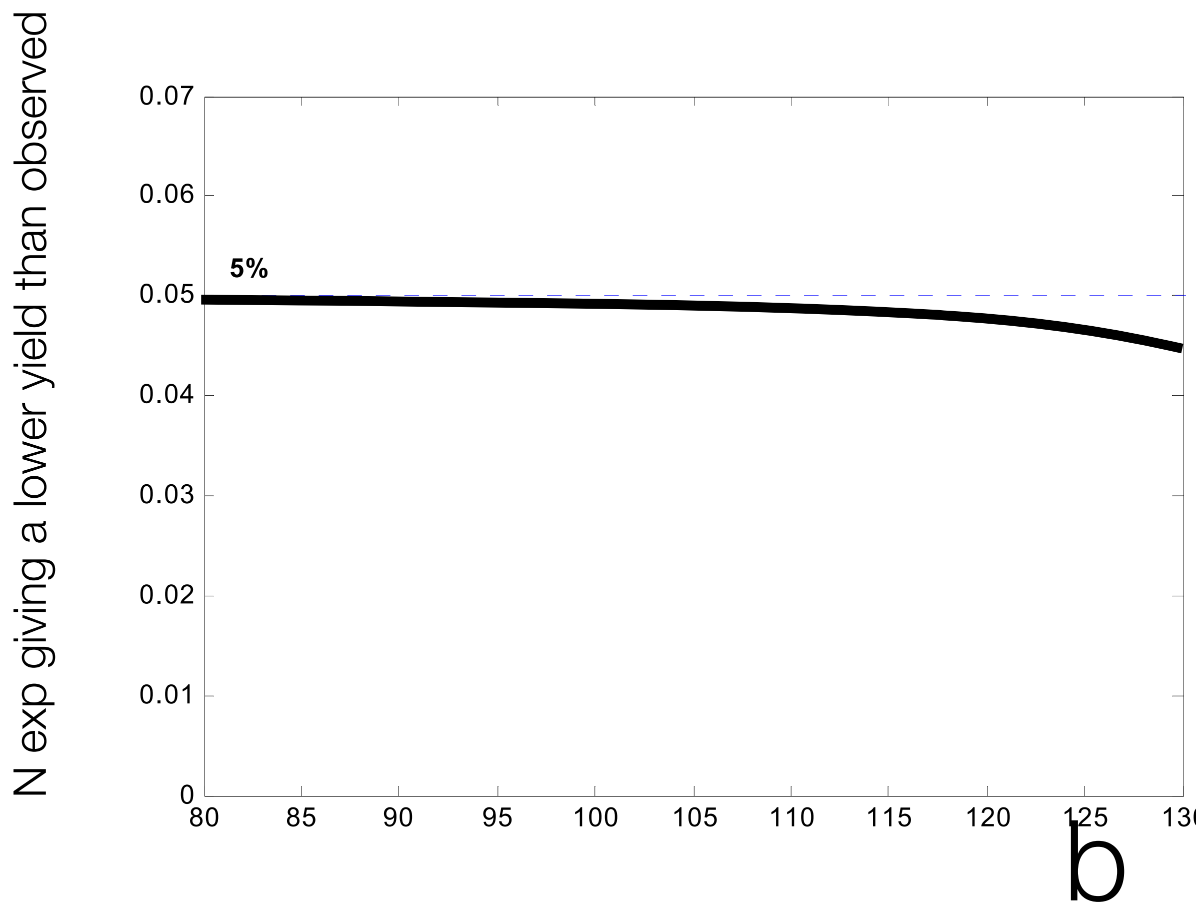
If s <<b $CL_{s+b}/CL_b \sim 1$ (no exclusion)

# Coverage

- If we exclude a signal  s at 95% C.L, we want that if we repeat the experiment may times in the s hypothesis, only 5% (false exclusion rate) of the times we get an event yield below the observed number of events, if such property holds we say that the C.L. is well covered;
- $CL_{s+b}$ is well covered by definition (we take the tail of the poisson distribution that integrates to 95% to set the 95% exclusion);
- $CL_s = CL_{s+b}/CL_b$ undercovers: if we set an exclusion at 95% C.L.; more than 95% of the experiments will give a number of events above the observed one for the excluded signal hypothesis s

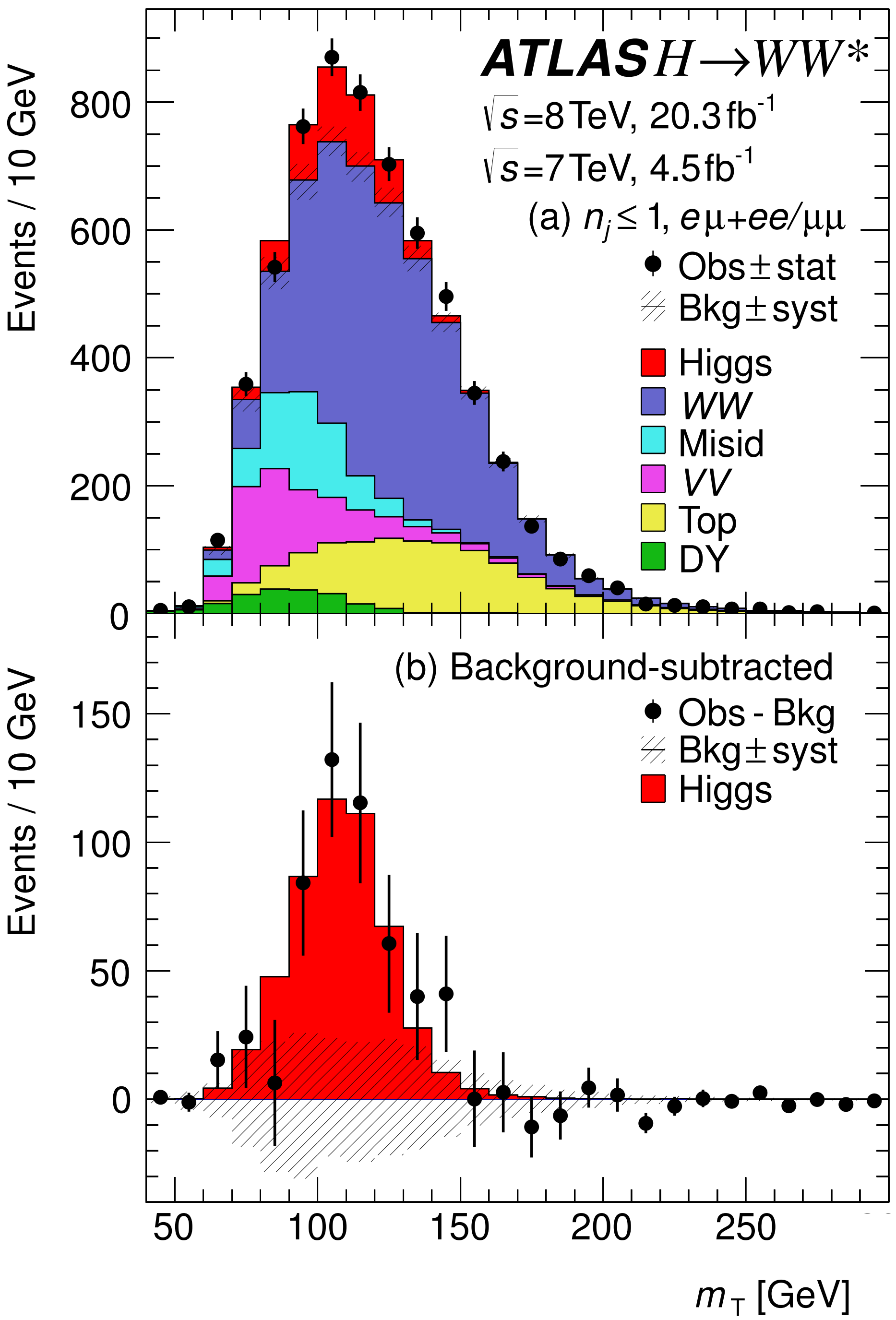
The problem: under coverage

for low σ signals the true false exclusion rate is below 5% (when quoting according to this recipe a 95% CL exclusion)



# p-values and limits on non-trivial analysis

- Typical Higgs search result is not a simple number counting experiment, but looks like this:




- Result is a distribution, not a single number

- Models for signal and background have intrinsic uncertainties

We have two hypotheses:

1. $H_s$ there is a signal;
2. $H_b$ there is only background

We have K bins, we know the acceptance in each bin i: $\varepsilon_i^b$ for background, $\varepsilon_i^s$ for signal: $<N_i(H_s)> = \varepsilon_i^b b + \varepsilon_i^s s$

$$L(N_1, \ldots, N_K) = \prod_{i=1}^{K} Poisson(N_i, \epsilon_i^b b + \epsilon_i^s s) = \prod_{i=1}^{k} \frac{(\epsilon_i^b b + \epsilon_i^s s)^{N_i}}{N_i!} e^{-\epsilon_i^b b - \epsilon_i^s s}$$

$$L(N_1, \ldots, N_K | H_b) = \prod_{i=1}^{K} Poisson(N_i | \epsilon_i^b b) = \prod_{i=1}^{K} \frac{(\epsilon_i^b b)^{N_i}}{N_i!} e^{-\epsilon_i^b b}$$

# Neyman-Pearson lemma

$$L(N_1, \ldots, N_K | H_s) = \prod_{i=1}^{K} Poisson(N_i, \epsilon_i^b b + \epsilon_i^s s) = \prod_{i=1}^{k} (\epsilon_i^b b + \epsilon_i^s s)^k e^{-\epsilon_i^b b - \epsilon_i^s s}$$

$$L(N_1, \ldots, N_K | H_b) = \prod_{i=1}^{K} Poisson(N_i | \epsilon_i^b) = \prod_{i=1}^{K} \frac{(\epsilon_i^b b)^{N_i}}{N_i!} e^{-\epsilon_i^b b}$$

The most powerful discriminant is the likelihood ratio

$$\lambda(N_1, \ldots, N_K | H_s, H_b) = \frac{L(N_1, \ldots, N_K | H_s)}{L(N_1, \ldots, N_K | H_b)}$$

A selection that maximises $\lambda$ is such that, for a given signal efficiency $\varepsilon_s$, it allows to have the lowest background efficiency $\varepsilon_b$

# Likelihood ratio for discovery

Discovery: what is the probability that the observed data are due to a background fluctuation?

Hypothesis 1: There is only background (we want to falsify this)
Hypothesis 2: There is a signal with arbitrary normalisation

If we expect **s** events from MC simulation of a signal with cross section **$\sigma_s$**, we test the **s** hypothesis with an arbitrary multiplicative factor **μ** (signal strength), I.e. we test an arbitrary signal yield **μ·s.**

This means that if data are better described by a signal, we prefer it to the background hypothesis (in this sense we increase the separation power)

Assuming *b* and *s* are known without uncertainties (no systematic uncertainties)

$$\lambda(N_1, \ldots, N_K|0) = \frac{L(N_1, \ldots, N_K|b)}{L(N_1, \ldots, N_K|b + \hat{\mu}s)}$$

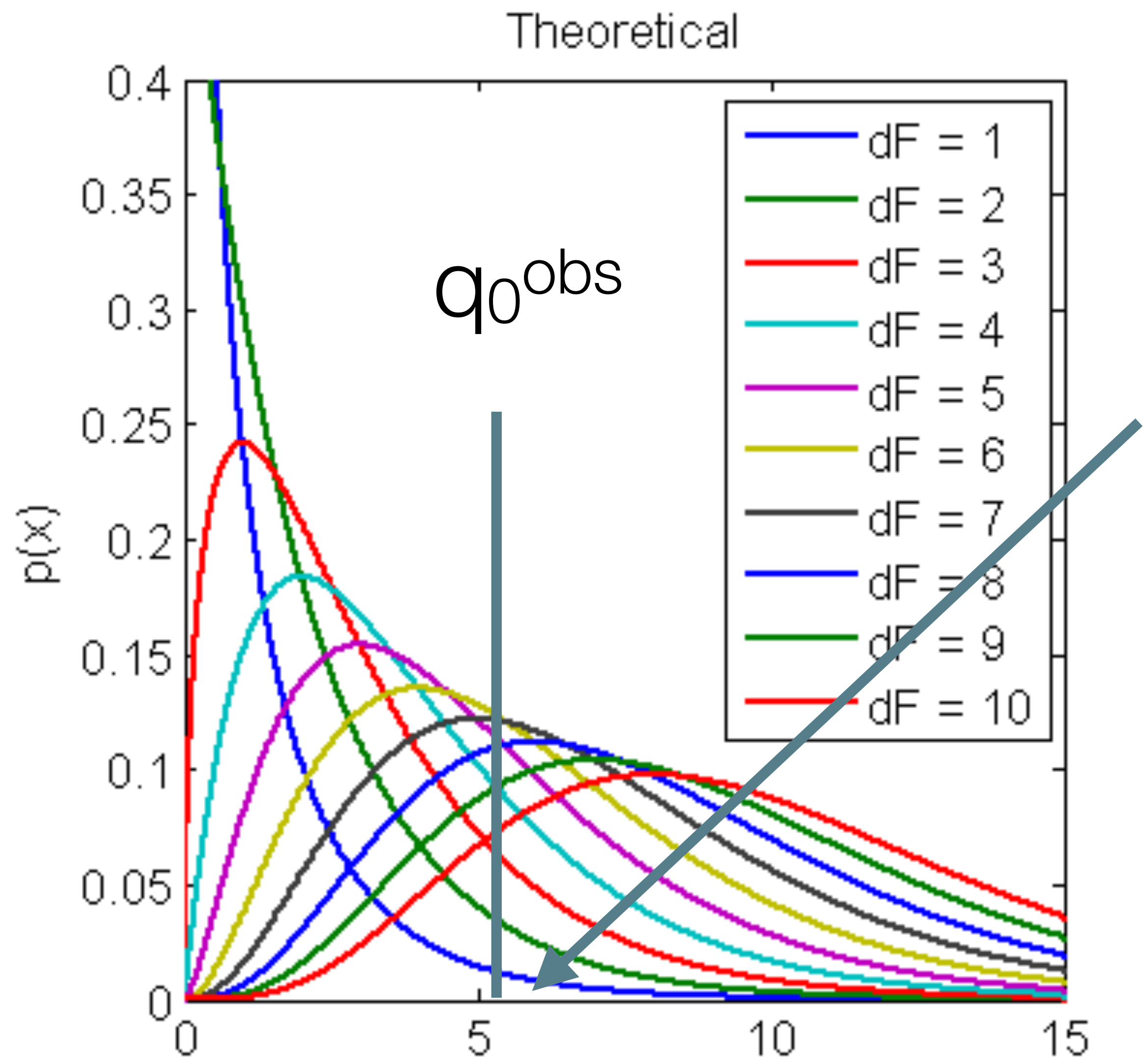
fixed number

$\hat{\mu}$ is obtained by maximising the denominator of λ

# Likelihood ratio for discovery (the test statistics)

$$q_0 = -2ln\left[\frac{L(N_1,...,N_K|b)}{L(N_1,...,N_K|b+\hat{\mu}s)}\right]$$

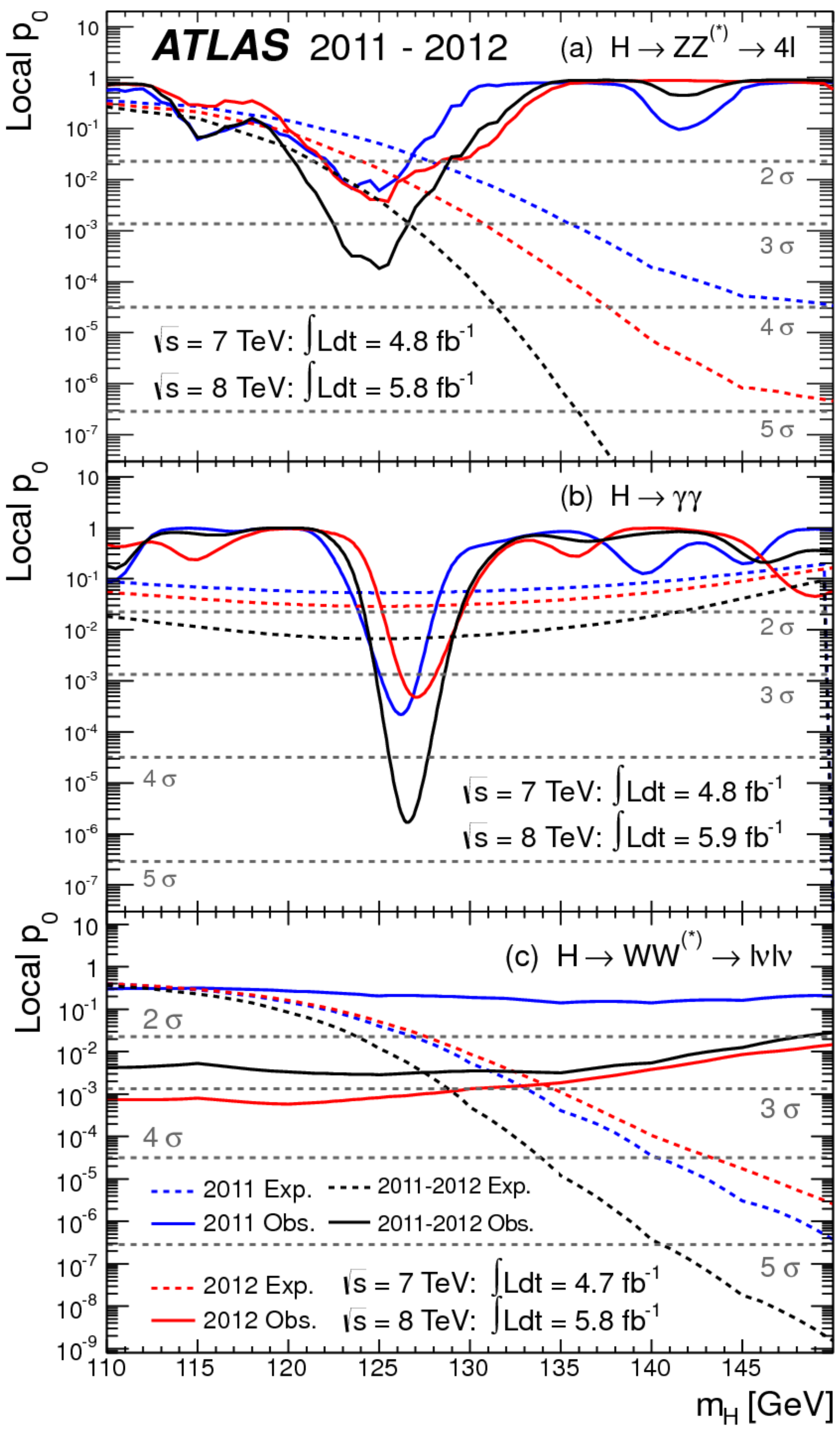
$q_0$ distributes according a $\chi^2$ distribution with 1 degree of freedom (dF)



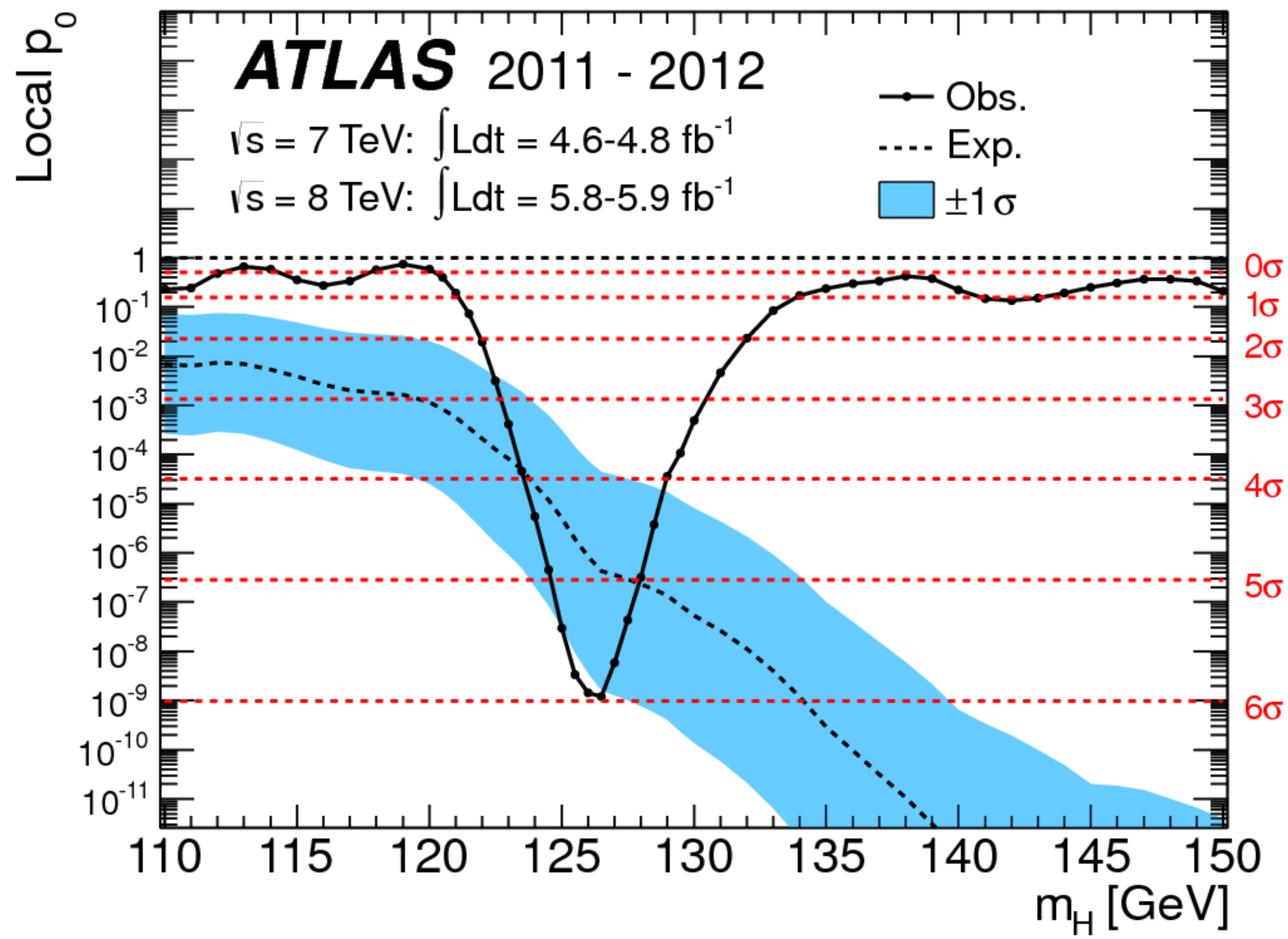

This area is the probability to have a $q_0$ value higher than the observed one (it is the $p_0$)

data are not background-like, L small, $q_0$ larger.

# Higgs boson discovery



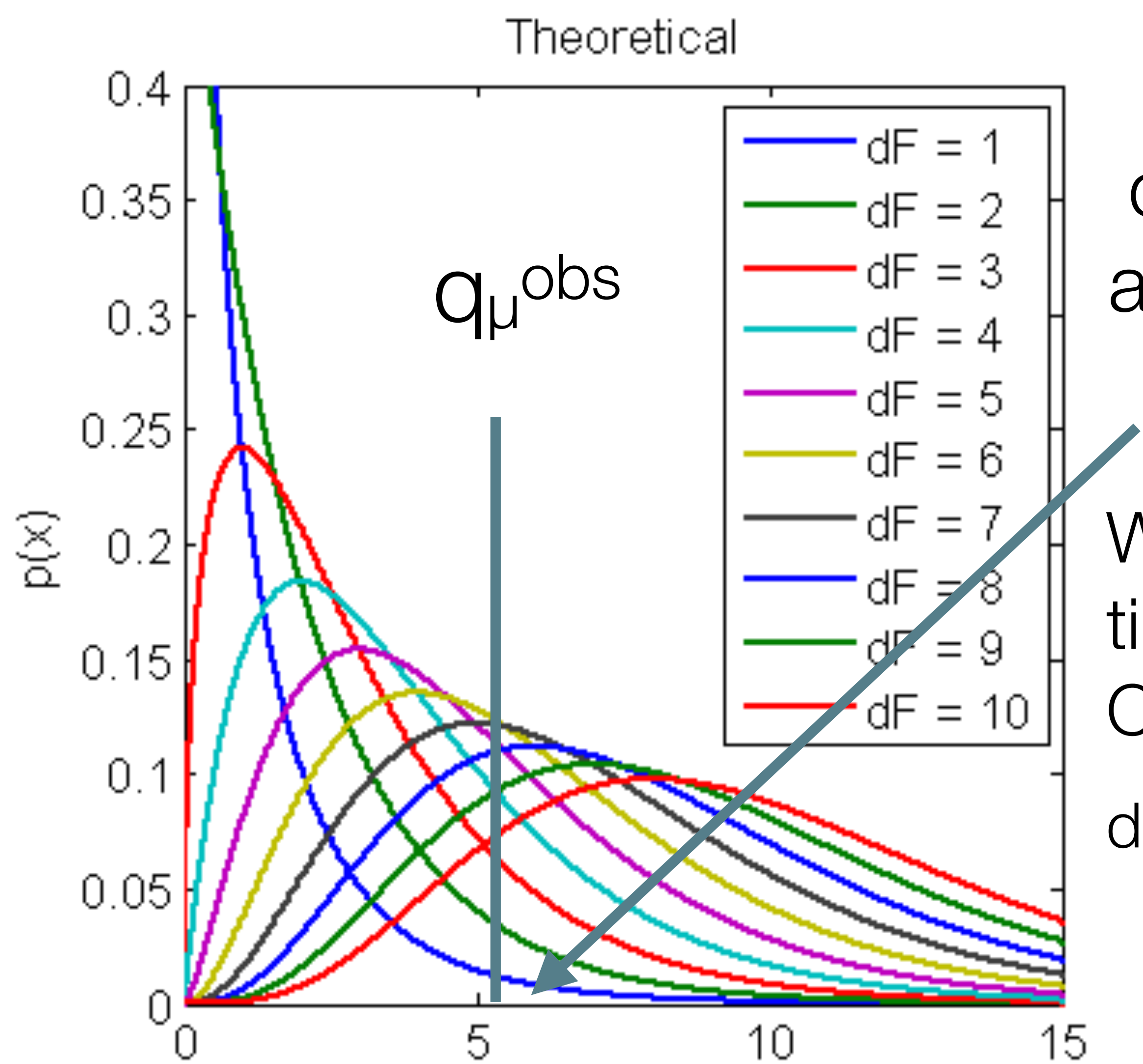

$p_0$ is computed for each mass hypothesis, the mass hypothesis changes the signal distributions (this plot would have no shape in case of a single count experiment)



# Likelihood ratio for exclusion of signal strength μ

1) $H_\mu$ hypothesis to have a signal that is μ times the SM expectation;
2) $H_{\hat{\mu}}$ hypothesis to have any signal strength

$$q_\mu = -2ln\left[\frac{L(N_1,\ldots,N_K|b+\mu s)}{L(N_1,\ldots,N_K|b+\hat{\mu}s)}\right]$$



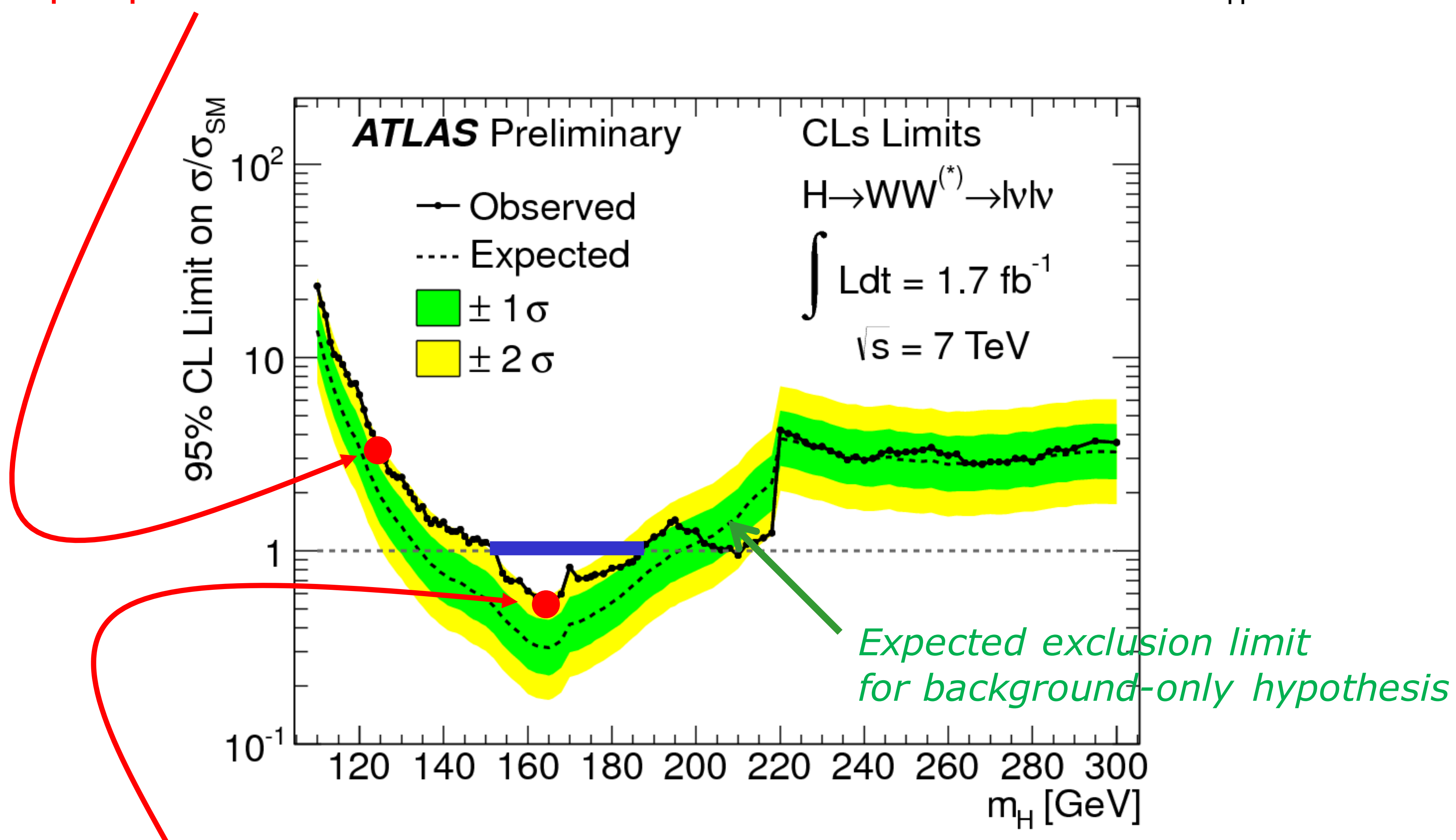

$q_\mu \geq 0$ and distributes according a $\chi^2$ distribution with 1 degree of freedom

We say that a signal with a cross section μ times larger than the SM is excluded at 95% C.L. $P(q_\mu > q_\mu^{obs}) < 5\%$, coverage is exact

dF: number of degree of freedom

# Example – 95% Exclusion limit vs $m_H$ for H→WW

Example point: ≈3 x SM H→WW cross-section excluded at $m_H$=125 GeV




Example point: ≈0.5 x SM H→WW cross-section excluded at $m_H$=165 GeV

Higgs with 1.0x SM cross-section excluded at 95% CL for $m_H$ in range [150,~187]

# How does likelihood ratio behave for small signals?

Let's assume to have 1 bin, and we want to test the μ = 1 hypothesis:

$$q_1 = -2ln\left[\frac{L(N_1, b+s)}{L(N_1, b+\hat{\mu}s)}\right] = -2ln\left[\frac{Poisson(N_1, b+s)}{Poisson(N_1, b+\hat{\mu}s)}\right]$$

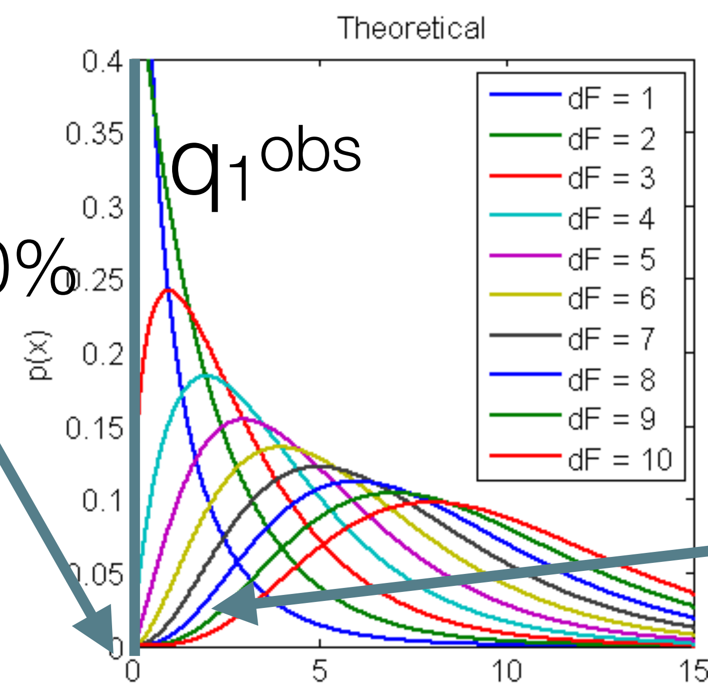
In order to evaluate $\hat{\mu}$

$$\frac{dL}{d\mu} = \frac{d}{d\mu}\frac{(b+\mu s)^{N_1}}{N_1!}e^{-b-\mu s} = \frac{s(b+\mu s)^{N_1-1}}{N_1!}e^{-b-\mu s}\,(N_1 - b - \mu s)$$

If data under fluctuate below b the derivative is negative, so L decreases with μ and its maximum is at $\mu = 0 \rightarrow \hat{\mu} = 0$

$$q_1 = -2ln\left[\frac{L(N_1, b+s)}{L(N_1, b)}\right]$$

$CL_s = 0\%$




100%

This term works like $1-p_b$ in the $CL_s$ method,
if s<<b $L(N_1,b+s) \sim L(N_1,b)$ and $q_1 = 0$, so we cannot exclude the signal at any confidence level.

# Summary

$CL_{s+b}$: coverage ok, but dangerous for s<<b;
$CL_s$: ok, but undercoverage
Likelihood ratio: coverage ok, protected for s<<b
can be used to test distributions

Up to now, discussed only about observation and exclusions, what about measurements?

Measurements are useful to look for deviations from SM, tune MC, check SM prediction: i.e. sin(2β), N.P. Kobayashi-Maskawa

I measure the Higgs mass $m_H$, what an error on $m_H$ means?

# Bayesian versus frequentist (the religious war)

1) the error on $m_H$ means that there is 68% probability that the true $m_H$ is between $m_H - \sigma_{mH}$ and $m_H + \sigma_{mH}$

What this probability is? $m_H$ has only one value… Do we mean that if we generate 100 universes in the 68% of cases $m_H$ will lay in that interval?

# Bayesian versus frequentist (the religious war)

## Bayesian

1) the error on $m_H$ means that there is 68% probability that the true $m_H$ is between $m_H - \sigma_{mH}$ and $m_H + \sigma_{mH}$

~~What this probability is? $m_H$ has only one value… Do we mean that if we generate 100 universes in the 68% of cases $m_H$ will lie in that interval?~~

2) it is our degree of believe…, it is like a bet: What is the probability that Juventus will win the Italian league?

In this case it is subjective, and it tries to estimate an objective number:

given the parameters I know about Juventus potentiality to win a match, if I take a sample of those parameters and try to simulate a match, what is the fraction of times Juventus will win?

There is always something subjective in this.

# Frequentist approach (Neyman construction of conf. belt)

If the Higgs mass is $m_H$, 68% of the experiments will measure an interval [$m_H^{meas}$ - σ, $m_H^{meas}$ + σ] that will contain the value $m_H$.

There is no subjective statement, the probability has a strictly frequentist definition

Neyman construction of confidence belt:

f(x;θ) distribution of x given θ

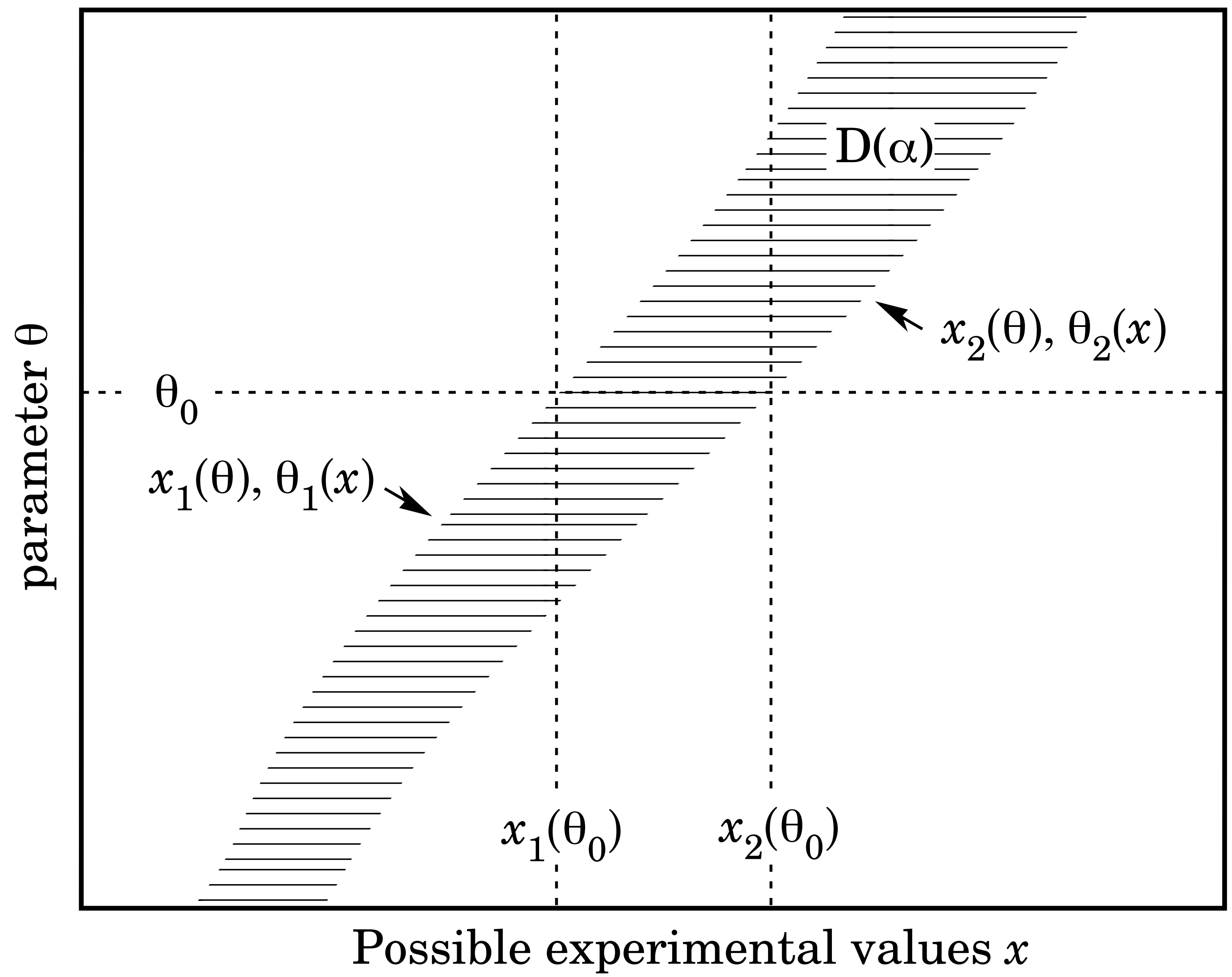
$$P(x_1 < x < x_2; \theta) = \int_{x_1}^{x_2} f(x;\theta)\, dx \geq 1 - \alpha$$

for 1σ 1-α = 0.68

when we change θ we get two curves for $x_1$ and $x_2$. We build the confidence belt using simulation.

The definition above doesn't define the belt fully, further conditions need to be applied: $x_1 = 0$ is a choice, or more often a symmetric condition is added:

$$x_1, x_2 : \int_{x_1}^{\bar{x}} f(x;\theta) dx = \int_{\bar{x}}^{x_2} f(x;\theta) dx$$



# Frequentist approach (Neyman construction of conf. belt)

If the Higgs mass is $m_H$, 68% of the experiments will measure an interval [$m_H^{meas}$ - σ, $m_H^{meas}$ + σ] that will contain the value $m_H$.

There is no subjective statement, the probability has a strictly frequentist definition

Neyman construction of confidence belt:

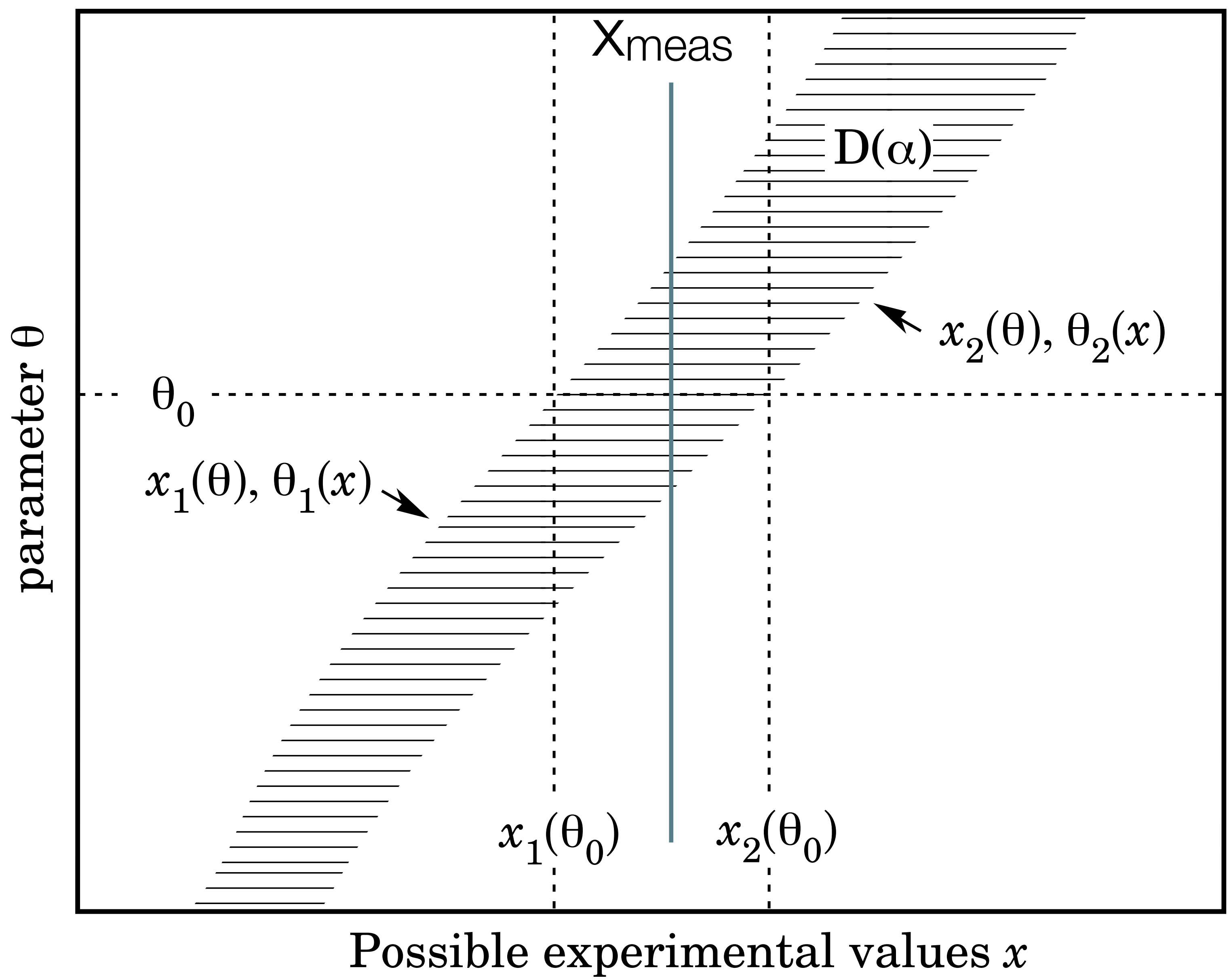
$f(x;\theta)$ distribution of x given θ

$$P(x_1 < x < x_2; \theta) = \int_{x_1}^{x_2} f(x;\theta)\, dx \geq 1 - \alpha$$

for 1σ 1-α = 0.68

when we change θ we get two curves for $x_1$ and $x_2$. We build the confidence belt using simulation.

Then we measure $x_{meas}$



# Frequentist approach (Neyman construction of conf. belt)

If the Higgs mass is $m_H$, 68% of the experiments will measure an interval $[m_H^{meas} - \sigma, m_H^{meas} + \sigma]$ that will contain the value $m_H$.

There is no subjective statement, the probability has a strictly frequentist definition

Neyman construction of confidence belt:

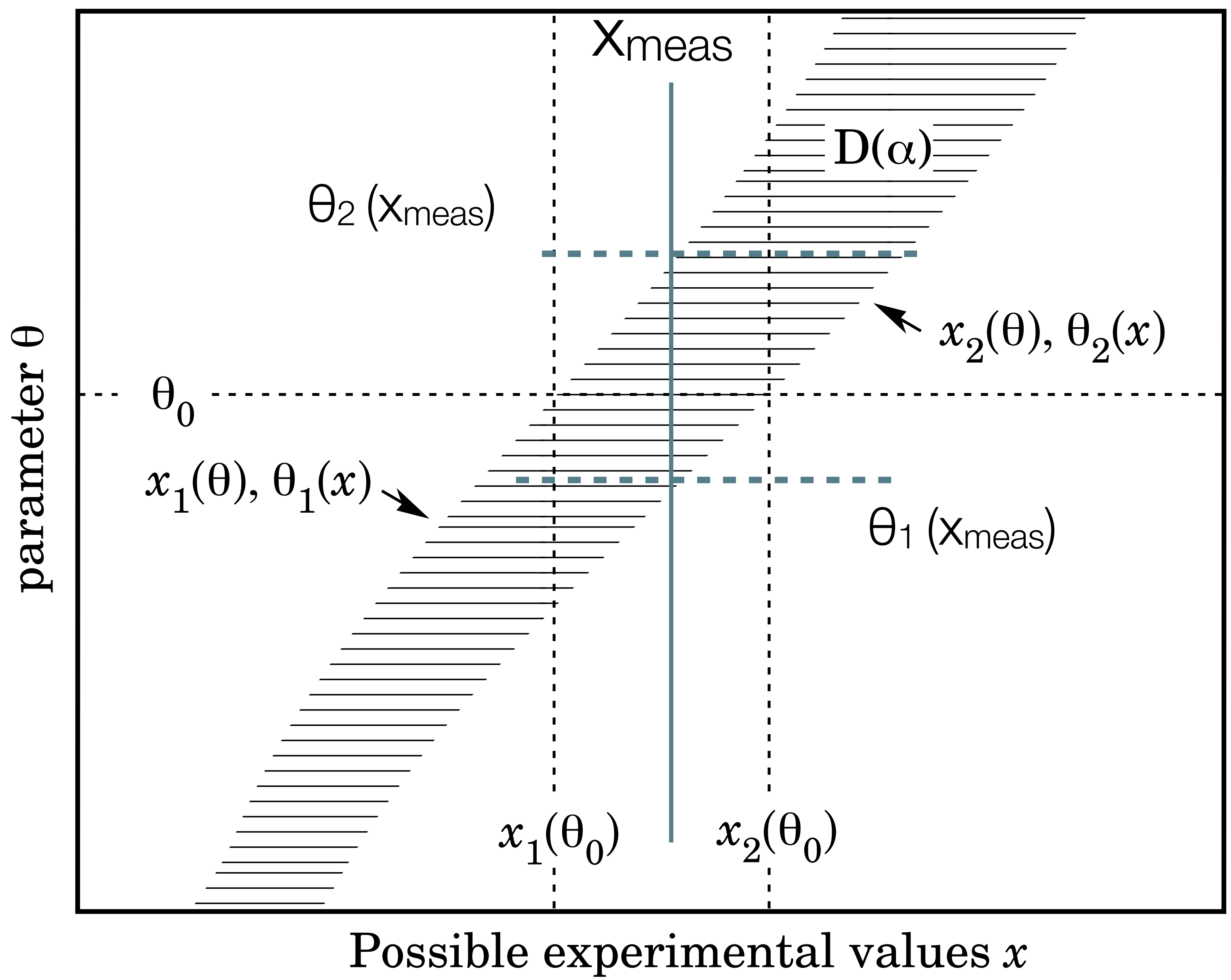
$f(x;\theta)$ distribution of x given $\theta$

$$P(x_1 < x < x_2; \theta) = \int_{x_1}^{x_2} f(x;\theta)\, dx \geq 1 - \alpha$$

for $1\sigma$ $1-\alpha = 0.68$

when we change $\theta$ we get two curves for $x_1$ and $x_2$. We build the confidence belt using simulation.

we set as interval for $\theta$ the range $[\theta_1, \theta_2]$.

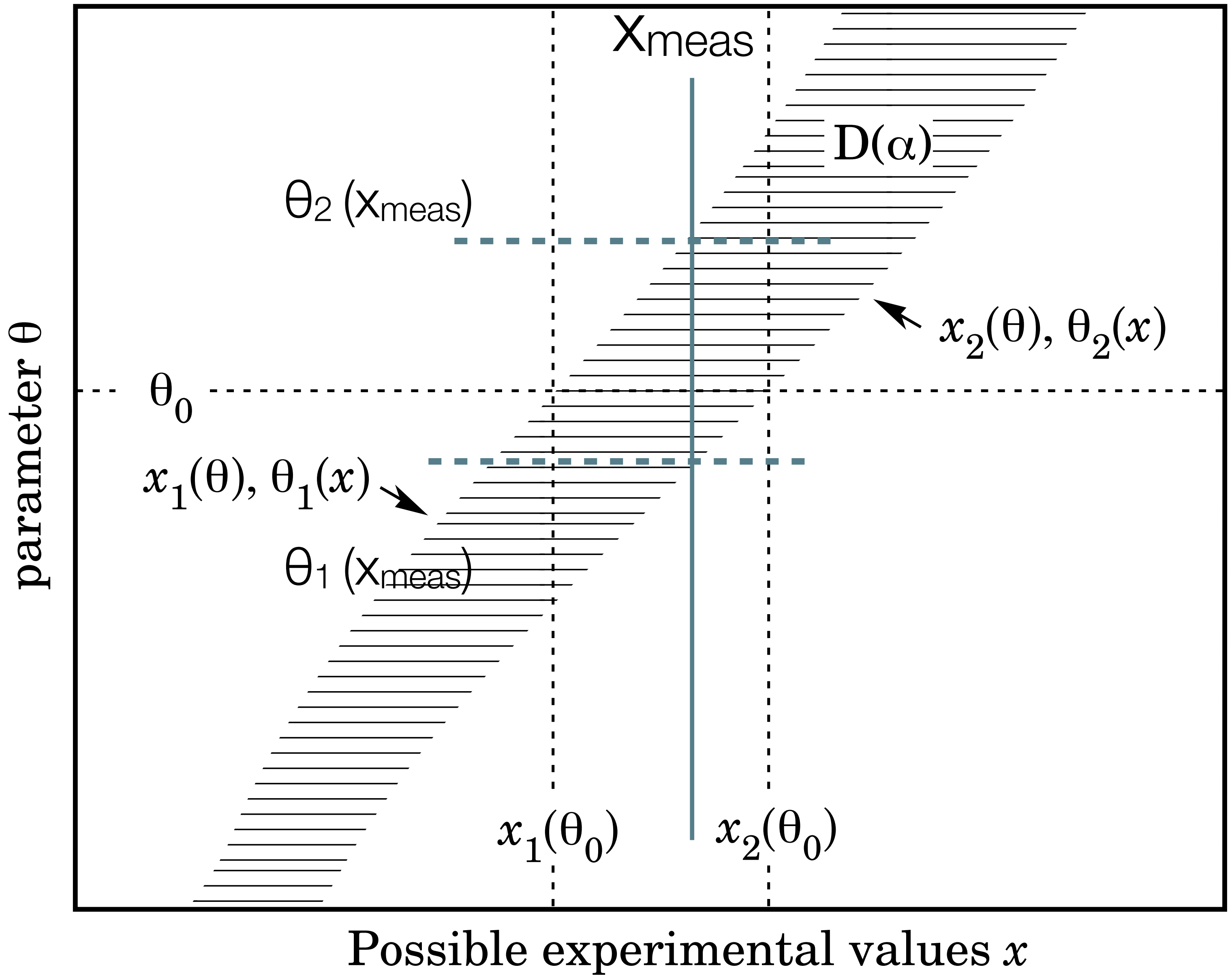


# Frequentist approach (Neyman construction of conf. belt)

If the Higgs mass is $m_H$, 68% of the experiments will measure an interval $[m_H^{meas} - \sigma, m_H^{meas} + \sigma]$ that will contain the value $m_H$.

There is no subjective statement, the probability has a strictly frequentist definition

if $\theta_0$ is the true value, we will have $x_1 < x_{meas} < x_2$ in 1-α of the cases (experiments) and consequently $\theta_1(X_{meas}) < \theta_0 < \theta_2(X_{meas})$ in the same fraction of cases, where $\theta_1$ and $\theta_2$ are random variables that is the outcome of the experiment.



# The Bayesian Way

likelihood of measured x given θ

new distribution for θ, improved after the measurement of x

a-priori distribution for θ

$$p(\theta \mid x) = \frac{L(x \mid \theta)\pi(\theta)}{\int L(x \mid \theta)\pi(\theta)d\theta}$$

if θ and x are random variables, this is a theorem otherwise it is the definition of $p(\theta,x)$

- Can the model have a probability?
- We assign a degree of belief in models parameterized by θ
- Instead of talking about confidence intervals we talk about credible intervals, where p(θ|x) is the credibility of θ given the data.

# Nuisance parameters (Systematics)

- **Nuisance** - something causing inconvenience or annoyance (Oxford Dictionary)
- **Systematic Errors** are equivalent to **Nuisance parameters** in the statistics jargon

  - D. Sinervo classified uncertainties into three classes classes:
    - **Class I**: Statistics like – uncertainties that are reduced with increasing statistics. Example: Calibration constants for a detector whose precision of (auxiliary) measurement is statistics limited
    - **Class II**: Systematic uncertainties that arise from one's limited knowledge of some data features and cannot be constrained by auxiliary measurements … One has to do some assumptions. Example: Background uncertainties due to fakes, isolation criteria in QCD events, shape uncertainties…. These uncertainties do not normally scale down with increasing statistics
    - **Class III**: The "Bayesian" kind… The theoretically motivated ones… Uncertainties in the model, Parton Distribution Functions, Hadronization Models…..

# Nuisance Parameters (Systematics)

- There are two related issues:
  – Classifying and estimating the systematic uncertainties
  – Implementing them in the analysis
- The physicist must make the difference between cross checks and identifying the sources of the systematic uncertainty.
  – Shifting cuts around and measure the effect on the observable…
    Very often the observed variation is dominated by the statistical uncertainty in the measurement.

# Treatment of Systematic Errors, the Bayesian Way

- Marginalization (Integrating) (The C&H Hybrid) **Cousins and Highland**
  - Integrate L over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others...)
  - Consistent Bayesian interpretation of uncertainty on nuisance parameters
- Note that in that sense MC “statistical” uncertainties (like background statistical uncertainty) are systematic uncertainties

# Integrating Out The Nuisance Parameters (Marginalization)

$$p(\theta,\lambda \mid x) = \frac{L(x \mid \theta,\lambda)\pi(\theta,\lambda)}{\int L(x \mid \theta,\lambda)\pi(\theta,\lambda)d\theta d\lambda} = \frac{L(x \mid \theta,\lambda)\pi(\theta,\lambda)}{Normalization}$$

- Our degree of belief in θ is the sum of our degree of belief in θ given λ (nuisance parameter), over "all" possible values of λ

$$p(\theta \mid x) = \int p(\theta,\lambda \mid x)d\lambda$$

# Priors

$$P(\theta \mid data) \sim \int L(data \mid \theta, \lambda)\pi(\lambda)d\theta d\lambda$$

- A prior probability is interpreted as a description of what we believe about a parameter preceding the current experiment
  - **Informative Priors**: When you have some information about $\lambda$ the prior might be informative (Gaussian or Truncated Gaussians…)
    - Most would say that subjective informative priors about the parameters of interest should be avoided ("….what's wrong with assuming that there is a Higgs in the mass range [115,140] with equal probability for each mass point?")
    - Subjective informative priors about the Nuisance parameters are more difficult to argue with
      - These Priors can come from our assumed model (Pythia, Herwig etc…)
      - These priors can come from subsidiary measurements of the response of the detector to the photon energy, for example.
      - Some priors come from subjective assumptions (theoretical, prejudice symmetries….) of our model

# Priors – Uninformative Priors

- **Uninformative Priors**: All priors on the parameter of interest should be uninformative….
  IS THAT SO?
  Therefore flat uninformative priors are most common in HEP.
  - When taking a uniform prior for the Higgs mass [115, $\infty$]… is it really uninformative? do uninformative priors exist?
  - When constructing an uninformative prior you actually put some information in it…

- **But** a prior flat in the coupling *g* will not be flat in $\sigma \sim g^2$
  Depends on the metric!
  (→ try Jeffrey Priors)
- Moreover, flat priors are improper and lead to serious problems of undercoverage (when one deals with >1 channel, i.e. beyond counting, one should AVOID them

–See Joel Heinrich Phystat 2005

# Choice of Priors

- A.W.F. Edwards: “Let me say at once that I can see no reason why it should always be possible to eliminate nuisance parameters. Indeed, one of the many objections to Bayesian inference is that is always permits this elimination.”

  Anonymous: “Who the ---- is A.W.F. Edwards…” http://en.wikipedia.org/wiki/A._W._F._Edwards

- But can you really argue with subjective informative priors about the Nuisance parameters
  (results of analysis are aimed at the broad scientific community.. See talk by Leszek Roszkowski constrained MSSM)

- Choosing the right priors is a science by itself

- Should we publish Bayesian (or hybrid ) results with various priors?
- Should we investigate the coverage of Bayesian (credible) intervals?

- Anyway, results should be given with the priors specified

# C&H Hybrid Method

- This method is coping with the Nuisance parameters by averaging on them weighted by a posterior.
- The Bayesian nature of the calculation is in the Nuisance parameters only….
- Say in a subsidiary measurement y of b, then the posterior is *p(b|y)*; $\mu$ is the *x* expectation.
- C&H will calculate the p-value of the observation $(x_o, y_o)$

$$p(x_o, y_o \mid \mu) = \int_0^{\infty} p(x_o \mid y_o, \mu) p(b \mid y_o) db$$

$$p(b \mid y_o) = \frac{p(y_o \mid b) p(b)}{p(y_o)}$$

$$p(y_o \mid b) = G(y_o \mid b, \sigma_b)$$

$$p(b) \ \ uniform$$

Note:
The original C&H used the Luminosity as the Nuisance parameter….

C&H Cousins & Highland

# The Profile Likelihood Method

$$\ell(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \Rightarrow Q(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \qquad -2\ln Q(s) = -2\ln\frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \to \chi^2(s)$$

$\hat{s}$ $\hat{b}$ obtained by maximizing the denominator

$\hat{\hat{b}}$ obtained by maximizing the numerator

$$\Delta\chi^2 = 2.7 \to 90\%\ C.I.$$

- The advantages of the Profile Likelihood
  - It has been with us for years….. (MINOS of MINUIT) (**Fred James**)
  - In the asymptotic limit it is approaching a $\chi^2$ distribution

**F. James**, e.g. Computer Phys. Comm. 20 (1980) 29 -35
W. Rolke, A. Lopez, J.Conrad. Nucl. Inst.Meth A 551 (2005) 493-503

# The Profile Likelihood for Significance Calculation

- A counting experiment with background uncertainty

$$L(n, b_{meas} \mid \mu, s, b) = Poiss(n \mid \mu s + b) G(b_{meas} \mid b, \sigma_b)$$

- The Likelihood-ratio

$$\lambda(\mu, b) = \frac{L(n, b_{meas} \mid \mu, s, b)}{L(n, b_{meas} \mid \hat{\mu}, s, \hat{b})}$$

Where $\hat{s}, \hat{b}$ are MLE

MLE: Maximum Likelihood Estimators

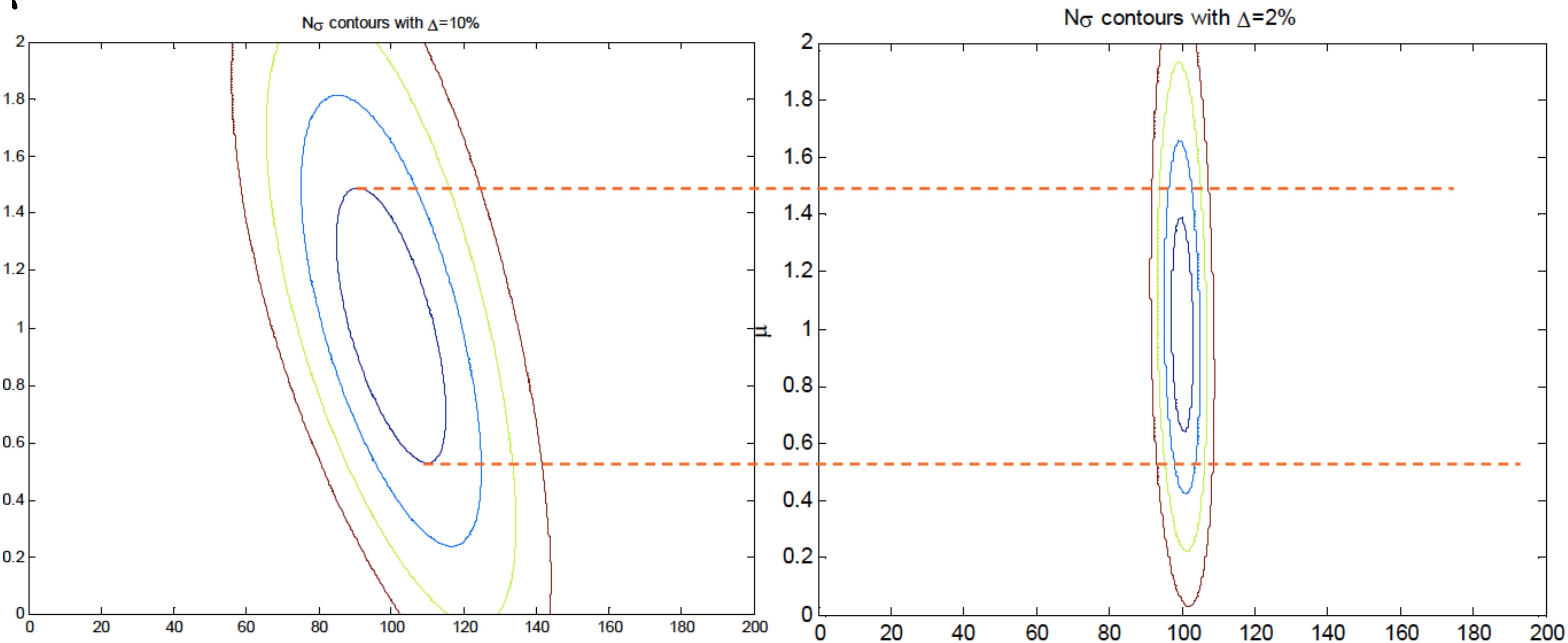
$-2 \log \lambda(\mu)$ is distributed as a $\chi^2$ with N degrees of freedom , N being the number of free parameters (parameters of interest)

(in this case N=2)

# Confidence intervals

μ

Δ fractional error on b

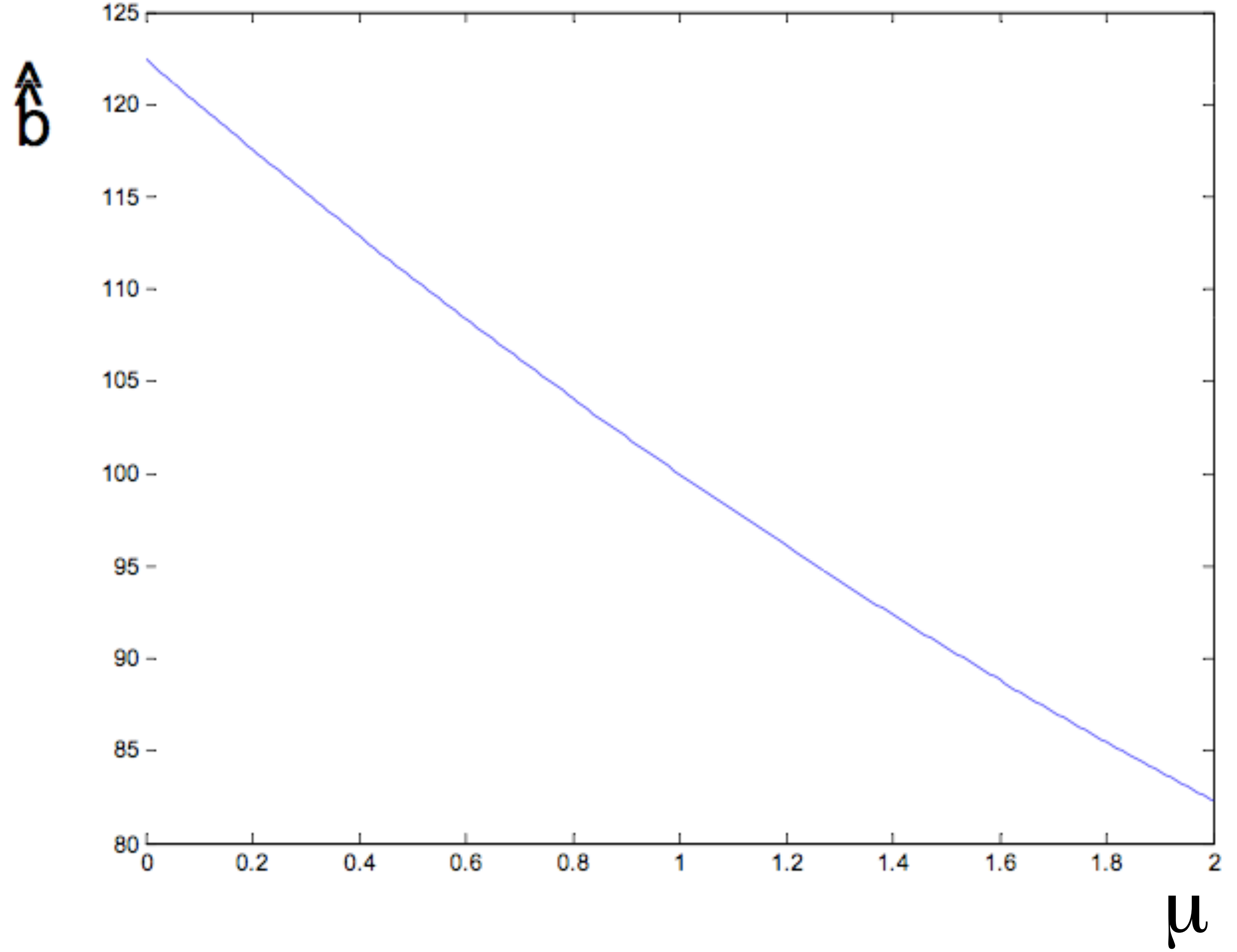



b b

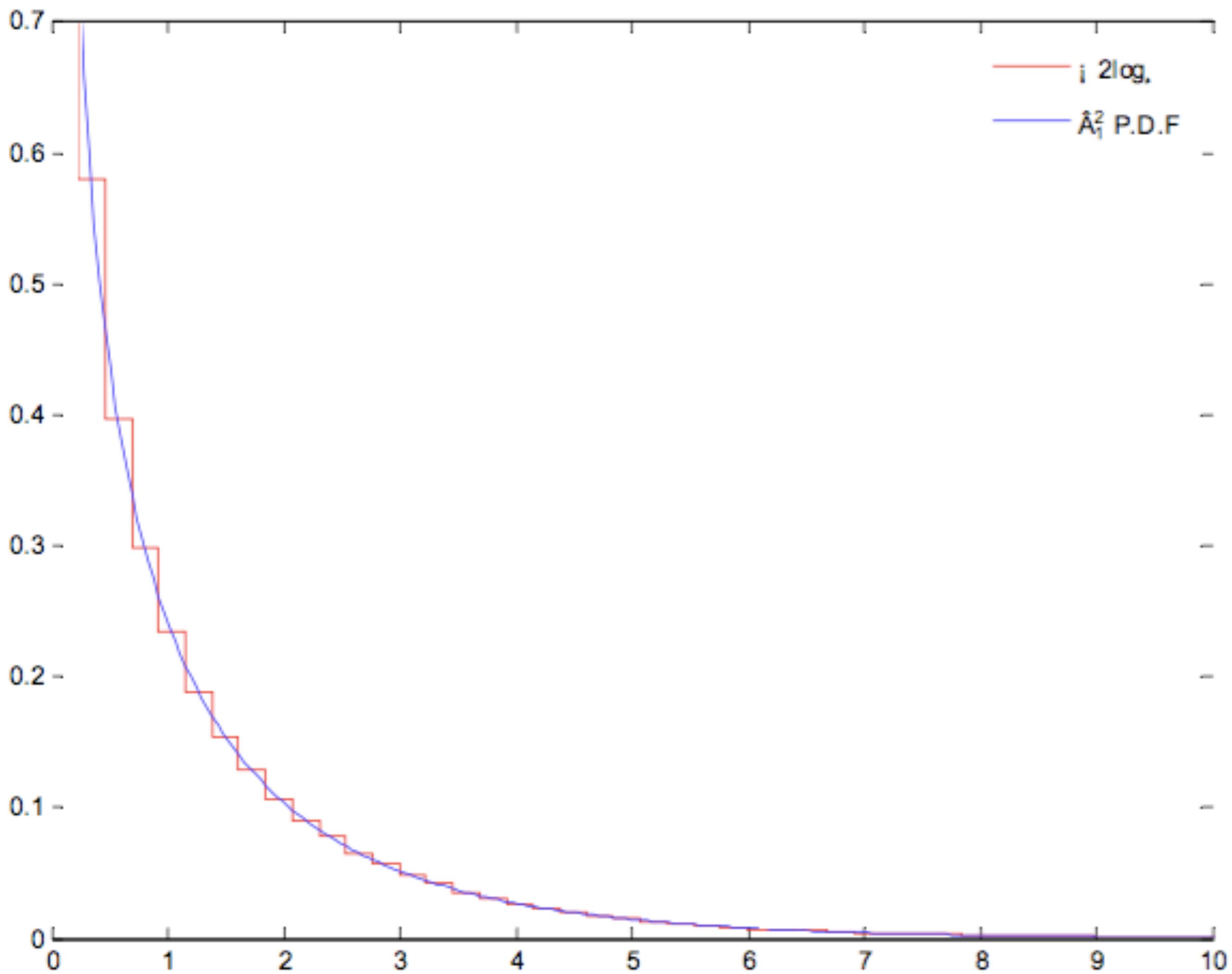
For $N_{obs}$=150 , $b_{meas}$ =100 , s=50

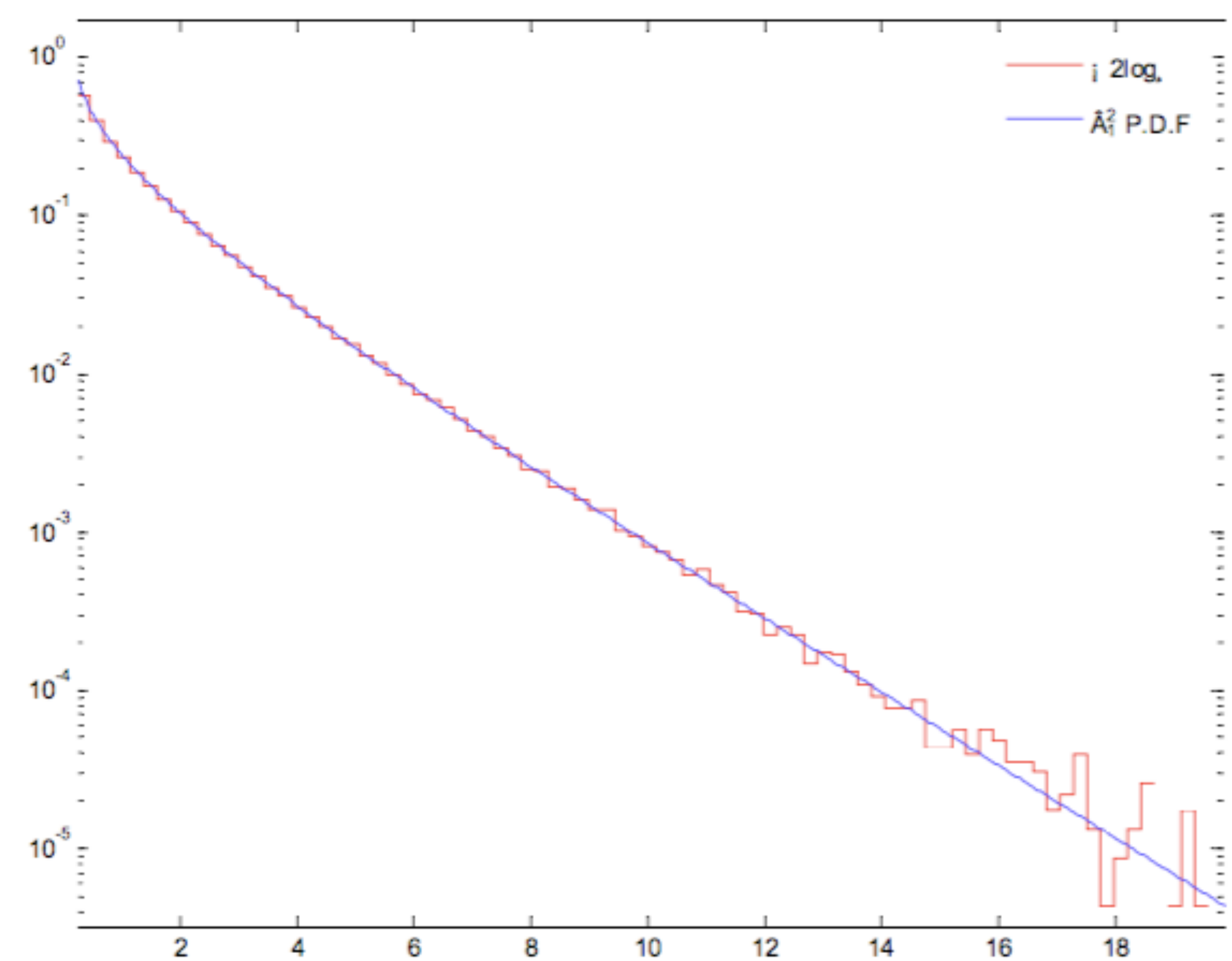
# Profile likelihood

$$\hat{\hat{b}}(\mu) = \frac{1}{2}\left\{b_{meas} - \mu s - \sigma_b^2 + \sqrt{(b_{meas} + \mu s - \sigma_b^2)^2 + 4n\sigma_b^2}\right\}$$



- distributes as a $\chi^2$ with 1 d.o.f
- this ensures simplicity, coverage, speed



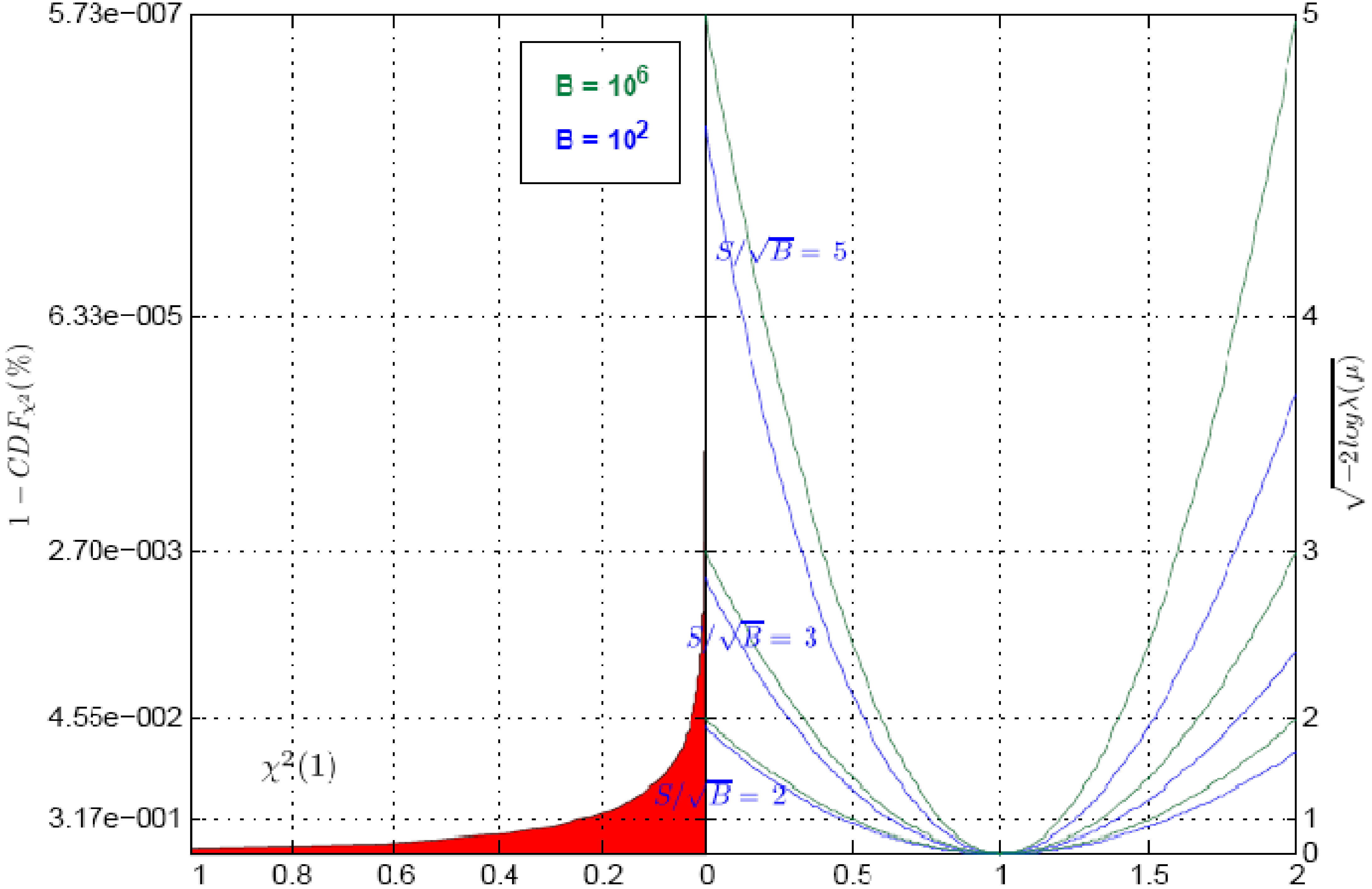
# The Profile Likelihood for Significance Calculation

$$-2\log\lambda(\hat{\mu} \pm N\sigma_{\hat{\mu}}) = N^2$$

$$N = \sqrt{-2\log\lambda(\mu)}$$

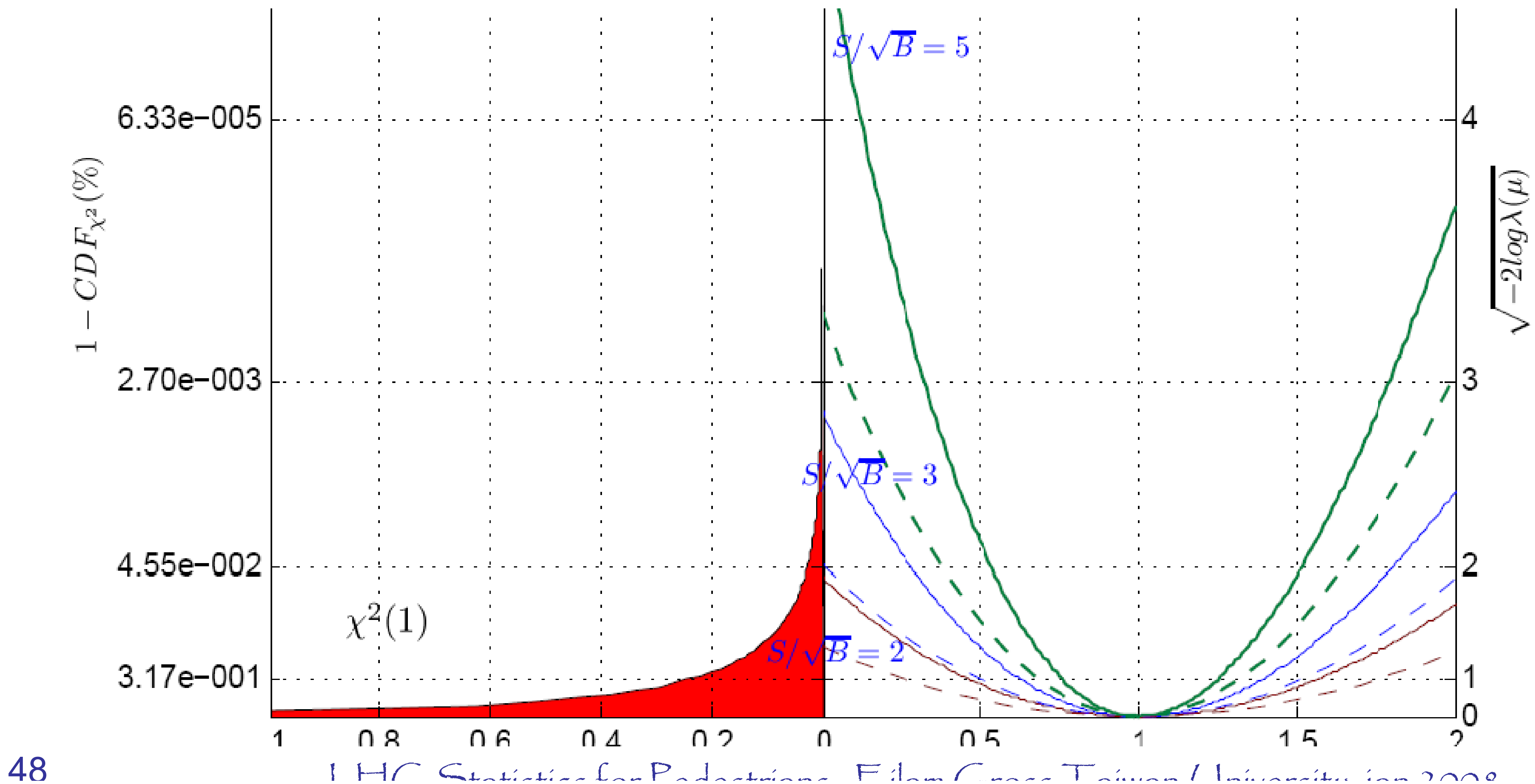
- In particular if we generate background only experiments, $\lambda(\mu=0)$ is distributed as $\chi^2$ with 1 d.o.f
- Discovery has to do with a low probability of the background only experiment to fluctuate and give us a signal like result….
- To estimate a discovery sensitivity we simulate a data compatible with a signal (s+b) and evaluate for this data $\lambda(\mu=0)$. For this data, the MLE of $\mu$ is 1

# 0% BG Systematics



# With 10% Background Systematics

For b=100 with 10% systematics, significance for S/√B=5 drops to ~3.6



# A lesson on systematics

- in absence of systematics the significance can be approximated to be $s/\sqrt{b}$
- however, if there is a fractional systematic error on b given by Δ

$$\frac{s}{\sqrt{(\sqrt{b})^2 + (\Delta \cdot b)^2}} = \frac{s}{\sqrt{b + \Delta^2 \cdot b^2}} \approx \frac{s}{\Delta \cdot b}$$

- to reach 5σ one needs  s/b > 5Δ
- with 10% systematics this implies: s/b > 0.5

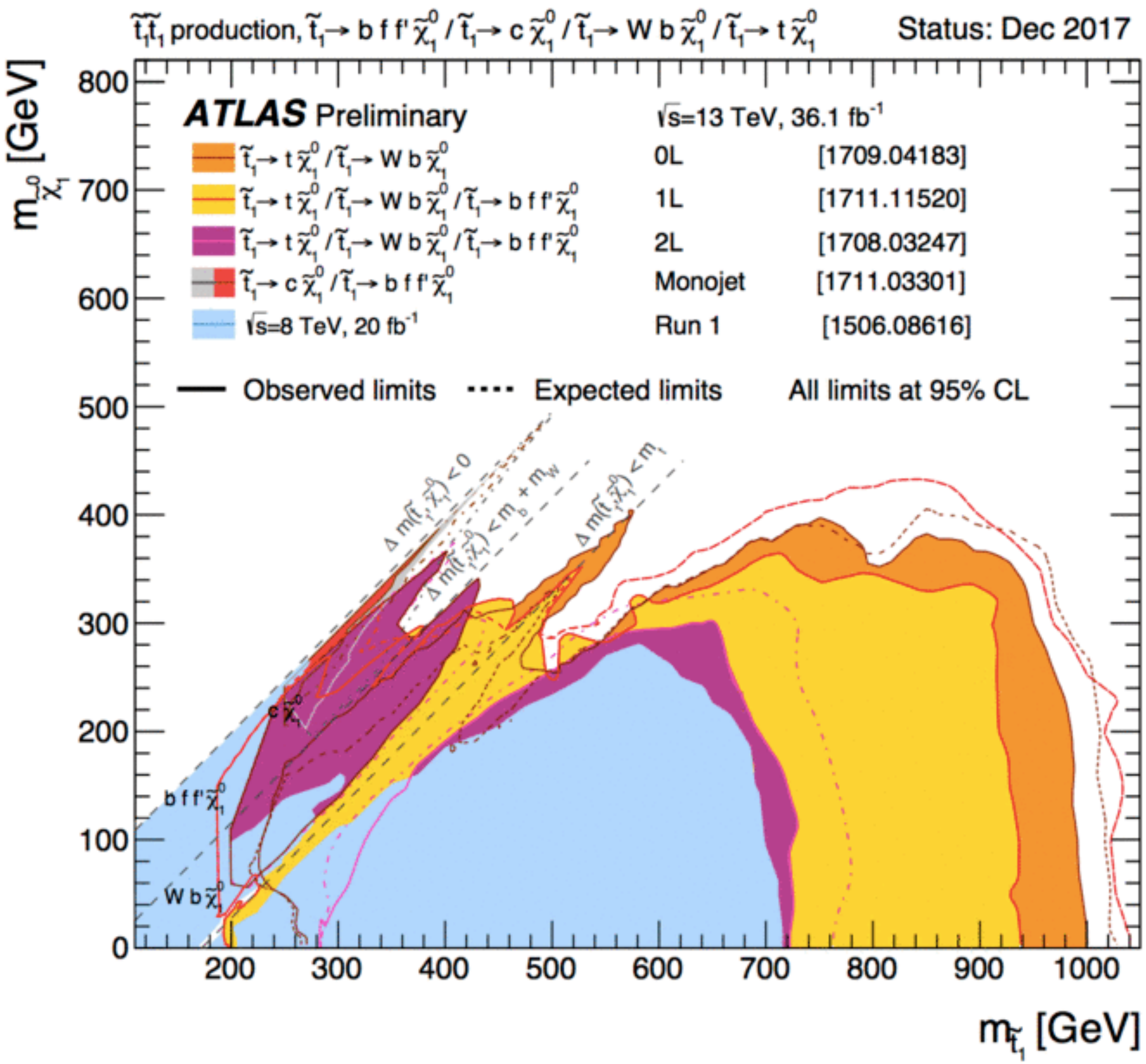
# Profile Likelihood

- The speed and ease allow us to produce all sorts of views in seconds!

- No numerical problems, can go up to any significance

# Why Profile Likelihood?

- For SUSY interpretations you usually have results in a grid (i.e. tg$\beta$,$m_A$)
- Each point is a different experiment
- There are 10s-100000s of possible points per channel
- In a shape-based analysis each bin is treated like a channel….
- The difference between O(minutes) per point and O(0.1 seconds) per point is critical!



$\tilde{t}_1\tilde{t}_1$ production, $\tilde{t}_1\rightarrow b f f' \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow c \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow W b \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow t \tilde{\chi}_1^0$
Status: Dec 2017
ATLAS Preliminary
$\sqrt{s}$=13 TeV, 36.1 fb$^{-1}$
$\tilde{t}_1\rightarrow t \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow W b \tilde{\chi}_1^0$
0L
[1709.04183]
$\tilde{t}_1\rightarrow t \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow W b \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow b f f' \tilde{\chi}_1^0$
1L
[1711.11520]
$\tilde{t}_1\rightarrow t \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow W b \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow b f f' \tilde{\chi}_1^0$
2L
[1708.03247]
$\tilde{t}_1\rightarrow c \tilde{\chi}_1^0$ / $\tilde{t}_1\rightarrow b f f' \tilde{\chi}_1^0$
Monojet
[1711.03301]
$\sqrt{s}$=8 TeV, 20 fb$^{-1}$
Run 1
[1506.08616]
Observed limits
Expected limits
All limits at 95% CL
$\Delta m(\tilde{t}_1,\tilde{\chi}_1^0) < 0$
$\Delta m(\tilde{t}_1,\tilde{\chi}_1^0) < m_b + m_W$
$\Delta m(\tilde{t}_1,\tilde{\chi}_1^0) < m_t$
$c\tilde{\chi}_1^0$
$b f f' \tilde{\chi}_1^0$
$W b \tilde{\chi}_1^0$
$m_{\tilde{\chi}_1^0}$ [GeV]
0
100
200
300
400
500
600
700
800
200
300
400
500
600
700
800
900
1000
$m_{\tilde{t}_1}$ [GeV]

# Exclusion with Profile Likelihood

- Exclusion is related to the probability of the "would be" signal to fluctuate down to the background only region (i.e. the p-value of the s+b "observation" )

- Here we suppose the data is the background only and the exclusion sensitivity is given by

$$N = \sqrt{-2\lambda(\mu = 1)}$$

- Exclusion at the 95% C.L. means N=2

# Signal Efficiencies Uncertainties

$$L(\mu \varepsilon s + b)$$

- How to cope with
  with background and efficiency systematics

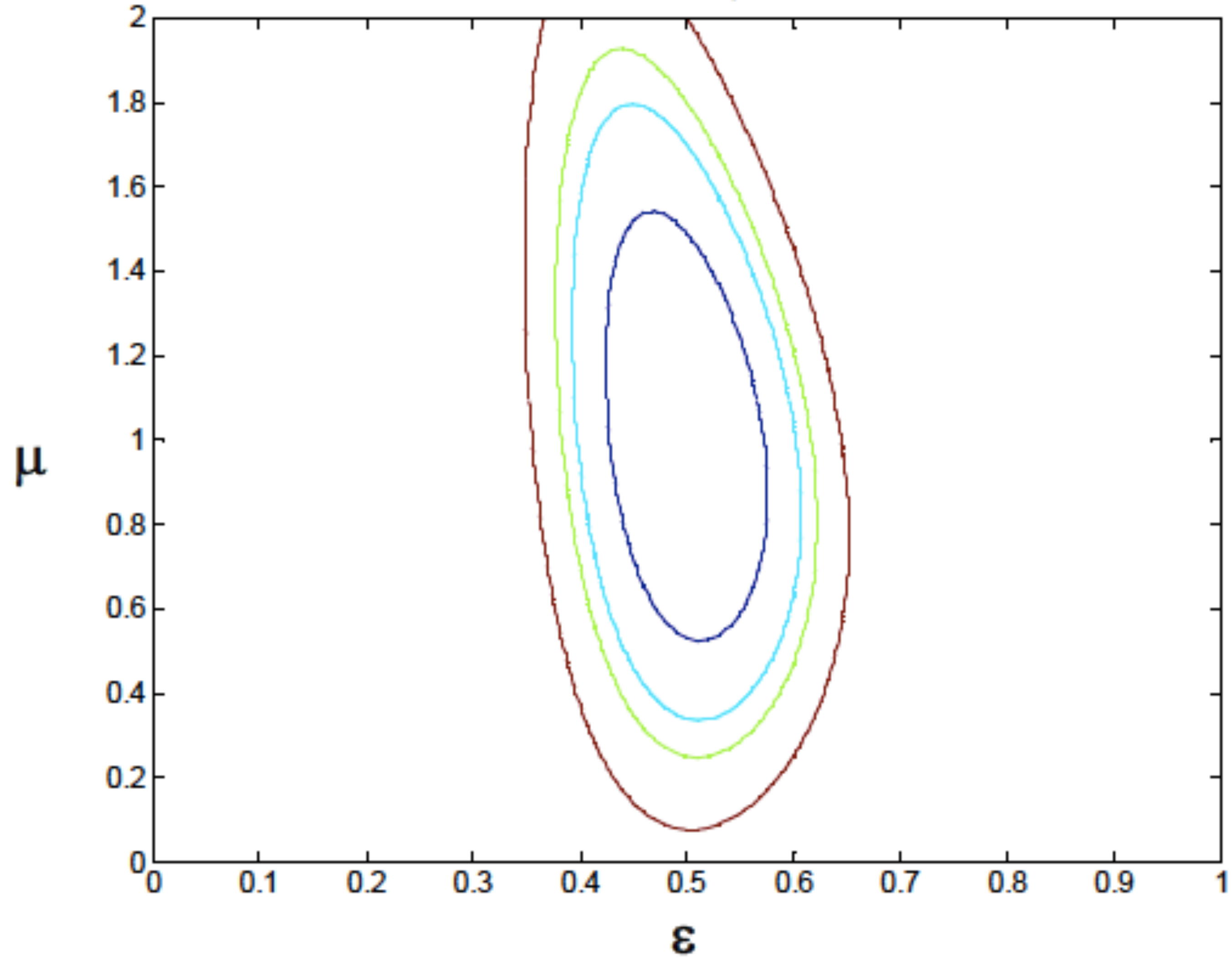
- Efficiency systematics have no effect on discovery sensitivity but can have large effects on exclusion sensitivity

# Including error on signal efficiency

$$L(\mu) = Poiss(n \mid \mu\varepsilon s + b)G(b_{meas} \mid b)G(\varepsilon_{meas} \mid \varepsilon)$$

$-2\log\lambda(\mu,\varepsilon)$

10% error on ε

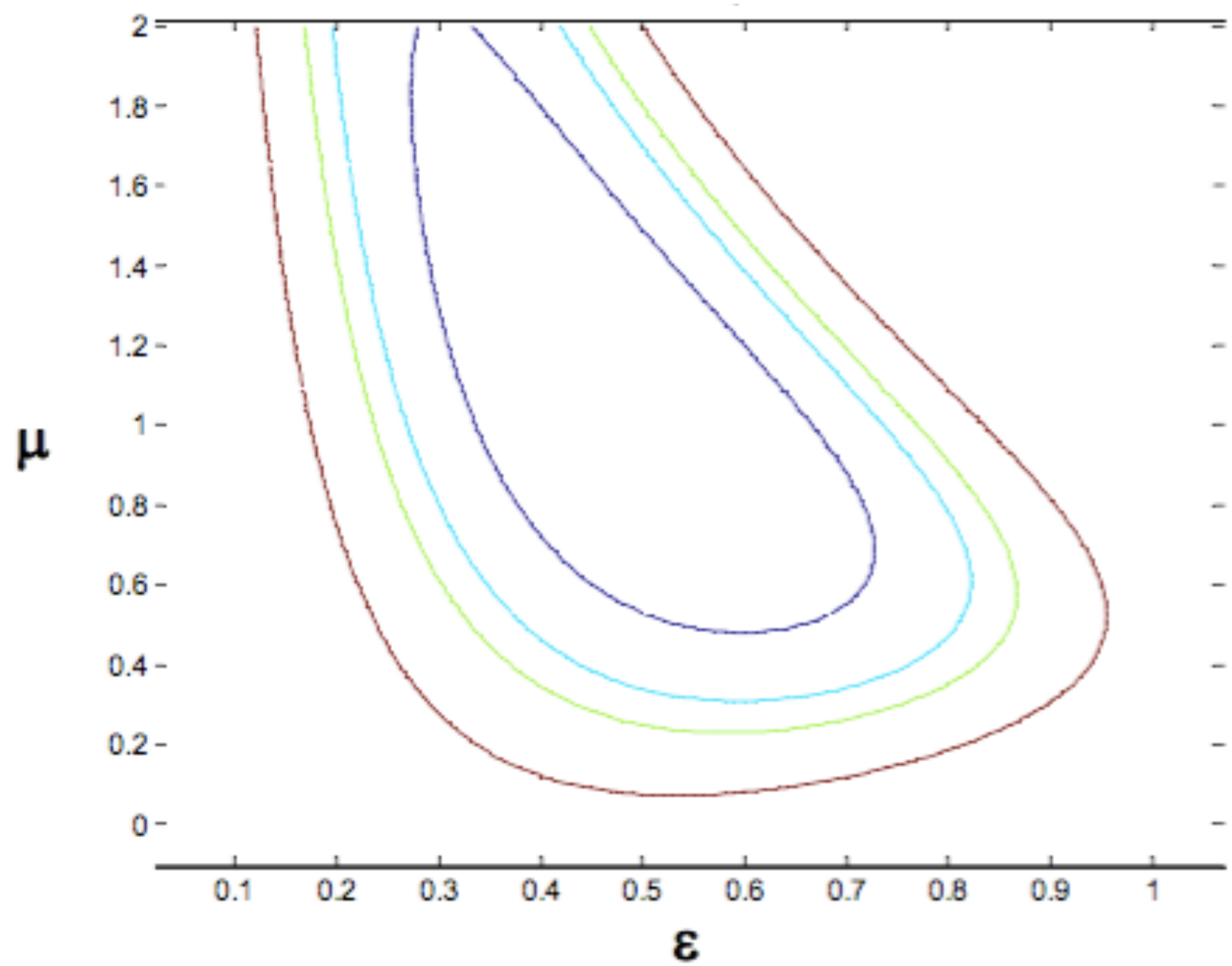


# Including error on signal efficiency

$$L(\mu) = Poiss(n \mid \mu\varepsilon s + b)G(b_{meas} \mid b)G(\varepsilon_{meas} \mid \varepsilon)$$

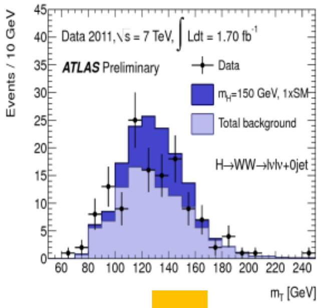
30% error on ε

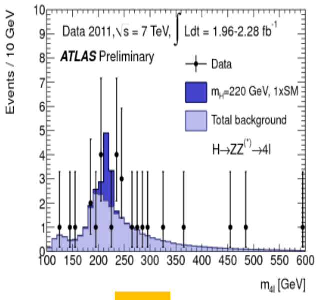
$$-2\log\lambda(\mu,\varepsilon)$$

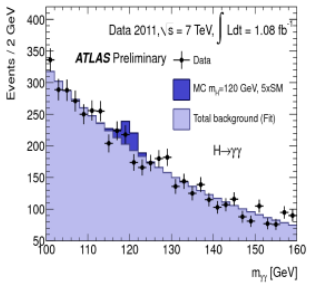


# Combining Higgs channels (and experiments)

- Procedure: define joint likelihood




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

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

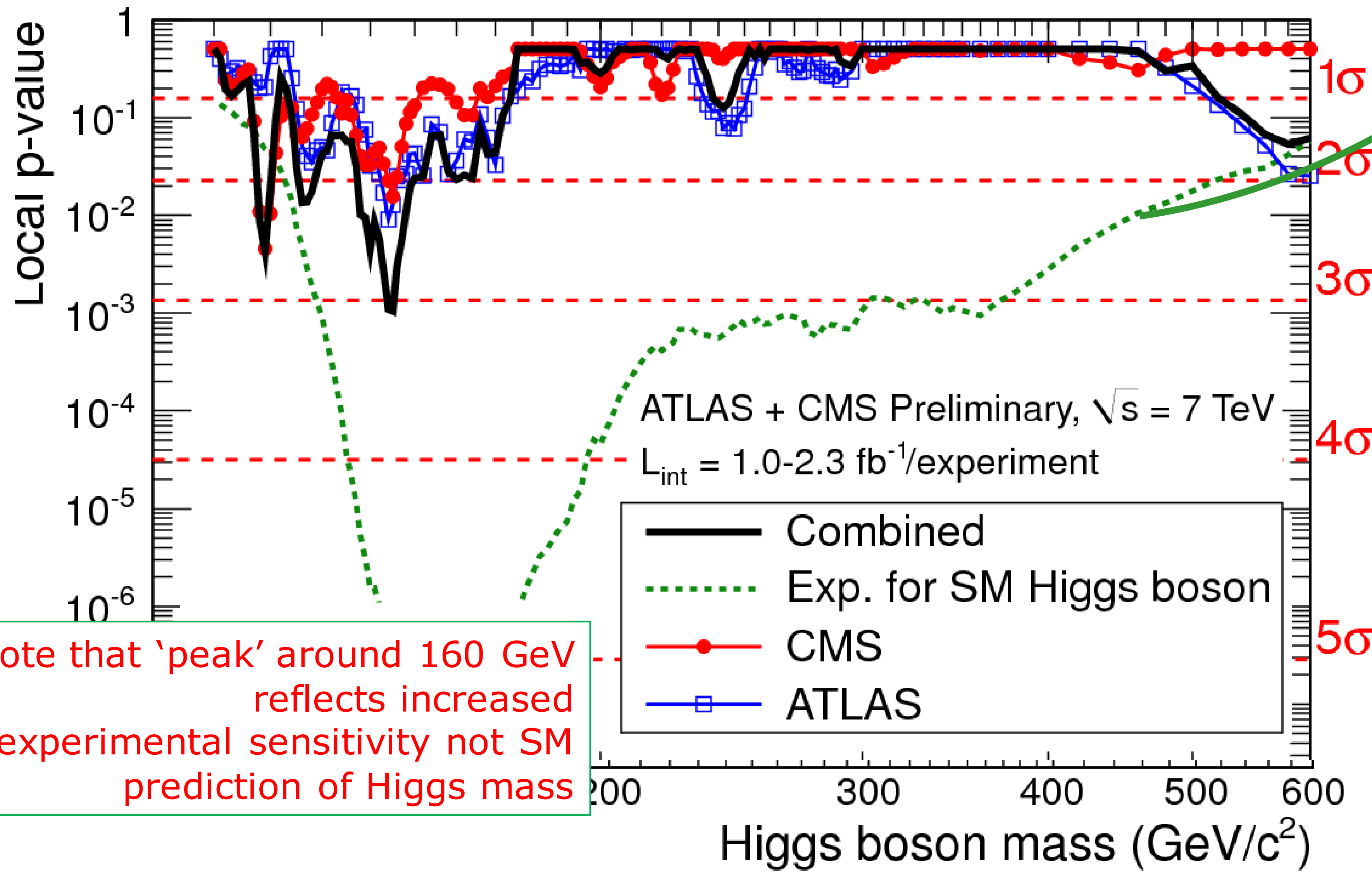
- Correlations between $\theta_{WW}$,$\theta_{\gamma\gamma}$ etc and between $\theta_{ATLAS}$,$\theta_{CMS}$ requires careful consideration!
- The construction profile likelihood ratio test statistic from joint likelihood and proceed as usual

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

Wouter Verkerke, NIKHEF

## Comb: p-value of background-only hypothesis (‘discovery’)

*Expected* p-value for background hypothesis of a data sample containing SM Higgs boson




Note that ‘peak’ around 160 GeV reflects increased experimental sensitivity not SM prediction of Higgs mass

# Conclusions

## Pros and Cons Profile Likelihood

- CONS:
  - The only disadvantage I see is its incapability to take the Look Elsewhere Effect in a built-in way….
  - One has to take the Look Elsewhere Effect in the LEP way
    (Using MC and factorize the resulting significance- need to be studied)

- PROS:
  - It is simple and easy to understand and apply
  - It is statistically reliable and a frequentists favorite
  - It can cope with Systematics and has the proper coverage
  - It is FAST!!!!!! O(0.1 Sec) vs O(Minutes).
  - Its probably the only method that can cope with as many as SUSY scenarios one wants!

# Basic Definitions

- Normally, we make **one** experiment and try to **estimate from this one experiment** the **confidence interval** at a specified CL% Confidence Level….

- In simple cases like Gaussians PDFs G(s,$s_{true}$) the Confidence Intrerval can be calculated analytically and ensures a complete coverage For example 68% coverage is precise for $\hat{s} \pm \sigma_{\hat{s}}$

# Frequentist approach (Neyman constr. of conf. belt)

$$P(x_1 < x < x_2; \theta) = \int_{x_1}^{x_2} f(x;\theta)\, dx \geq 1 - \alpha$$

it is not enough to define $x_1$ and $x_2$, need to add further informations: i.e. central values $x_c$ is such that $P(x < x_1) = P( x > x_2) = \alpha/2$