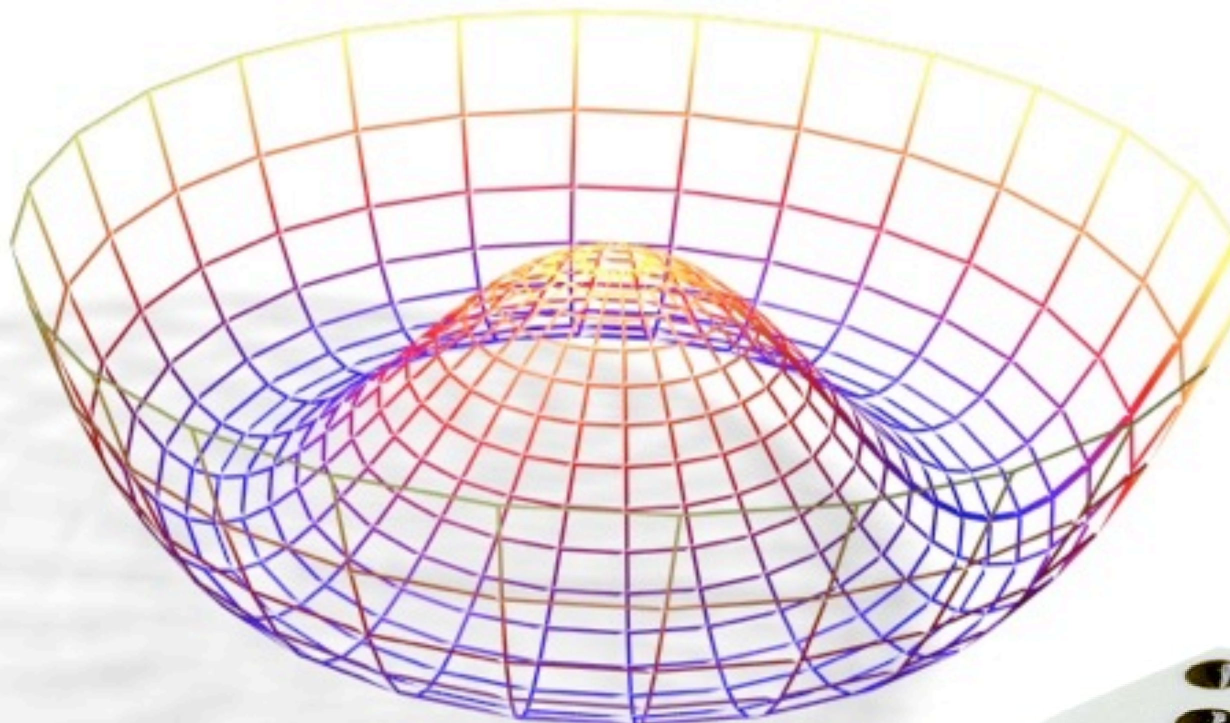
Center for
Cosmology and
Particle Physics

Practical Statistics for Particle Physics

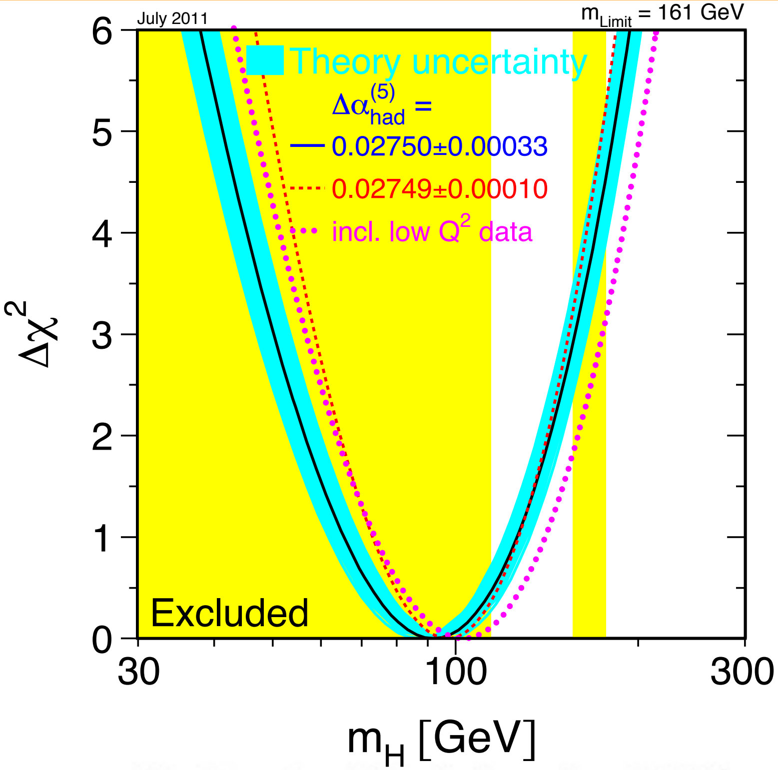


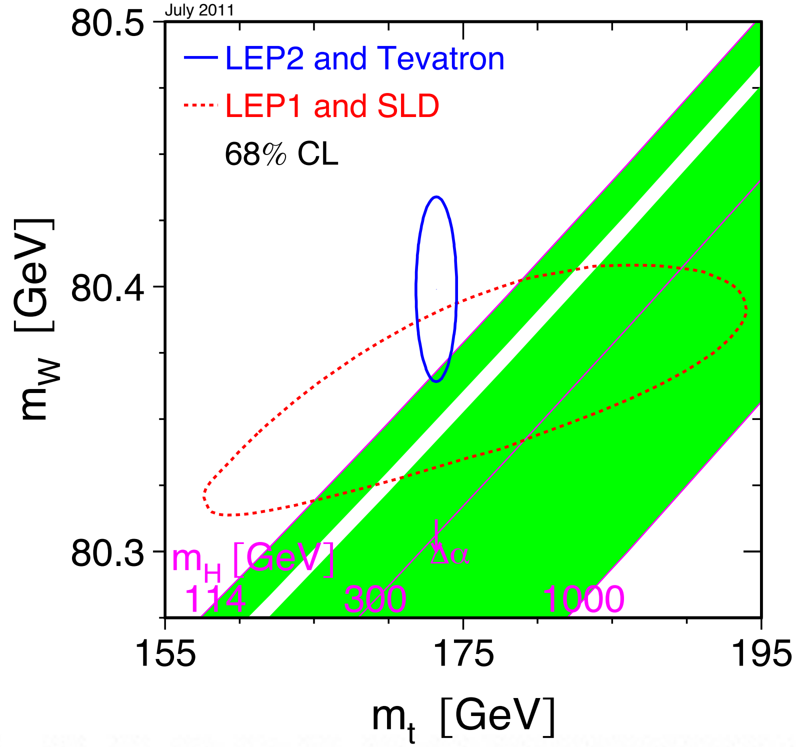
Kyle Cranmer,
New York University

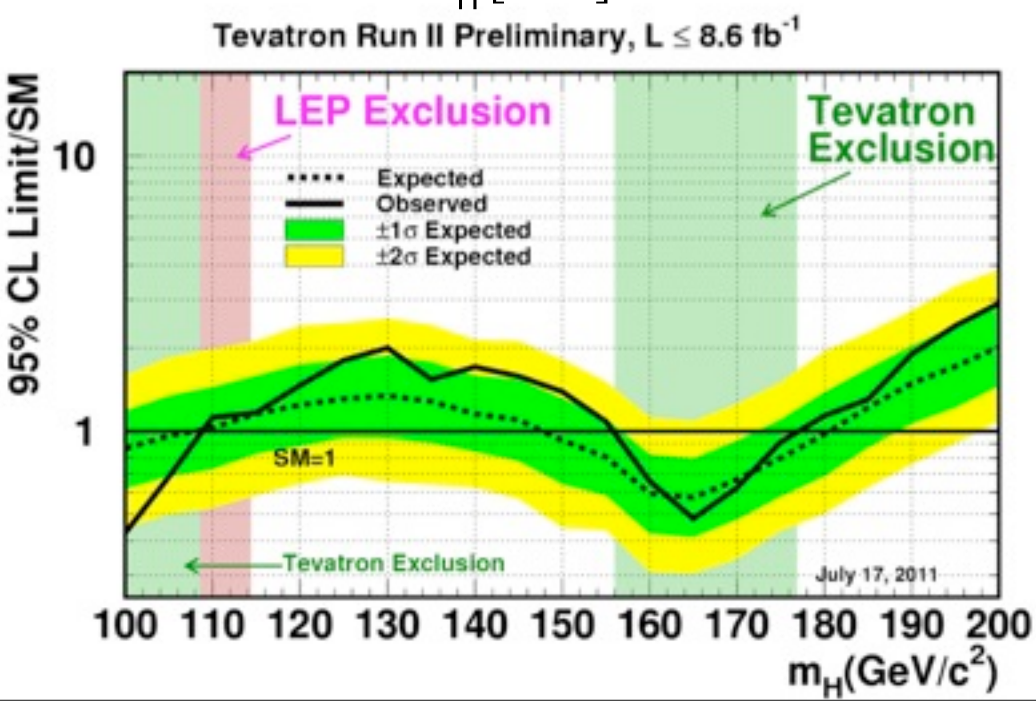
Limits
&
Confidence Intervals

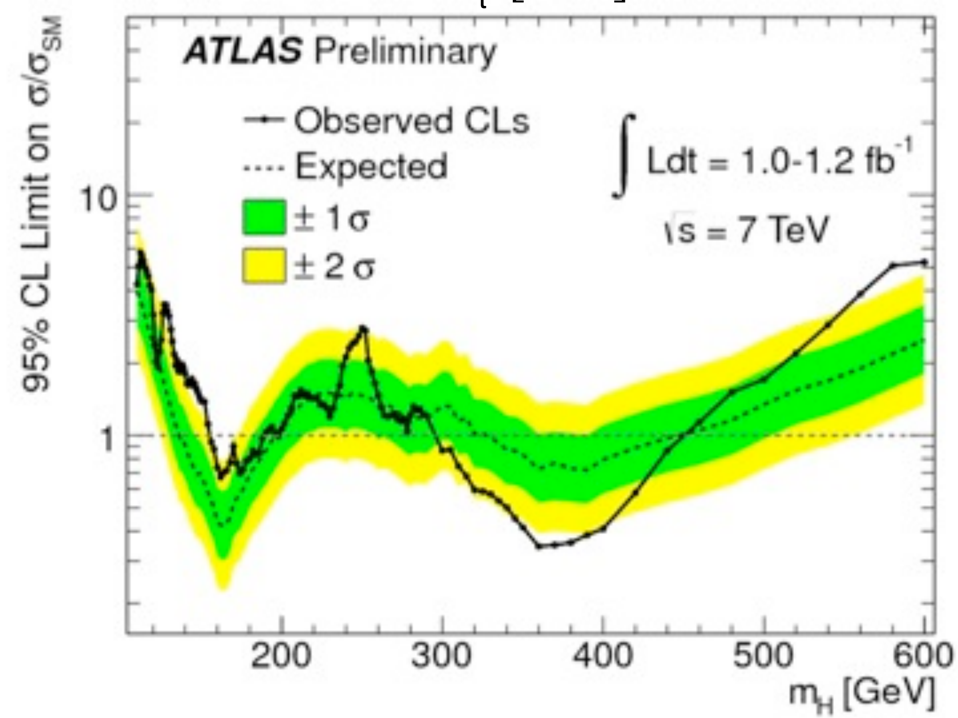
What do these plots mean?

Center for Cosmology and Particle Physics

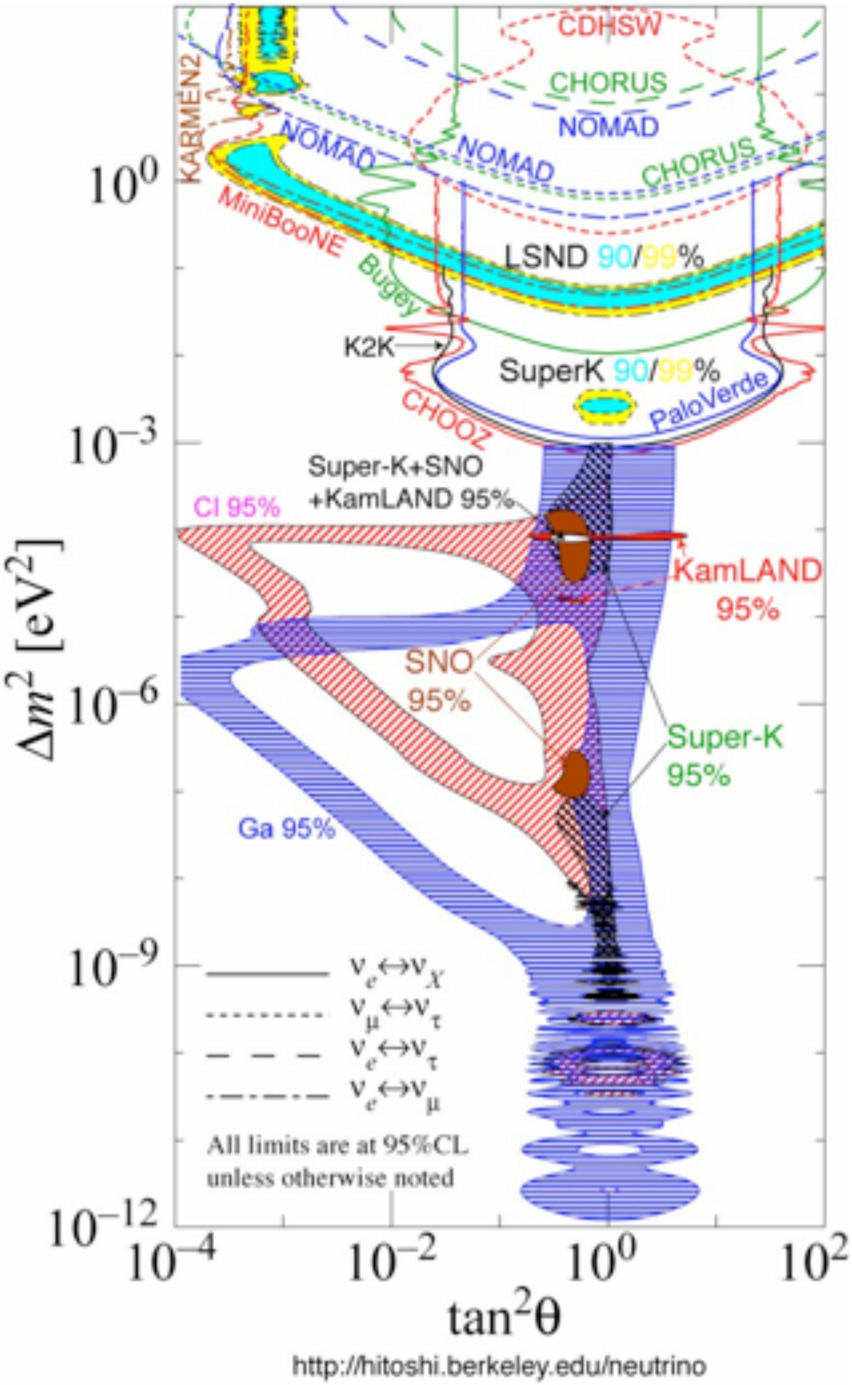


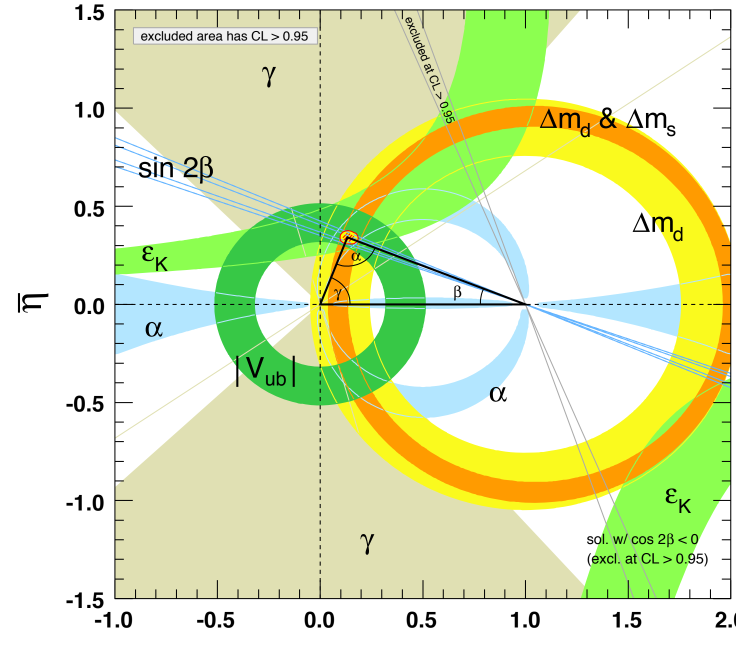


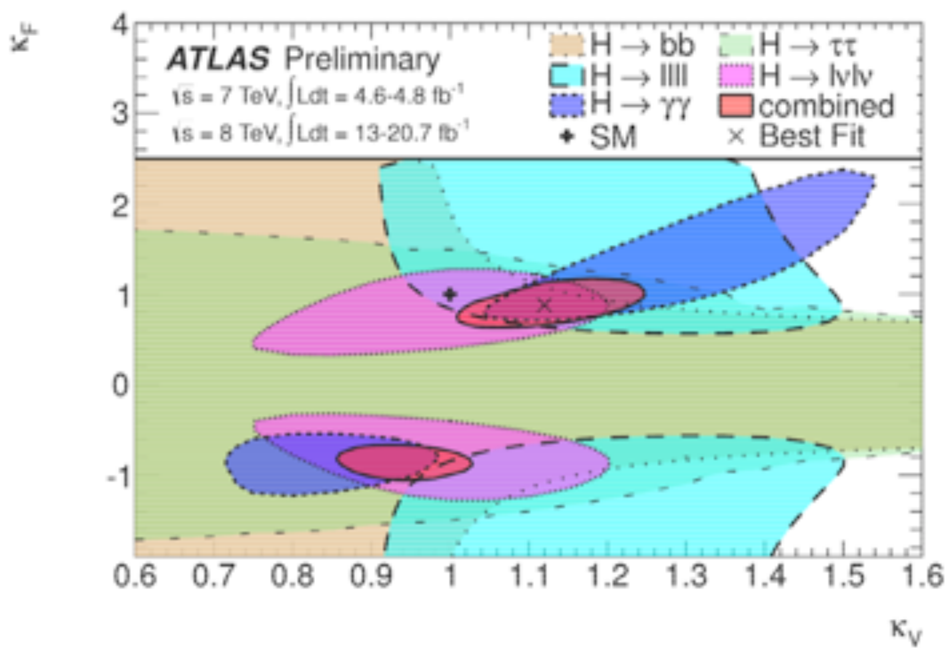




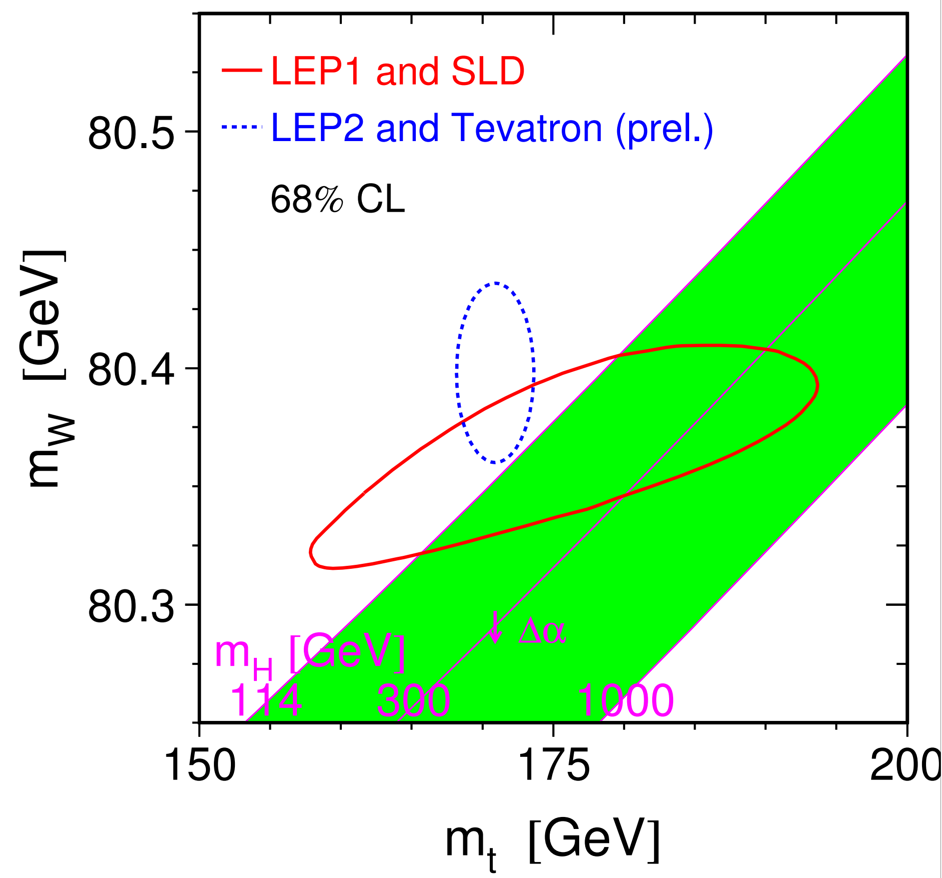
Other examples of Confidence Intervals





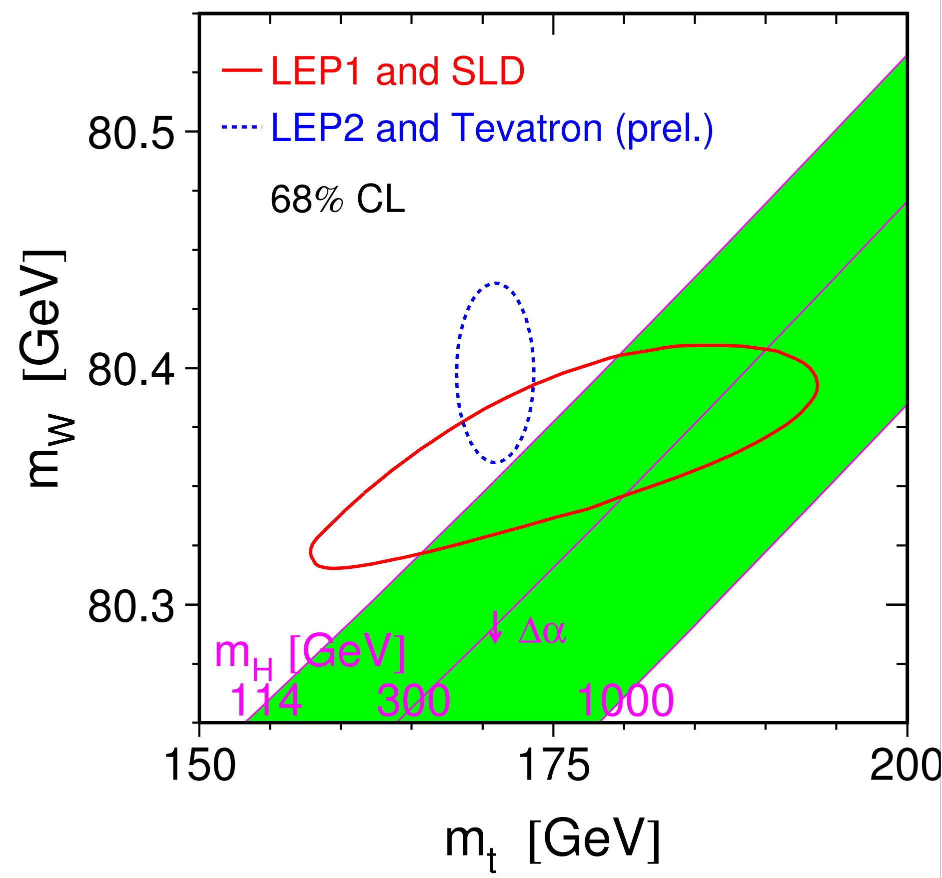


Confidence Interval



Confidence Interval

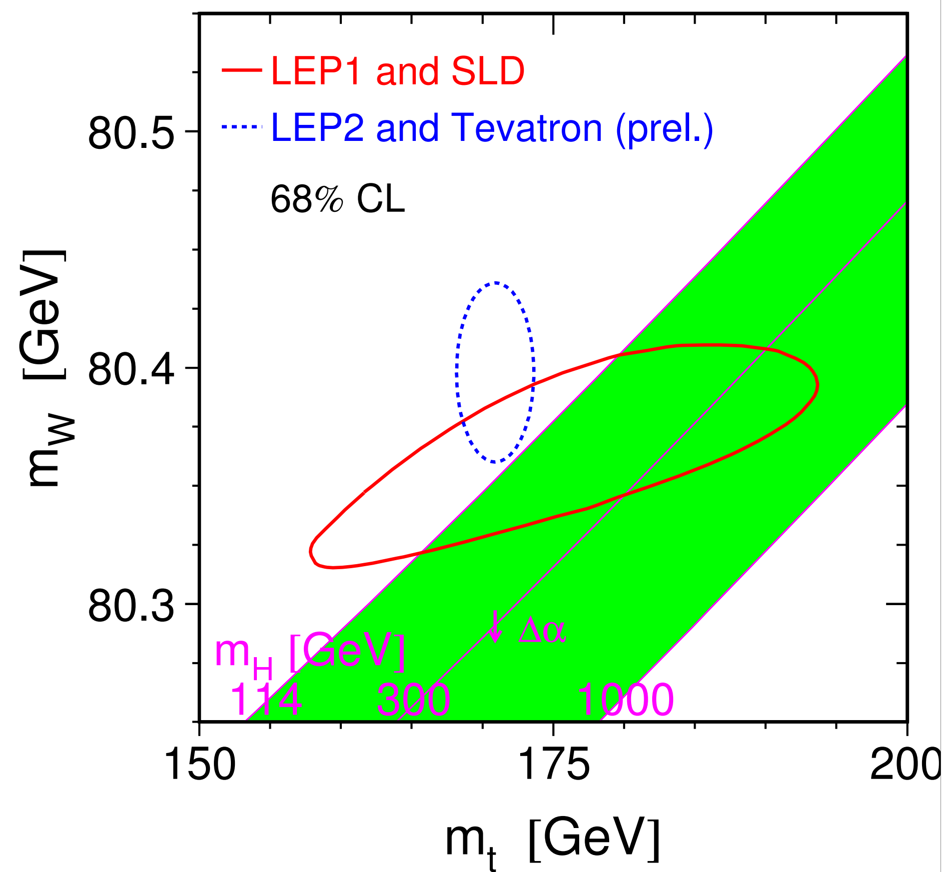
What is a “Confidence Interval?



Center for Cosmology and Particle Physics

What is a "Confidence Interval?

- you see them all the time:

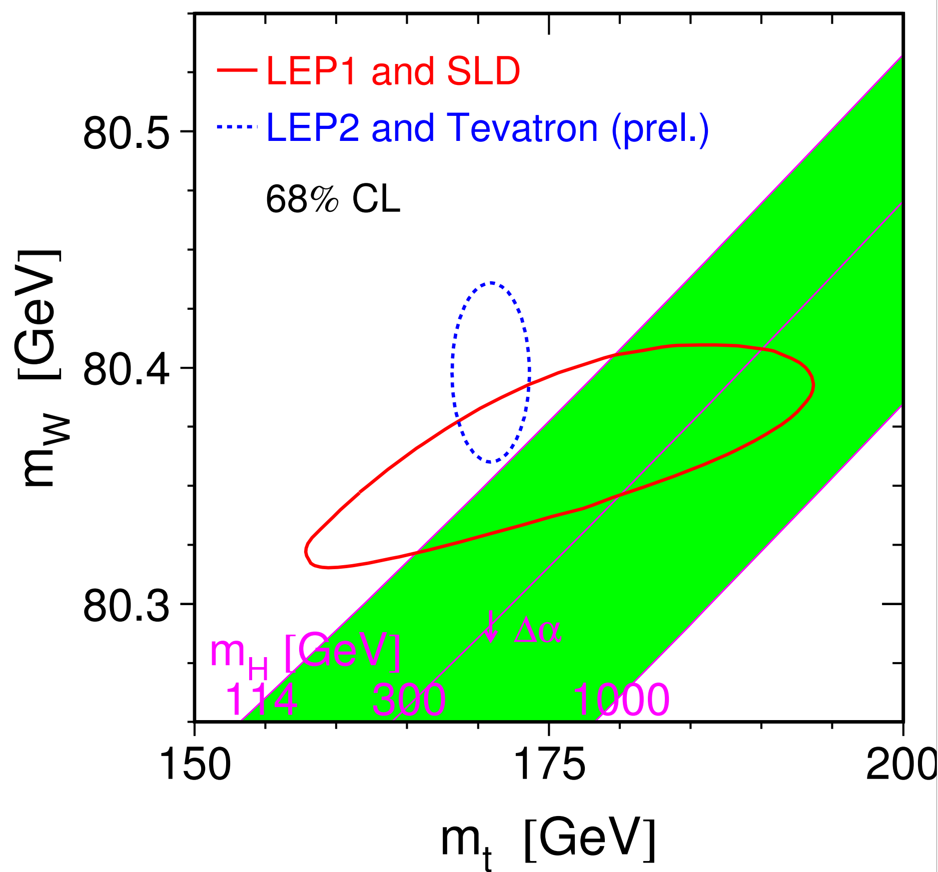


Confidence Interval

What is a “Confidence Interval?

- you see them all the time:

Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval



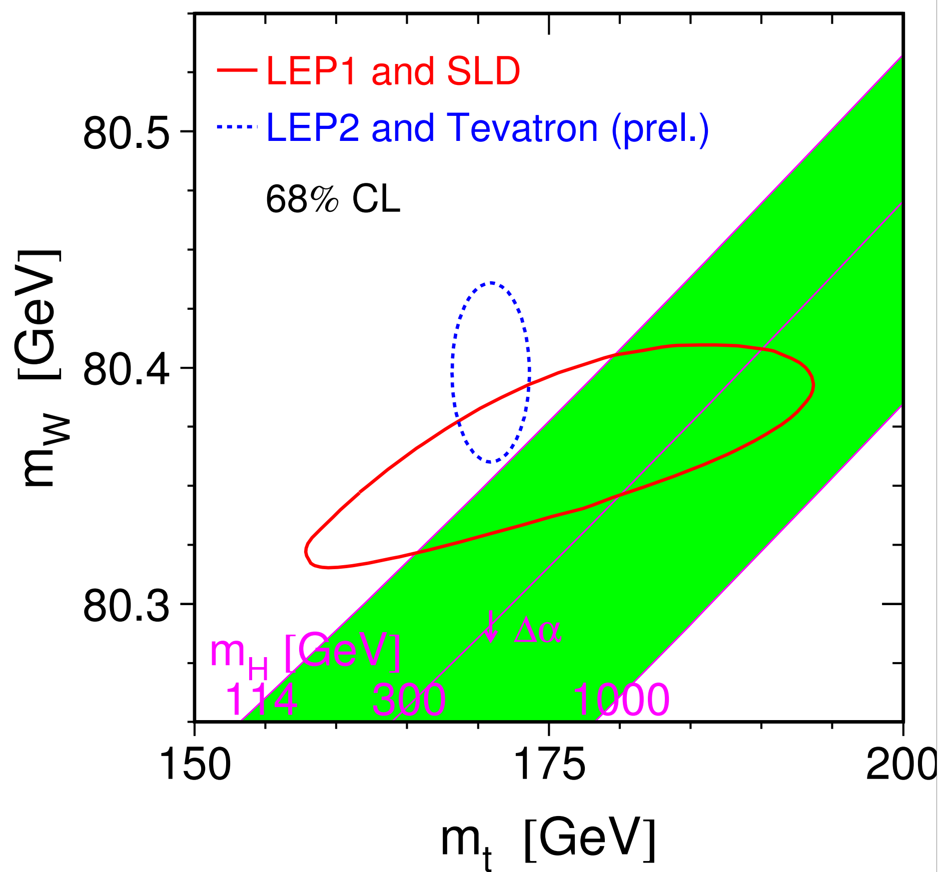
Confidence Interval

What is a “Confidence Interval?

- you see them all the time:

Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval

- but that’s P(theory|data)!



Confidence Interval

Center for Cosmology and Particle Physics

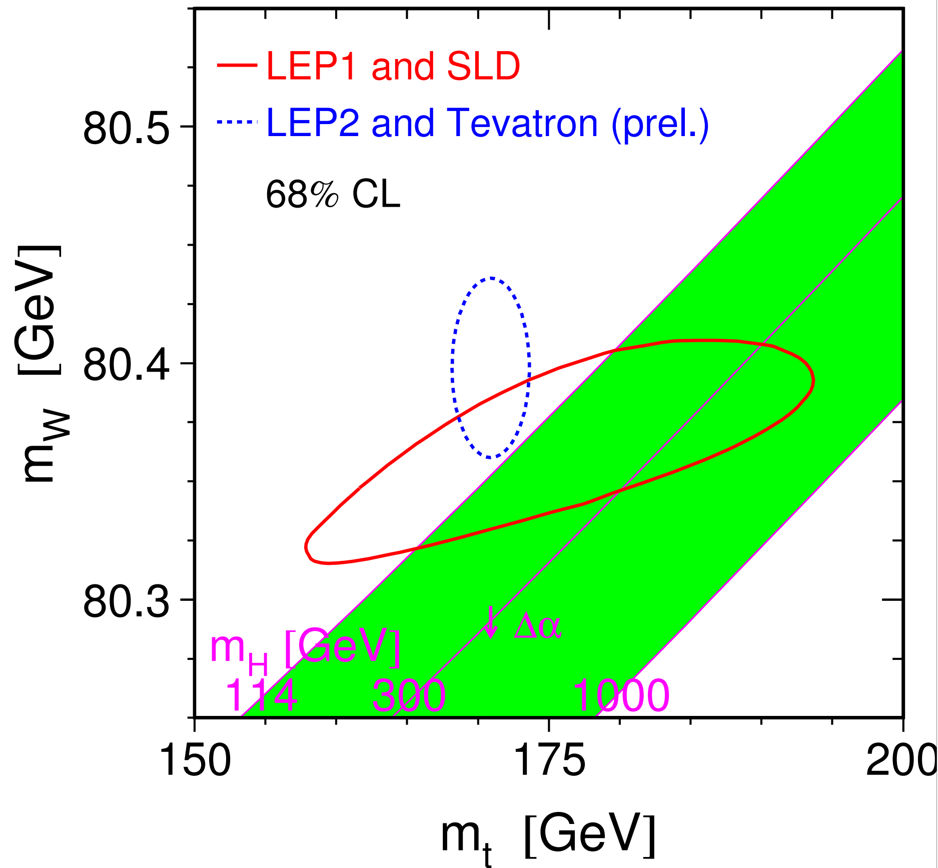
What is a "Confidence Interval?

- you see them all the time:

Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval

- but that's P(theory|data)!

Correct frequentist statement is that the interval **covers** the true value 68% of the time



Confidence Interval

What is a "Confidence Interval?

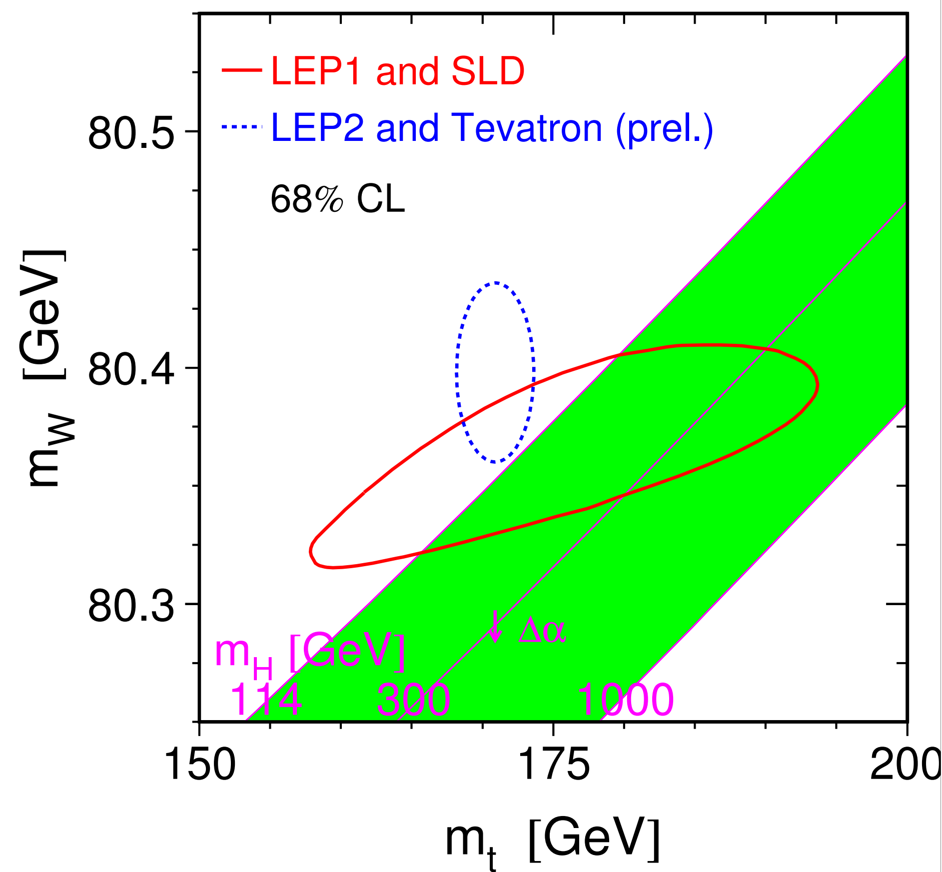
- you see them all the time:

Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval

- but that's P(theory|data)!

Correct frequentist statement is that the interval **covers** the true value 68% of the time

- remember, the contour is a function of the data, which is random. So it moves around from experiment to experiment



Confidence Interval

Center for Cosmology and Particle Physics

What is a "Confidence Interval?

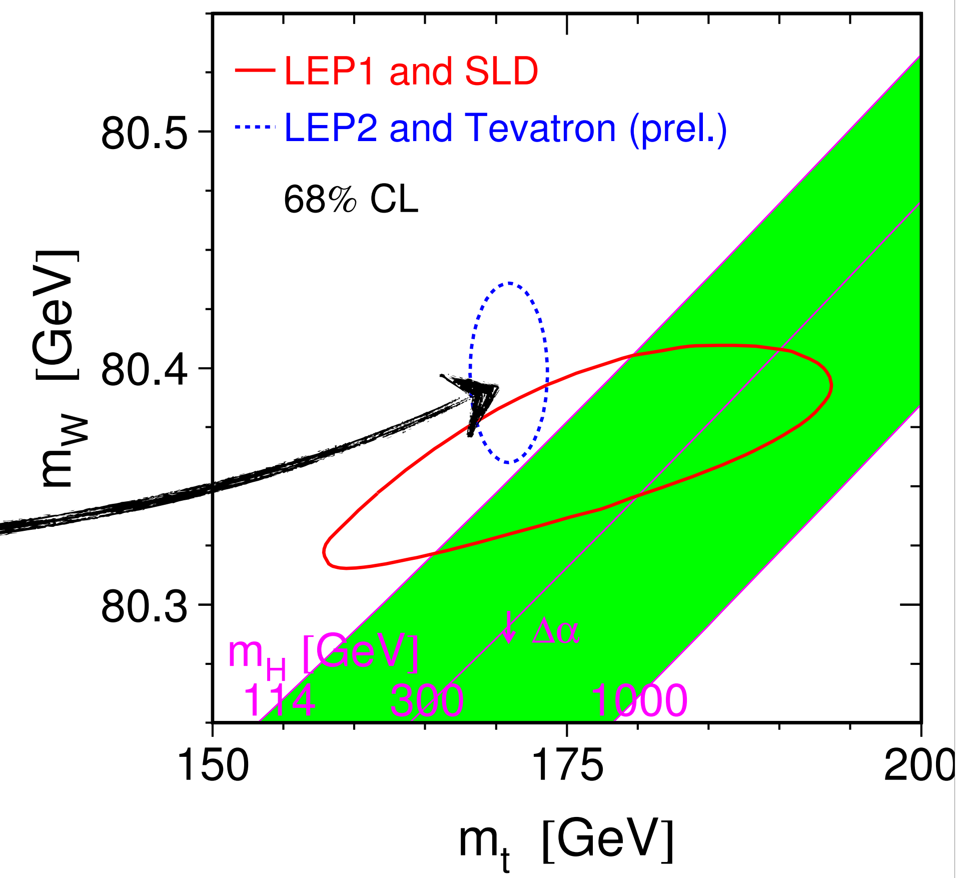
- you see them all the time:

Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval

- but that's P(theory|data)!

Correct frequentist statement is that the interval **covers** the true value 68% of the time

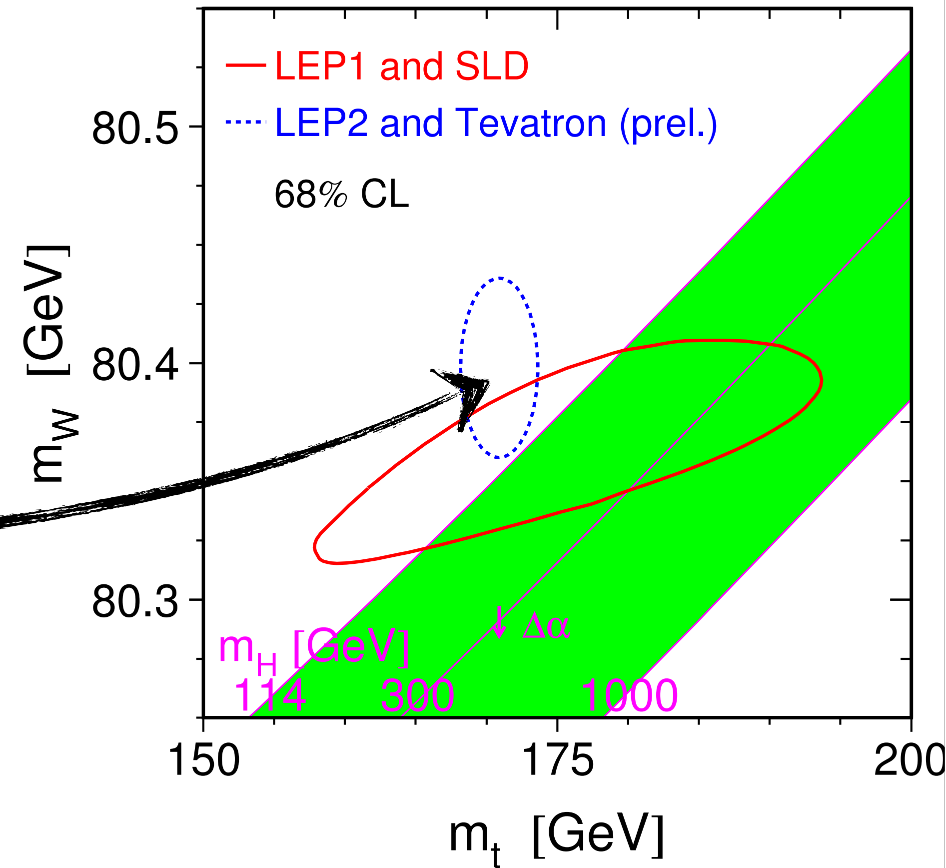
- remember, the contour is a function of the data, which is random. So it moves around from experiment to experiment



Confidence Interval

What is a "Confidence Interval?

- you see them all the time:



Want to say there is a 68% chance that the true value of (m_W, m_t) is in this interval

- but that's P(theory|data)!

Correct frequentist statement is that the interval **covers** the true value 68% of the time

- remember, the contour is a function of the data, which is random. So it moves around from experiment to experiment

- Bayesian "credible interval" does mean probability parameter is in interval. The procedure is very intuitive:

$$P(\theta \in V) = \int_V \pi(\theta|x) = \int_V d\theta \frac{f(x|\theta)\pi(\theta)}{\int d\theta f(x|\theta)\pi(\theta)}$$

Joke

"Bayesians address the question everyone is interested in, by using assumptions no-one believes"

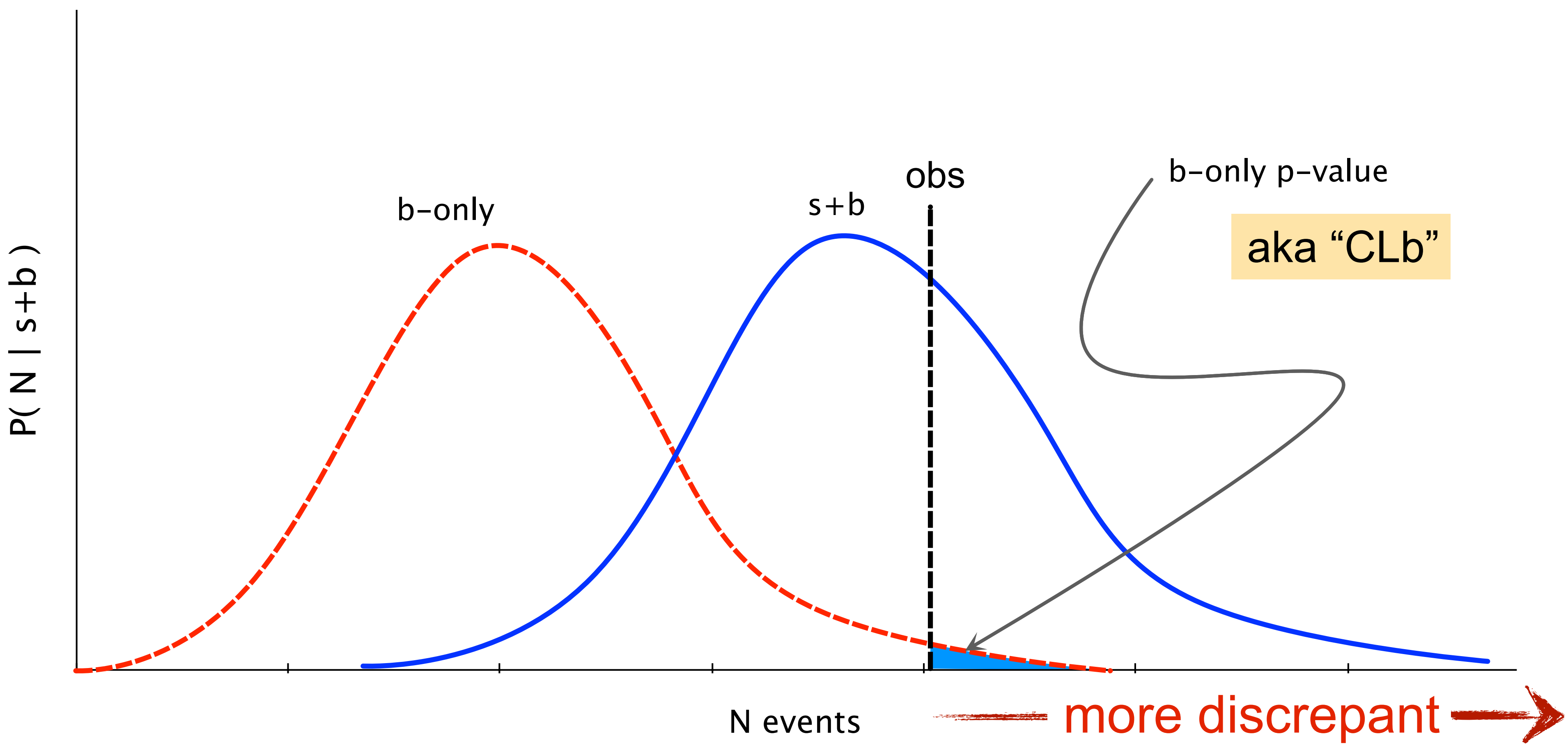
"Frequentists use impeccable logic to deal with an issue of no interest to anyone"

-L. Lyons

Discovery in pictures

Discovery: test b-only (null: s=0 vs. alt: s>0)

- note, **one-sided** alternative. larger N is "more discrepant"

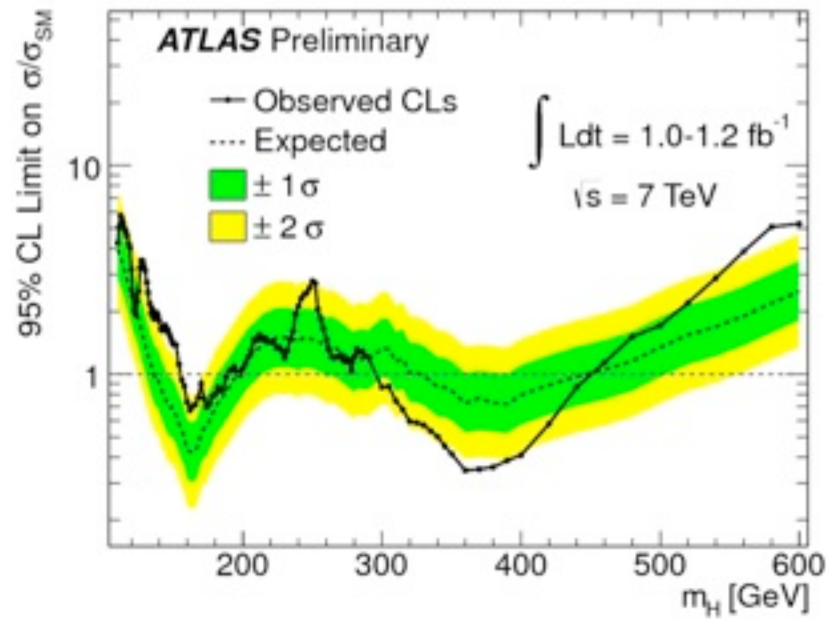


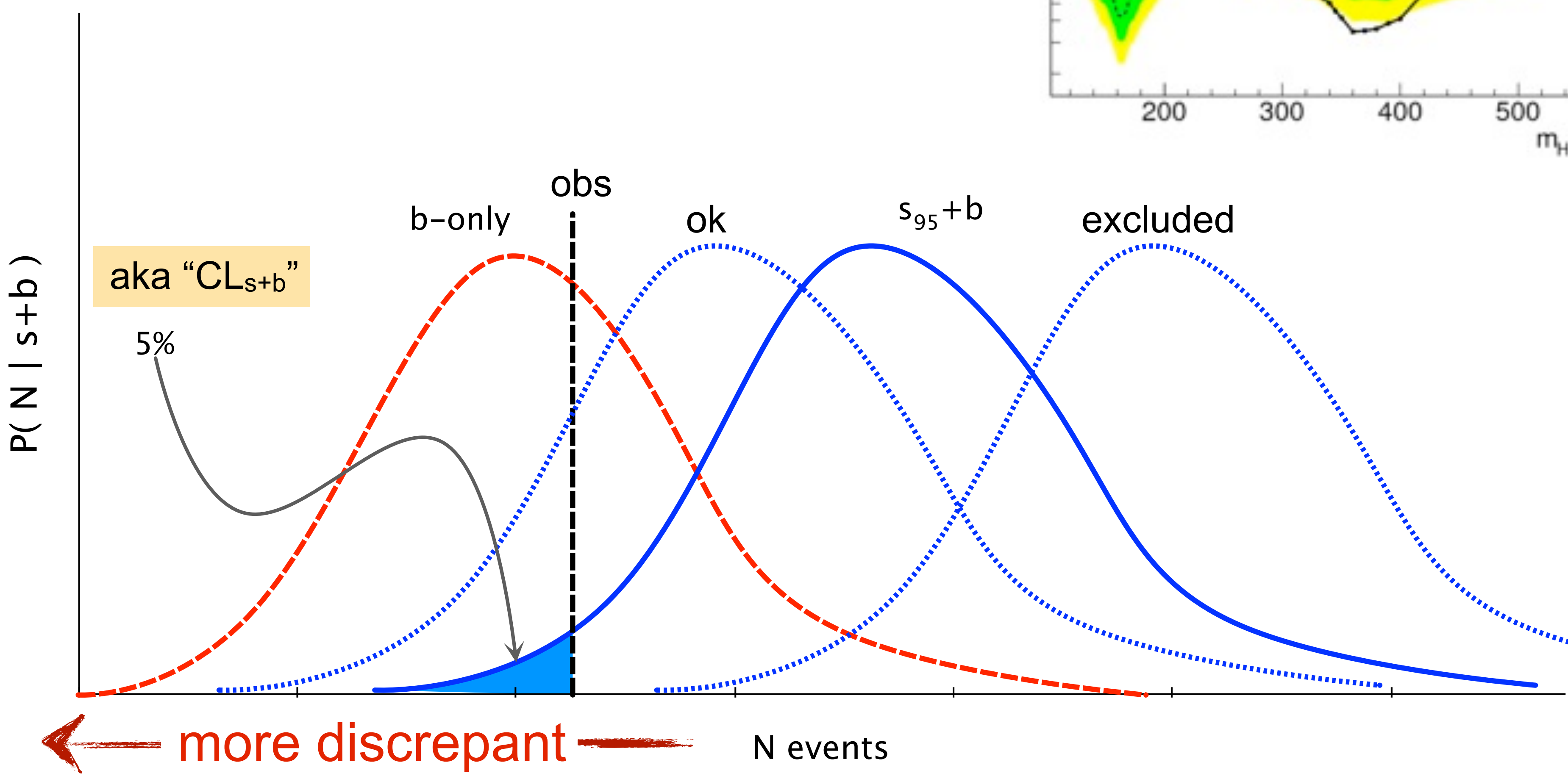
Upper limits in pictures

What is meant by “95% upper limit” ?

See the picture below?

- ie. increase s, until the probability to have data “more discrepant” is < 5%

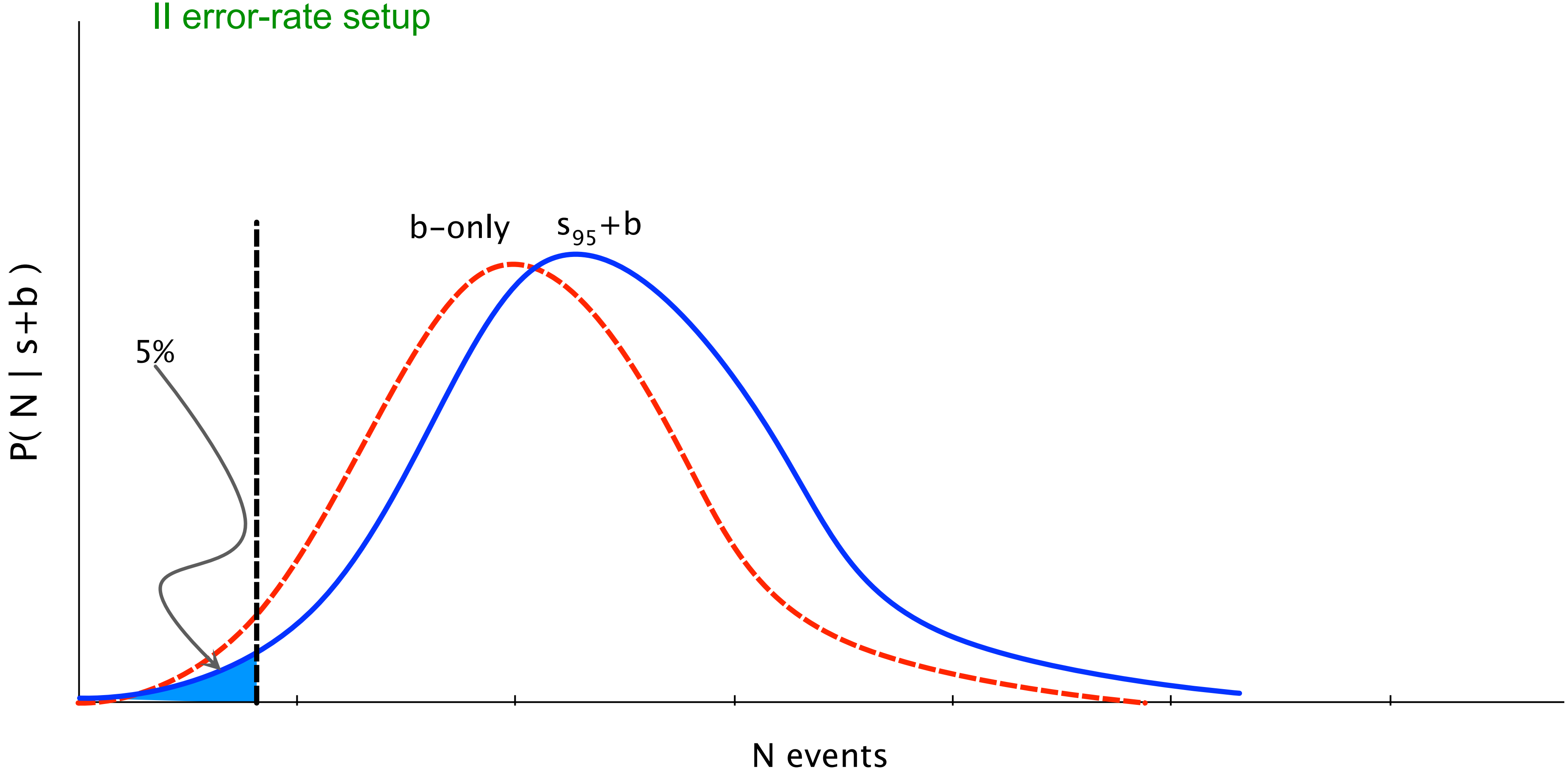




The sensitivity problem

The physicist's worry about limits in general is that if there is a strong downward fluctuation, one might exclude arbitrarily small values of *s*

- with a procedure that produces proper frequentist 95% confidence intervals, one should expect to exclude the true value of *s* 5% of the time, no matter how small *s* is!
  - This is not a problem with the procedure, but an undesirable consequence of the Type I / Type II error-rate setup



CL_s

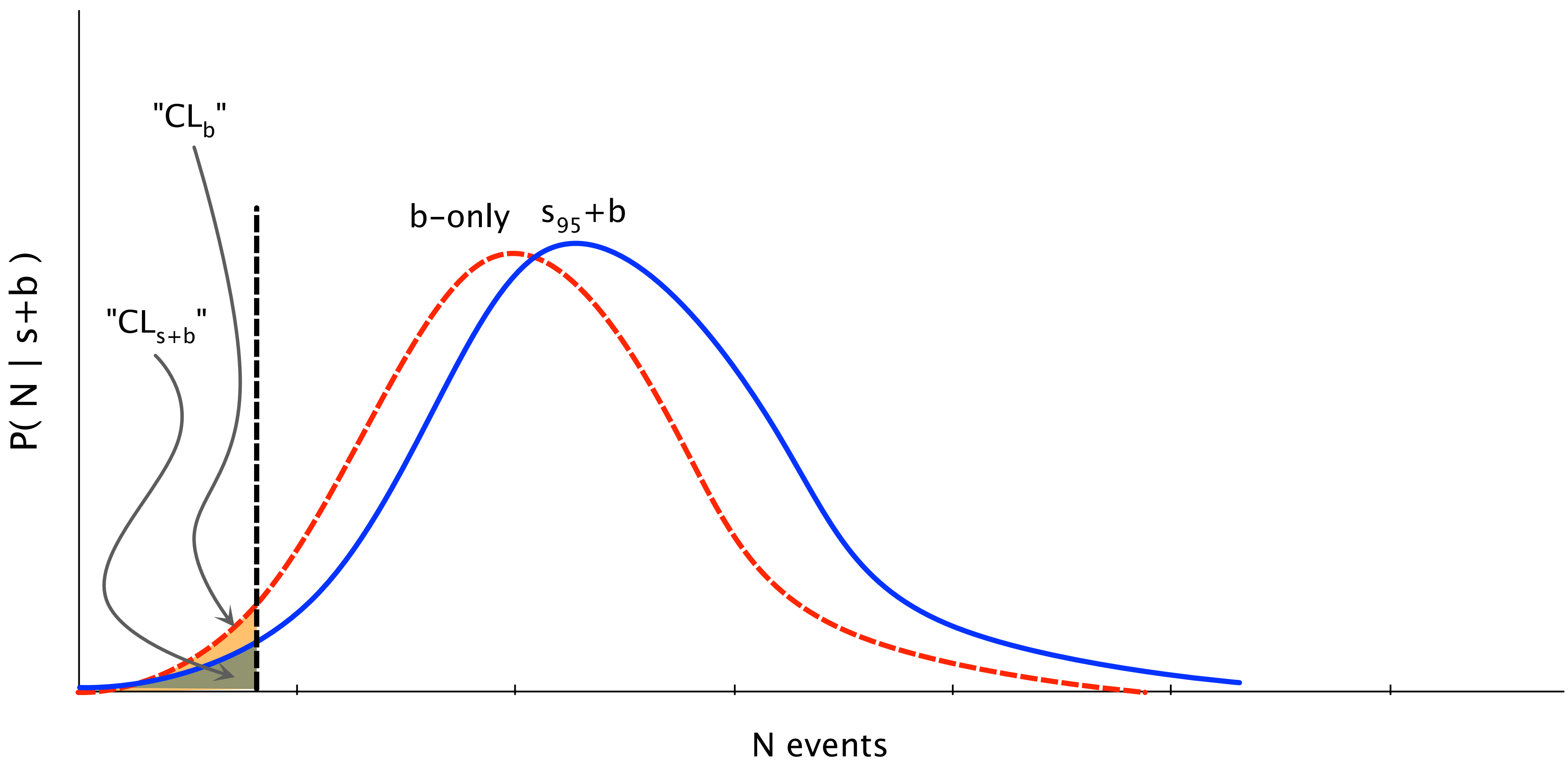
http://inspirehep.net/record/599622

To address the sensitivity problem, CLs was introduced

- common (misused) nomenclature: $CL_s = CL_{s+b}/CL_b$
- idea: only exclude if CL_s<5% (if CL_b is small, CL_s gets bigger)

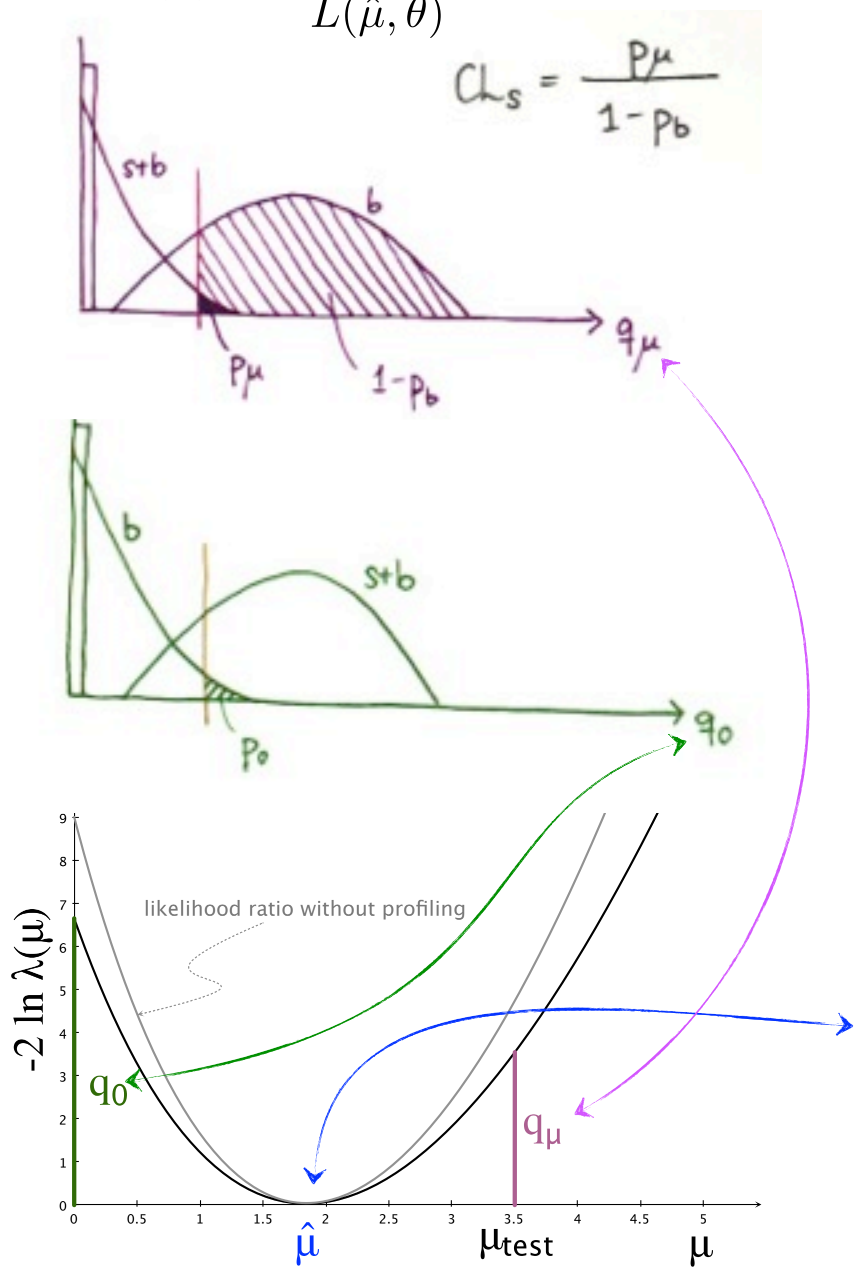
CL_s is known to be "conservative" (over-cover): expected limit covers with 97.5%

- Note: CL_s is NOT a probability



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

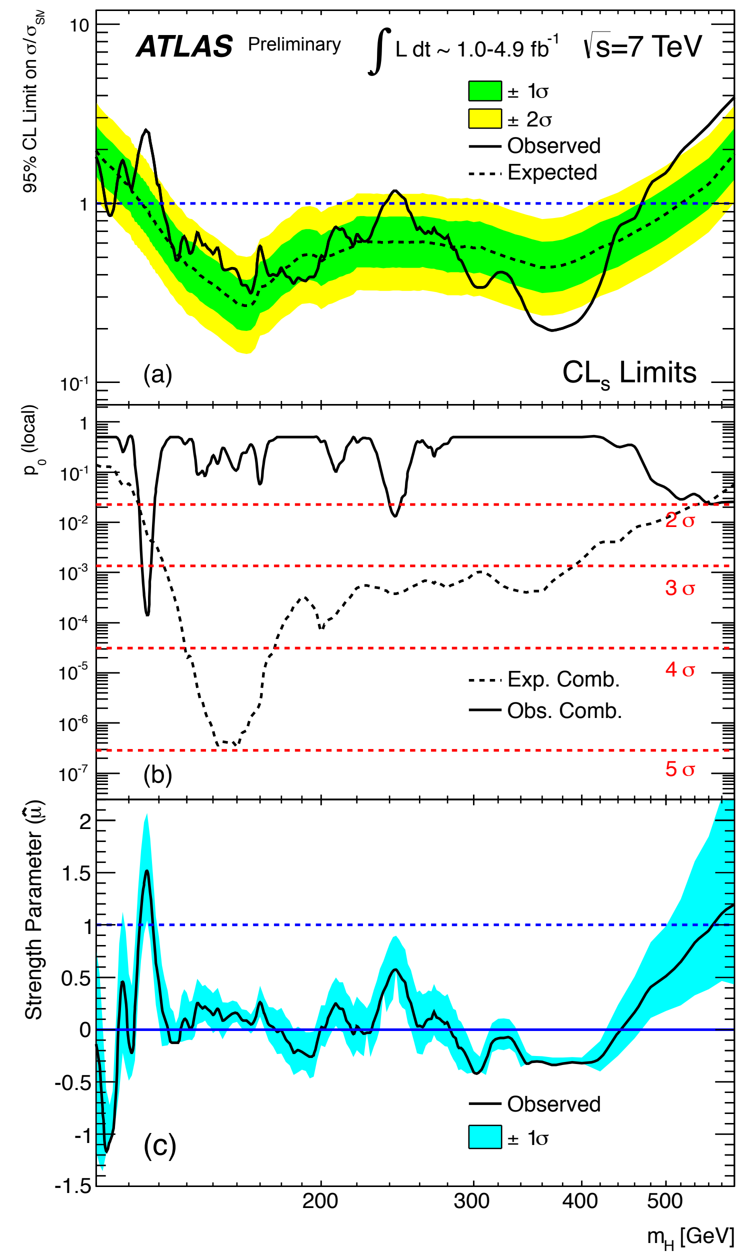
Follow LHC-HCG Combination Procedures



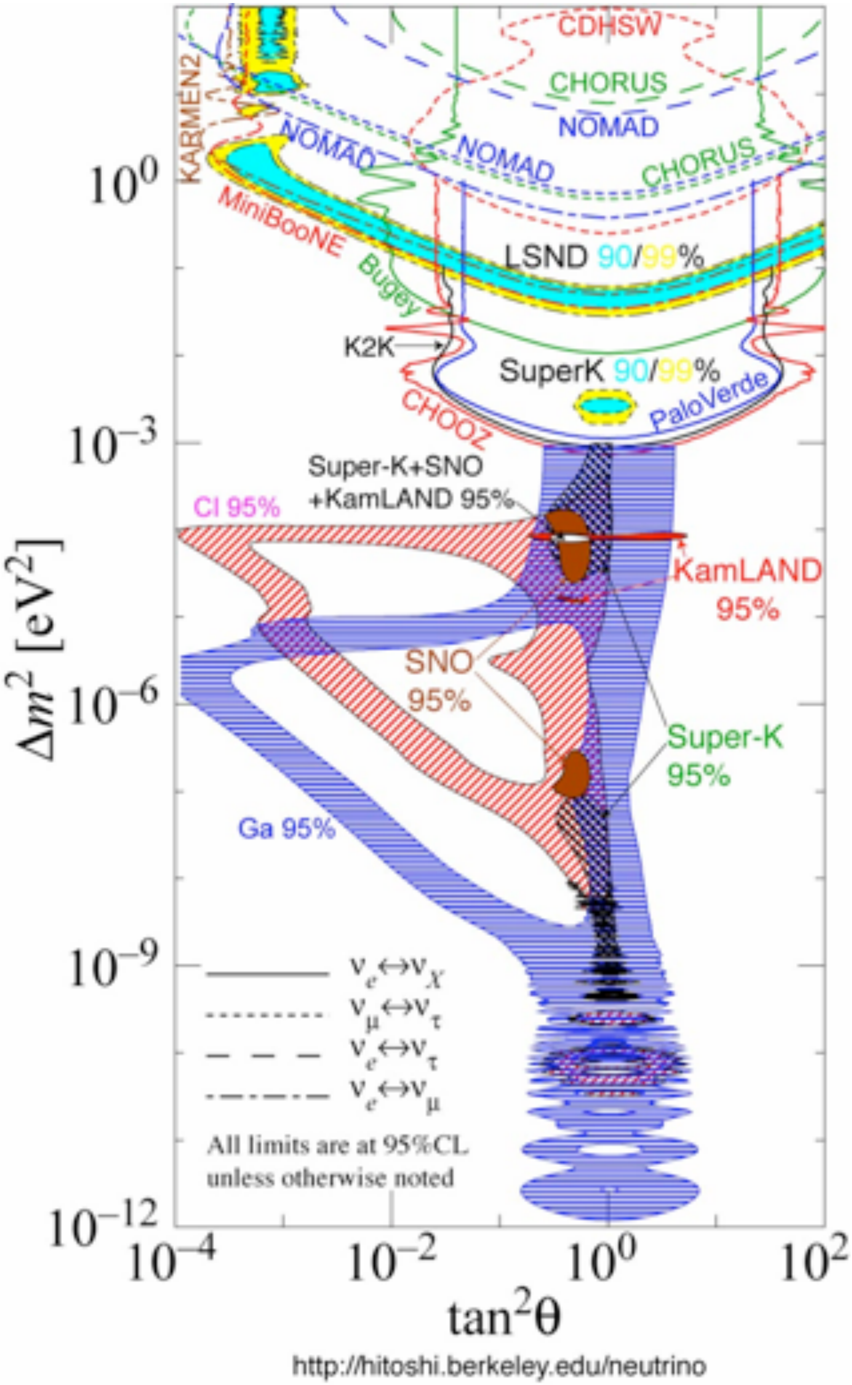
CL_s to test signal hypothesis

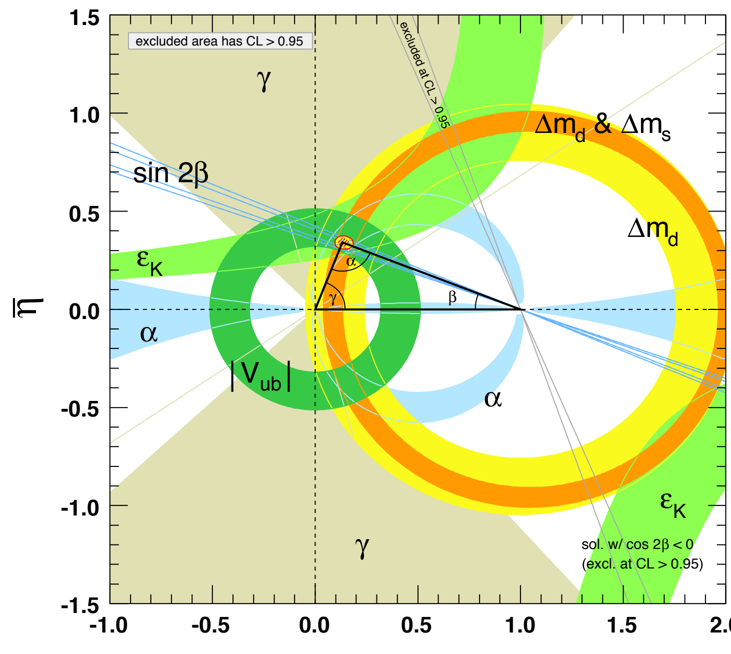
p_0 to test background hypothesis

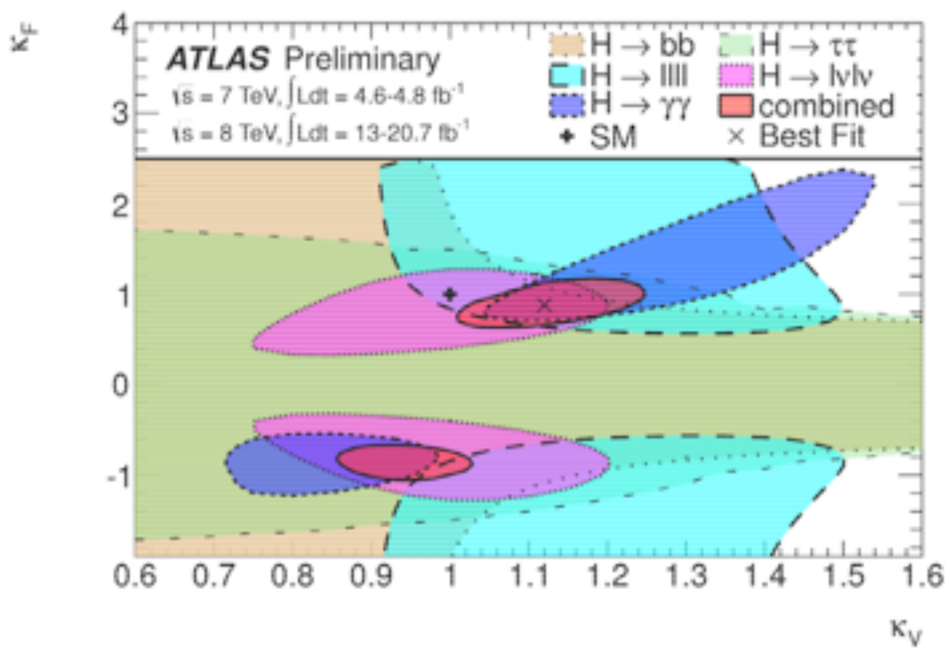
$\hat{\mu}$ to estimate signal strength



How do we generalize?

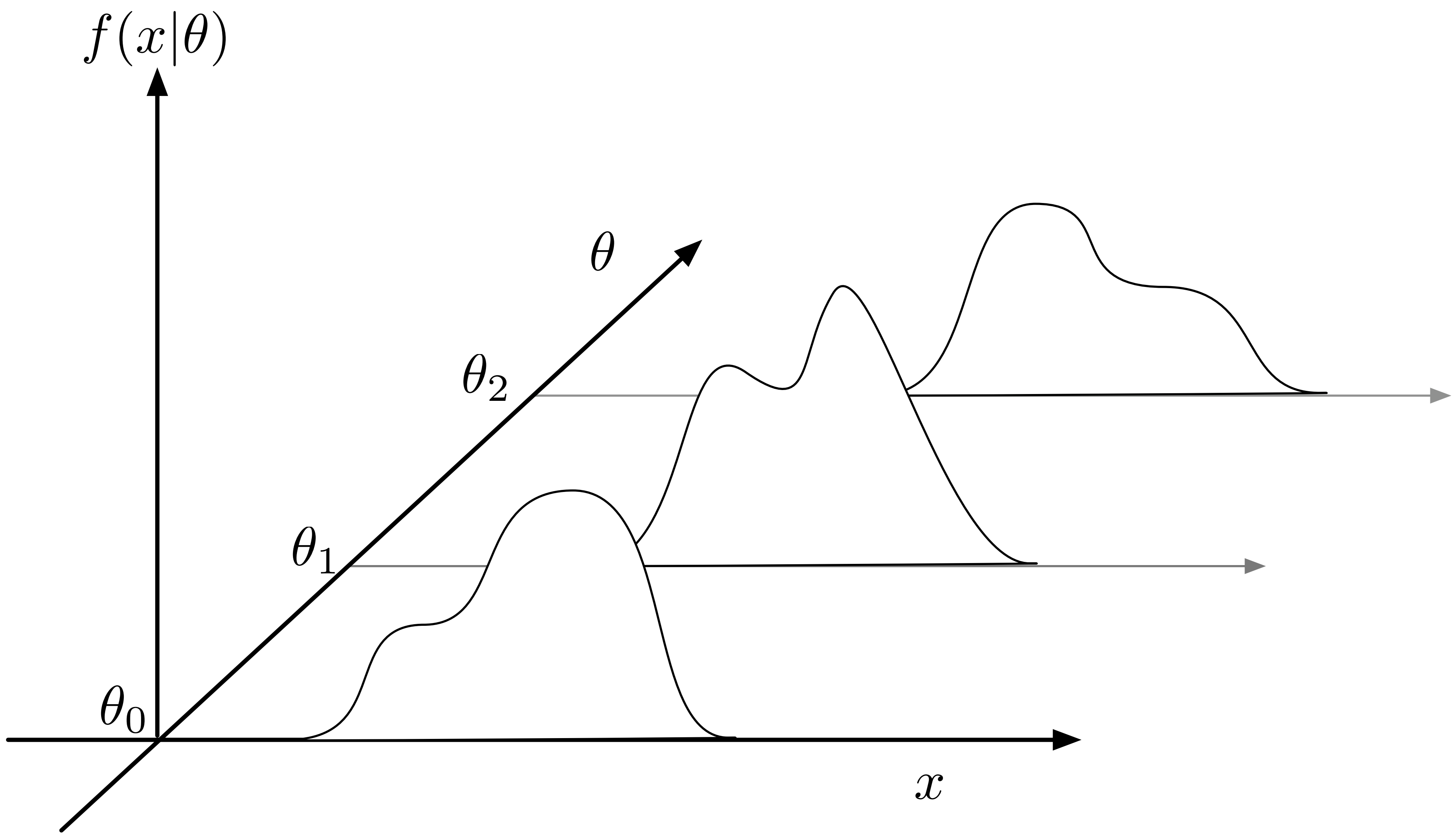






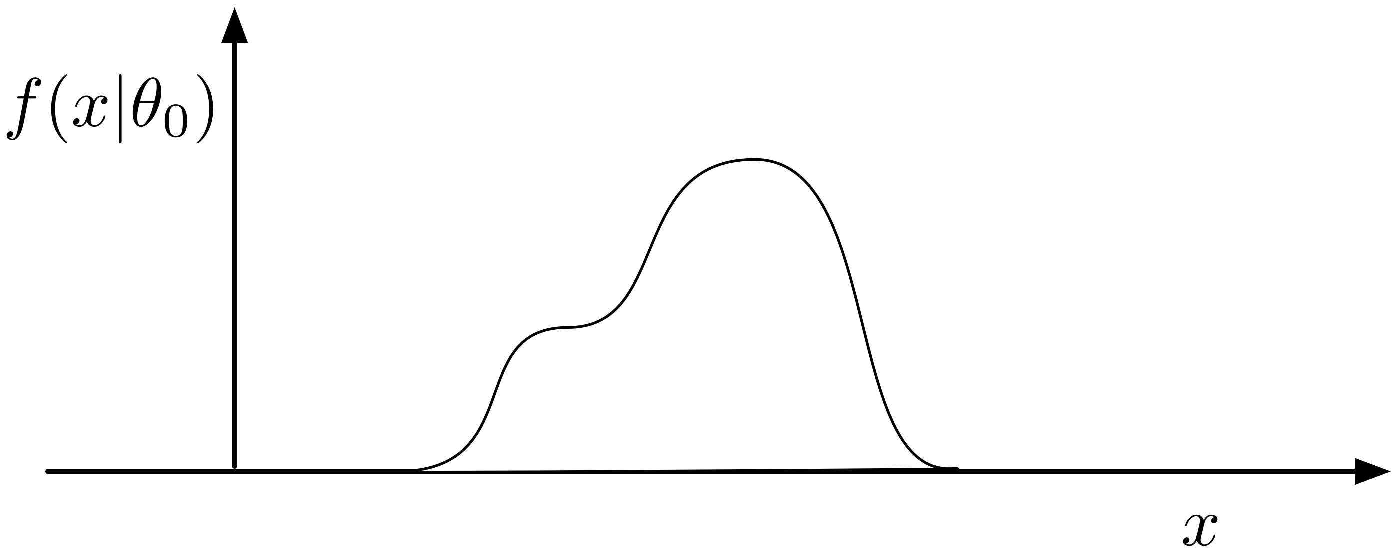
Neyman Construction example

For each value of θ consider $f(x|\theta)$



Neyman Construction example

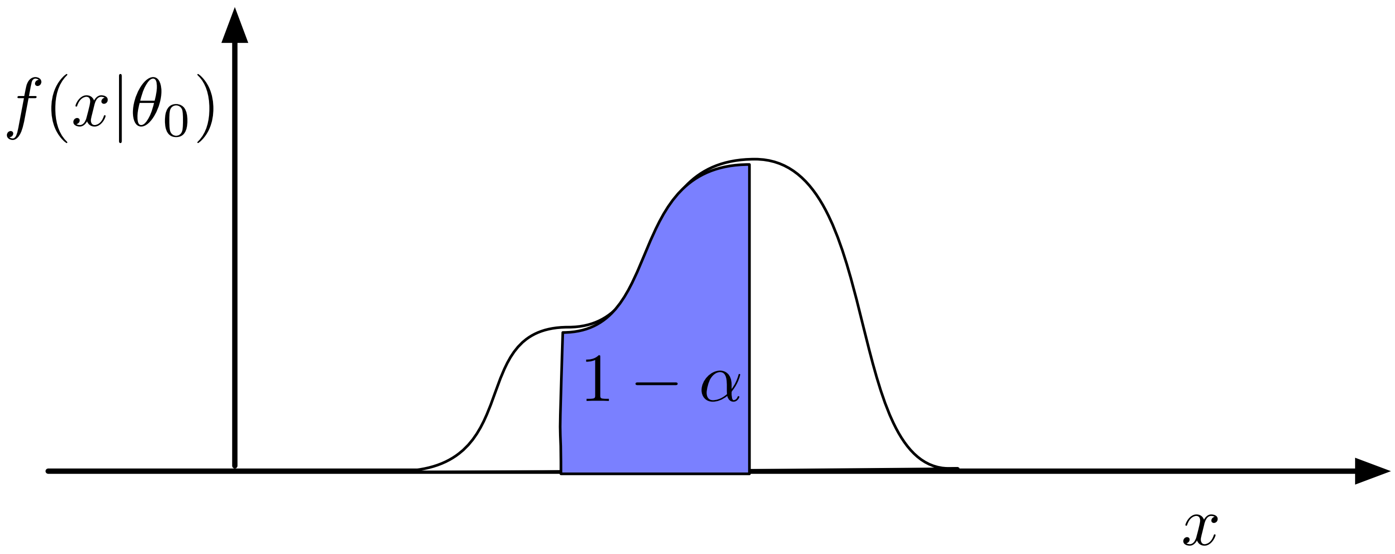
Let's focus on a particular point $f(x|\theta_o)$



Neyman Construction example

Let's focus on a particular point $f(x|\theta_o)$

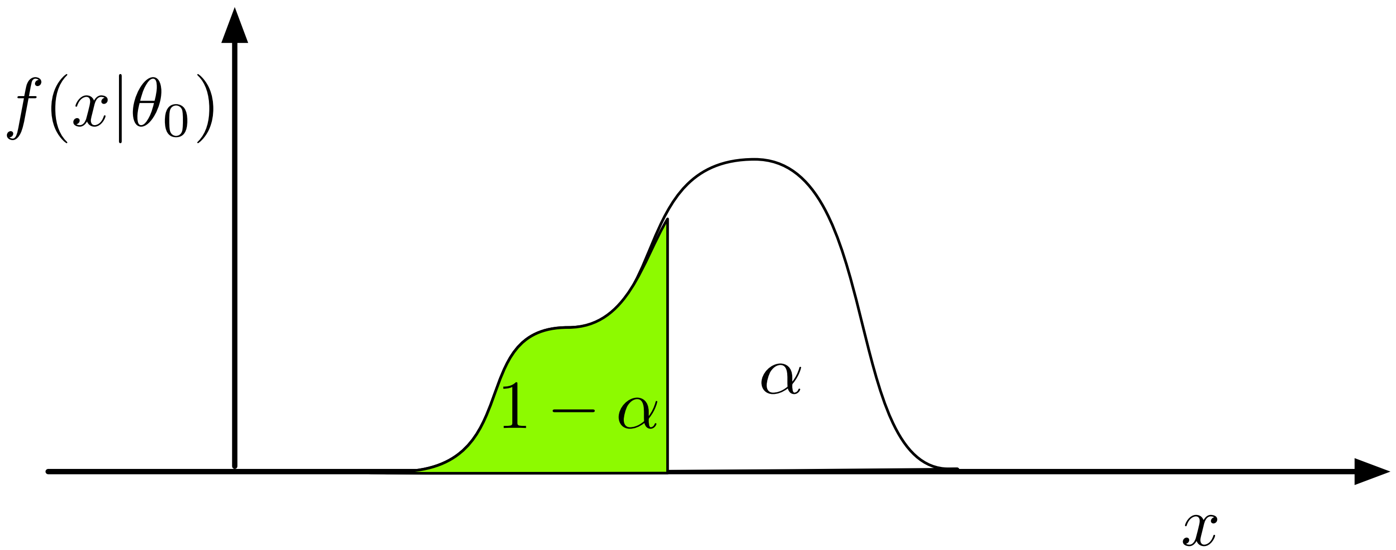
- we want a test of size α
- equivalent to a $100(1-\alpha)\%$ confidence interval on θ
- so we find an **acceptance region** with $1-\alpha$ probability



Neyman Construction example

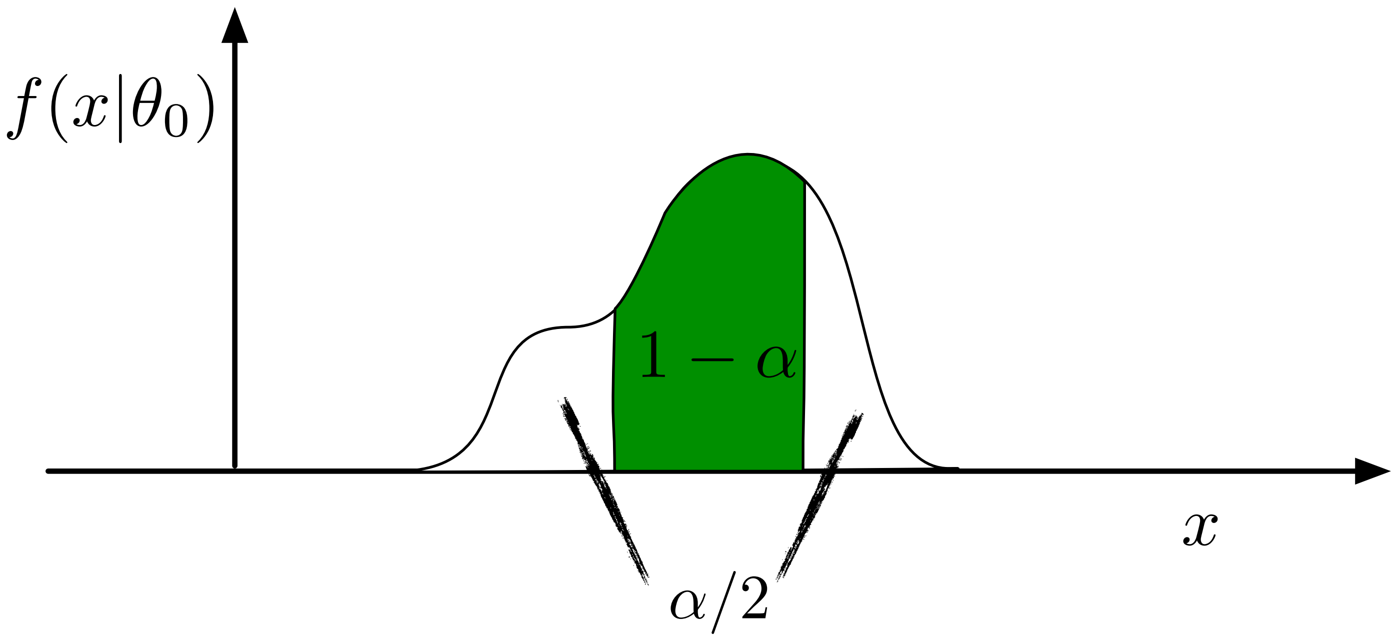
Let's focus on a particular point $f(x|\theta_o)$

- No unique choice of an acceptance region
- here's an example of a lower limit



Let's focus on a particular point $f(x|\theta_o)$

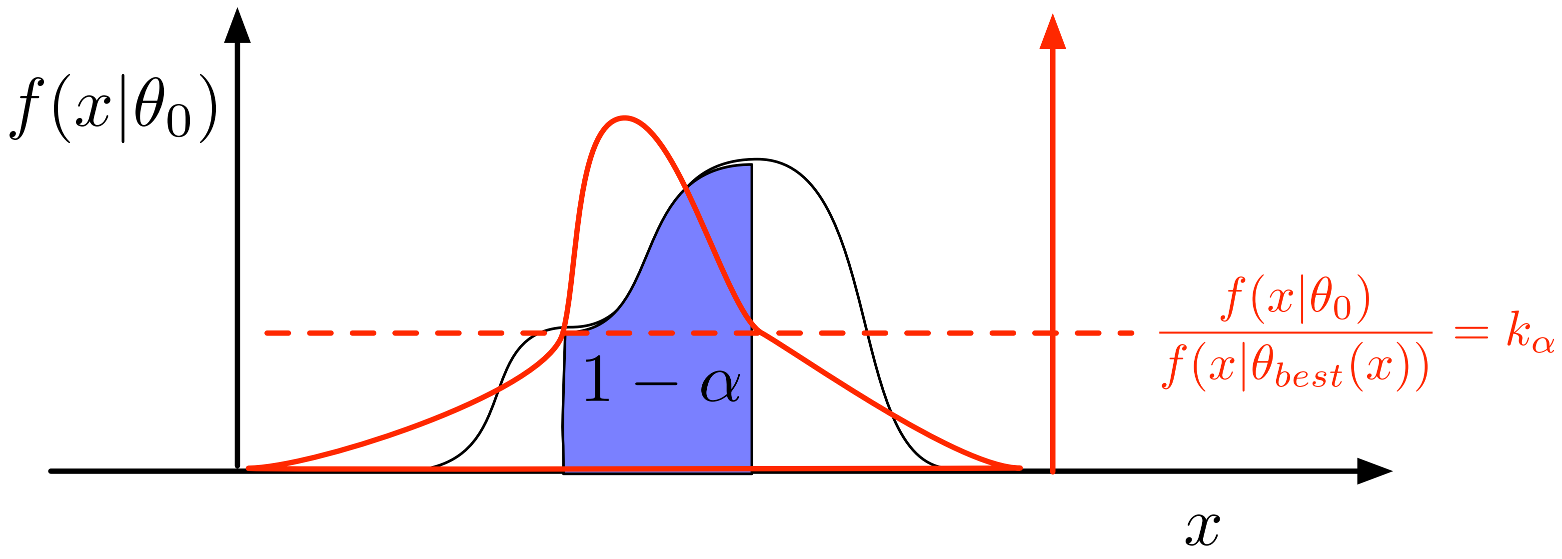
- No unique choice of an acceptance region
- and an example of a central limit



Neyman Construction example

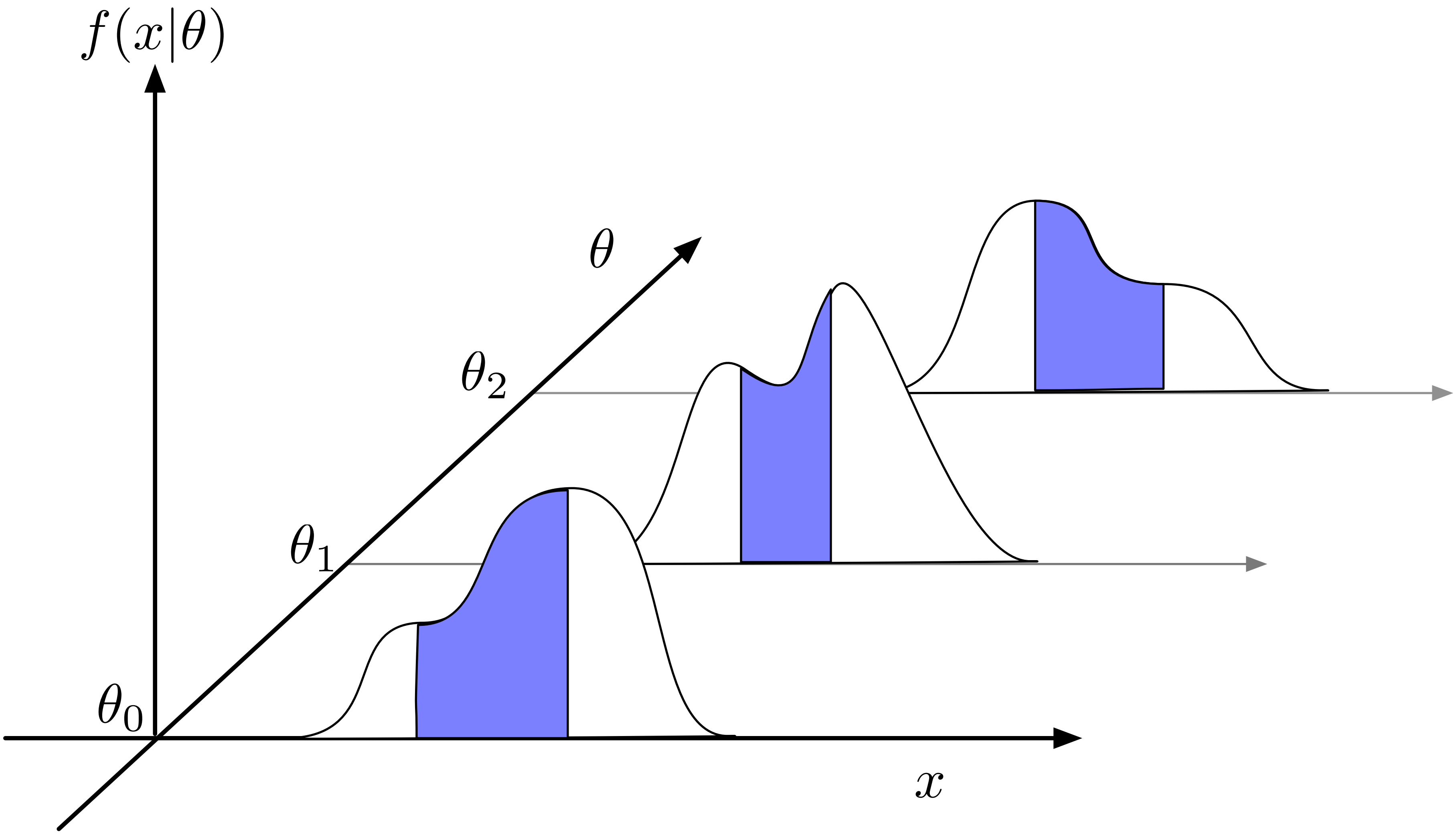
Let's focus on a particular point $f(x|\theta_o)$

- choice of this region is called an **ordering rule**
- In Feldman-Cousins approach, ordering rule is the likelihood ratio. Find contour of L.R. that gives size α



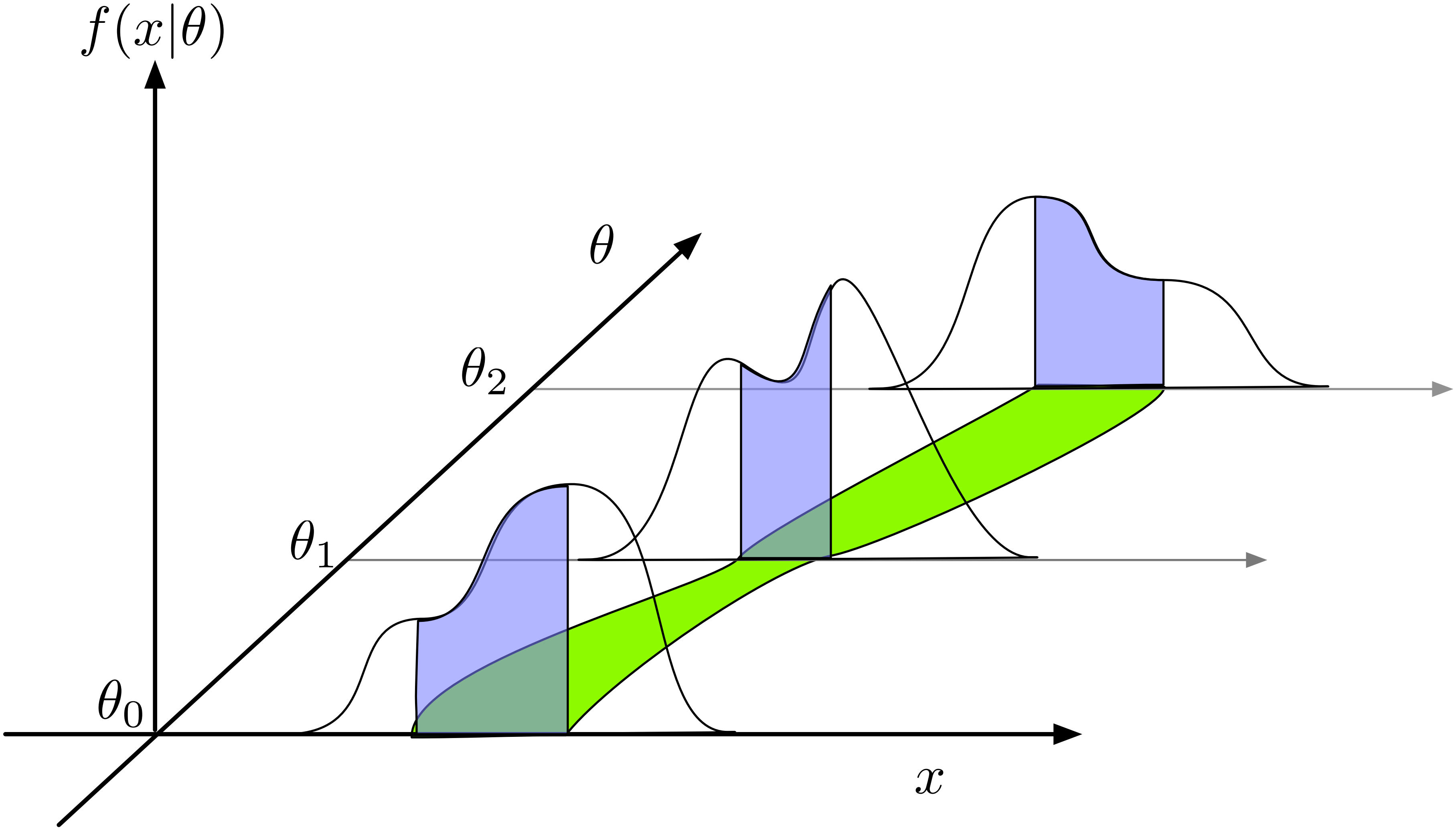
Neyman Construction example

Now make acceptance region for every value of θ



Neyman Construction example

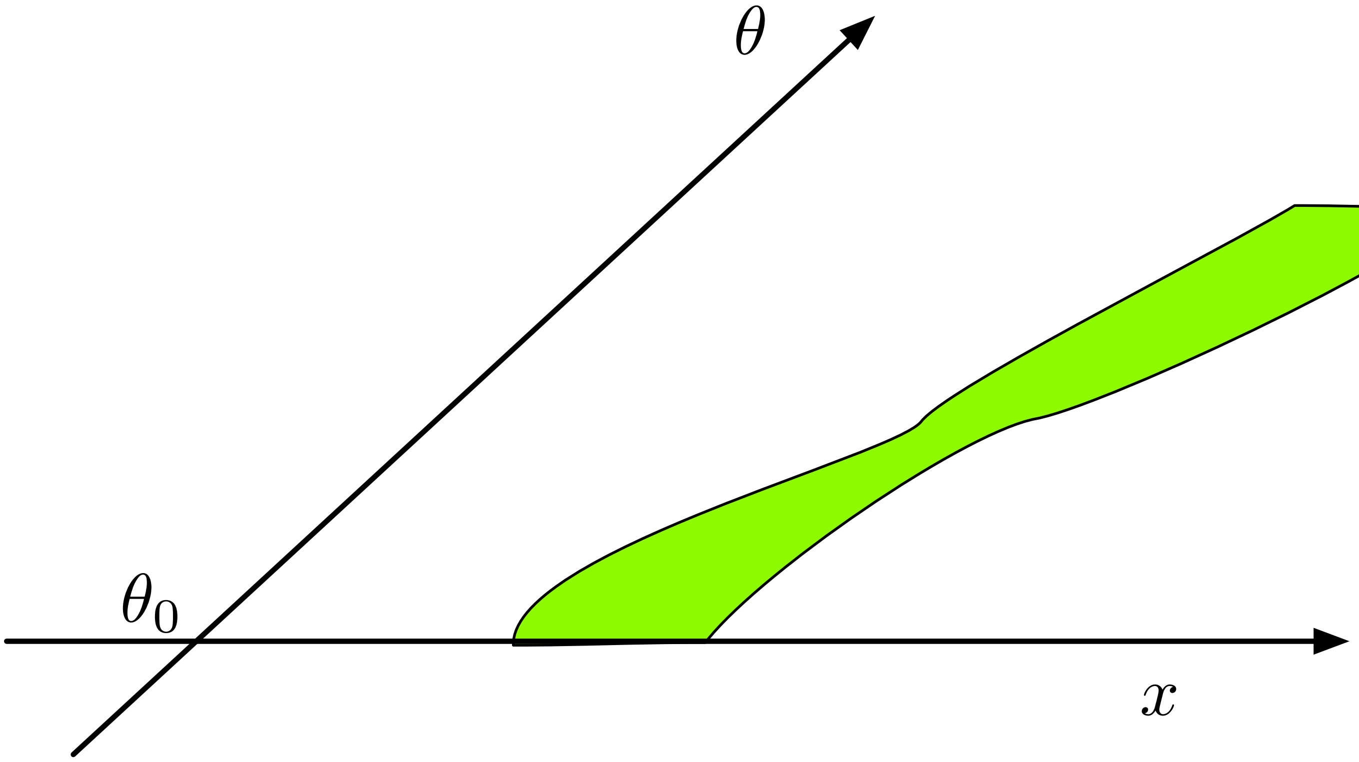
This makes a **confidence belt** for θ



Neyman Construction example

This makes a **confidence belt** for θ

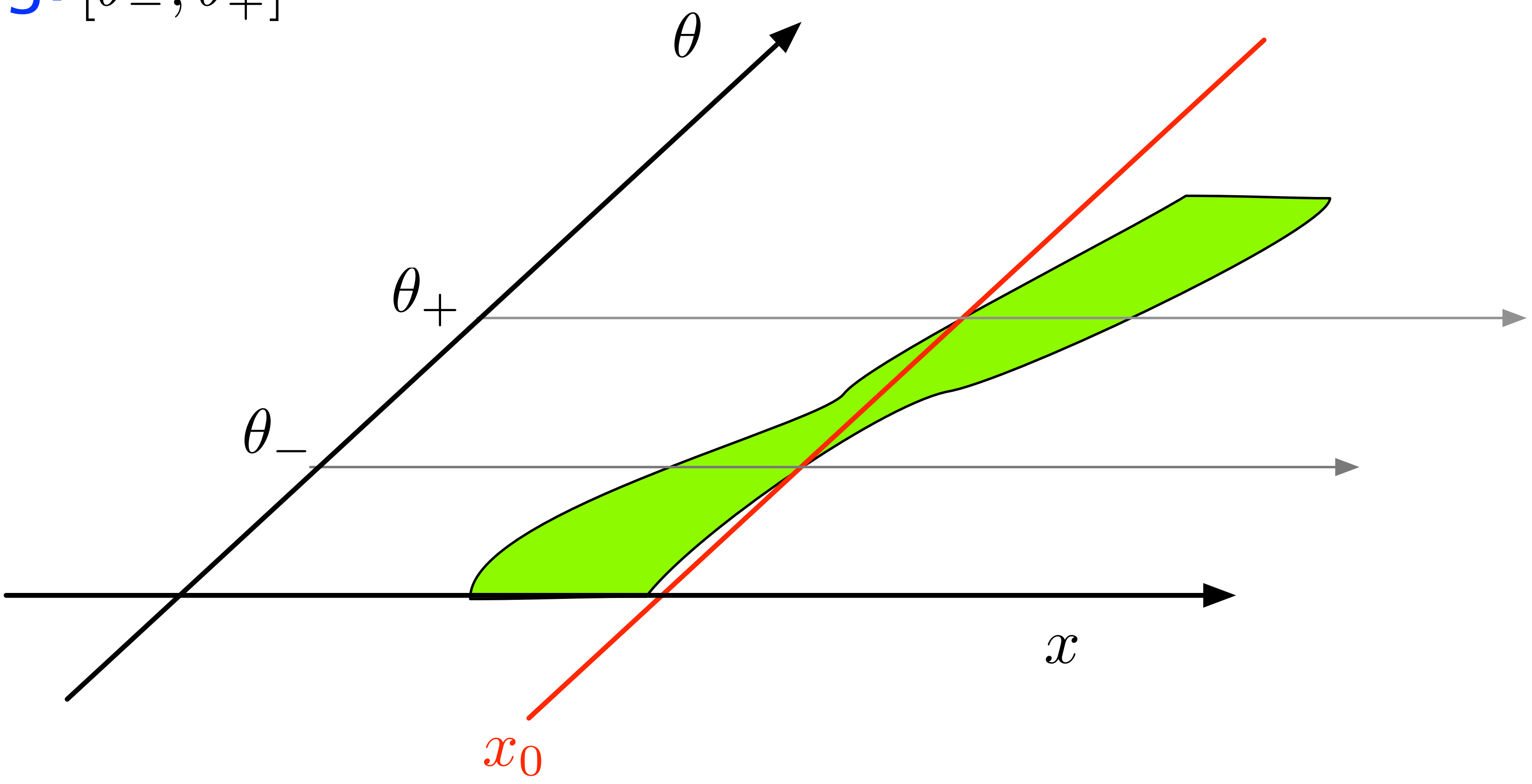
- the regions of **data** in the confidence belt can be considered as **consistent** with that value of θ



Now we make a measurement x_0

the points θ where the belt intersects x_0 a part of the **confidence interval** in θ for this measurement

eg. $[\theta_-, \theta_+]$

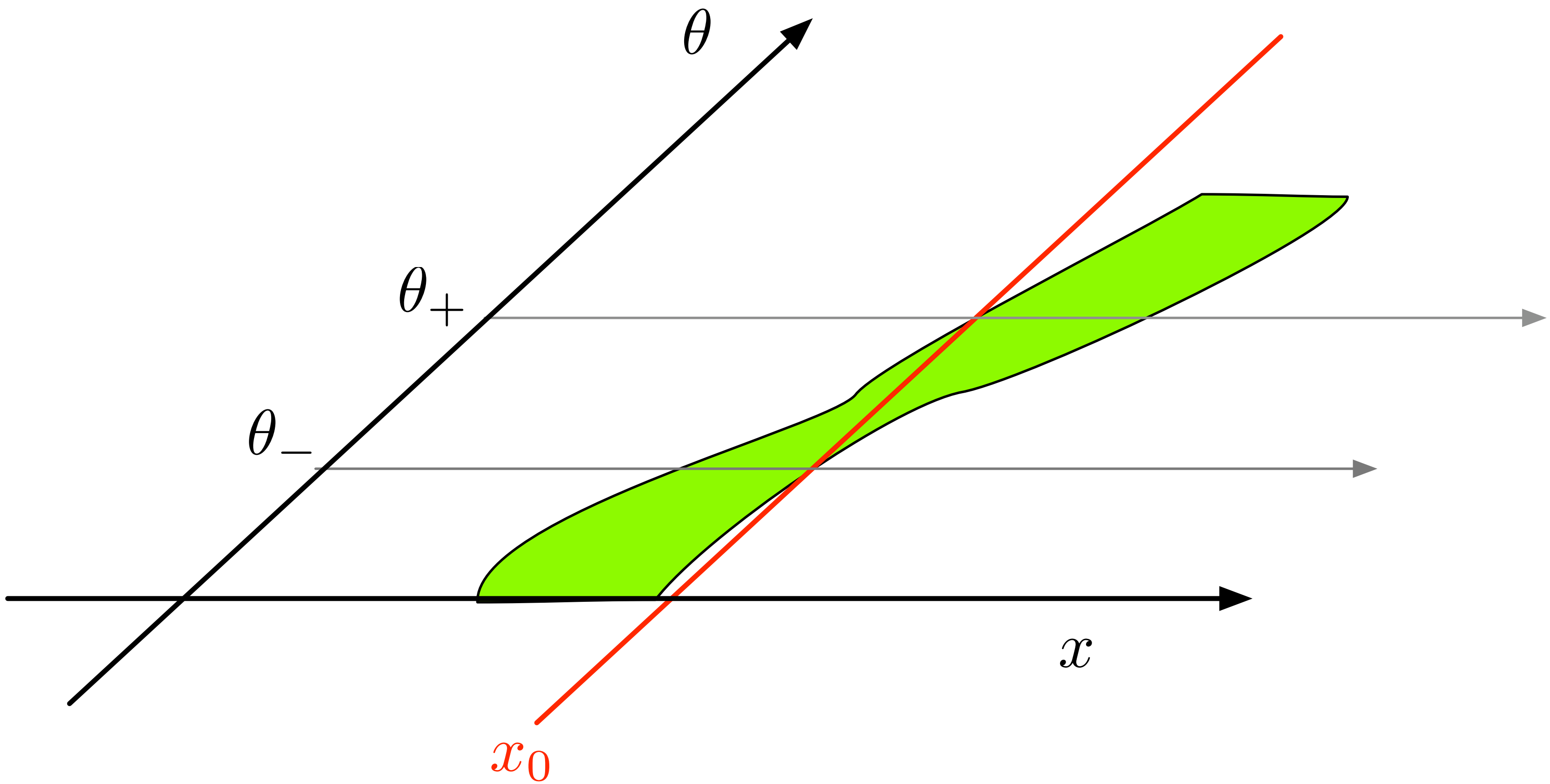


Neyman Construction example

For every point θ, if it were true, the data would fall in its acceptance region with probability $1-\alpha$

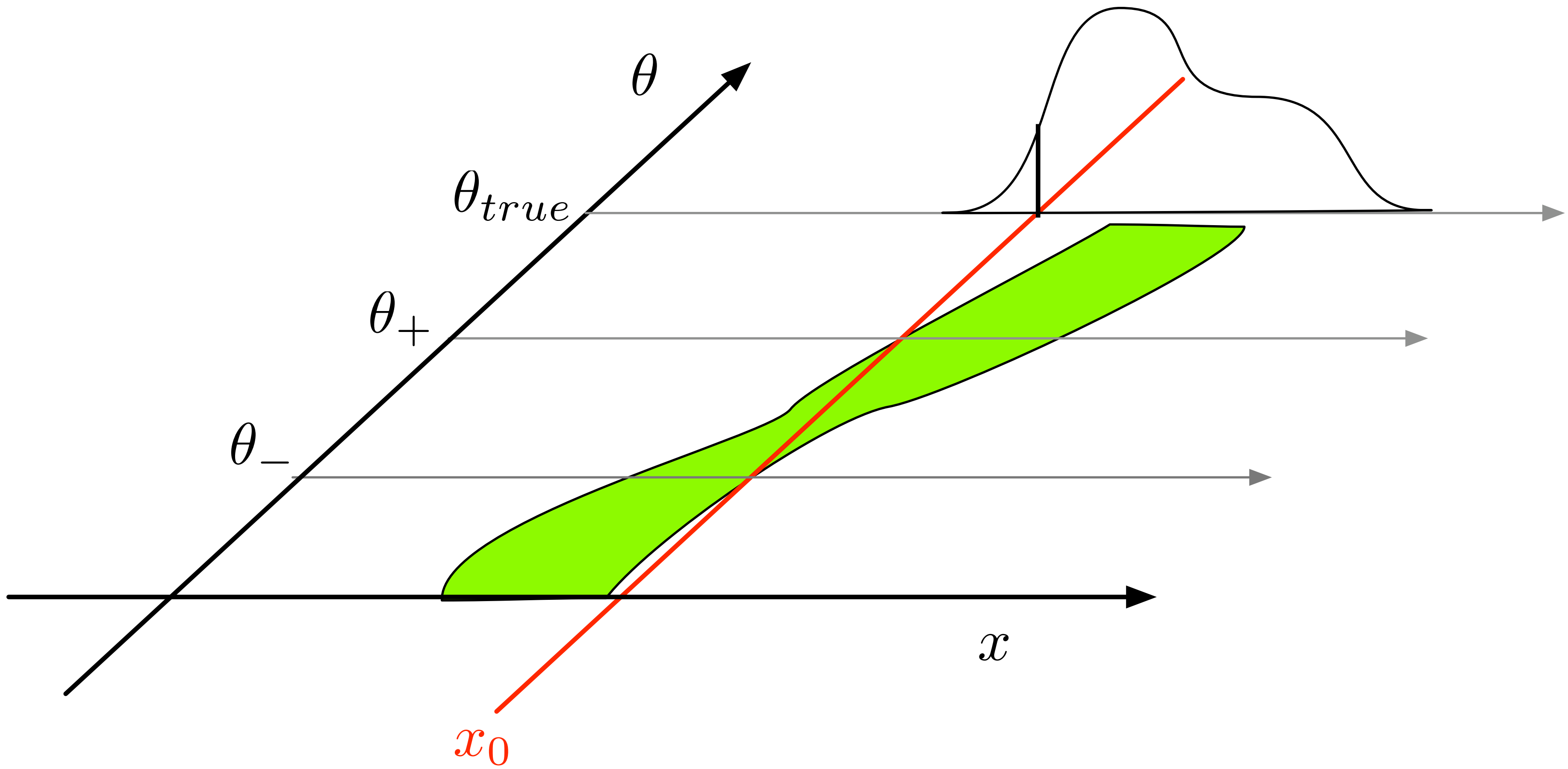
If the data fell in that region, the point θ would be in the interval $[\theta_-, \theta_+]$

So the interval $[\theta_-, \theta_+]$ covers the true value with probability $1-\alpha$



A Point about the Neyman Construction

This is not Bayesian... it doesn't mean the probability that the true value of θ is in the interval is $1-\alpha$!



Inverting Hypothesis Tests

There is a precise dictionary that explains how to move from from hypothesis testing to confidence intervals

- **Type I error:** probability interval does not cover true value of the parameters (eg. it is now a function of the parameters)
- **Power** is probability interval does not cover a false value of the parameters (eg. it is now a function of the parameters)
  - We don't know the true value, consider each point θ_0 as if it were true

What about null and alternate hypotheses?

- when testing a point θ_0 it is considered the null
- all other points considered "alternate"

So what about the Neyman-Pearson lemma & Likelihood ratio?

- as mentioned earlier, there are no guarantees like before
- a common generalization that has good power is:

$$\frac{f(x|H_0)}{f(x|H_1)} \Rightarrow \frac{f(x|\theta_0)}{f(x|\theta_{best}(x))}$$

Coverage

Coverage is the probability that the interval covers the true value.

Methods based on the Neyman-Construction always cover.... by construction.

- sometimes they over-cover (eg. “conservative”)

Bayesian methods, do not necessarily cover

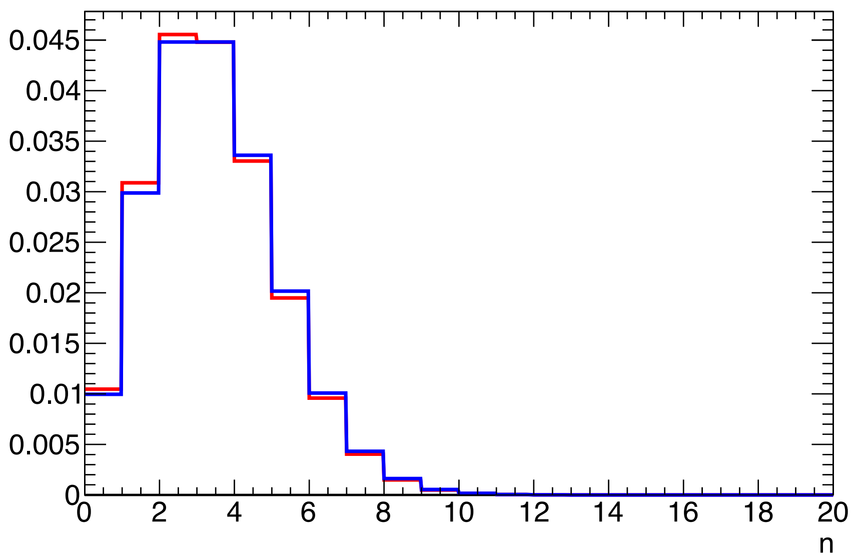
- but that’s not their goal.
- but that also means you shouldn’t interpret a 95% Bayesian “Credible Interval” in the same way

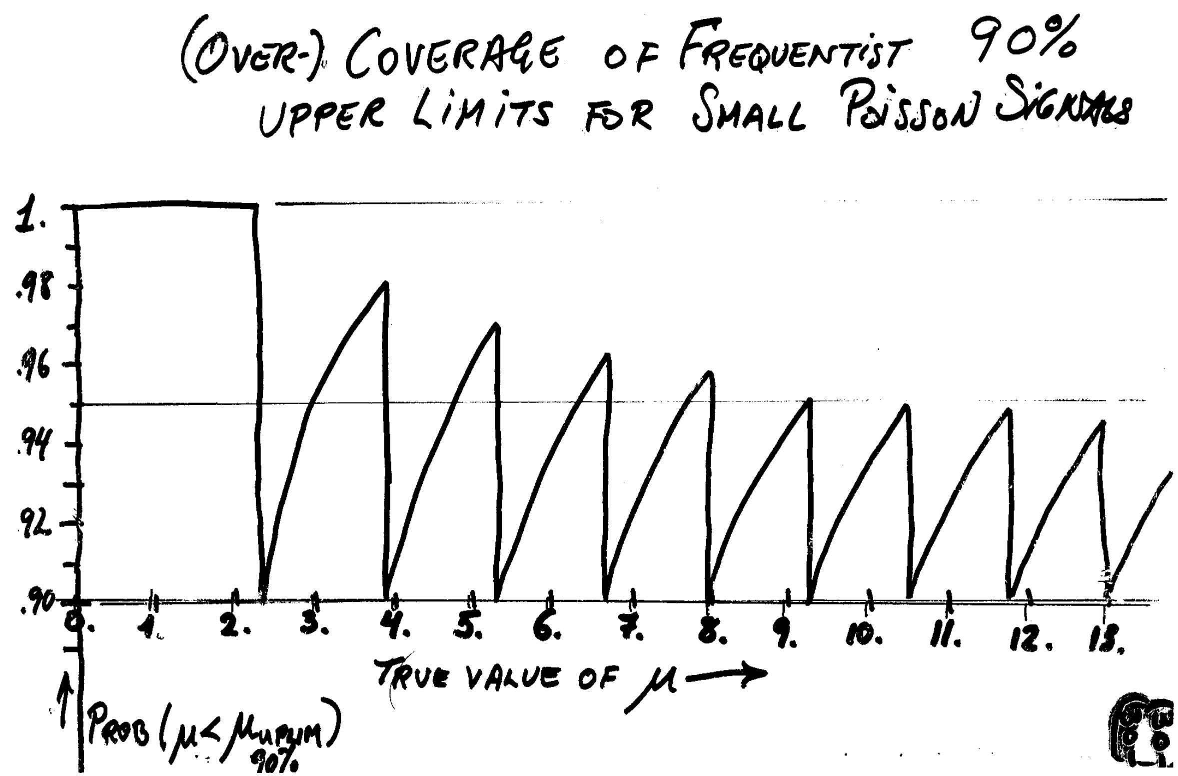
Coverage can be thought of as a **calibration of our statistical apparatus**. [explain under-/over-coverage]

Discrete Problems

In discrete problems (eg. number counting analysis with counts described by a Poisson) one sees:

- discontinuities in the coverage (as a function of parameter)
- over-coverage (in some regions)
- Important for experiments with few events. There is a lot of discussion about this, not focusing on it here

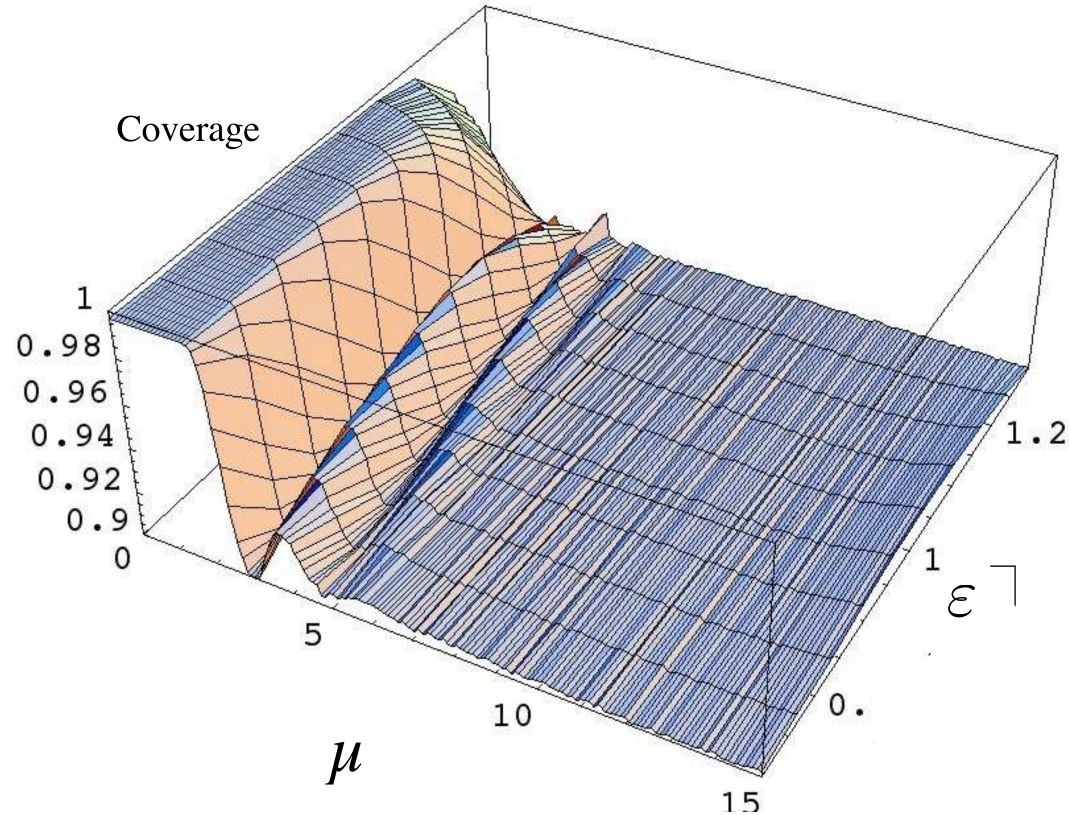




Coverage can be different at each point in the parameter space

Example:

G. Punzi - PHYSTAT 05 - Oxford, UK



Poisson(+background), with a systematic uncertainty on efficiency:

$$x \sim Pois(\varepsilon\mu + b) \quad e \sim G(\varepsilon, \sigma)$$

e is a measurement of the unknown efficiency ε, with resolution σ

ε is the efficiency (a “normalization factor”, can be larger than 1).

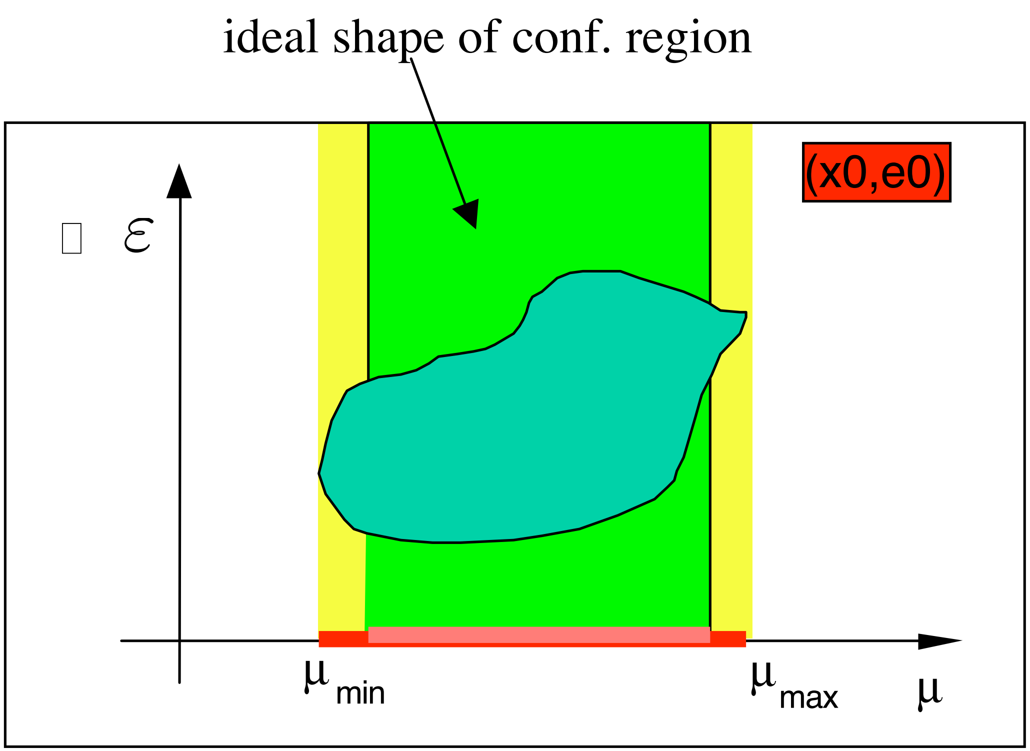
Neyman Construction with Nuisance parameters

In the strict sense, one wants coverage for μ **for all** values of the nuisance parameters (here ε)

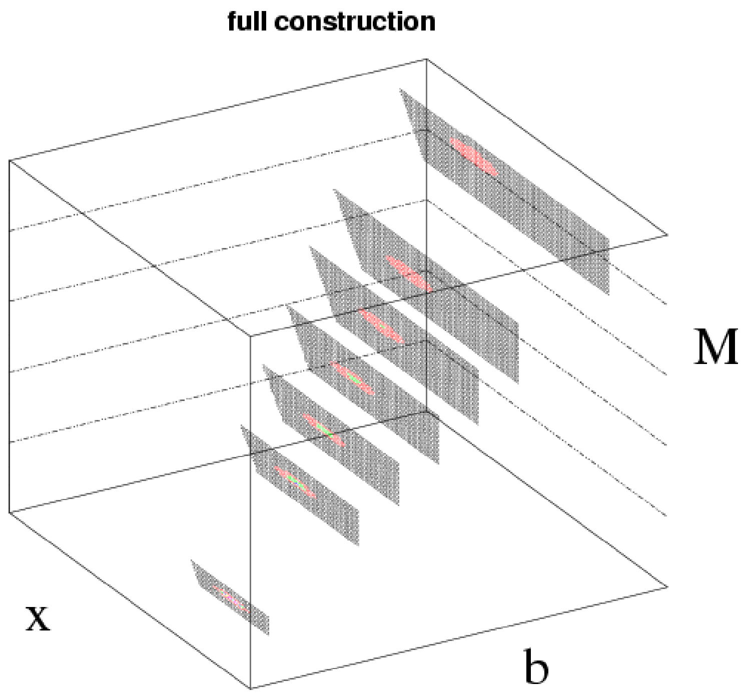
- The "full construction" one nuisance parameter

Challenge for full Neyman Construction is computational time (scan in 50-D isn't practical) and to avoid significant over-coverage

- note: projection of nuisance parameters is a union (eg. set theory) not an integration (Bayesian)



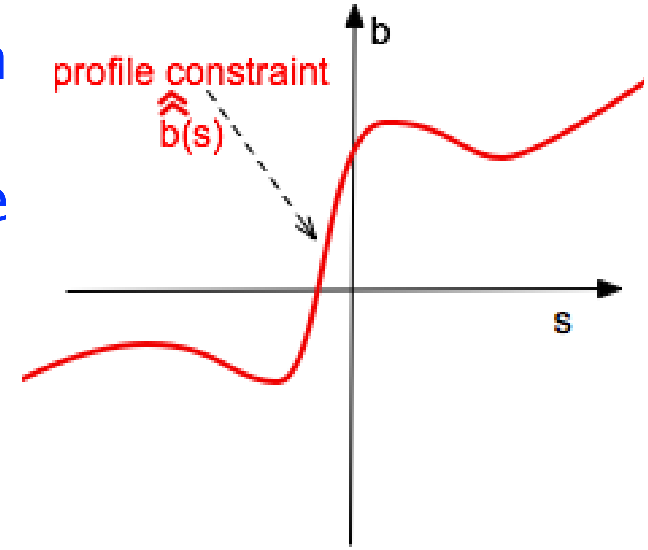
G. Punzi - PHYSTAT 05 - Oxford, UK

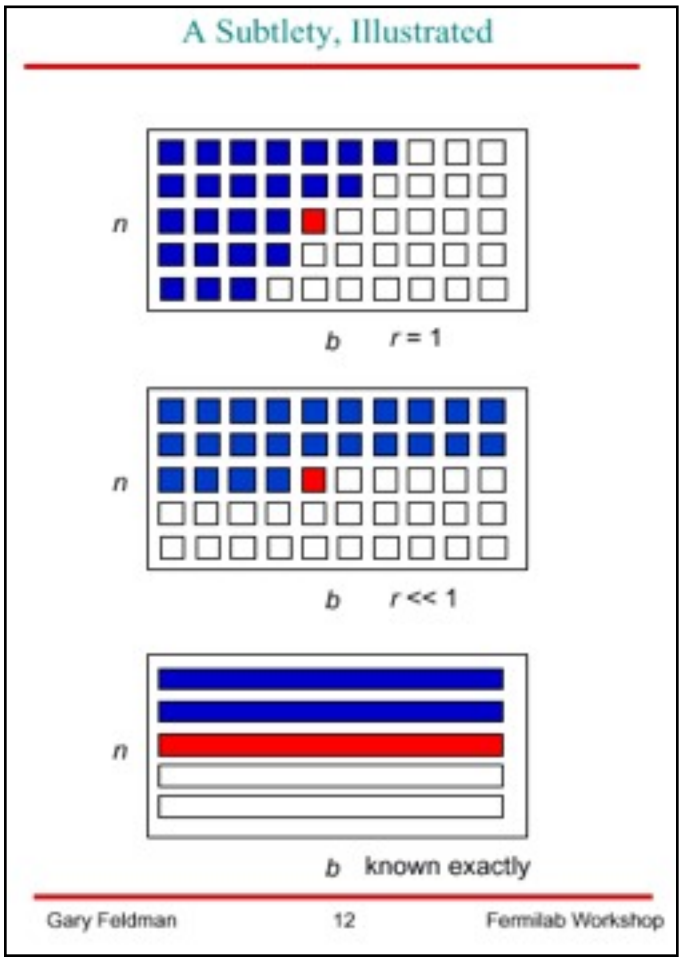


K. Cranmer - PHYSTAT 03 - SLAC

Profile Construction

Gary Feldman presented an approximate Neyman Construction, based on the profile likelihood ratio as an ordering rule, but only performing the construction on a subspace (eg. their conditional maximum likelihood estimate)





The **profile construction** means that one does not need to scan each nuisance parameter (keeps dimensionality constant)

- easier computationally (in RooStats)

This approximation does not guarantee exact coverage, but

- tests indicate impressive performance
- one can expand about the profile construction to improve coverage, with the limiting case being the full construction

Profile Construction: professional literature

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

While I have been calling it the “profile construction”, it have been called a “hybrid resampling” technique by professional statisticians

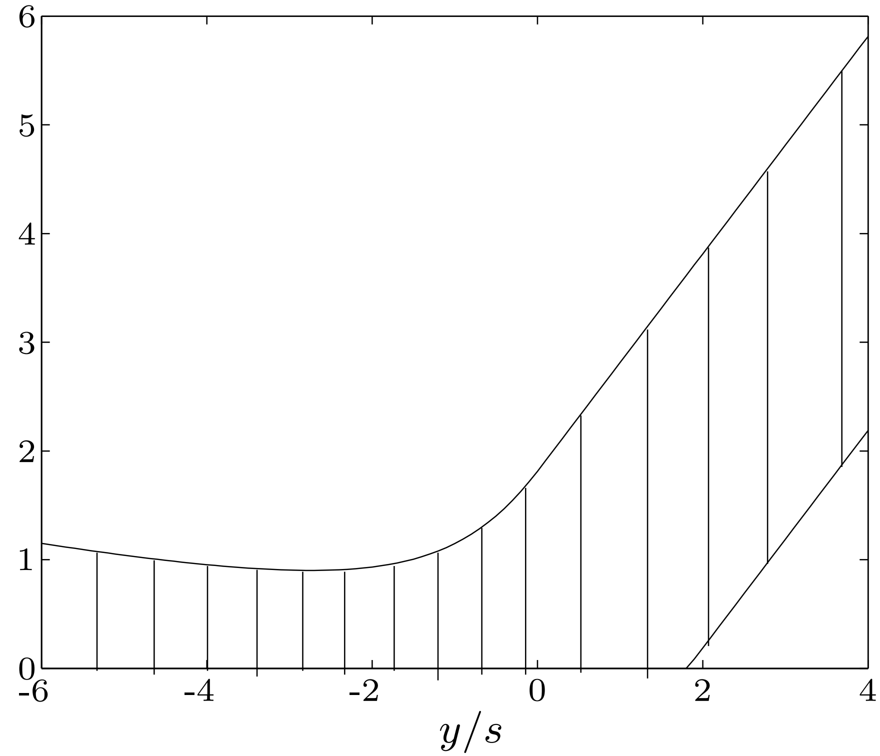
- Note: ‘hybrid’ here has nothing to do with Bayesian-Frequentist Hybrid, but a connection to “boot-strapping”

Statistica Sinica **19** (2009), 301-314

ON THE UNIFIED METHOD WITH NUISANCE PARAMETERS

Bodhisattva Sen, Matthew Walker and Michael Woodroofe

The University of Michigan



Resampling methods for confidence intervals in group sequential trials

BY CHIN-SHAN CHUANG

Department of Statistics, University of Wisconsin at Madison, Madison, Wisconsin 53706, U.S.A.

cchuang@stat.wisc.edu

AND TZE LEUNG LAI

Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.

lait@leland.stanford.edu

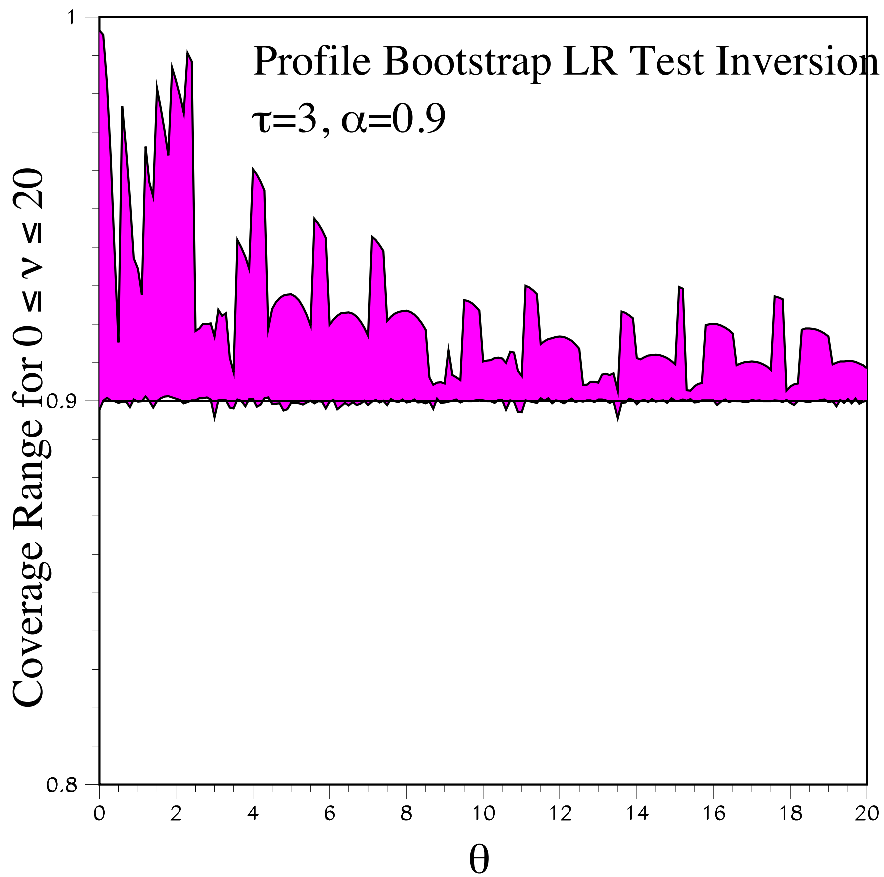
Chuang, C. and Lai, T. L. (1998). Resampling methods for confidence intervals in group sequential trials. *Biometrika* **85**, 317-332.

Chuang, C. and Lai, T. L. (2000). Hybrid resampling methods for confidence intervals. *Statist. Sinica* **10**, 1-50.

Quick Announcement

Luc Demortier has done first coverage study (that I have seen) of our standard approach (the profile construction) for dealing with nuisance parameters in the Neyman Construction when Asymptotics are not necessarily valid.

- results are very good: no significant undercoverage even for small counts. Good news for SUSY and exotics



Asymptotic Properties of likelihood based tests

&

Likelihood-based methods

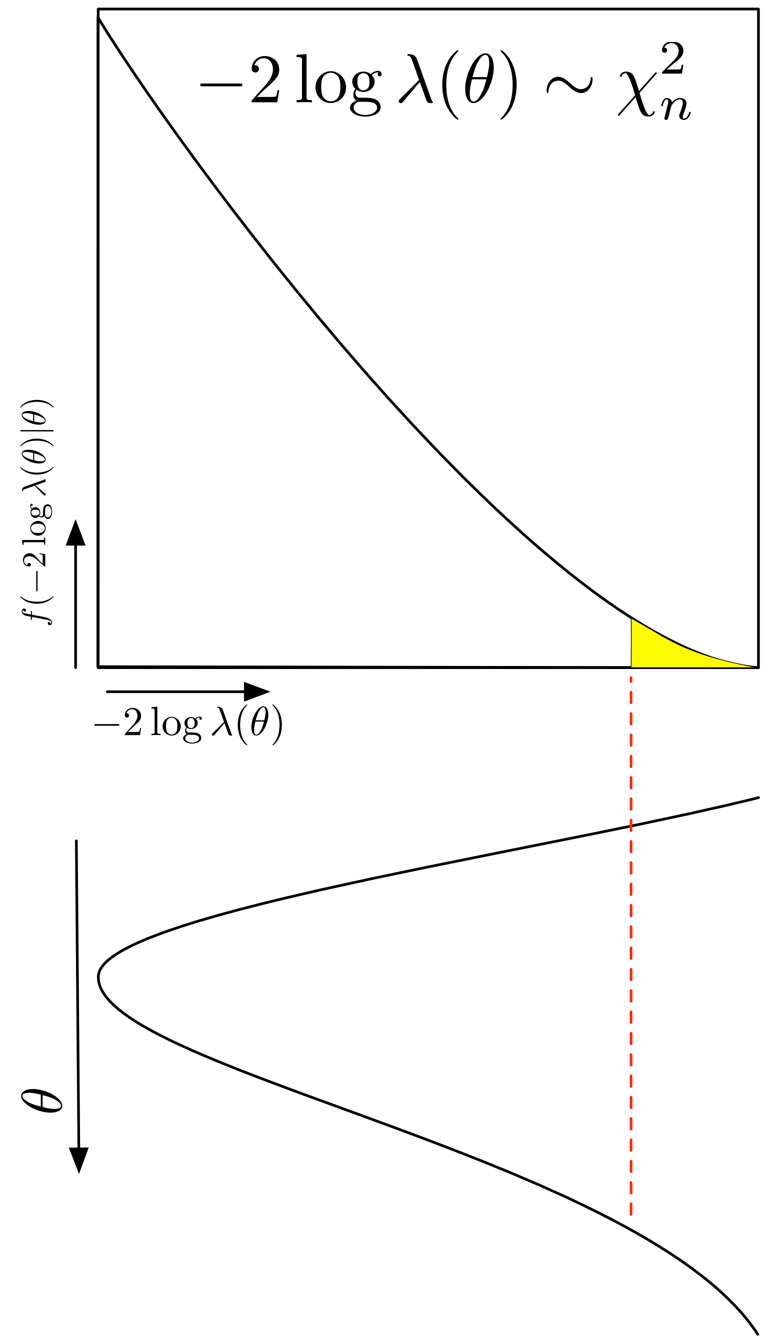
Likelihood-based Intervals

Wilks's theorem tells us how the profile likelihood ratio evaluated at θ is "asymptotically" distributed **when θ is true**

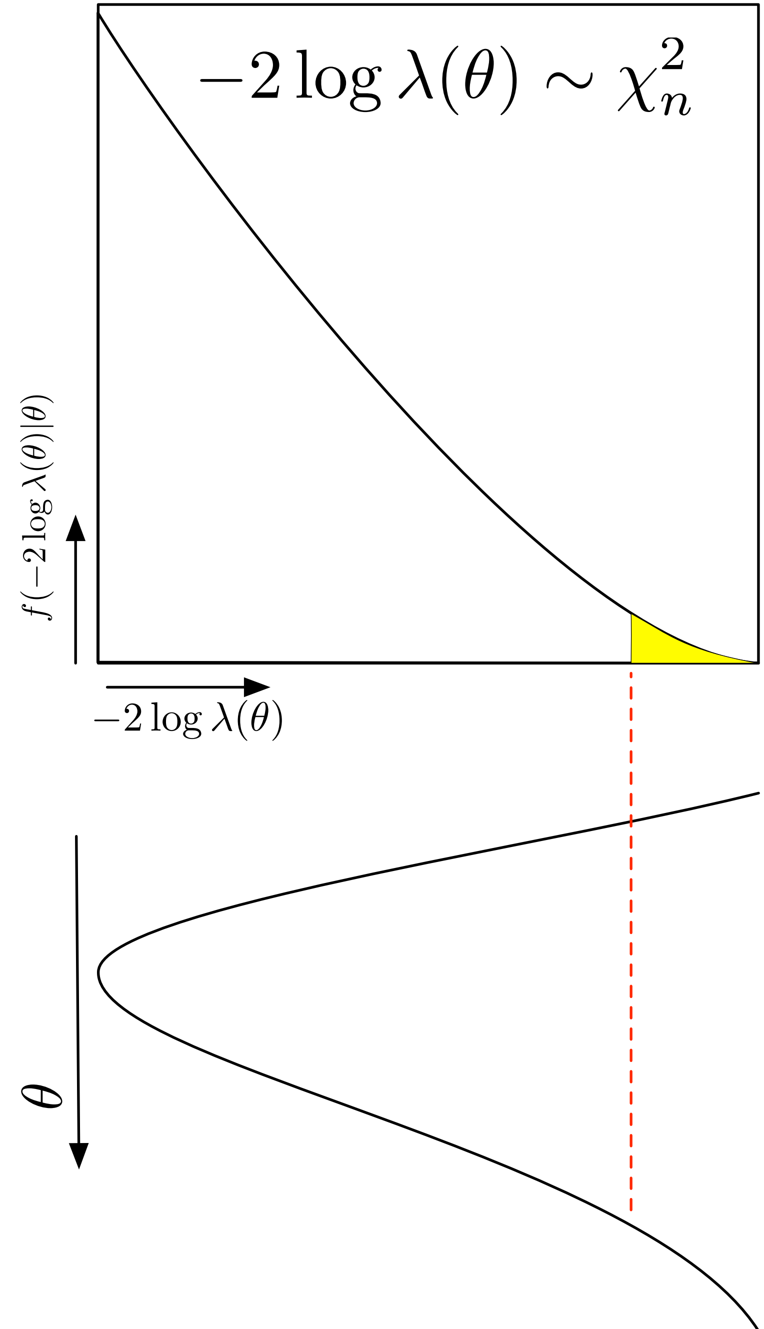
- asymptotically means there is sufficient data that the log-likelihood function is parabolic
- does NOT require the model **f(x|θ)** to be Gaussian

So we don't really need to go to the trouble to build its distribution by using Toy Monte Carlo or fancy tricks with Fourier Transforms

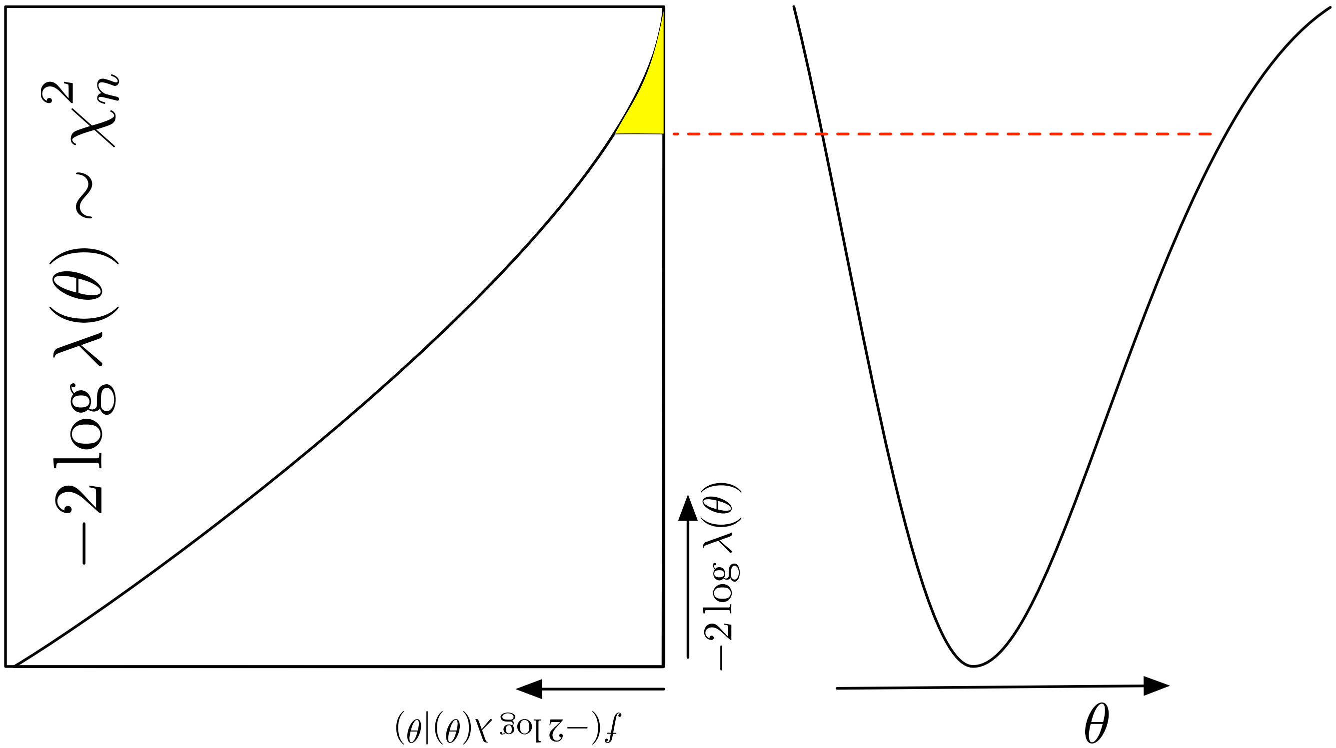
We can go immediately to the threshold value of the profile likelihood ratio



Likelihood-based Intervals

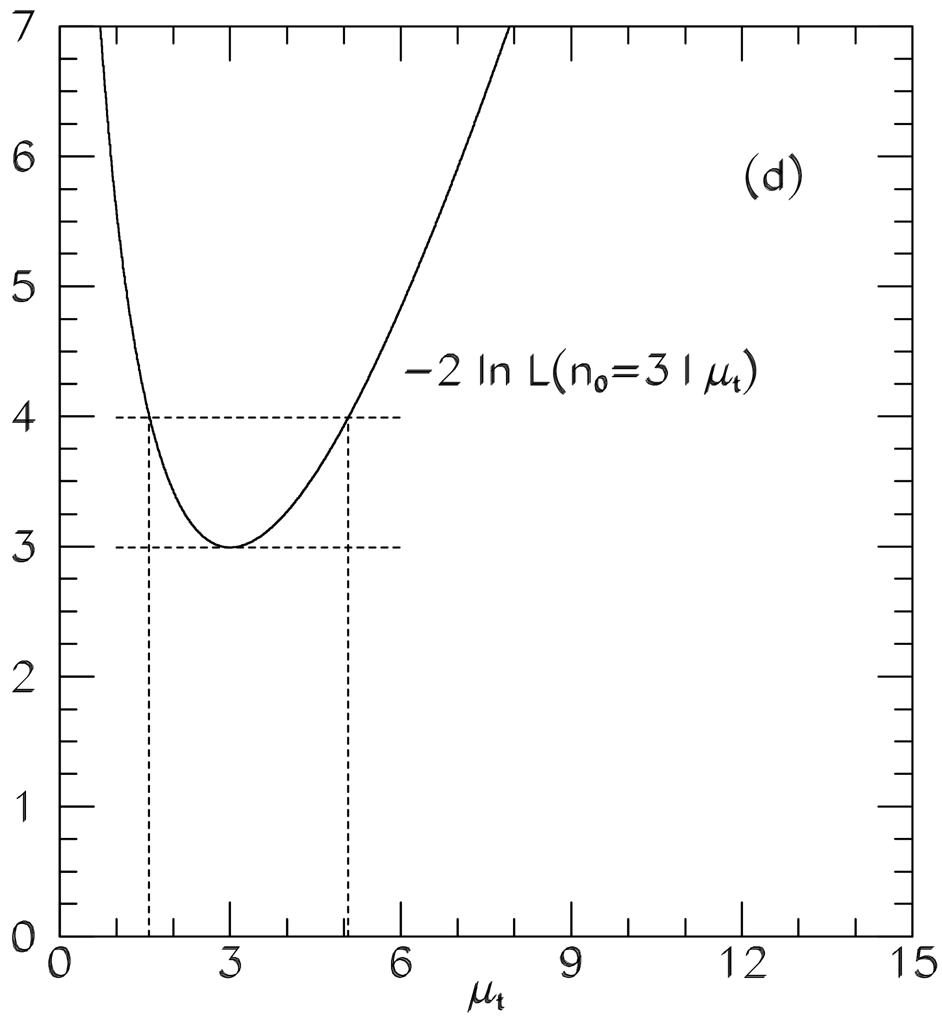


Likelihood-based Intervals



And typically we only show the likelihood curve and don't even bother with the implicit (asymptotic) distribution

Likelihood-based Intervals



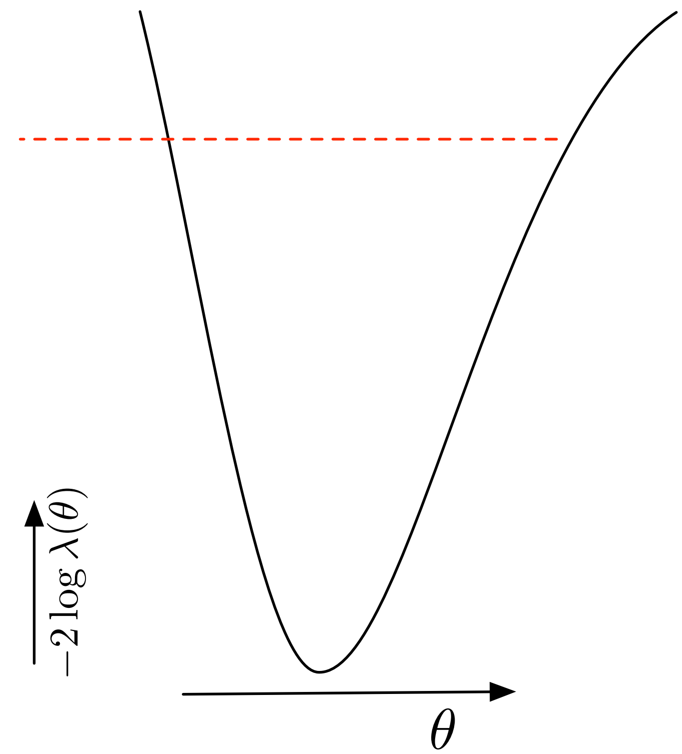
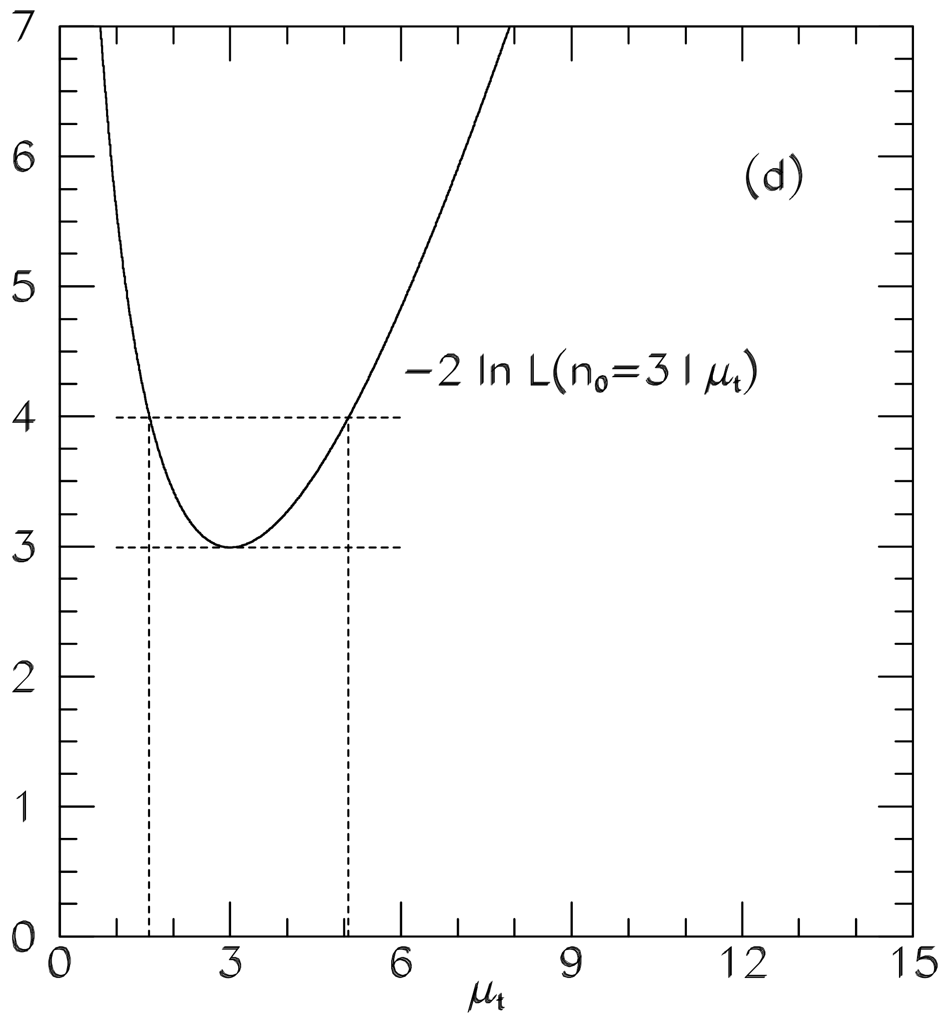


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

And typically we only show the likelihood curve and don't even bother with the implicit (asymptotic) distribution

Likelihood-based Intervals



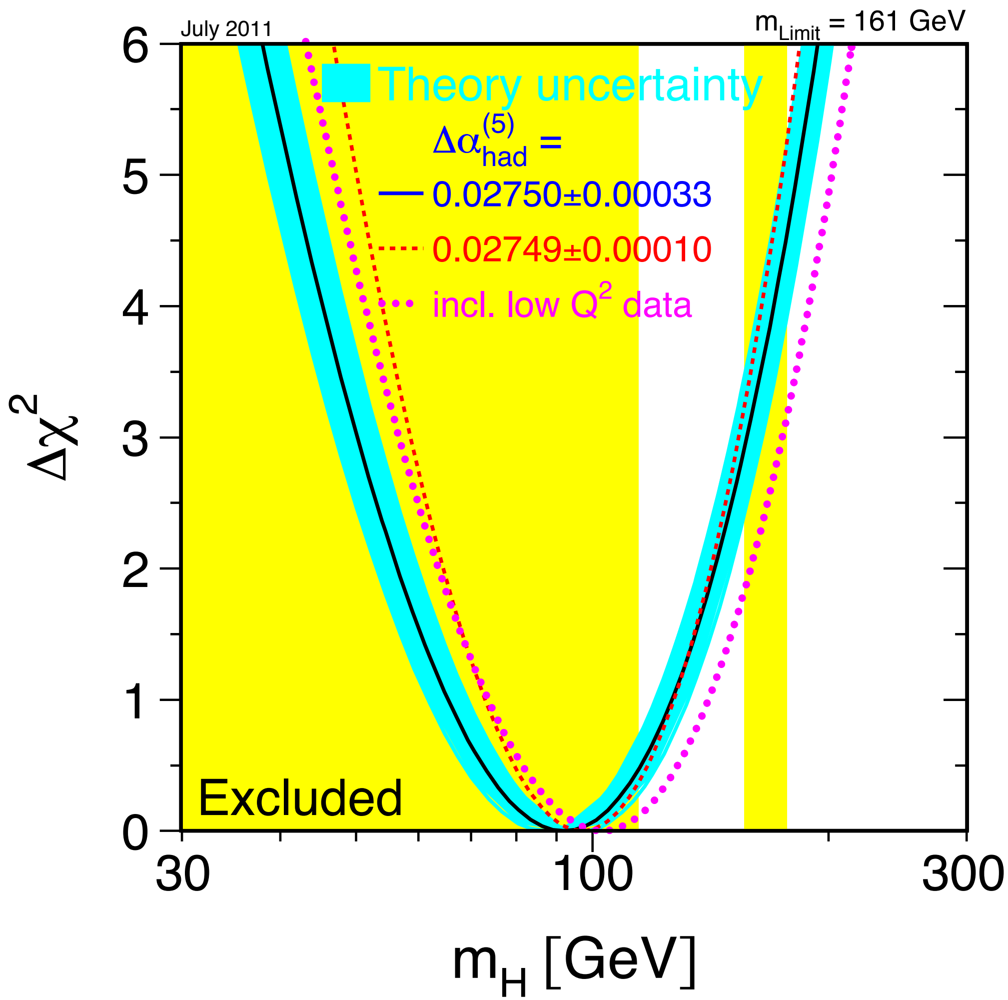


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

And typically we only show the likelihood curve and don't even bother with the implicit (asymptotic) distribution

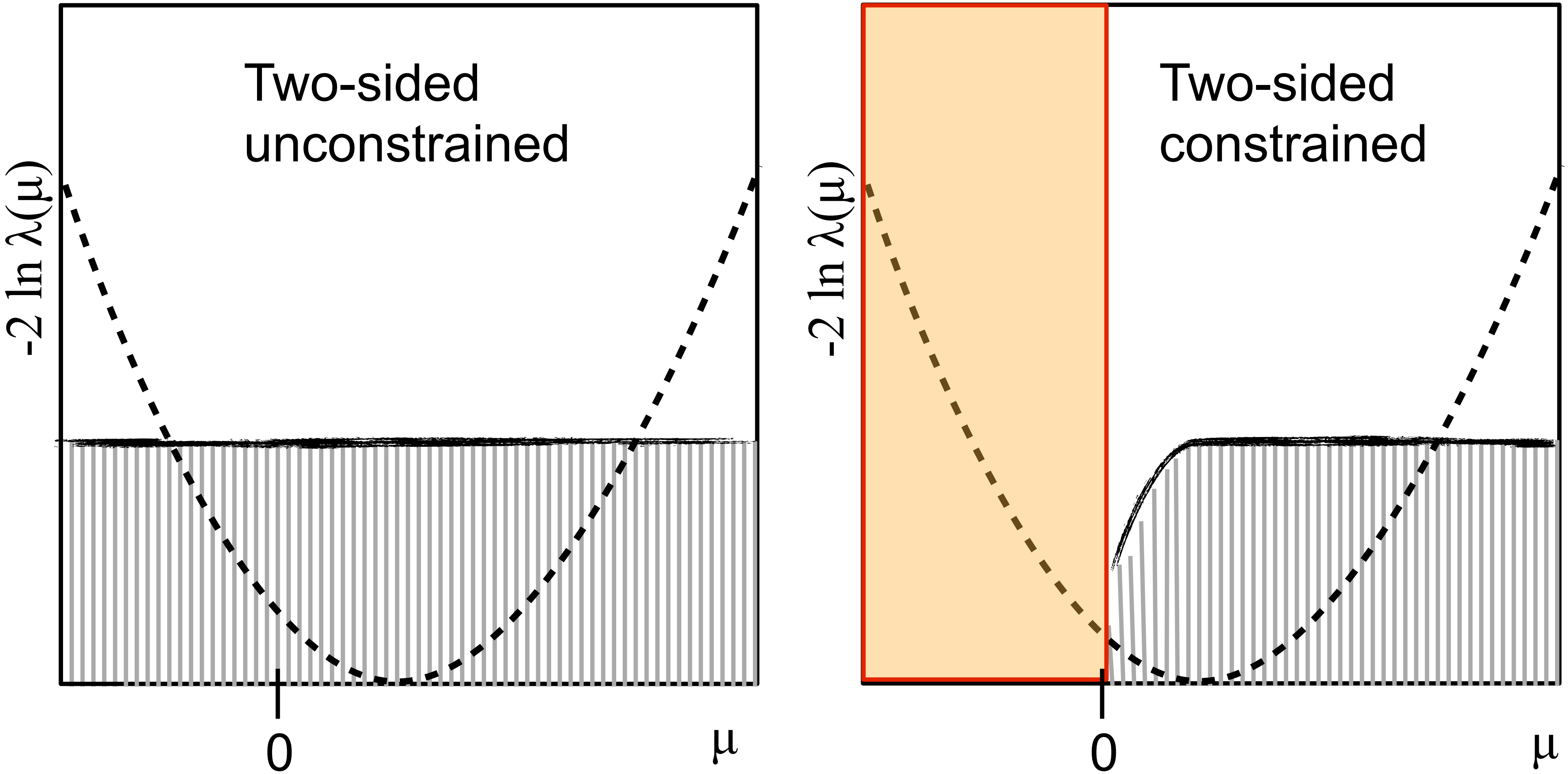
Feldman-Cousins with and without constraint

Wilks's theorem gives a short-cut for the Monte Carlo procedure used to find threshold on test statistic ⇒ MINOS is asymptotic approximation of Feldman-Cousins

- With a physical constraint (μ>0) the confidence band changes

$$t_\mu = -2\ln\lambda(\mu)$$

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0\,. \end{cases}$$



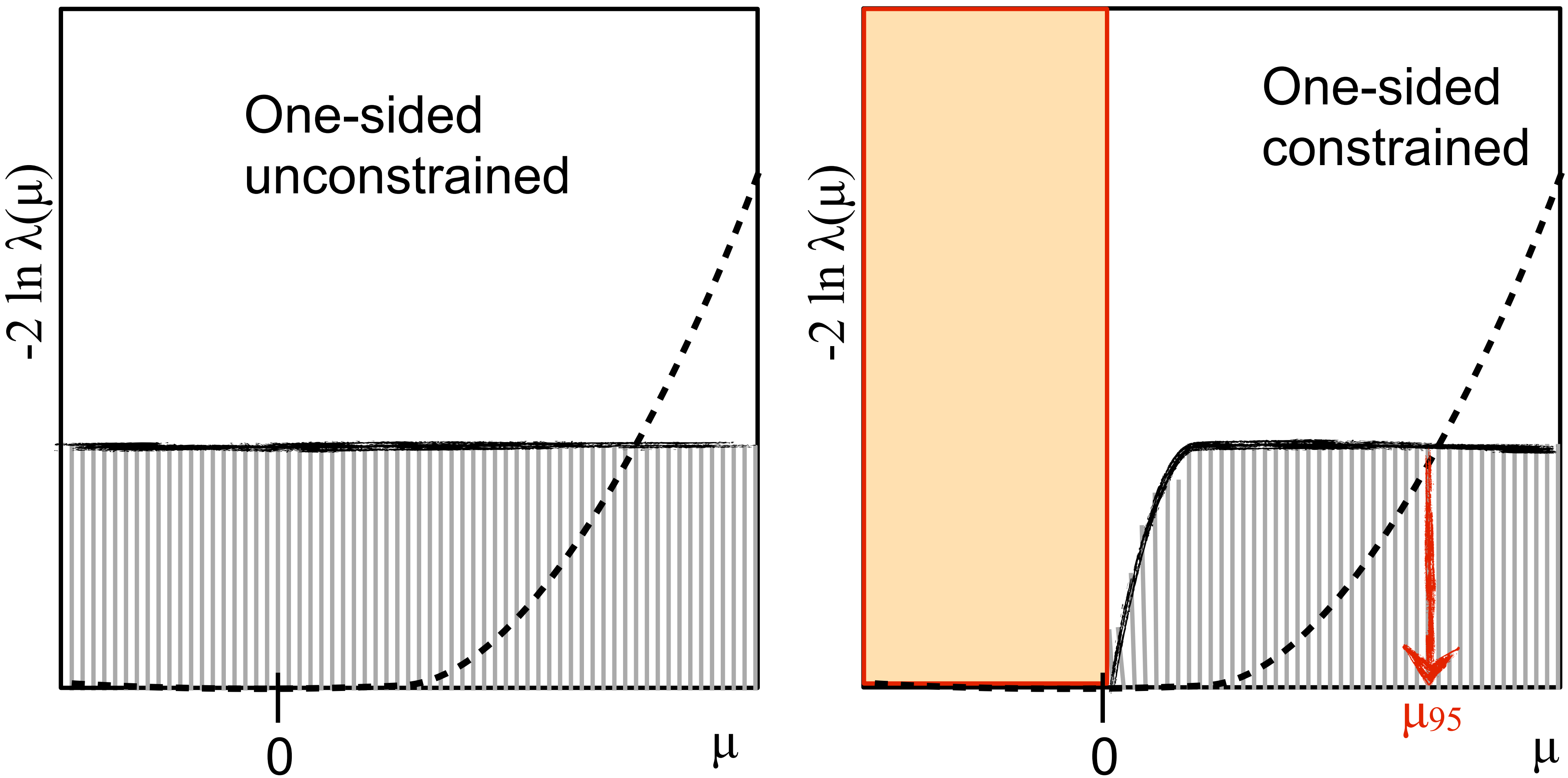
Modified test statistic for 1-sided upper limits

For 1-sided upper-limit the threshold on the test statistic is different

- and with physical boundaries, it is again more complicated

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu\,, \\ 0 & \hat{\mu} > \mu\,, \end{cases}$$

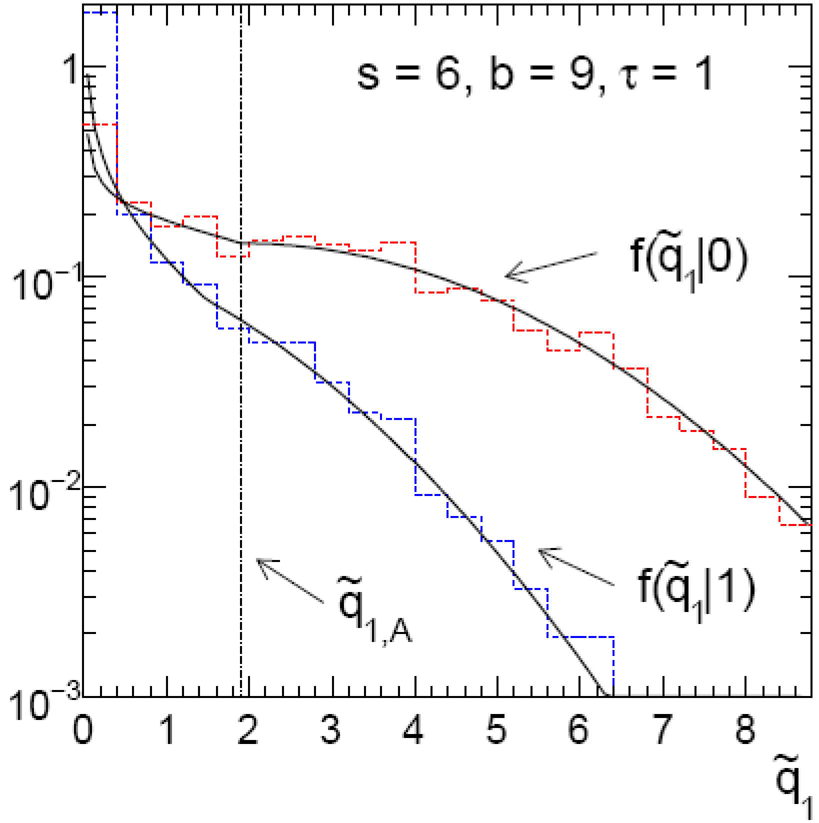
$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}})} & 0 \le \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu\,. \end{cases}$$



Monte Carlo test of asymptotic formulae

Same message for test based on $\tilde{q}_\mu$.

q_μ and $\tilde{q}_\mu$ give similar tests to the extent that asymptotic formulae are valid.



Center for Cosmology and Particle Physics

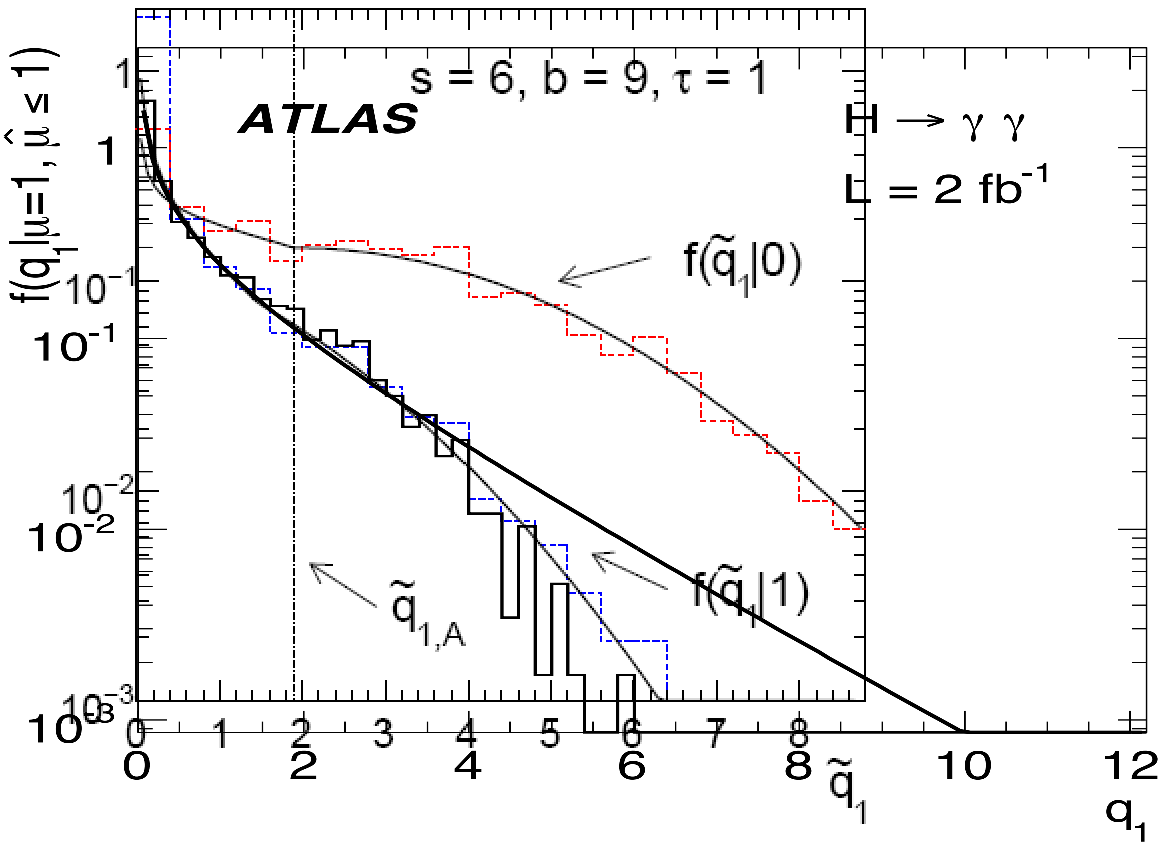
Monte Carlo test of asymptotic formulae

$$f(\tilde{q}_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(\tilde{q}_\mu)$$
$$+ \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\tilde{q}_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right] & 0 < \tilde{q}_\mu \le \mu^2/\sigma^2 , \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu-(\mu^2-2\mu\mu')/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2 . \end{cases}$$

Same message for test based on $\tilde{q}_\mu$.

q_μ and $\tilde{q}_\mu$ give similar tests to the extent that asymptotic formulae are valid.

We now can describe effect of the boundary on the distribution of the test statistic.



G. Cowan Using the Profile Likelihood in Searches for New Physics / Banff 2010 24

2-sided boundaries on physical parameter

Center for Cosmology and Particle Physics

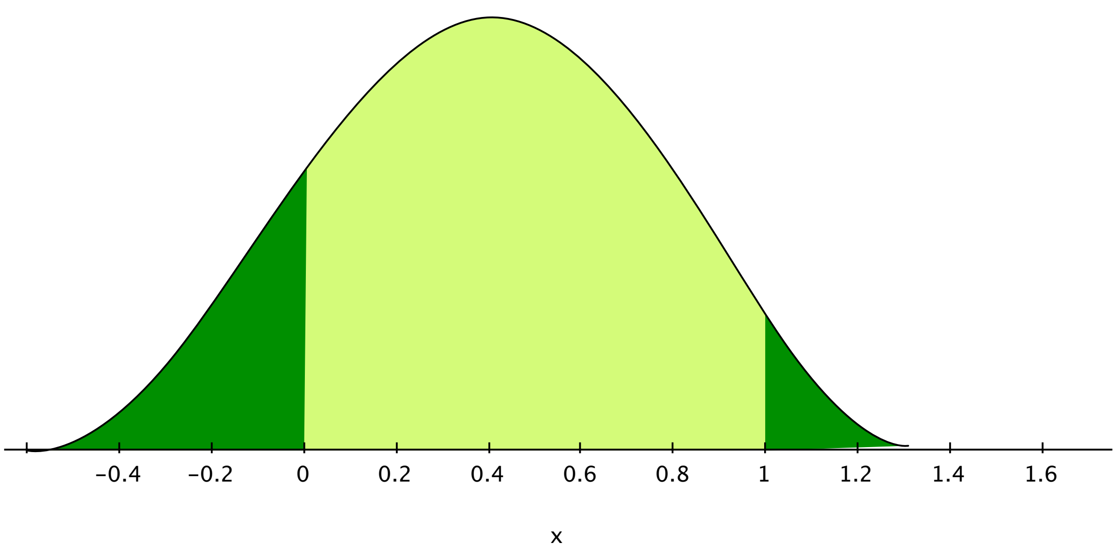
Consider a parameter of interest $\mu \in [0, 1]$

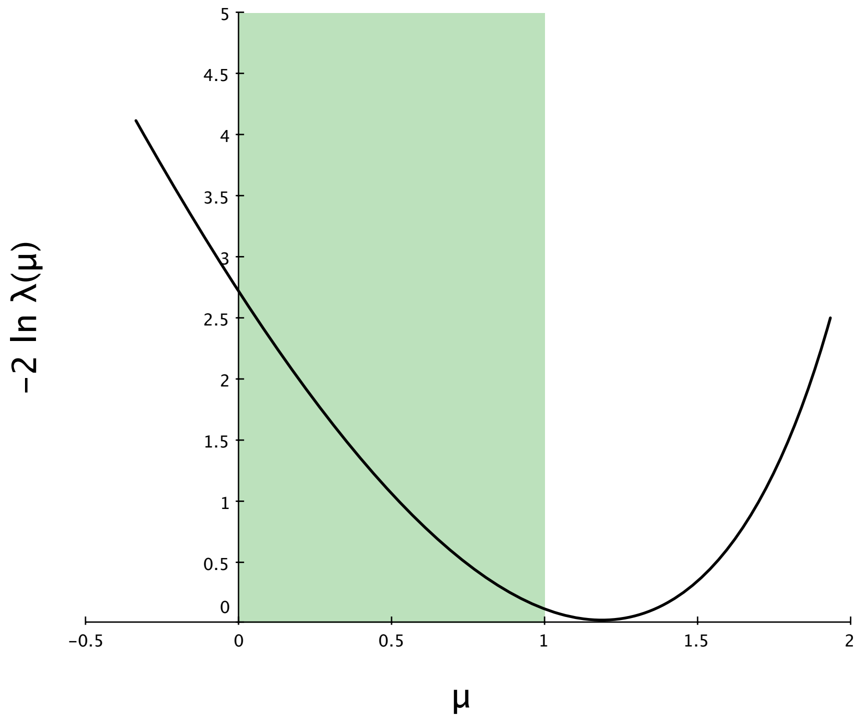
- branching ratio, CKM matrix element, etc.

And a measurement $x \in (-\infty, \infty)$

And a model relating the two $G(x|\mu, \sigma)$

- what do you do if $x < 0$ or $x > 1$





2-sided boundaries on physical parameter

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

Asymptotic distribution for two-sided tests with lower and upper boundaries on the parameter of interest

Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells

with

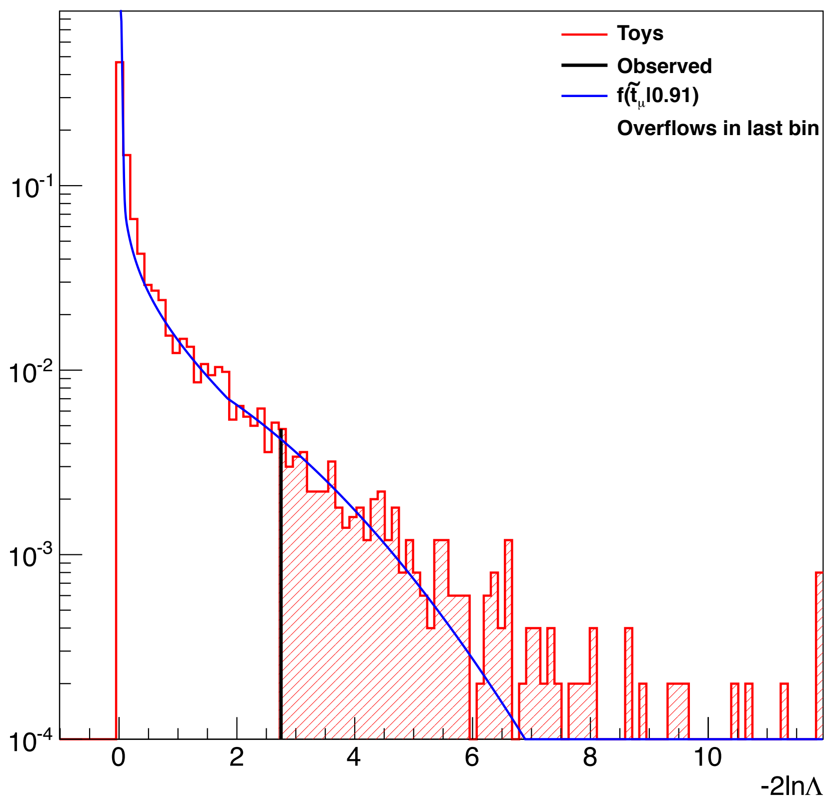
$$f_L(\tilde{t}_\mu|\mu') = \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}} \exp\left[-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right] & \tilde{t}_\mu \le \delta_-^2 \\ \frac{1}{\sqrt{2\pi}}\frac{1}{2\delta_-} \exp\left[-\frac{1}{2}\frac{(\tilde{t}_\mu - (\delta_-^2 - 2\delta_-\delta'_-))^2}{(2\delta_-)^2}\right] & \tilde{t}_\mu > \delta_-^2 \end{cases} \quad (4)$$

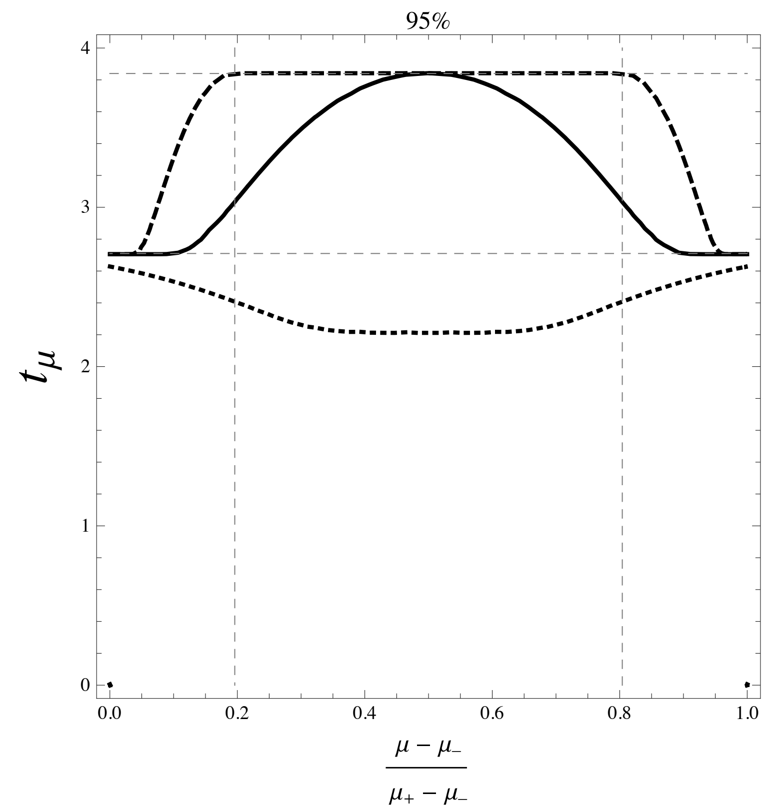
and

$$f_R(\tilde{t}_\mu|\mu') = \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_\mu}} \exp\left[-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2\right] & \tilde{t}_\mu \le \delta_+^2 \\ \frac{1}{\sqrt{2\pi}}\frac{1}{2\delta_+} \exp\left[-\frac{1}{2}\frac{(\tilde{t}_\mu + (\delta_+^2 - 2\delta_+\delta'_+))^2}{(2\delta_+)^2}\right] & \tilde{t}_\mu > \delta_+^2\ , \end{cases} \quad (5)$$

where the dimensionless variables $\delta_- = (\mu - \mu_-)/\sigma$, $\delta'_- = (\mu' - \mu_-)/\sigma$, $\delta_+ = (\mu - \mu_+)/\sigma$, and $\delta'_+ = (\mu' - \mu_+)/\sigma$ are used to simplify the expressions.

arXiv:1210.6948





Bayesian methods

Coverage & Likelihood principle

Methods based on the Neyman-Construction always cover.... by construction.

- this approach violates the likelihood principle

Bayesian methods obey likelihood principle, but do not necessarily cover

- that doesn't mean Bayesians shouldn't care about coverage

Coverage can be thought of as a **calibration of our statistical apparatus**. [explain under-/over-coverage]

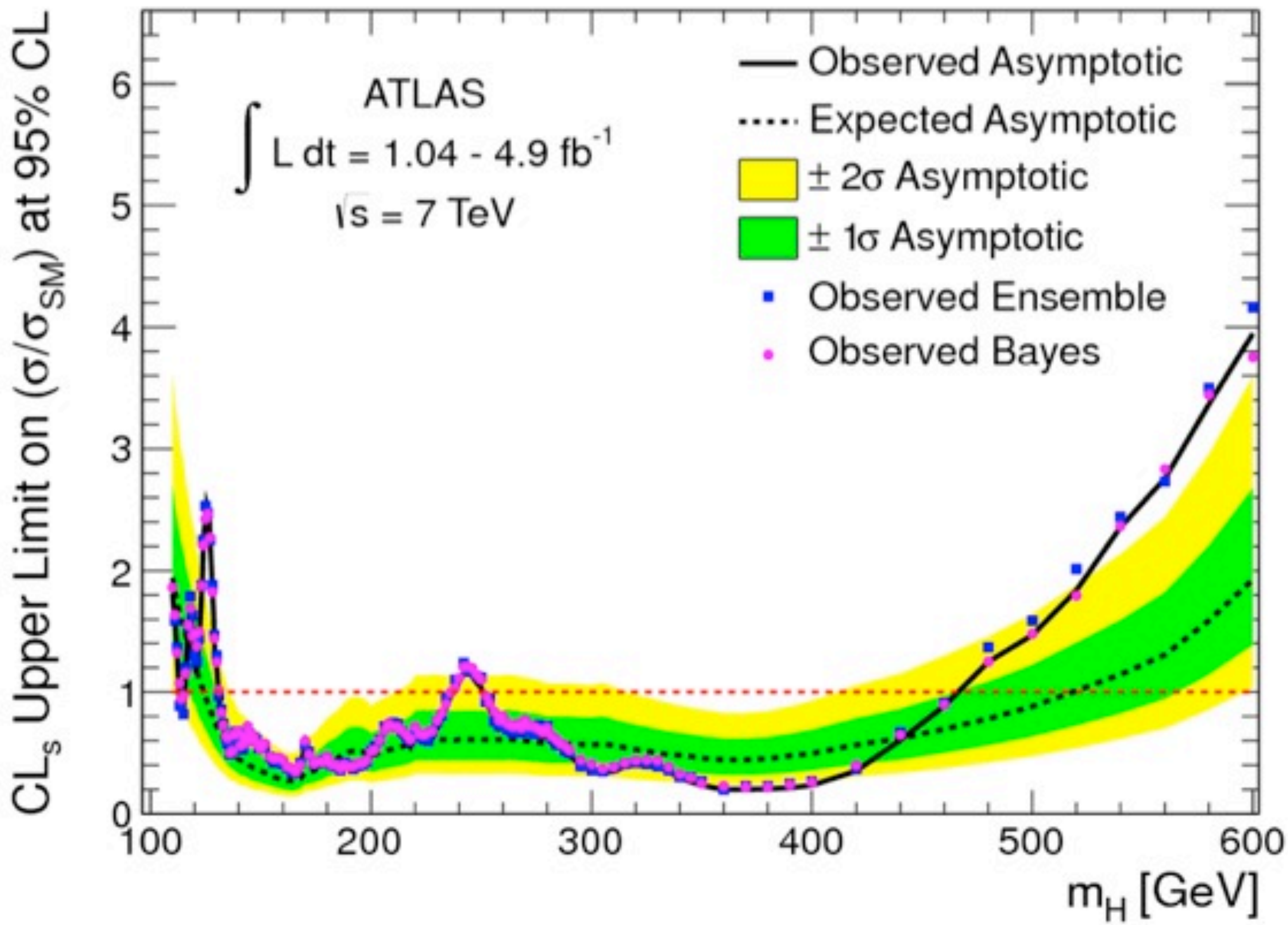
> what should be the view today;
> Objective Bayesian analysis is the
> best frequentist tool around. -Jim Berger

Bayesian and Frequentist results answer different questions

- major differences between them may indicate severe coverage problems and/or violations of the likelihood principle

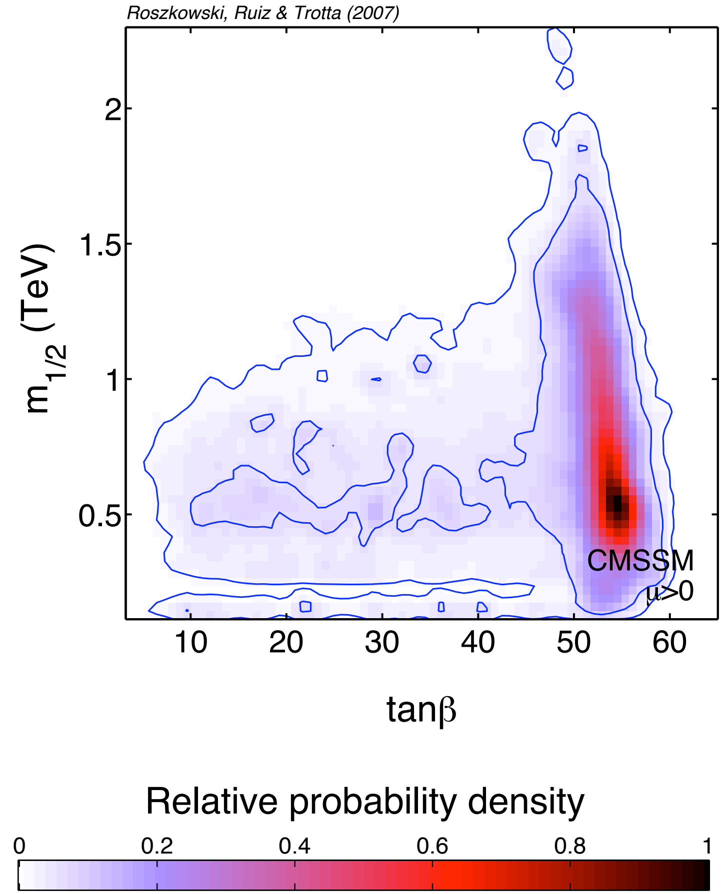
Monte Carlo, asymptotic, Bayesian

Here we see comparisons of explicit ensembles generated with Monte Carlo techniques, the asymptotic results, and Bayesian results using MCMC and nested sampling with a uniform prior on μ



Bayesian credible intervals

Bayesian "credible interval" V does mean that there is a 95% that the probability parameter is in interval.



The procedure is very intuitive:

$$P(\theta \in V) = \int_V \pi(\theta|x) = \int_V d\theta \frac{f(x|\theta)\pi(\theta)}{\int d\theta f(x|\theta)\pi(\theta)}$$

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a nice technique which will produce a sampling of a parameter space which is proportional to a posterior

- it works well in high dimensional problems
- Metropolis-Hastings Algorithm: generates a sequence of points $\{\vec{\alpha}^{(t)}\}$
  - Given the likelihood function $L(\vec{\alpha})$ & prior $P(\vec{\alpha})$, the posterior is proportional to $L(\vec{\alpha}) \cdot P(\vec{\alpha})$
  - propose a point $\vec{\alpha}'$ to be added to the chain according to a proposal density $Q(\vec{\alpha}'|\vec{\alpha})$ that depends only on current point $\vec{\alpha}$
  - if posterior is higher at $\vec{\alpha}'$ than at $\vec{\alpha}$, then add new point to chain
  - else: add $\vec{\alpha}'$ to the chain with probability

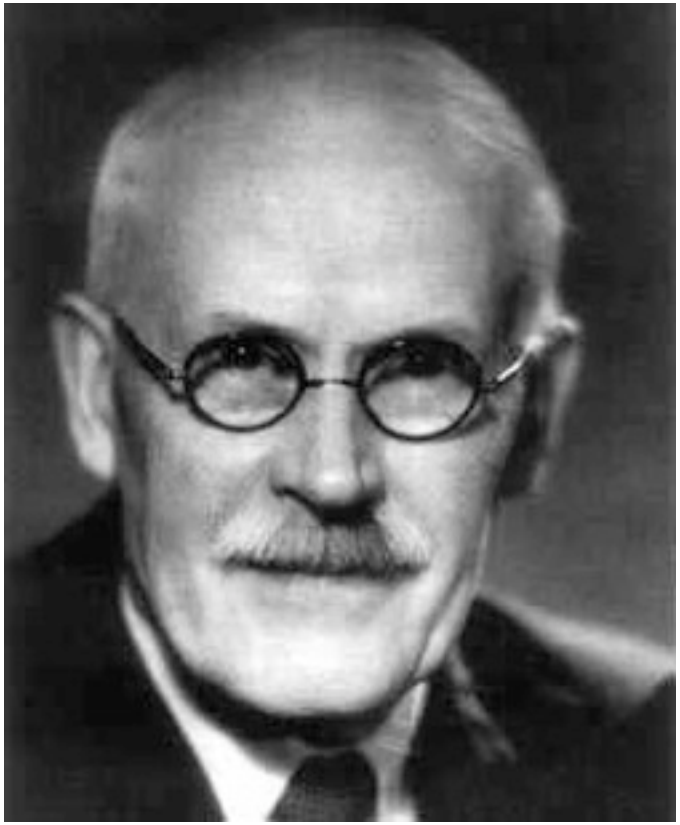
$$\rho = \frac{L(\vec{\alpha}') \cdot P(\vec{\alpha}')}{L(\vec{\alpha}) \cdot P(\vec{\alpha})} \cdot \frac{Q(\vec{\alpha}|\vec{\alpha}')}{Q(\vec{\alpha}'|\vec{\alpha})}$$

  - (appending original point $\vec{\alpha}$ with complementary probability)
- RooStats works with any $L(\vec{\alpha})$, $P(\vec{\alpha})$
- can use any RooFit PDF as proposal function $Q(\vec{\alpha}'|\vec{\alpha})$
  - Helper for forming custom multivariate Gaussian, Bank of Clues, etc.
  - New Sequential Proposal function similar to BAT

The Jeffreys Prior

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

Physicist Sir Harold Jeffreys had the clever idea that we can "**objectively**" create a flat prior uniform in a metric determined by $I(\theta)$



Adds "minimal information" in a precise sense, and results in: $p(\vec{\theta}) \propto \sqrt{I(\vec{\theta})}.$

It has the key feature that it is invariant under reparameterization of the parameter vector $\vec{\varphi}$ In particular, for an alternate parameterization $\vec{\theta}$ we can derive

$$\begin{aligned} p(\vec{\varphi}) &= p(\vec{\theta}) \left| \det\left(\frac{\partial \theta_i}{\partial \varphi_j}\right) \right| \\ &\propto \sqrt{I(\vec{\theta}) \det^2\left(\frac{\partial \theta_i}{\partial \varphi_j}\right)} \\ &= \sqrt{\det\left(\frac{\partial \theta_k}{\partial \varphi_i}\right) \det\left(E\left[\frac{\partial \ln L}{\partial \theta_k}\frac{\partial \ln L}{\partial \theta_l}\right]\right) \det\left(\frac{\partial \theta_l}{\partial \varphi_j}\right)} \\ &= \sqrt{\det\left(E\left[\sum_{k,l} \frac{\partial \theta_k}{\partial \varphi_i}\frac{\partial \ln L}{\partial \theta_k}\frac{\partial \ln L}{\partial \theta_l}\frac{\partial \theta_l}{\partial \varphi_j}\right]\right)} \\ &= \sqrt{\det\left(E\left[\frac{\partial \ln L}{\partial \varphi_i}\frac{\partial \ln L}{\partial \varphi_j}\right]\right)} = \sqrt{I(\vec{\varphi})}. \end{aligned}$$

Unfortunately, the Jeffreys prior in multiple dimensions causes some problems, and in certain circumstances gives undesirable answers.

Reference Priors

Reference priors are another type of "objective" priors, that try to save Jeffreys' basic idea.

Noninformative priors have been studied for a long time and most of them have been found defective in more than one way. Reference analysis arose from this study as the only *general* method that produces priors that have the required *invariance* properties, deal successfully with the *marginalization* paradoxes, and have consistent *sampling* properties.

Ideally, such a method should be very general, applicable to all kinds of measurements regardless of the number and type of parameters and data involved. It should make use of *all* available information, and coherently so, in the sense that if there is more than one way to extract all relevant information from data, the final result will not depend on the chosen way. The desiderata of generality, exhaustiveness and coherence are satisfied by Bayesian procedures, but that of objectivity is more problematic due to the Bayesian requirement of specifying prior probabilities in terms of degrees of belief. Reference analysis[2], an objective Bayesian method developed over the past twenty-five years, solves this problem by replacing the question "what is our prior degree of belief?" by "what would our posterior degree of belief be, if our prior knowledge had a minimal effect, relative to the data, on the final inference?"

See Luc Demortier's PhyStat 2005 proceedings

http://physics.rockefeller.edu/luc/proceedings/phystat2005_refana.ps

Jeffreys's Prior

Jeffreys's Prior is an "objective" prior based on formal rules
(it is related to the Fisher Information and the Cramér-Rao bound]

$$\pi(\vec{\theta}) \propto \sqrt{\det \mathcal{I}\left(\vec{\theta}\right)}.$$

$$\left(\mathcal{I}\left(\theta\right)\right)_{i,j} = -\,\mathrm{E}\left[\left.\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\ln f(X;\theta)\right|\theta\right].$$

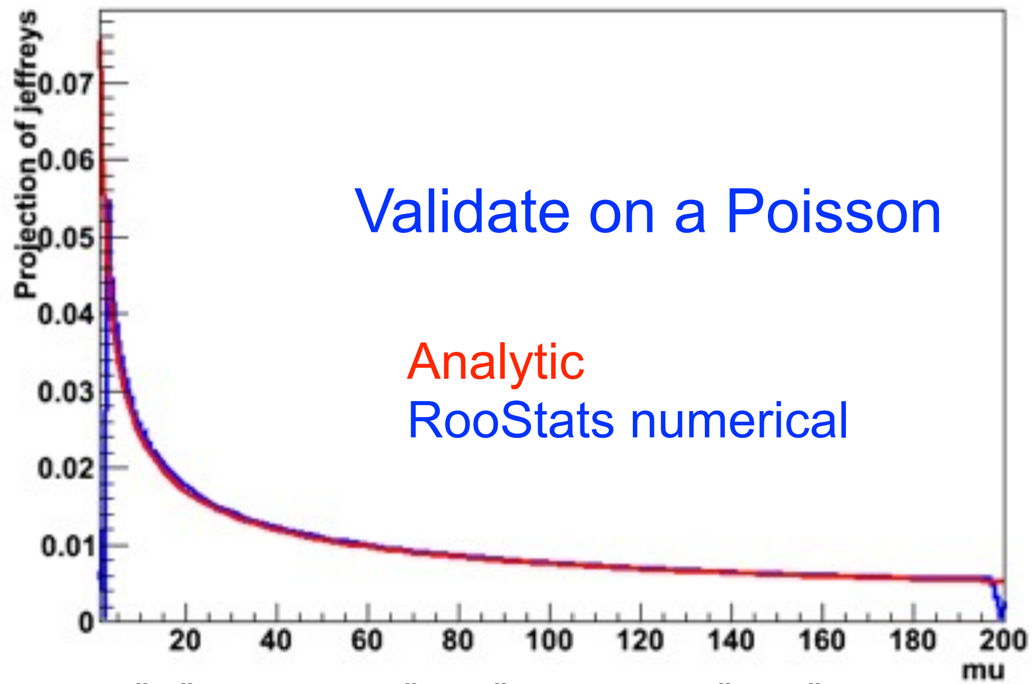
Eilam, Glen, Ofer, and I showed in arXiv:1007.1727 that the Asimov data provides a fast, convenient way to calculate the Fisher Information

$$V_{jk}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_j \partial\theta_k}\right] = -\frac{\partial^2 \ln L_{\mathrm{A}}}{\partial\theta_j \partial\theta_k} = \sum_{i=1}^{N} \frac{\partial\nu_i}{\partial\theta_j}\frac{\partial\nu_i}{\partial\theta_k}\frac{1}{\nu_i} + \sum_{i=1}^{M} \frac{\partial u_i}{\partial\theta_j}\frac{\partial u_i}{\partial\theta_k}\frac{1}{u_i}$$

Use this as basis to calculate
Jeffreys's prior for an arbitrary PDF!

```
RooWorkspace w("w");
w.factory("Uniform::u(x[0,1])");
w.factory("mu[100,1,200]");
w.factory("ExtendPdf::p(u,mu)");

w.defineSet("poi","mu");
w.defineSet("obs","x");
//  w.defineSet("obs2","n");
```




```
RooJeffreysPrior pi("jeffreys","jeffreys",*w.pdf("p"),*w.set("poi"),*w.set("obs"));
```

The Bayesian Solution

Center for Cosmology and Particle Physics

Bayesian solution generically have a prior for the parameters of interest as well as nuisance parameters

- 2010 recommendations largely echoes the PDG's stance.

Recommendation: When performing a Bayesian analysis one should separate the objective likelihood function from the prior distributions to the extent possible.

Recommendation: When performing a Bayesian analysis one should investigate the sensitivity of the result to the choice of priors.

Warning: Flat priors in high dimensions can lead to unexpected and/or misleading results.

Recommendation: When performing a Bayesian analysis for a single parameter of interest, one should attempt to include Jeffreys's prior in the sensitivity analysis.

The End

Thank You!