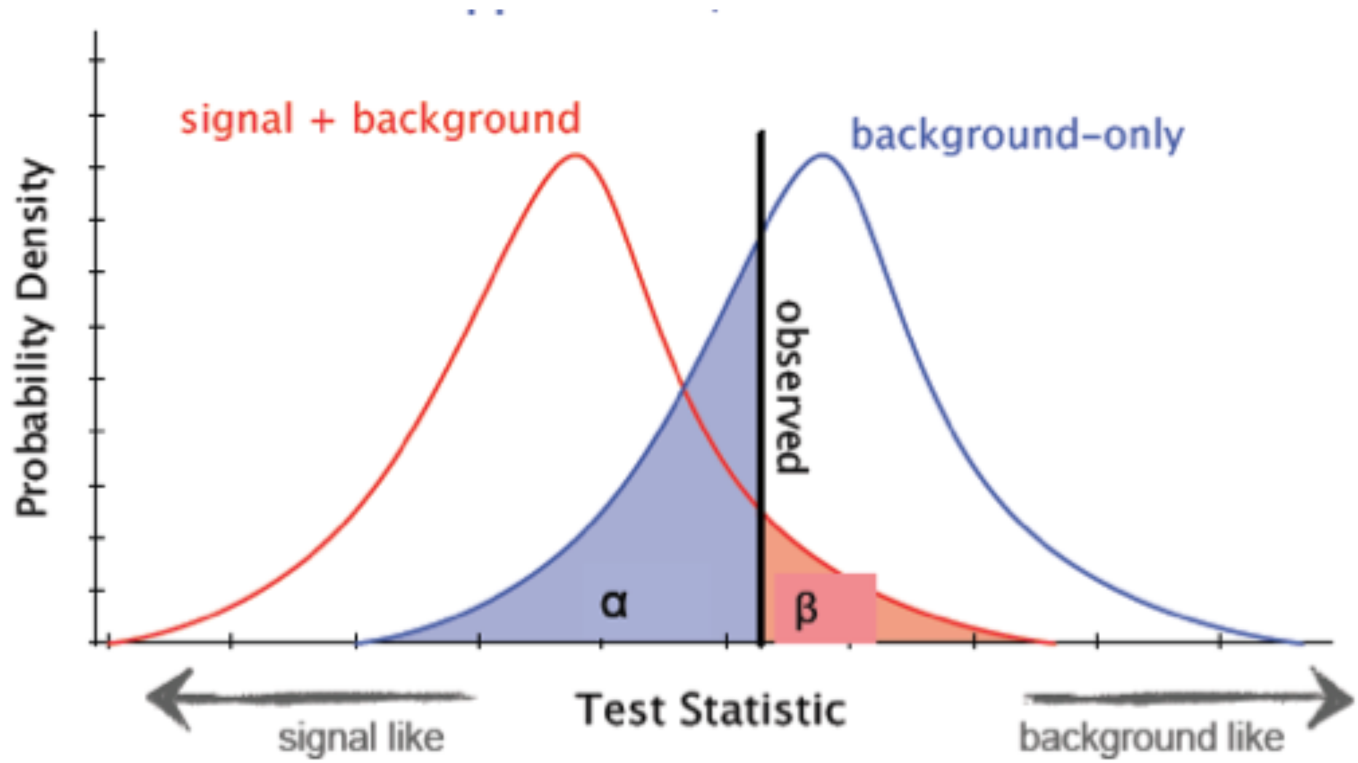
Practical Statistics for Particle Physics

Daniel Whiteson, UC Irvine

HCPSS, 2014: Lecture 3

Test statistic

Reduce vector of observables to 1 number



How to build distribution of TS? (Usually MC)

How to choose TS?

(K. Cranmer)

Test statistic

Define μ to be signal strength,
μ=0 is no signal
μ=1 is theory prediction

At LEP, this was used:

$$Q_{LEP} = L_{s+b}(\mu = 1)/L_b(\mu = 0)$$

Where the nuisance parameters are fixed to their nominal values

Test statistic

Define μ to be signal strength,
μ=0 is no signal
μ=1 is theory prediction

At LEP, this was used:

$$Q_{LEP} = \frac{L(data|\mu = 1, b, \nu)}{L(data|\mu = 0, b, \nu)}$$

This also means the background estimate doesn't vary.

Tevatron

Still consider two points (0,1)
but now float the NPs at those points

$$Q_{TEV} = L_{s+b}(\mu = 1, \hat{\hat{\nu}})/L_b(\mu = 0, \hat{\hat{\nu}}')$$

Ratio of profiled likelihoods:
the model is adapted to the data
even in the signal region

LHC

Profile likelihood

$$\lambda(\mu=0) = \frac{L(data|\mu=0,\hat{\hat{b}}(\mu=0),\hat{\hat{\nu}}(\mu=0))}{L(data|\hat{\mu},\hat{b},\hat{\nu})}$$

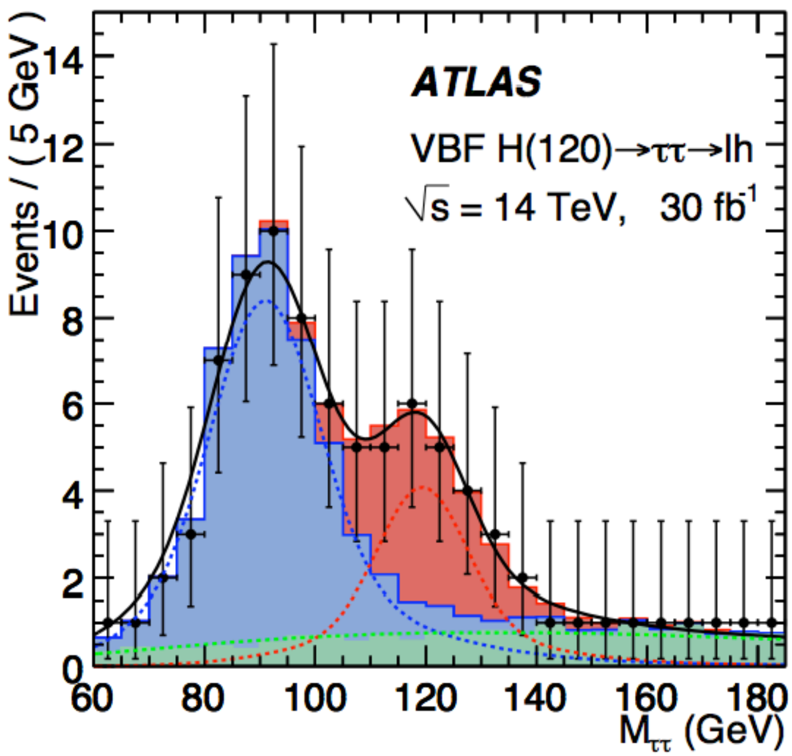
fit best value of NPs at μ=0

and at **best fit value of μ**

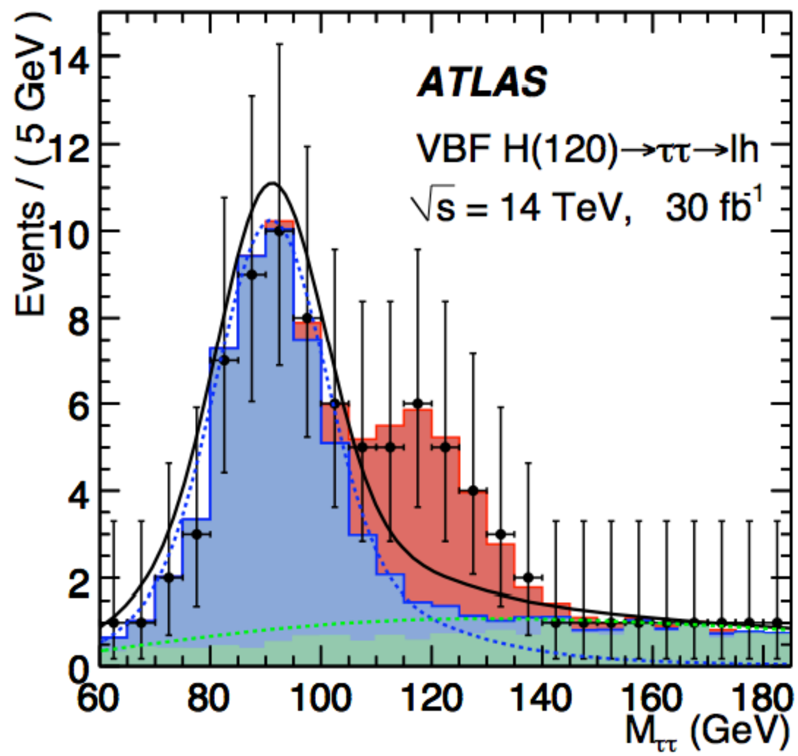
Two fits to data

$$\lambda(\mu=0) = \frac{L(data|\mu=0, \hat{\hat{b}}(\mu=0), \hat{\hat{\nu}}(\mu=0))}{L(data|\hat{\mu}, \hat{b}, \hat{\nu})},$$

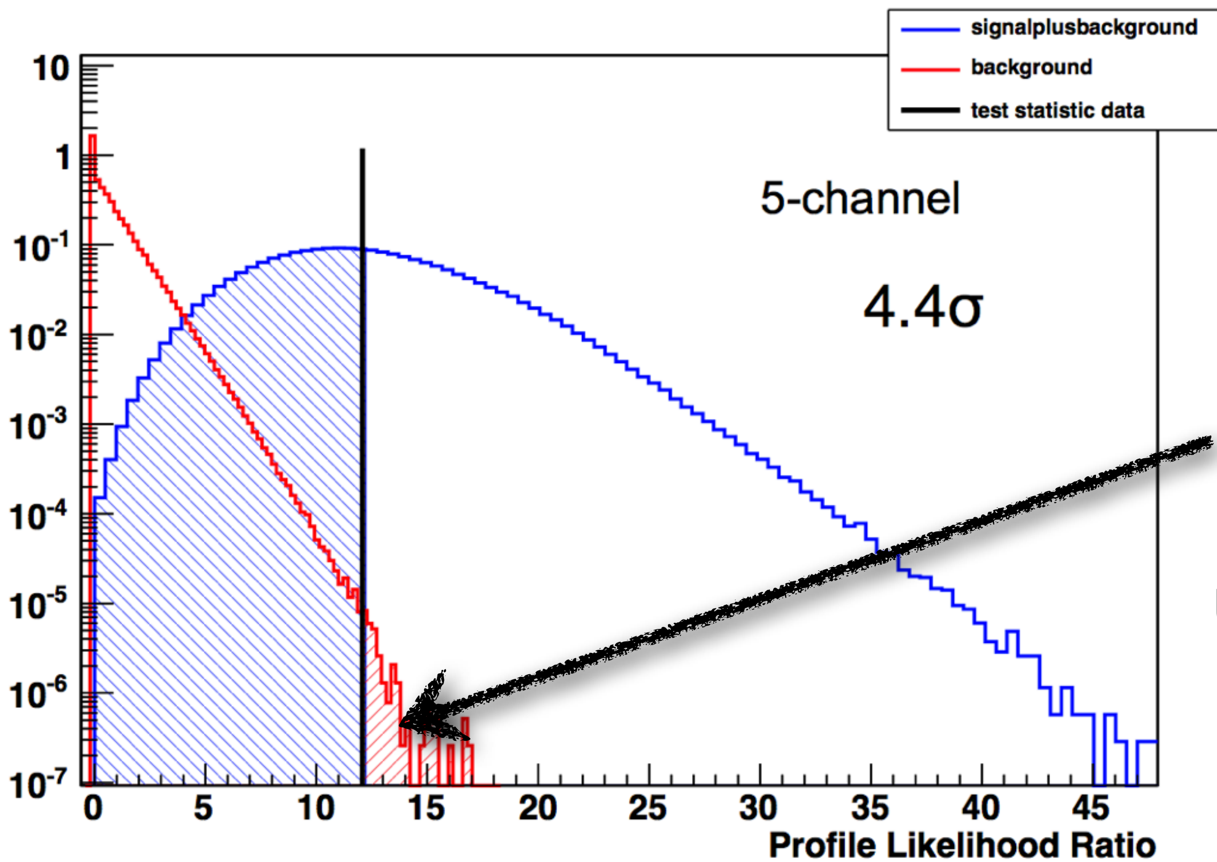
$L(data|\hat{\mu}, \hat{b}, \hat{\nu})$



$L(data|\mu=0, \hat{\hat{b}}, \hat{\hat{\nu}})$



p values

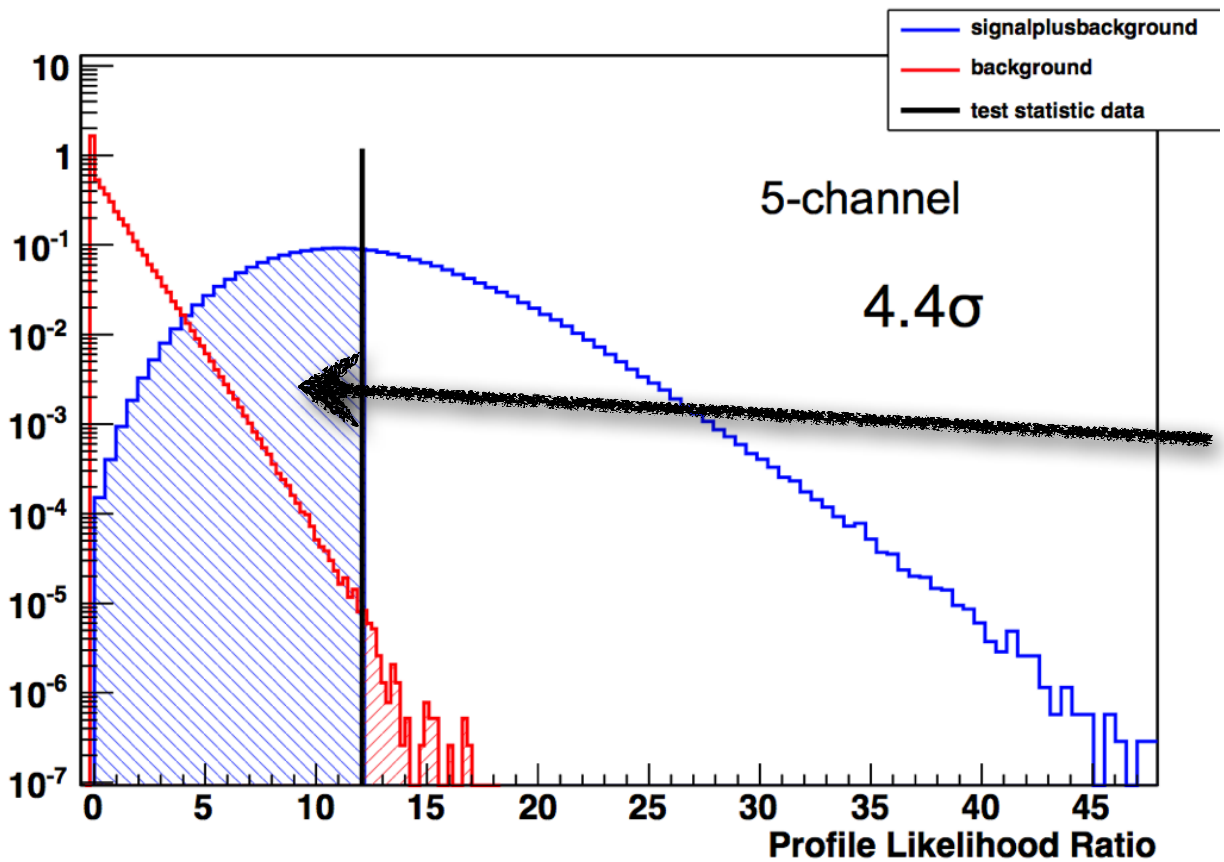


p_0 =
probability
to observe data
or **more** signal-like
under **background**
hypothesis

$$p_0 = P\left(q_0 \geq q_o^{obs}\right)$$

(K. Cranmer)

p values



p_μ =
probability
to observe data
or **less** signal-like
under **signal+b**
hypothesis

(K. Cranmer)

Philosophy

Bayesian

&

Frequentist

Bayesian

Data: fixed

Parameter values: unknown

Probability: our lack of knowledge

PDFs over parameters: sensible

Frequentist

Data: one example from ens.
Parameter values: fixed (even if unknown)
Probability: rate of occurance
PDFs over parameters: not sensible

Bayesian Prob.

Bayes theorem:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

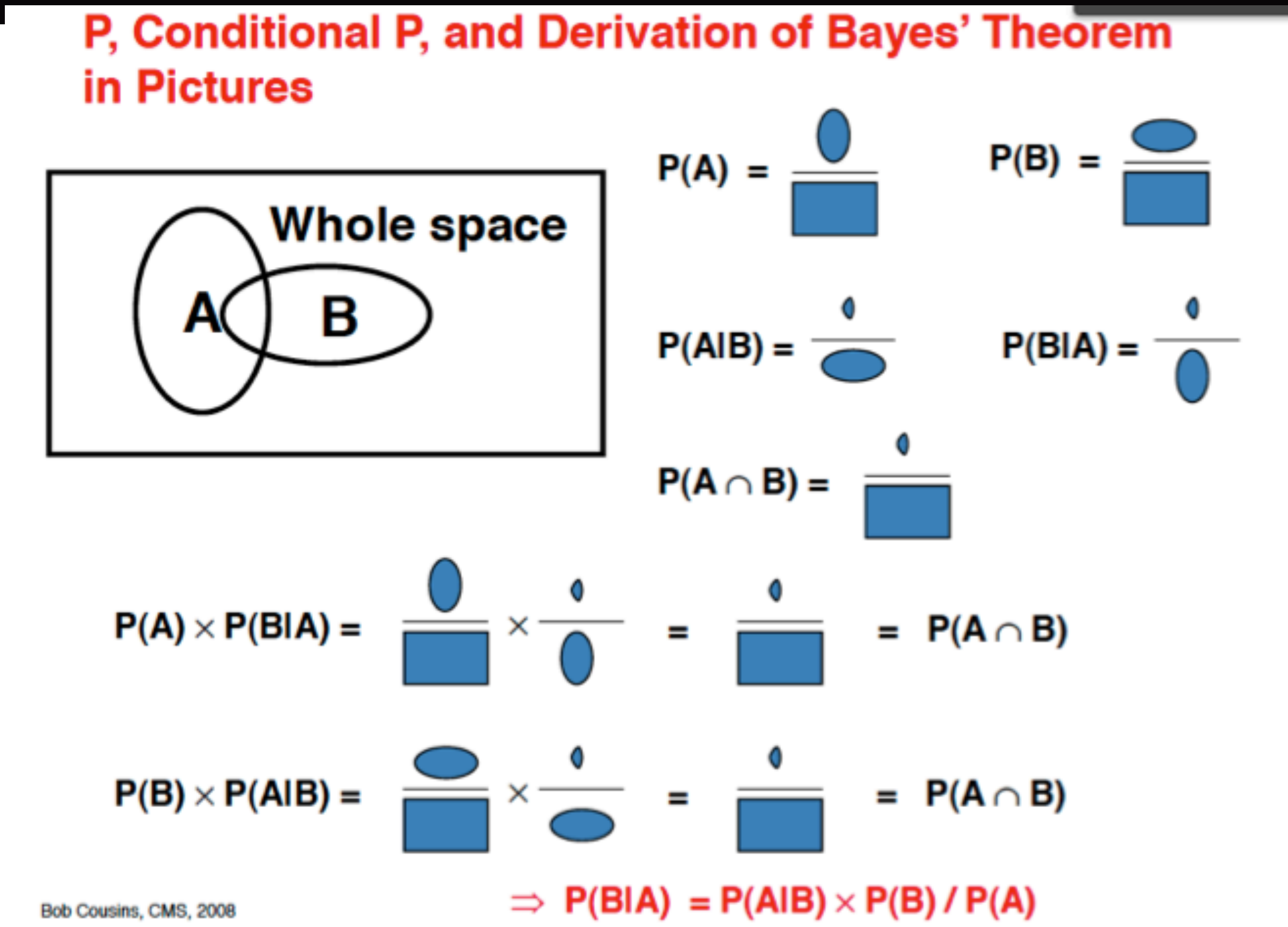
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

rearrange:

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In Pictures


P, Conditional P, and Derivation of Bayes' Theorem in Pictures
Whole space
A
B
P(A) =
P(B) =
P(AIB) =
P(BIA) =
P(A ∩ B) =
P(A) × P(BIA) =
×
=
= P(A ∩ B)
P(B) × P(AIB) =
×
=
= P(A ∩ B)
⇒ P(BIA) = P(AIB) × P(B) / P(A)
Bob Cousins, CMS, 2008

Example 1

P(data|theory) != P(theory|data)

Theory = (male or female)

Data = (pregnant | not pregnant)

P(pregnant | female) ~ 3%

BUT

P(female | pregnant) >99%

Example 2

Higgs search

Expected bg = 0.1

Expected signal = 10

P(N| no Higgs) = 0.1

P(N| Higgs) = 10.1

What is P(Higgs|N=8)?

$$P(H|N=8) = \frac{P(N=8|H)P(H)}{P(N=8)}$$

Depends on P(H)!

(K Cranmer)

Parameter estimation

<u>Bayesian parameter estimation:</u>

Want to know the probability that some parameter θ is in some range $[\theta_0, \theta_1]$

- or -

Want to find a range $[\theta_0, \theta_1]$ that has probability of 0.95

Parameter estimation

Bayesian parameter estimatio

Want to know the probab

parameter θ is in som

- or -

Want t $\theta_0,\theta_1]$ that has probability

of

Remember:

Probability reflects lack of knowledge, and prior information.

How?

The probability that the true value is inside an interal is:

$$1 - \alpha = \int_{\theta_{lo}}^{\theta_{hi}} p(\theta|x) d\theta$$

For lower or upper limits, choose zero or infinity as boundaries.
where we integrate out the nuisance parameters:

$$p(\theta|x) = \int d\nu p(\theta, \nu|x)$$

where

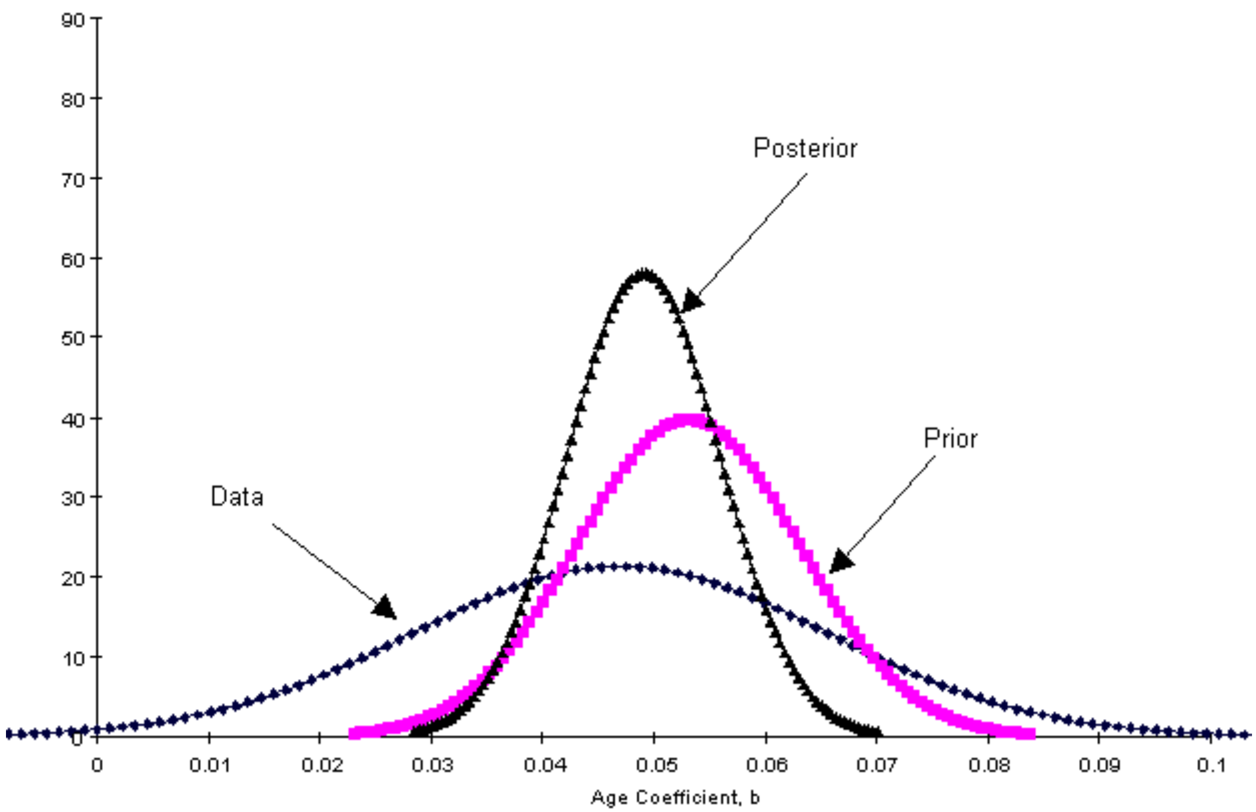
$$p(\theta, \nu|x) = \frac{p(x|\theta, \nu) p(\theta, \nu)}{p(x)}$$

These integrals can be very hard to do if the space is high dimensional.

Priors

<u>Choice of prior p(θ)</u>

- important but subjective choice



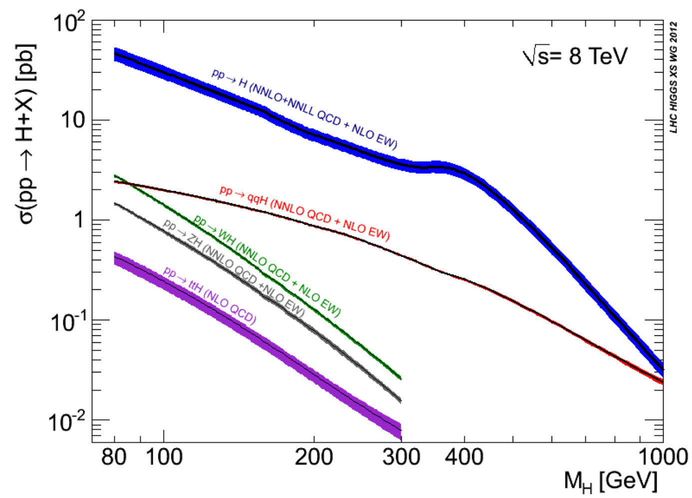
Priors

Choice of prior $p(\theta)$

- Example: measuring Higgs cross-section
- Want to be unbiased: choose uniform prior?

$$\sigma = [0, \Lambda] \rightarrow \; P = k$$

- But σ and mass relationship makes this prior **not flat in mass**



- no uninformative prior across all transformations

Parameter estimation

<u>Frequentist parameter estimation:</u>

Want to know in what fraction of experiments the true value of some parameter θ is in that experiments range $[\theta_{0i},\theta_{1i}]$

- or -

Want a range-finding strategy such that $[\theta_0,\theta_1]$ contains the true value in 95/100 experiments.

Parameter estimation

Frequentist parameter estimation:

Want to know in what fraction of experiments the true value of some parameter θ is in that experiments range $[\theta_{0i},\theta_{1i}]$

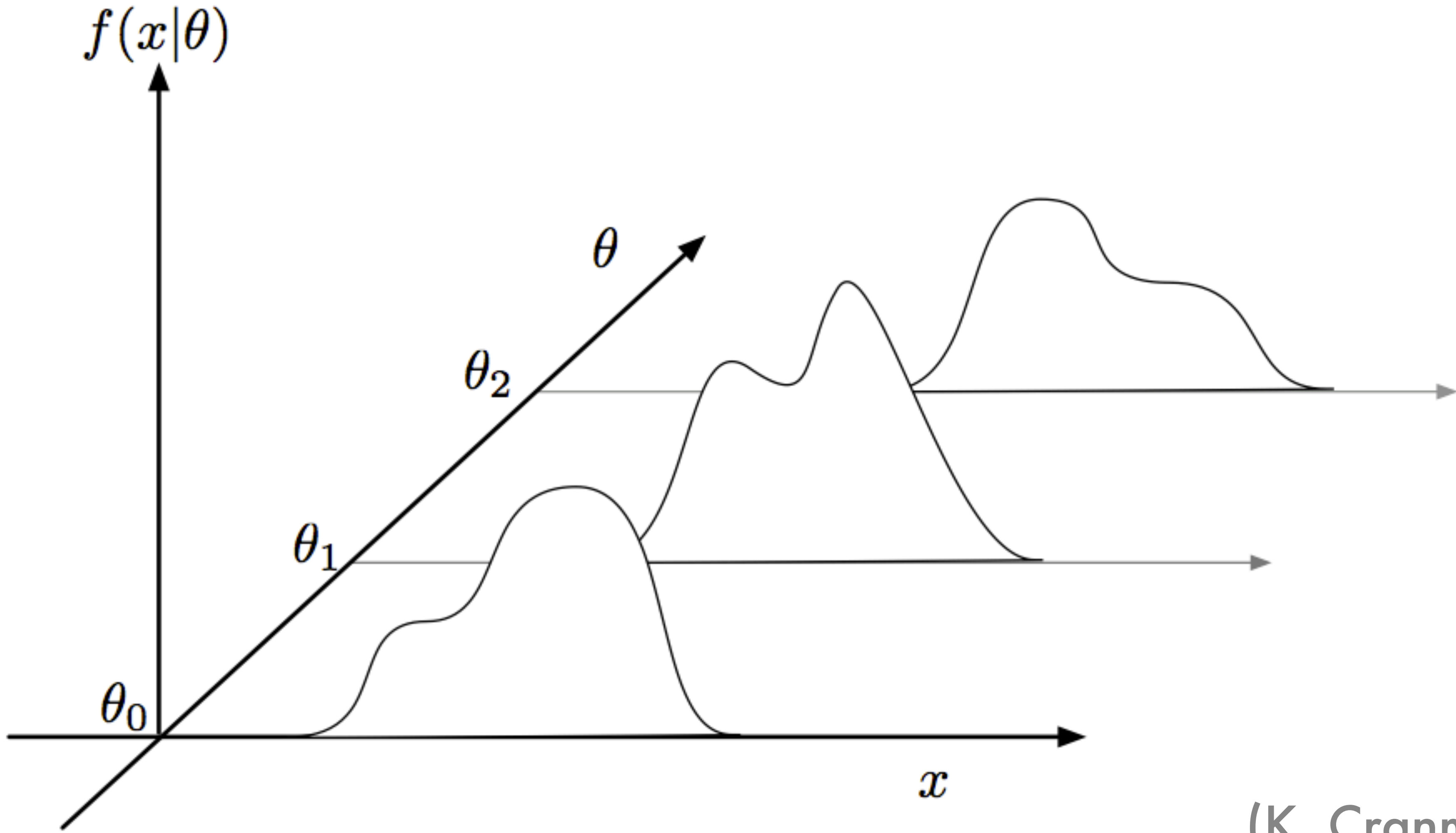
Different for every experiment

- or -

Want a range-finding strategy such that $[\theta_0,\theta_1]$ contains the true value in 95/100 experiments.

Neyman Construction

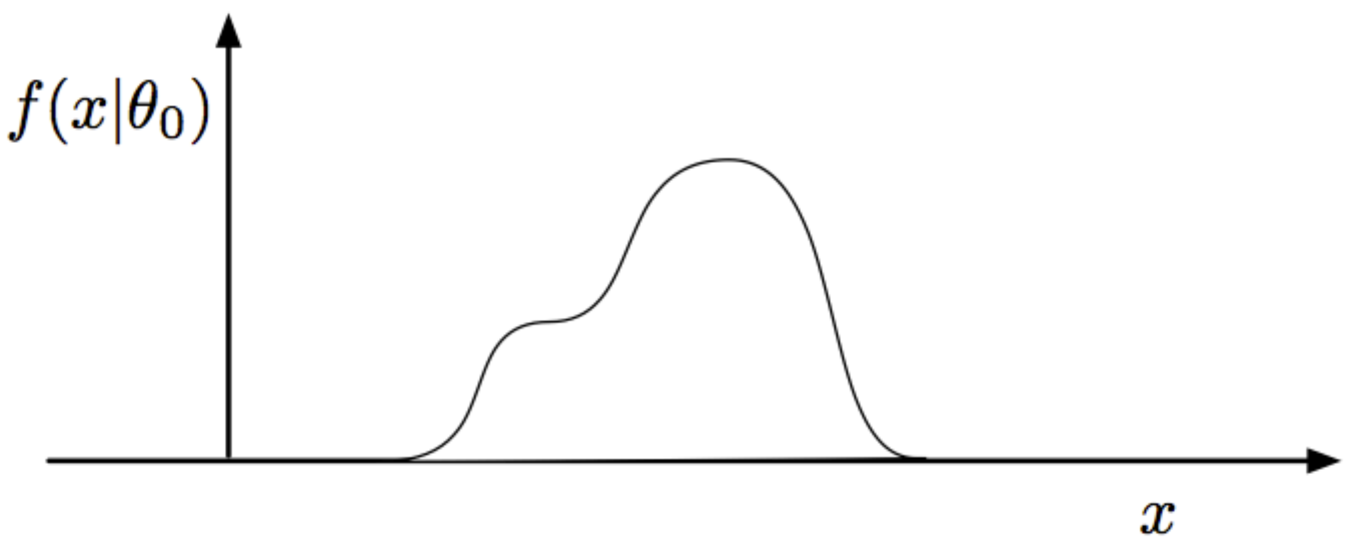
For each value of θ consider $f(x|\theta)$



(K. Cranmer)

Neyman Construction

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

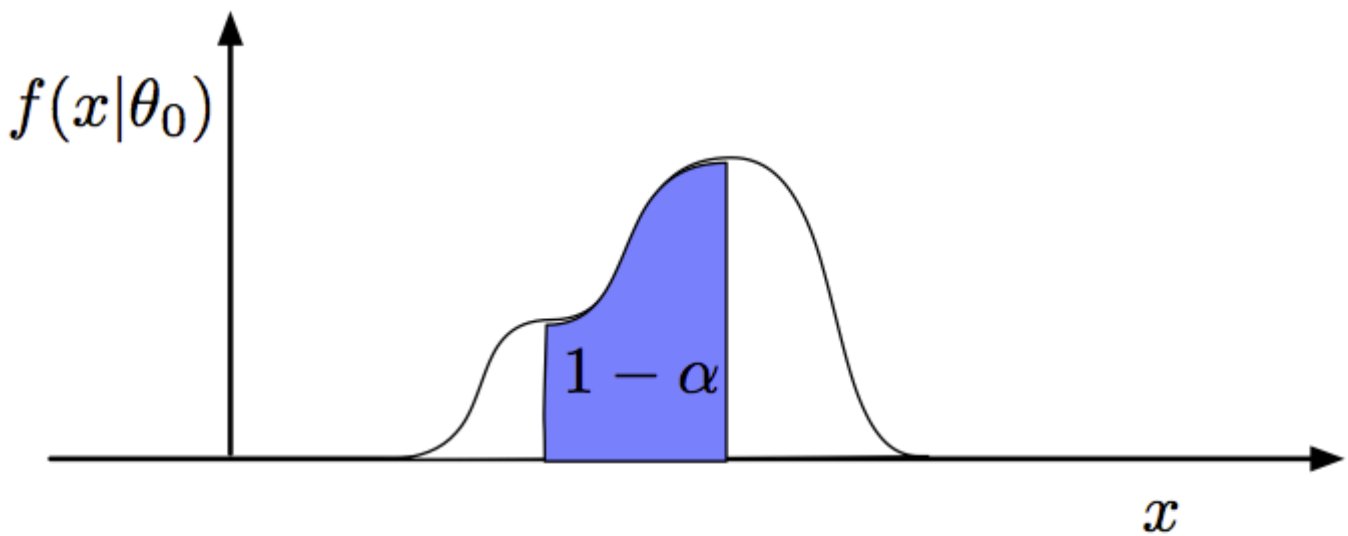


(K. Cranmer)

Neyman Construction

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



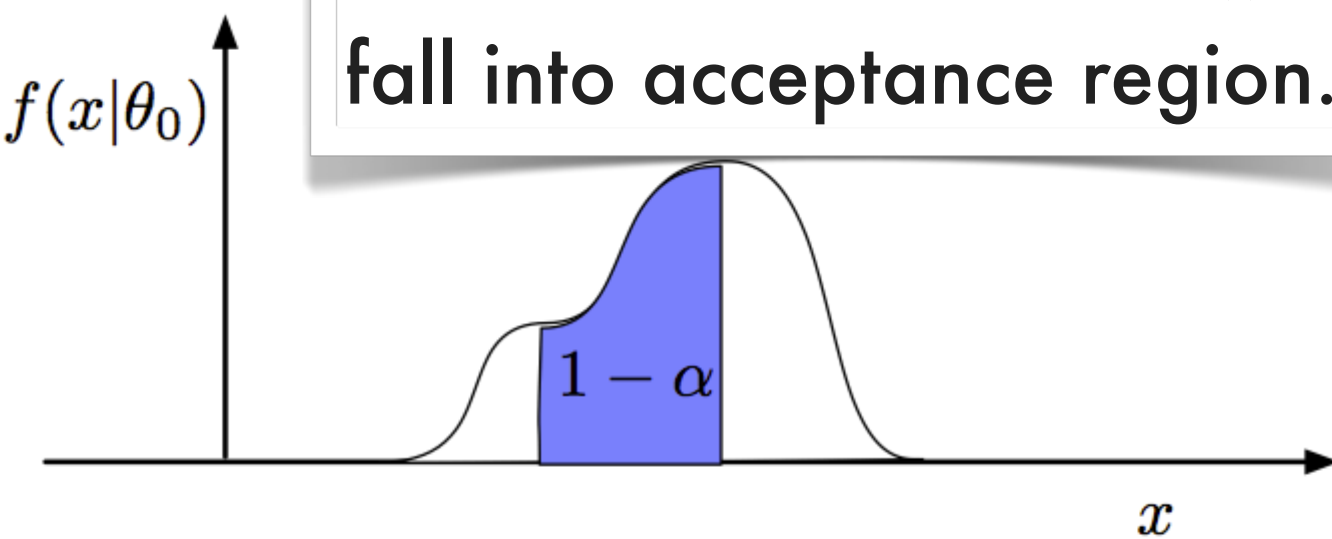
(K. Cranmer)

Neyman Construction

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

- we want a test of size α
- equivale ... nterval on θ
- so we fi ... $-\ \alpha$ probability

Constructed to satisfy requirement that if $\theta = \theta_0$ then $1-\alpha$ measurements will fall into acceptance region.

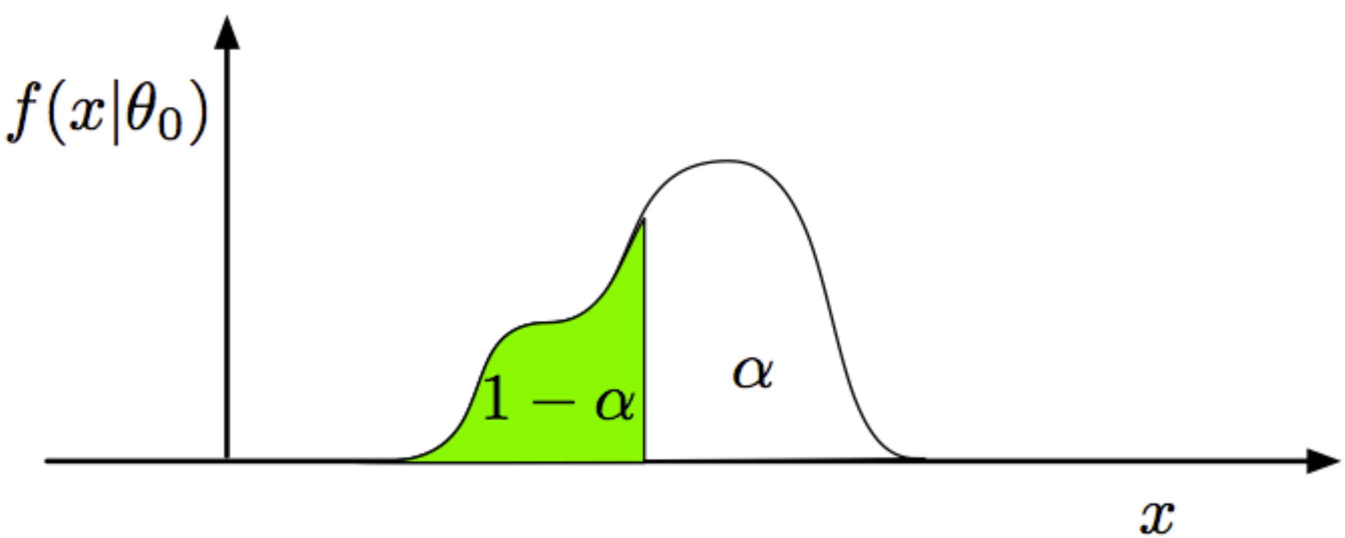


(K. Cranmer)

Neyman Construction

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

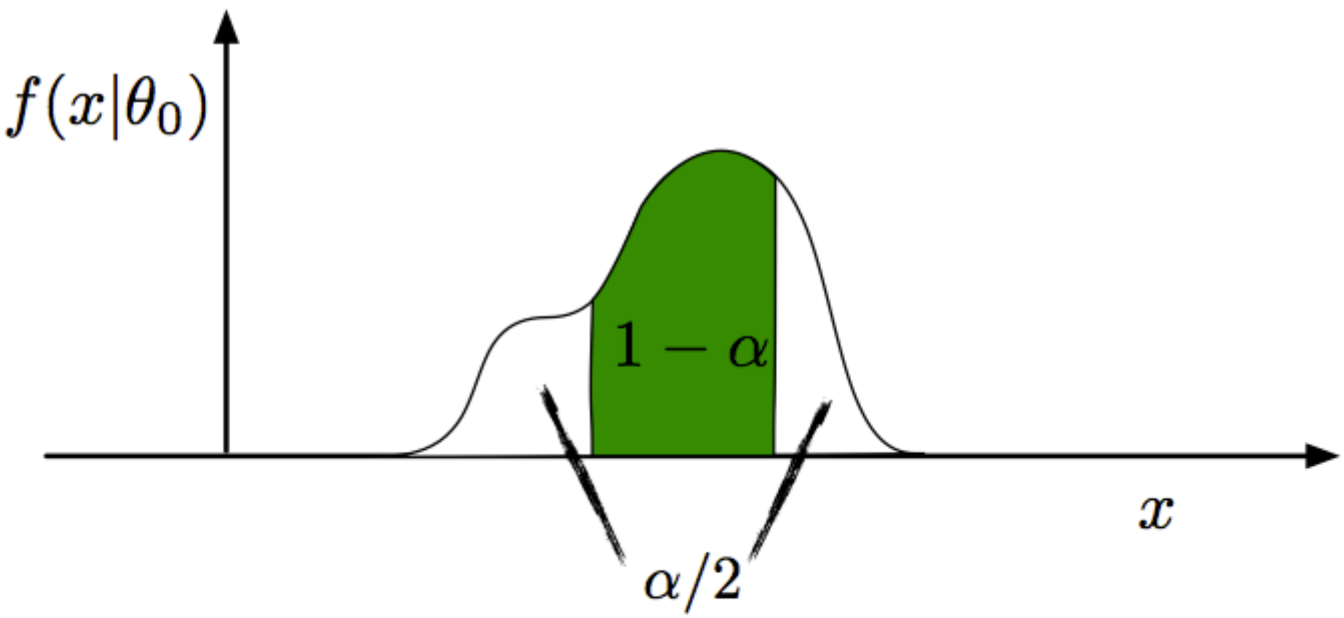


(K. Cranmer)

Neyman Construction

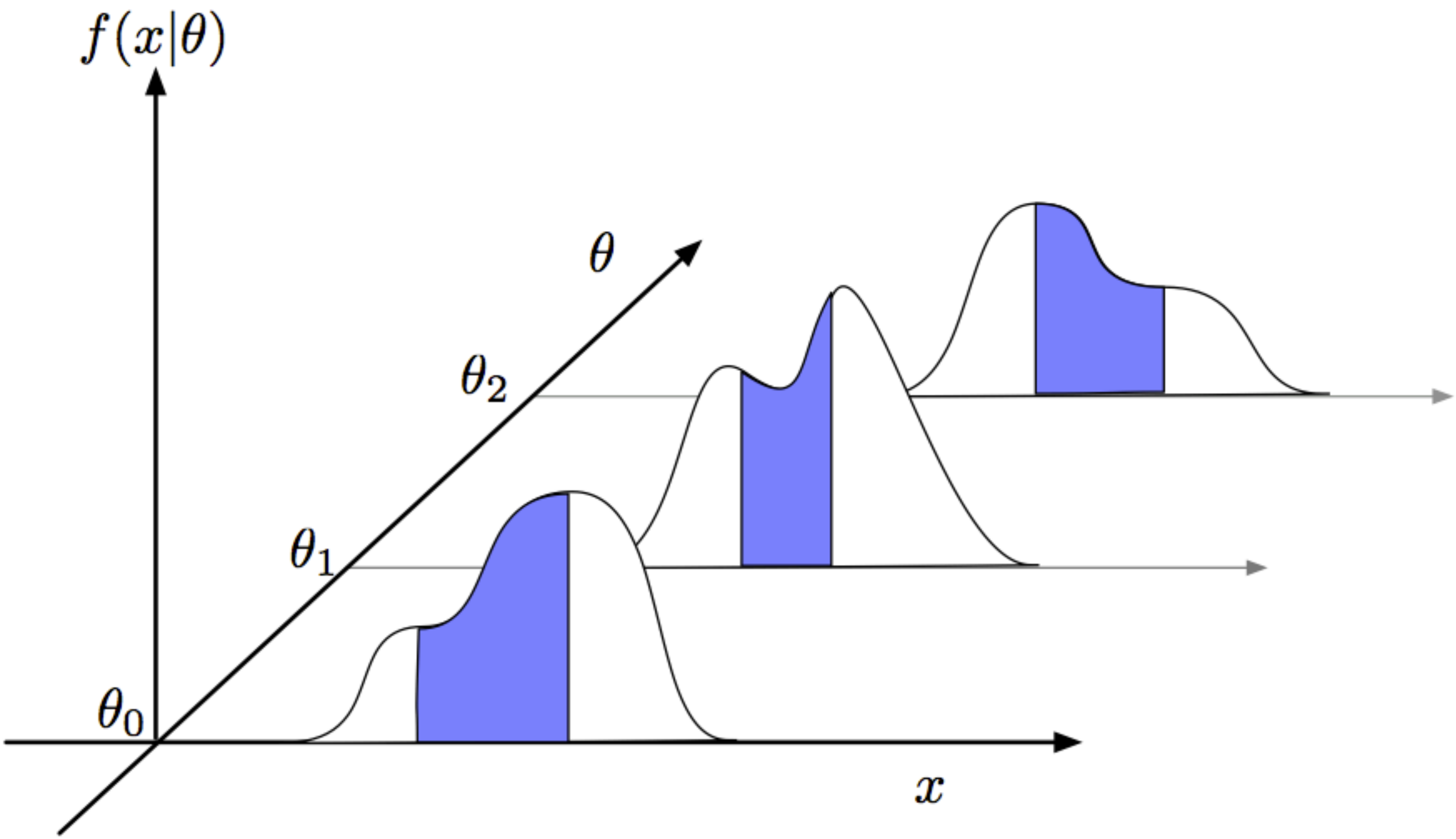
Let's focus on a particular point $f(x|\theta_o)$

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



(K. Cranmer)

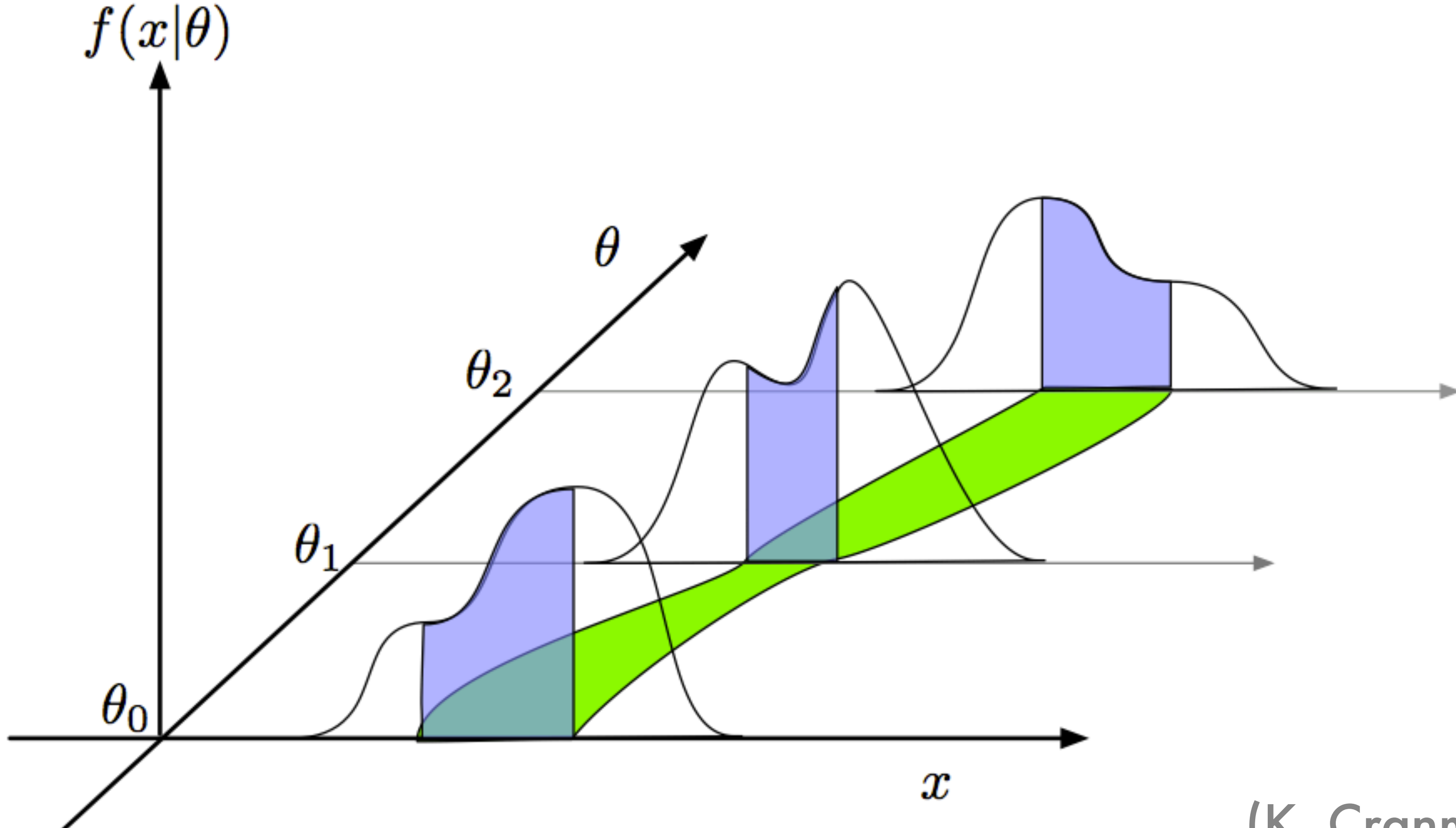
Neyman Construction



(K. Cranmer)

Neyman Construction

This makes a **confidence belt** for θ

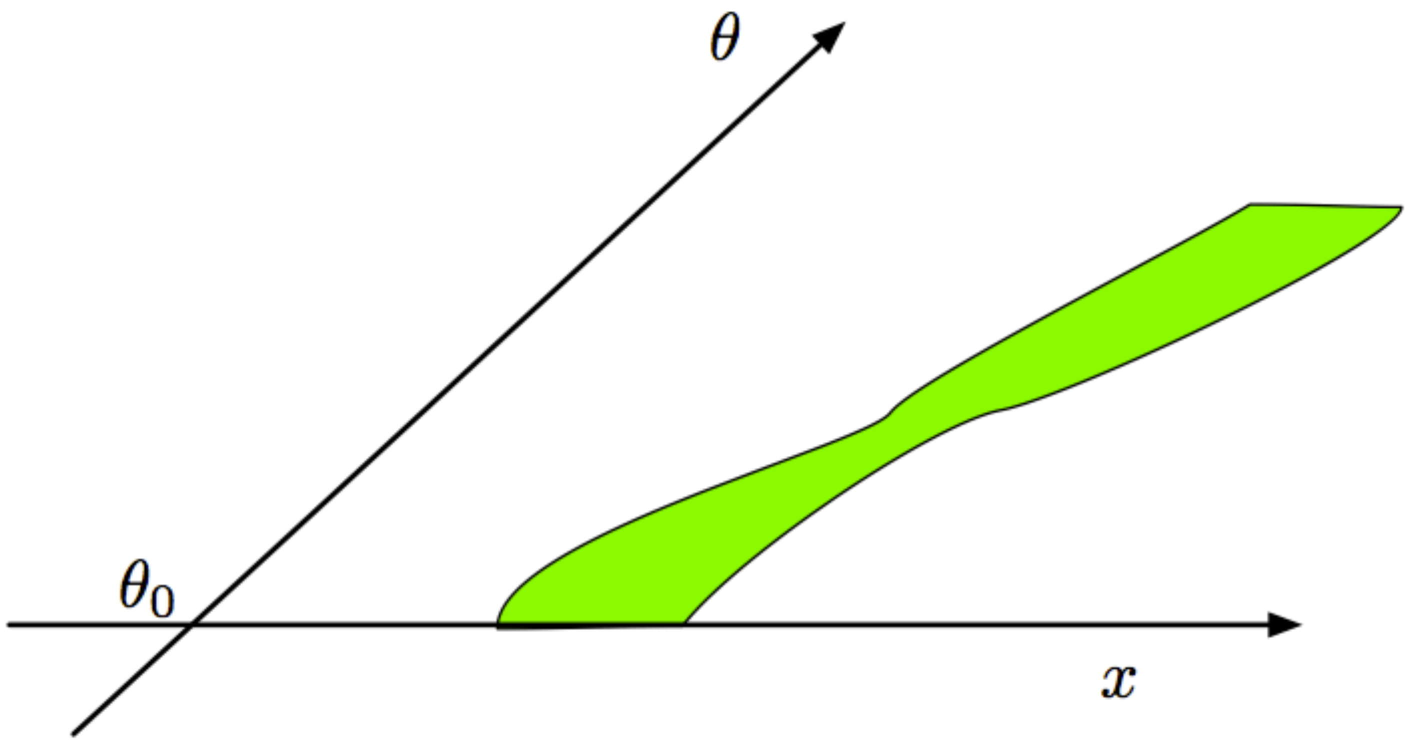


(K. Cranmer)

Neyman Construction

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



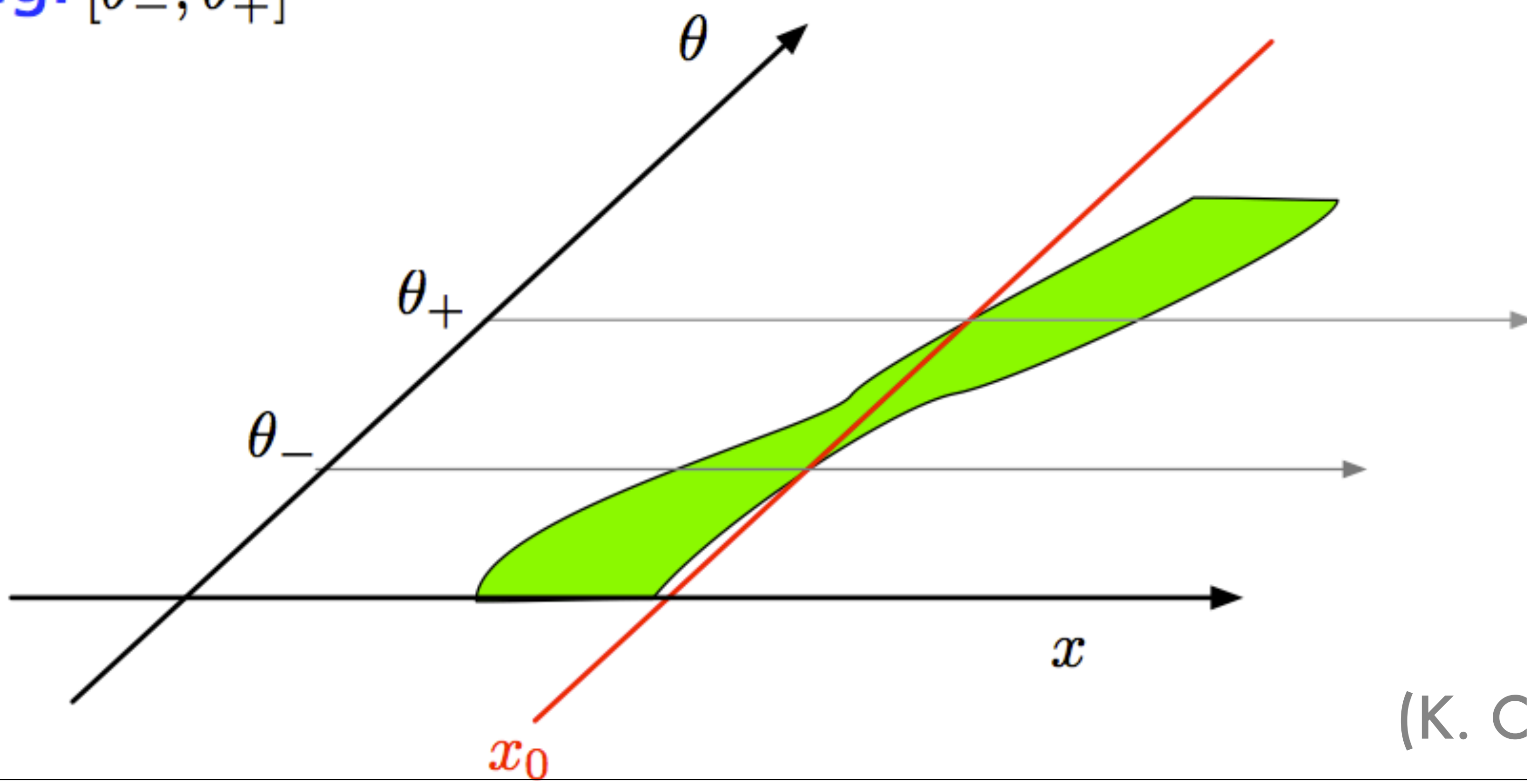
(K. Cranmer)

Neyman Construction

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



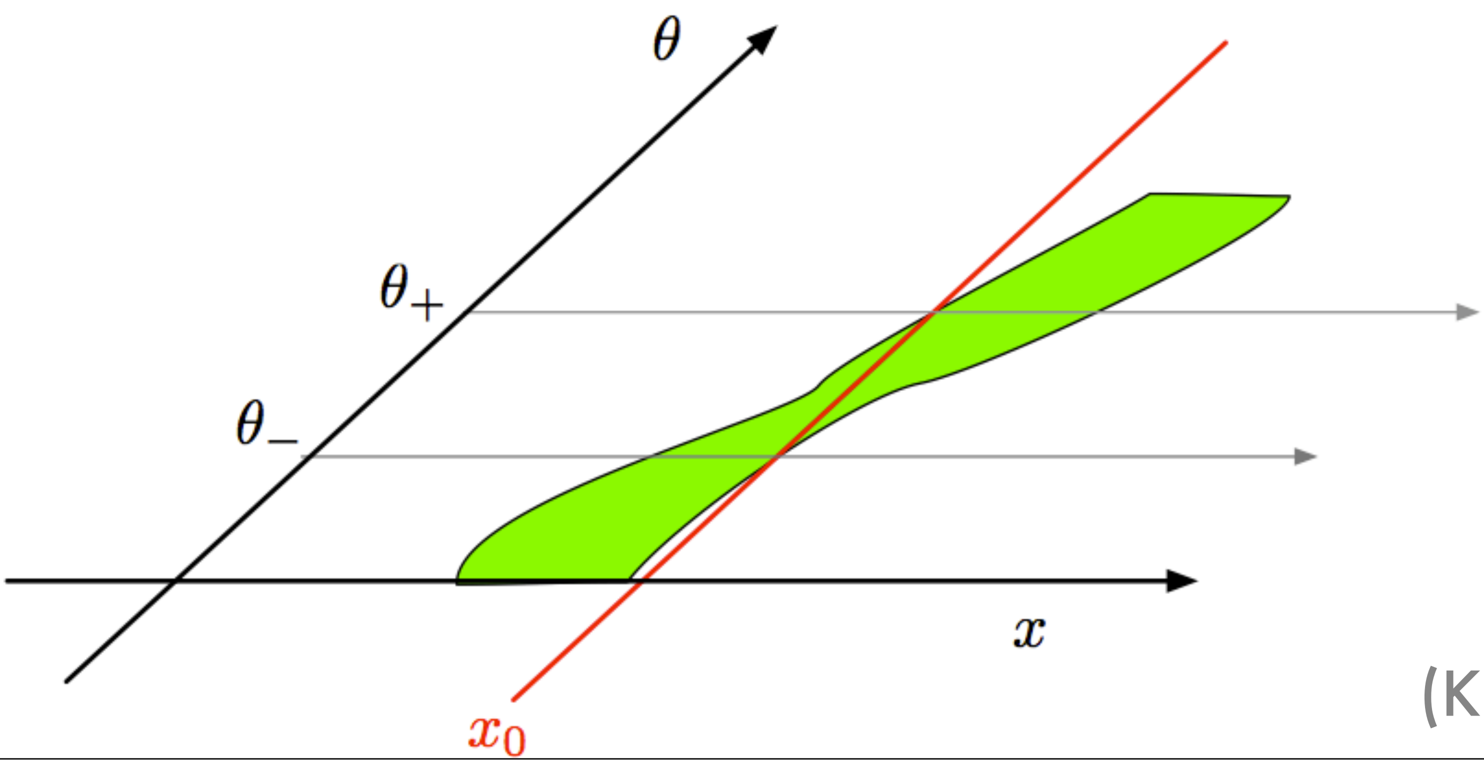
(K. Cranmer)

Neyman Construction

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 $[\theta_-, \theta_+]$

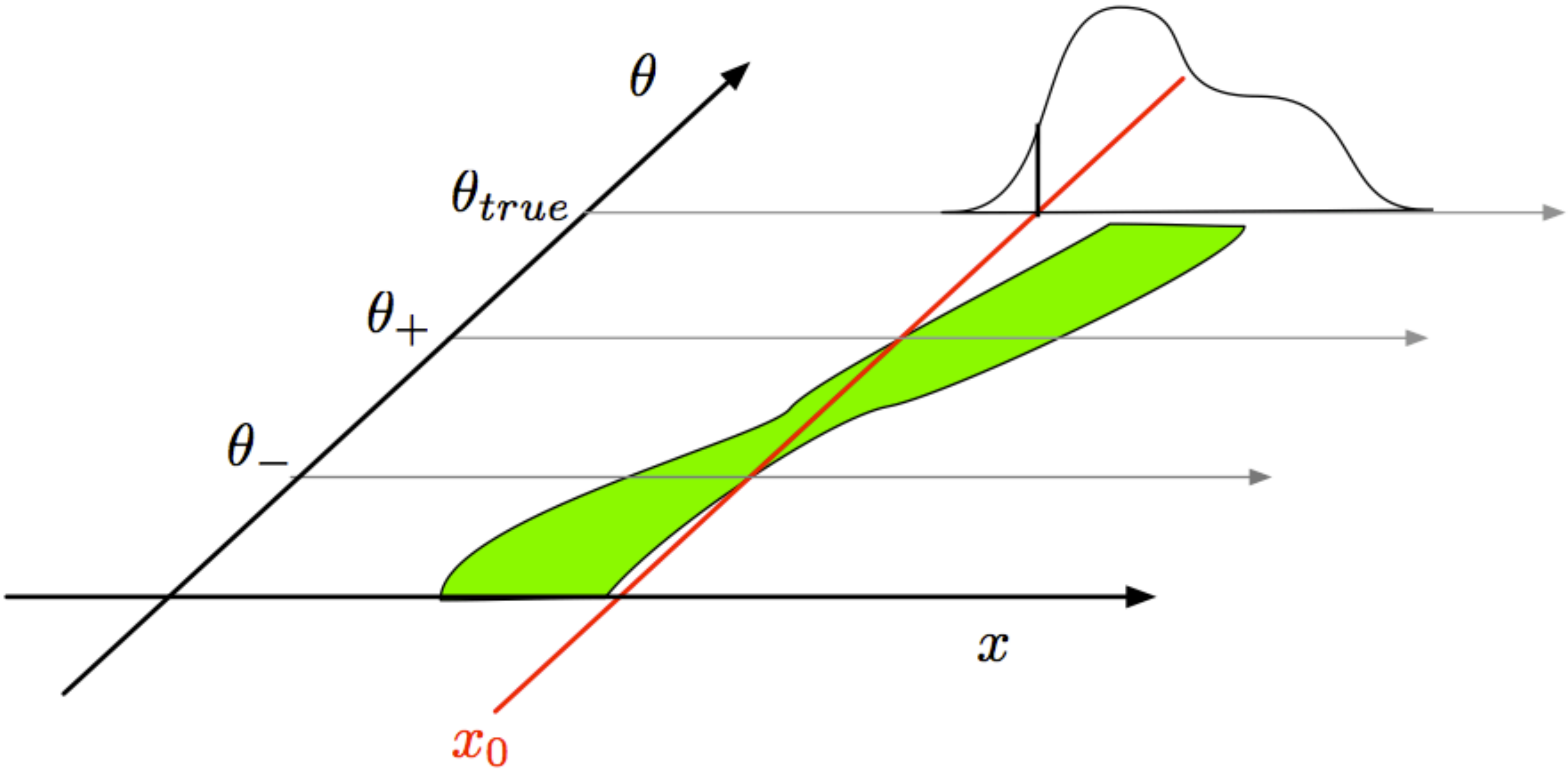
So the interval $[\theta_-, \theta_+]$ covers the true value with probability $1-\alpha$



(K. Cranmer)

[redacted]

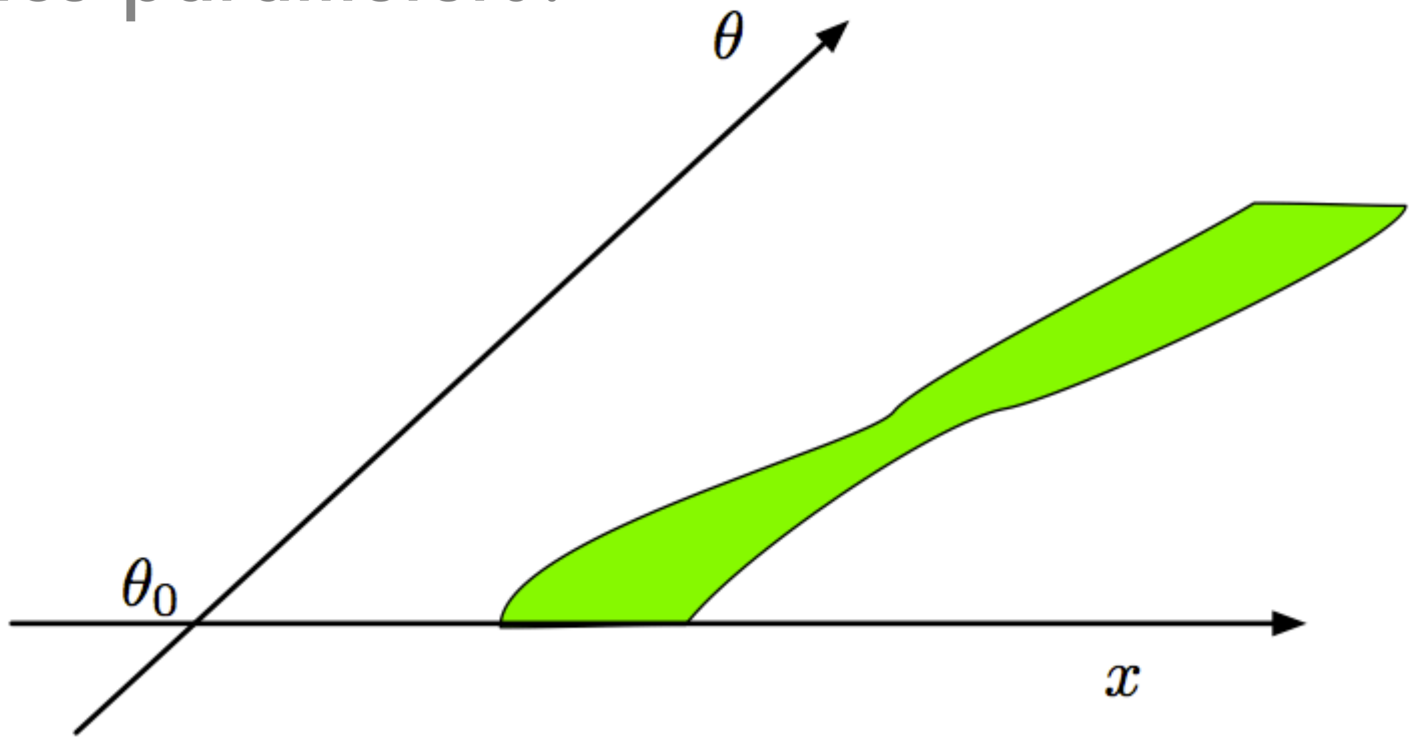
This is not Bayesian... it doesn't mean the probability that the true value of θ is in the interval is $1-\alpha$!



(K. Cranmer)

Nuisance Params

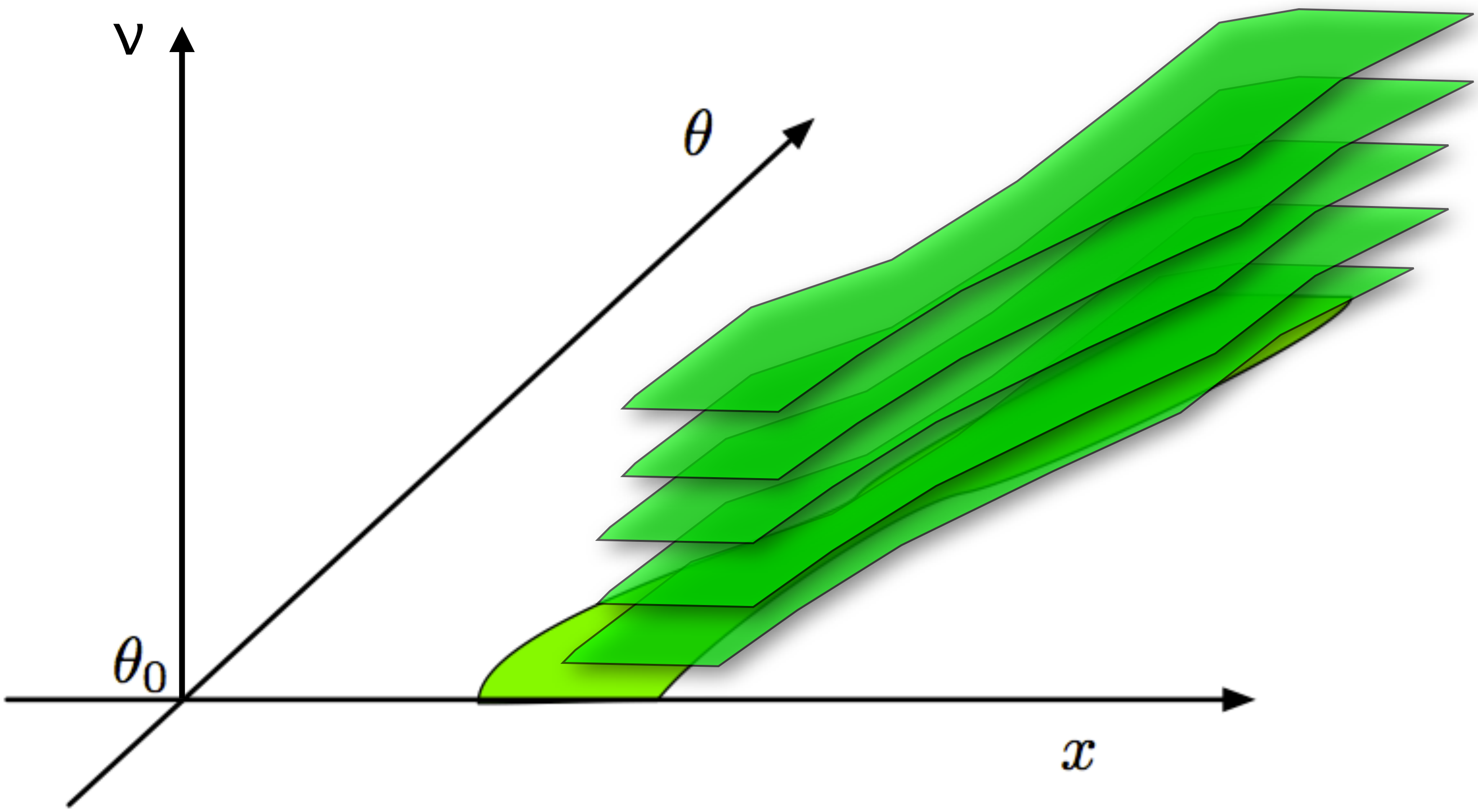
How do we handle nuisance parameters?



(K. Cranmer)

Nuisance Params

Make acceptance regions for each value



asymptotic approximation

After a close look at the profile likelihood ratio

$$\lambda(\mu=0) = \frac{L(data|\mu=0, \hat{\hat{b}}(\mu=0), \hat{\hat{\nu}}(\mu=0))}{L(data|\hat{\mu}, \hat{b}, \hat{\nu})},$$

one can see the function is independent of true values of ν

- though its distribution might depend indirectly

Wilks's theorem states that under certain conditions the distribution of the profile likelihood ratio has an asymptotic form

$$-2\log\lambda(\mu=0) \sim \chi^2_1$$

Thus, we can calculate the p-value for the background-only hypothesis by calculating

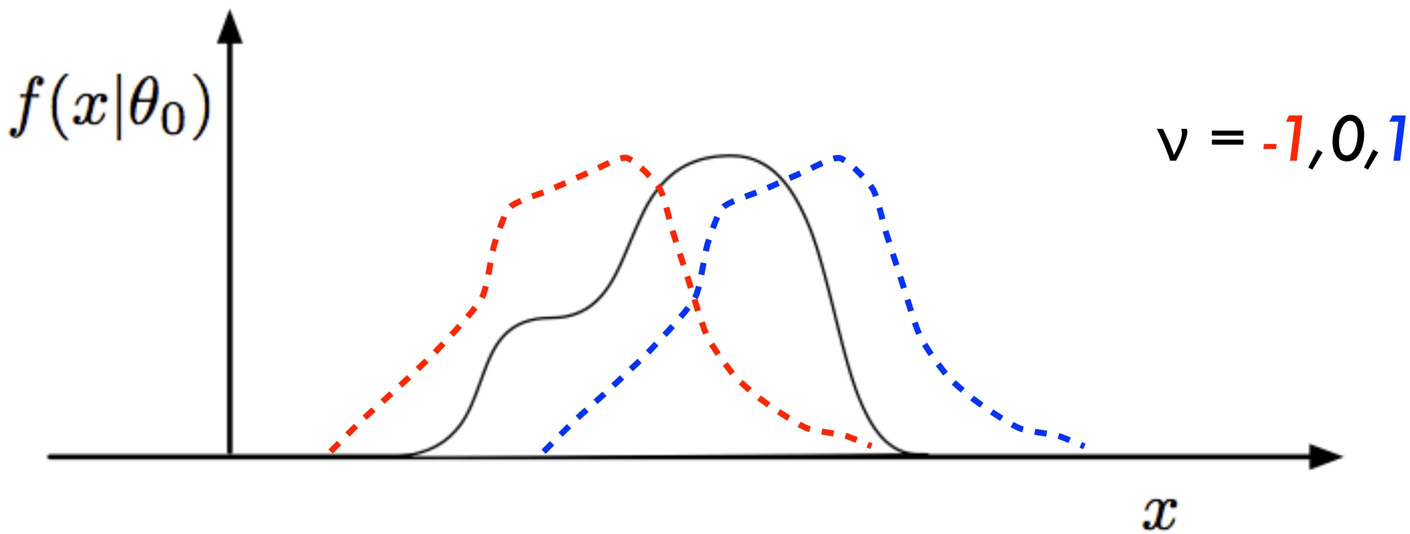
$$-2\log\lambda(\mu=0)$$

or equivalently:

$$Z = \sqrt{-2\log\lambda(\mu=0)}$$

hybrid solutions

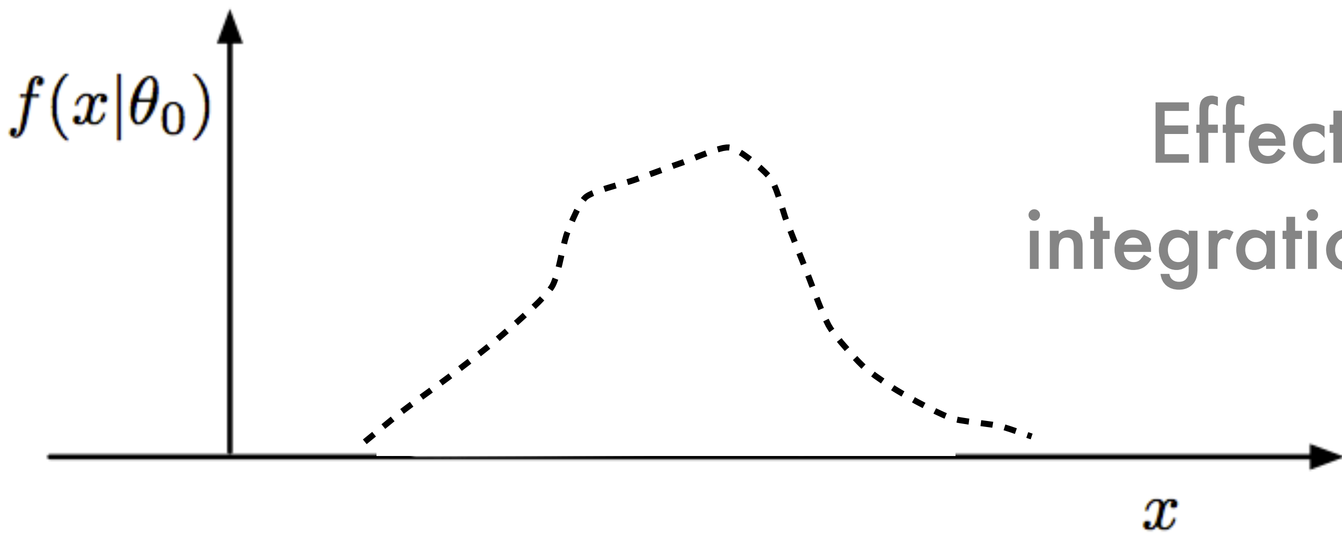
Fold nuisance parameter variation



into pseudo-experiments
used to create acceptance region

hybrid solutions

Fold nuisance parameter variation

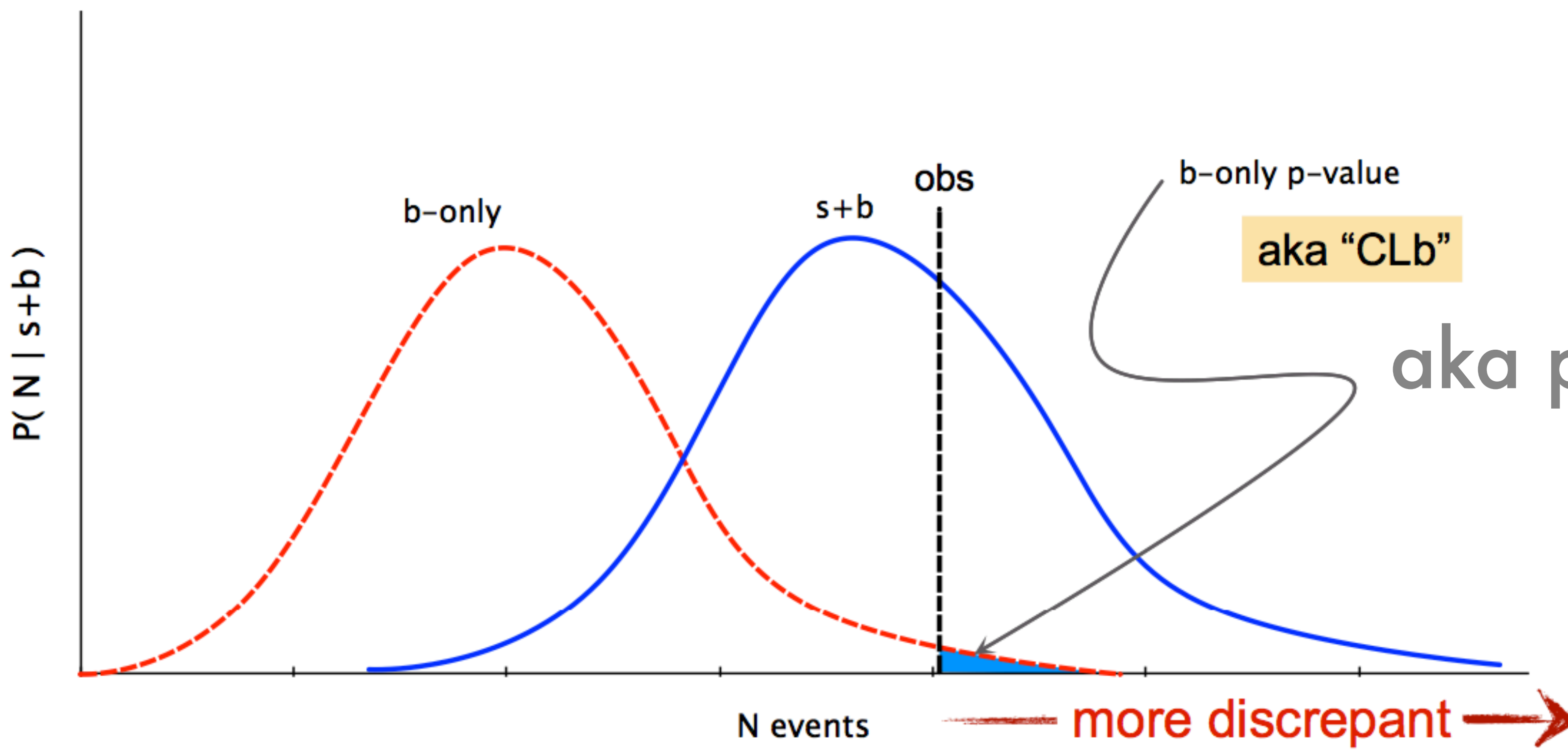


Effective MC integration over NP

Required to specify prior on NP
This is a Bayesian procedure!

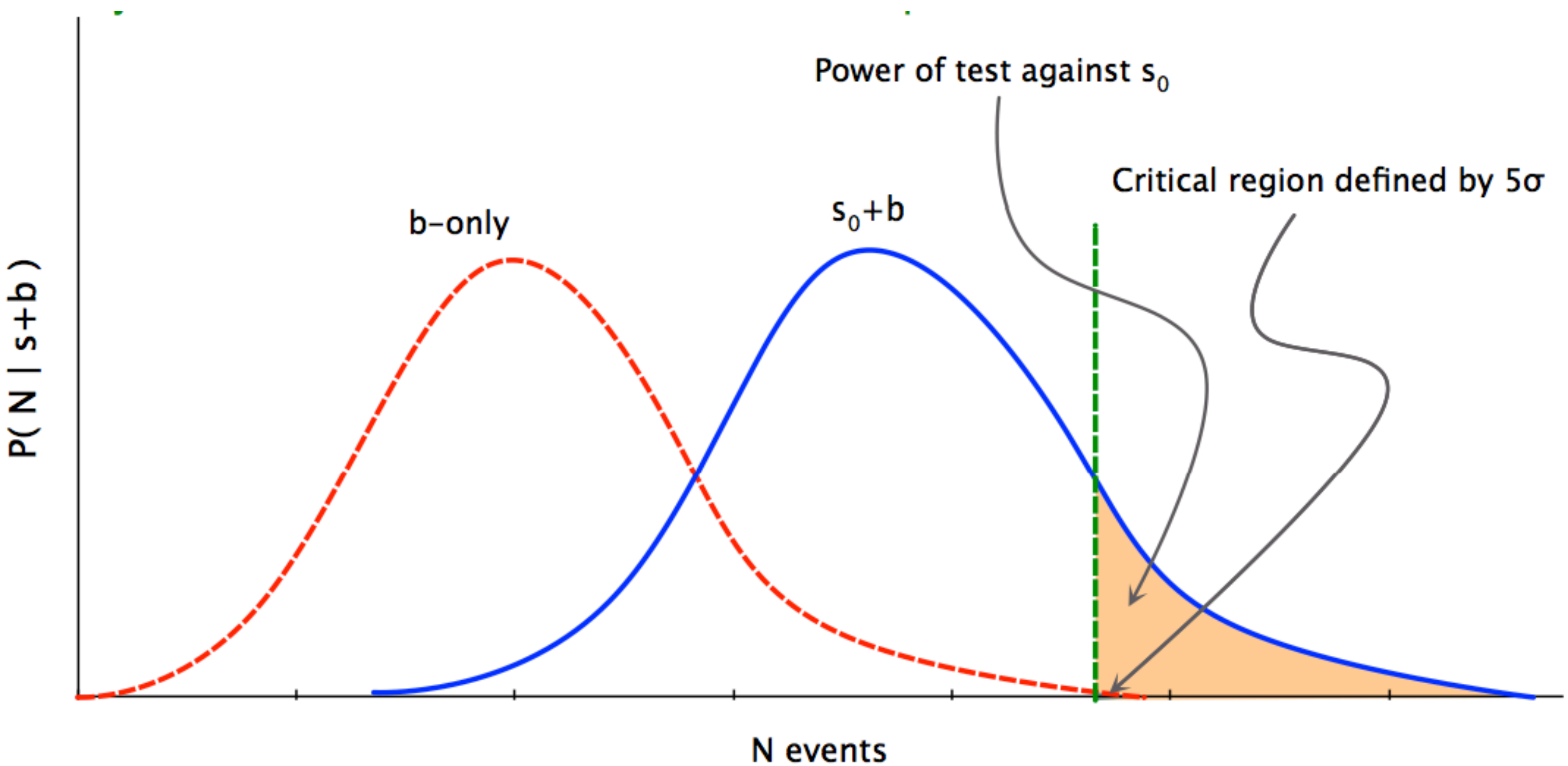
More on p-values

To reject background hypothesis



(K. Cranmer)

Power



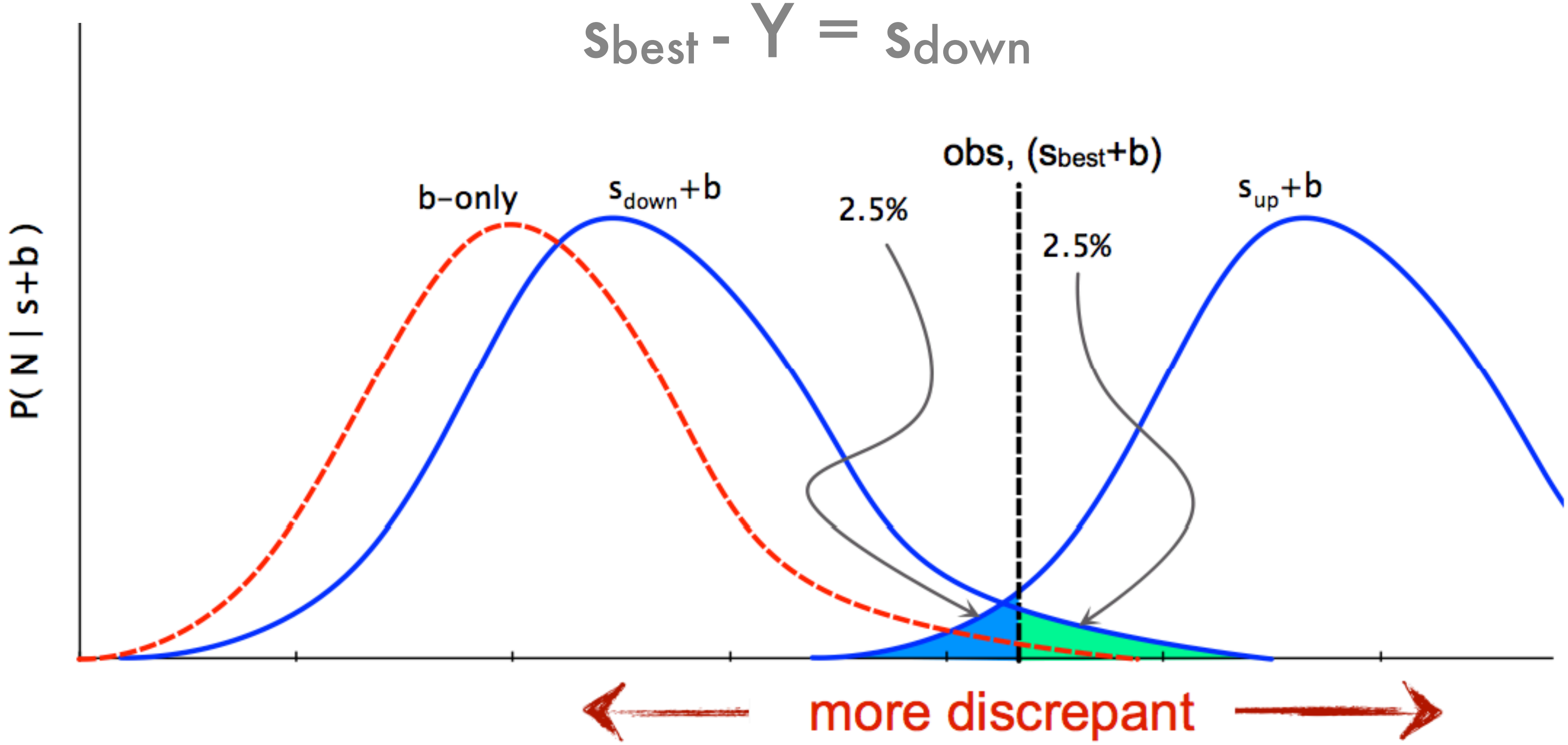
(K. Cranmer)

Measurement

Measure: s_{best} +X - Y (95%=2σ errors)

$s_{best} + X = s_{up}$

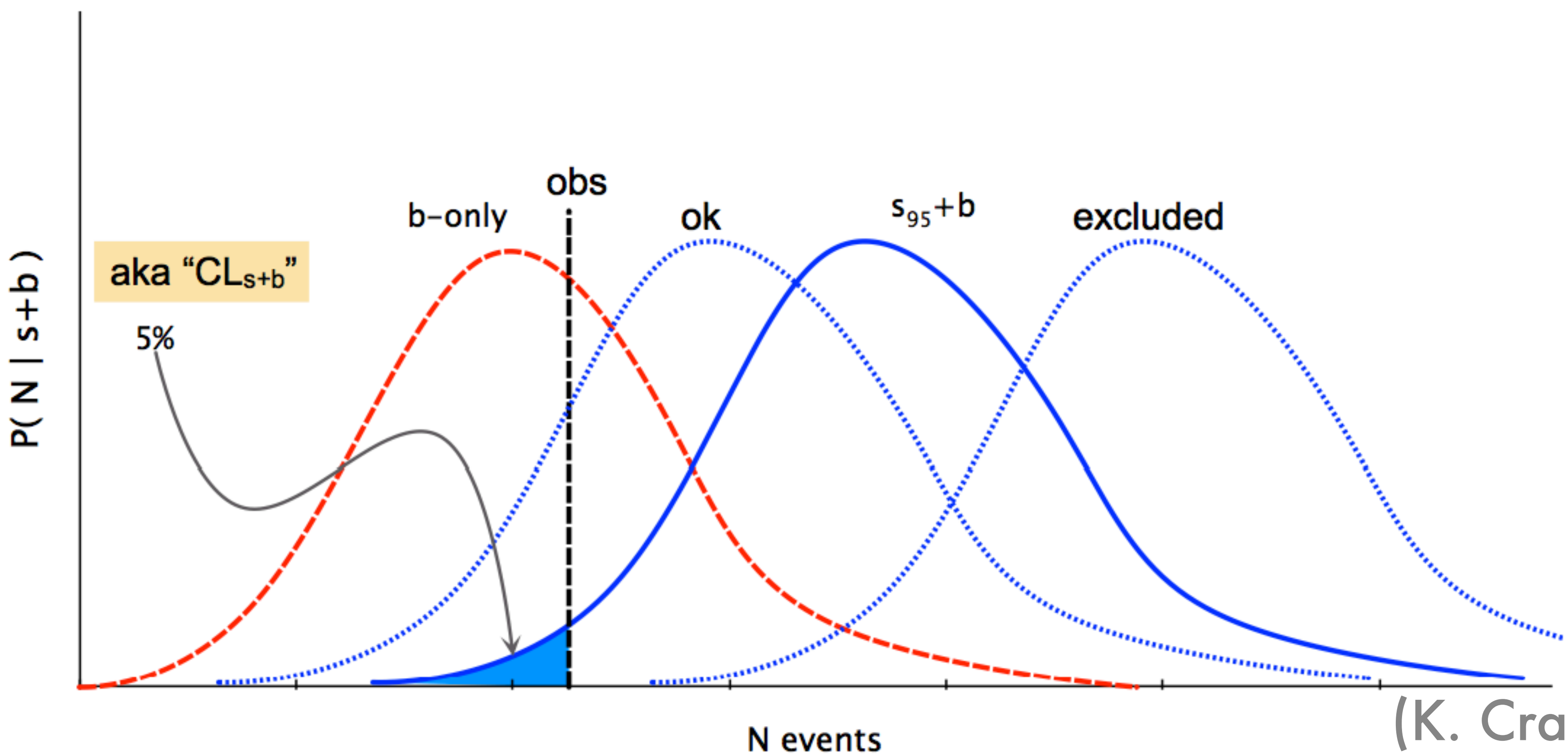
$s_{best} - Y = s_{down}$



(K. Cranmer)

Upper limits

Find value s95 such that CLs+b = 5%

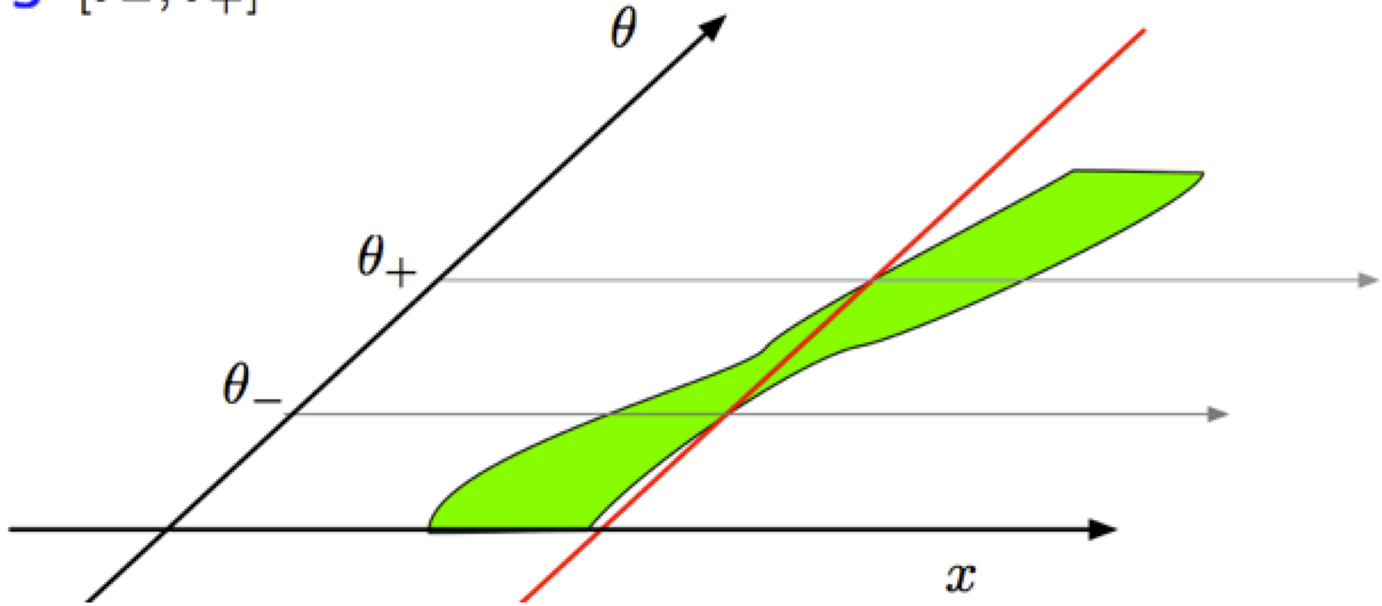


(K. Cranmer)

Neyman construction

Remember this picture

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



Finding s95:

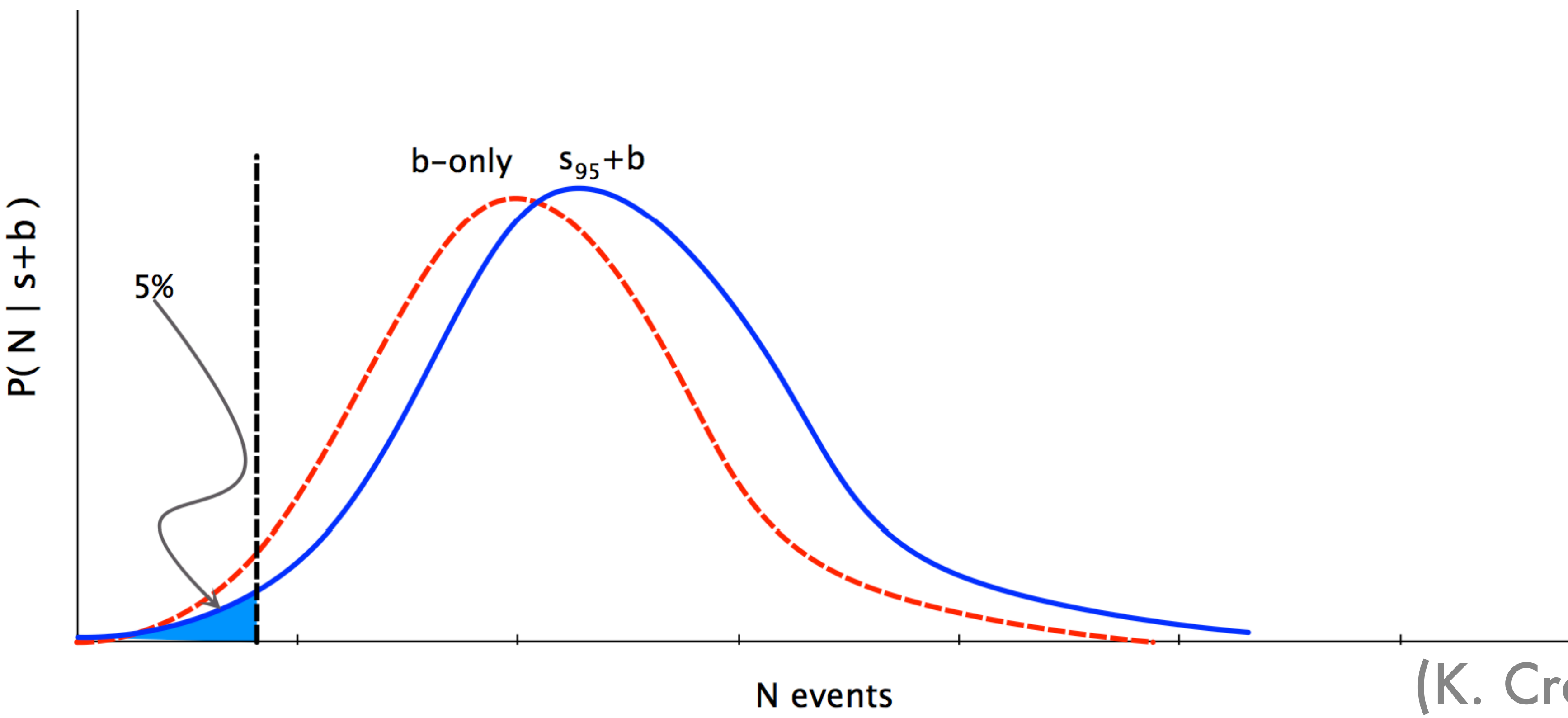
$\theta_- = 0 \quad \theta_+ = s95$

Low power

<u>What happens if</u>

- S+B looks a lot like B
- downward fluctuation

We asked for Type I error= 0.05 that means in 5% of experiments, the interval we get [0,s] **will not contain the true value**



(K. Cranmer)

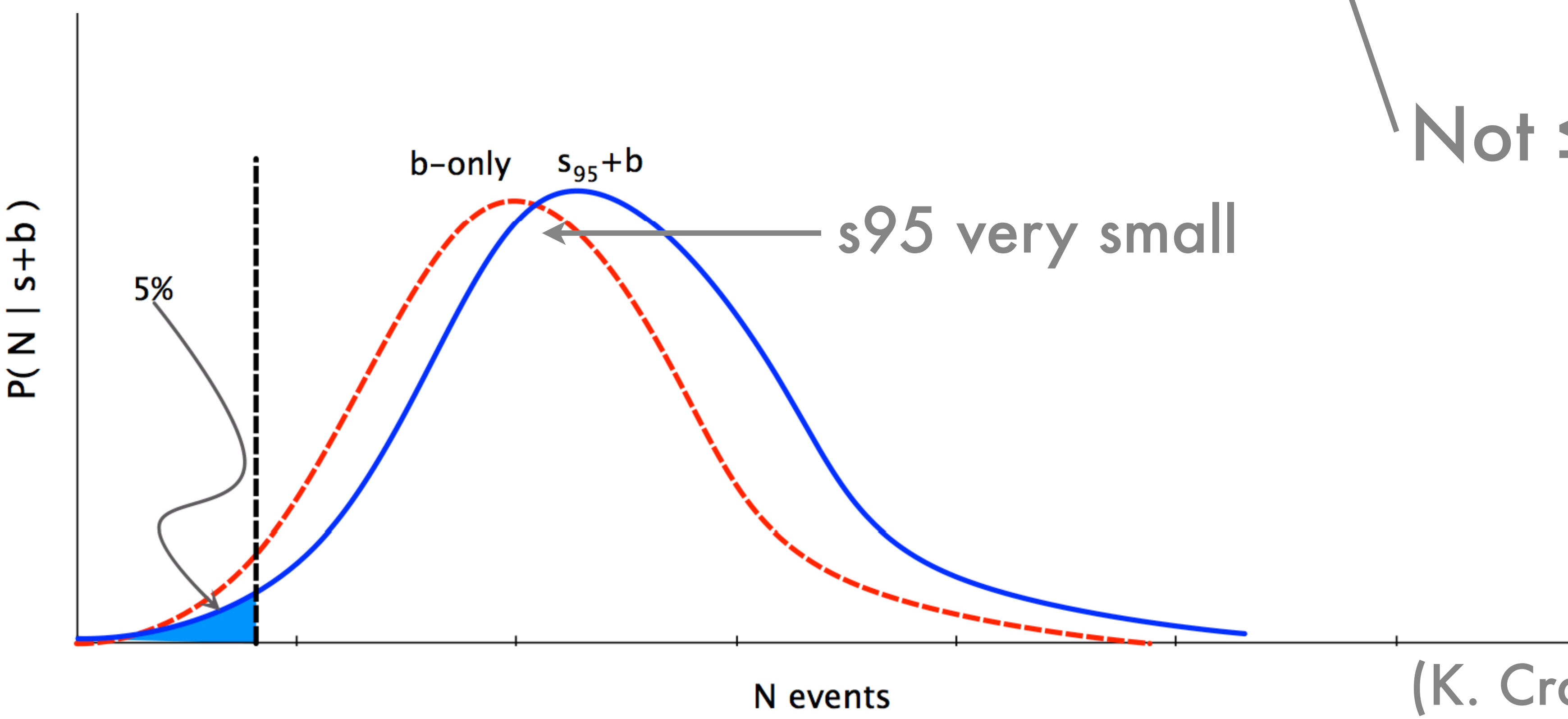
Low power

What happens if
- S+B looks a lot like B
- downward fluctuation

We asked for Type I error= 0.05 that means in 5% of experiments, the interval we get [0,s] will not contain the true value

Not ≤ 5%

s95 very small



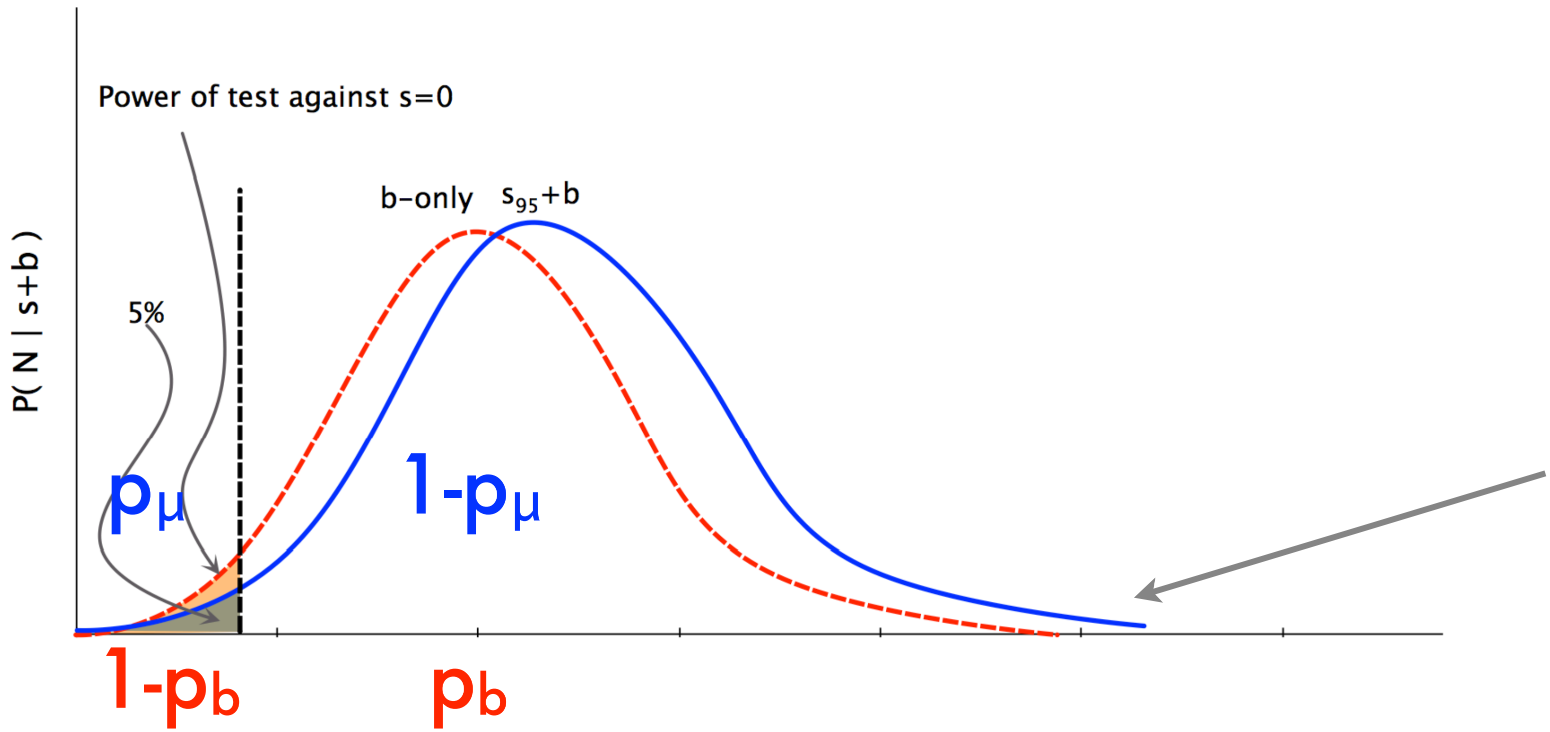
(K. Cranmer)

CLs

CLs = p_μ/ $1-p_b$

Exclude at 95% if CLs < 0.05 *(CLs not a prob)*

weaken if p_b is large ($1-p_b$ is small)



(K. Cranmer)

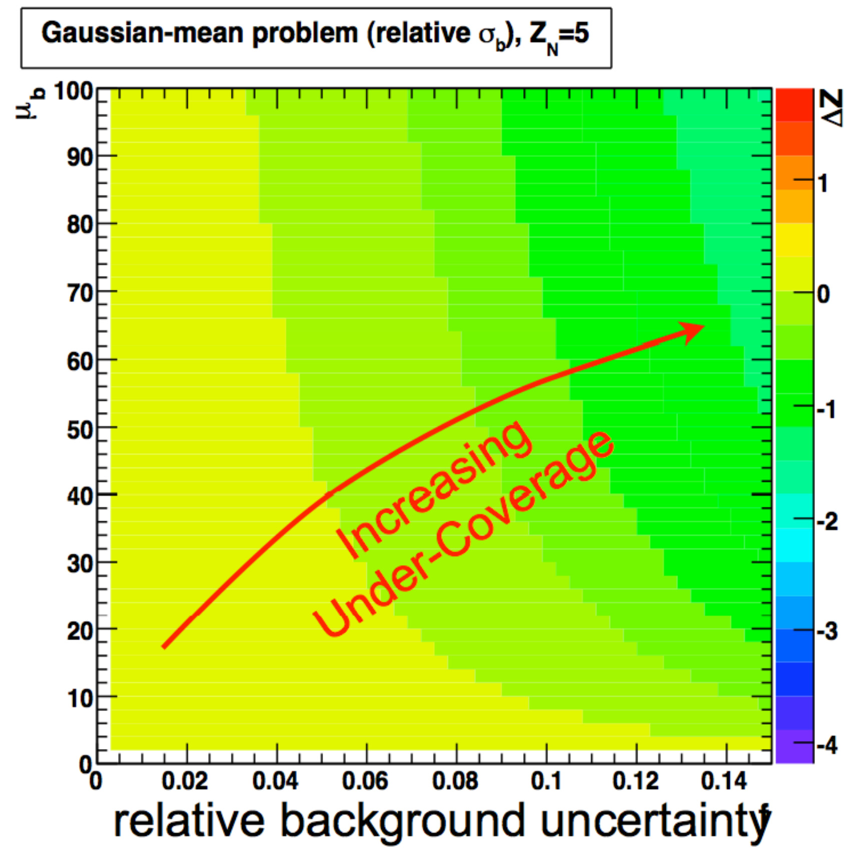
coverage

Expect 5σ to mean specific Type I error α

Expect 95%CL intervals
to have α=0.05

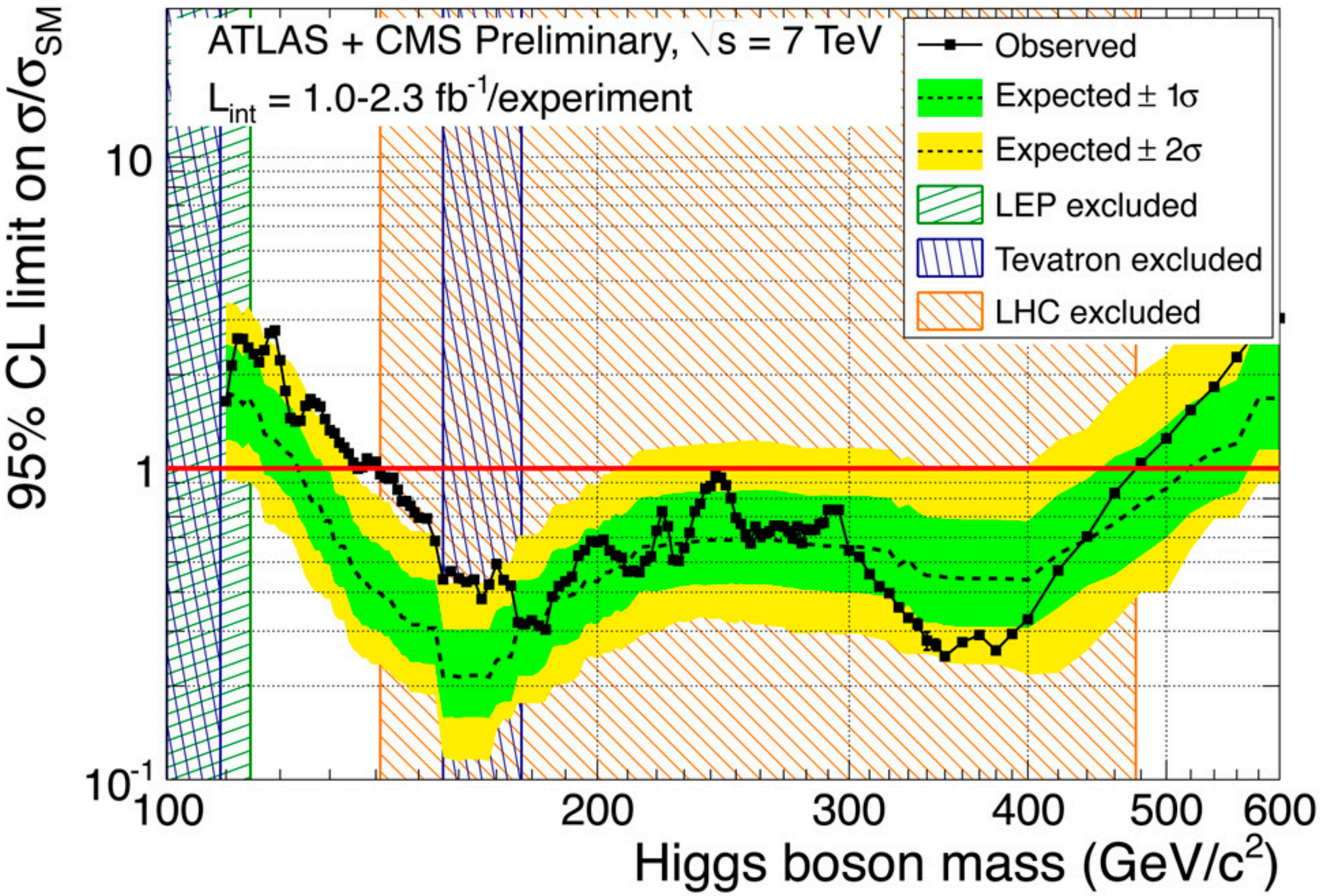
Go measure it, make sure it is.

Coverage is a calibration of
your statistical tools.

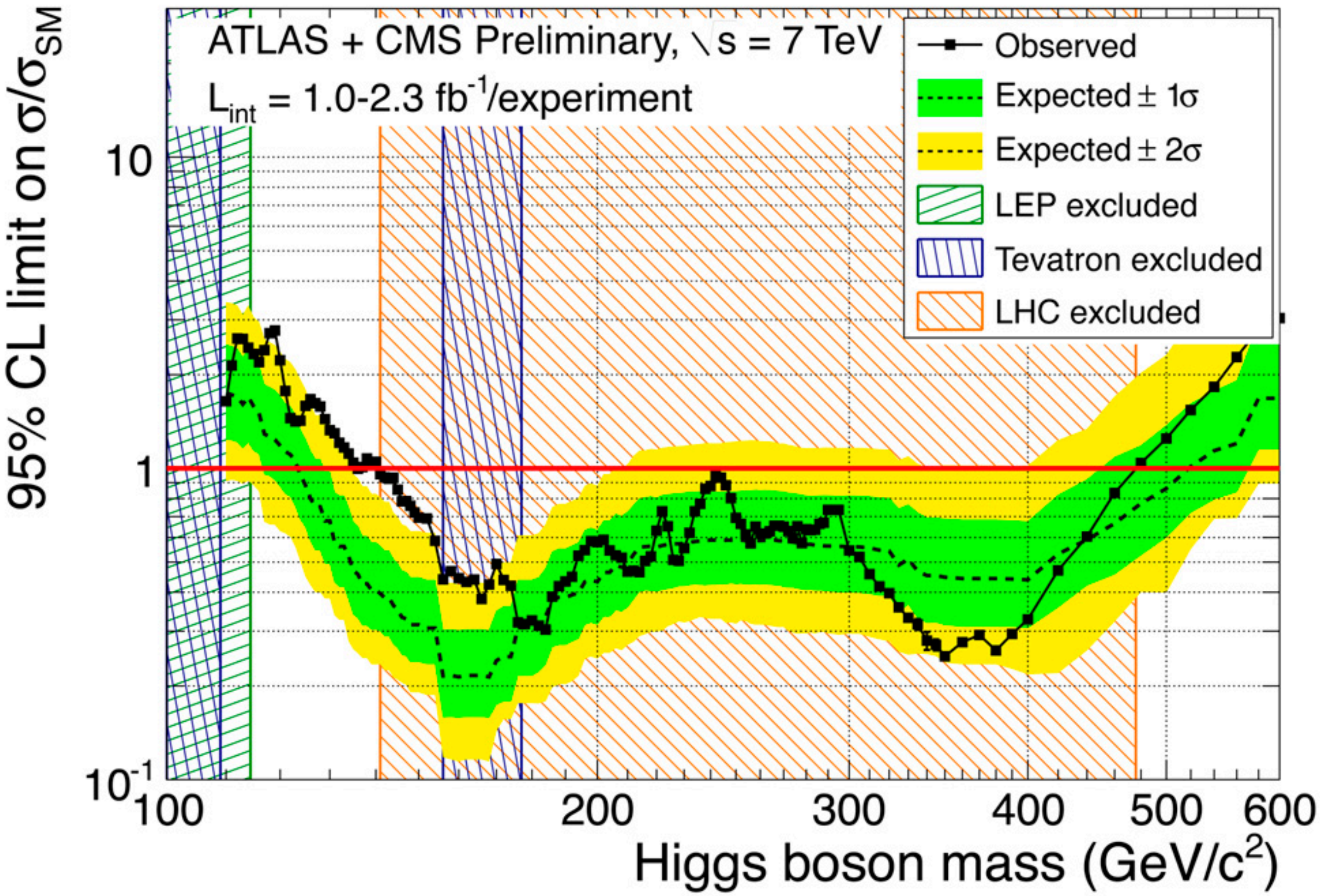


Recent work by Bob Cousins & Jordan Tucker, [physics/0702156]

What does my interval mean?

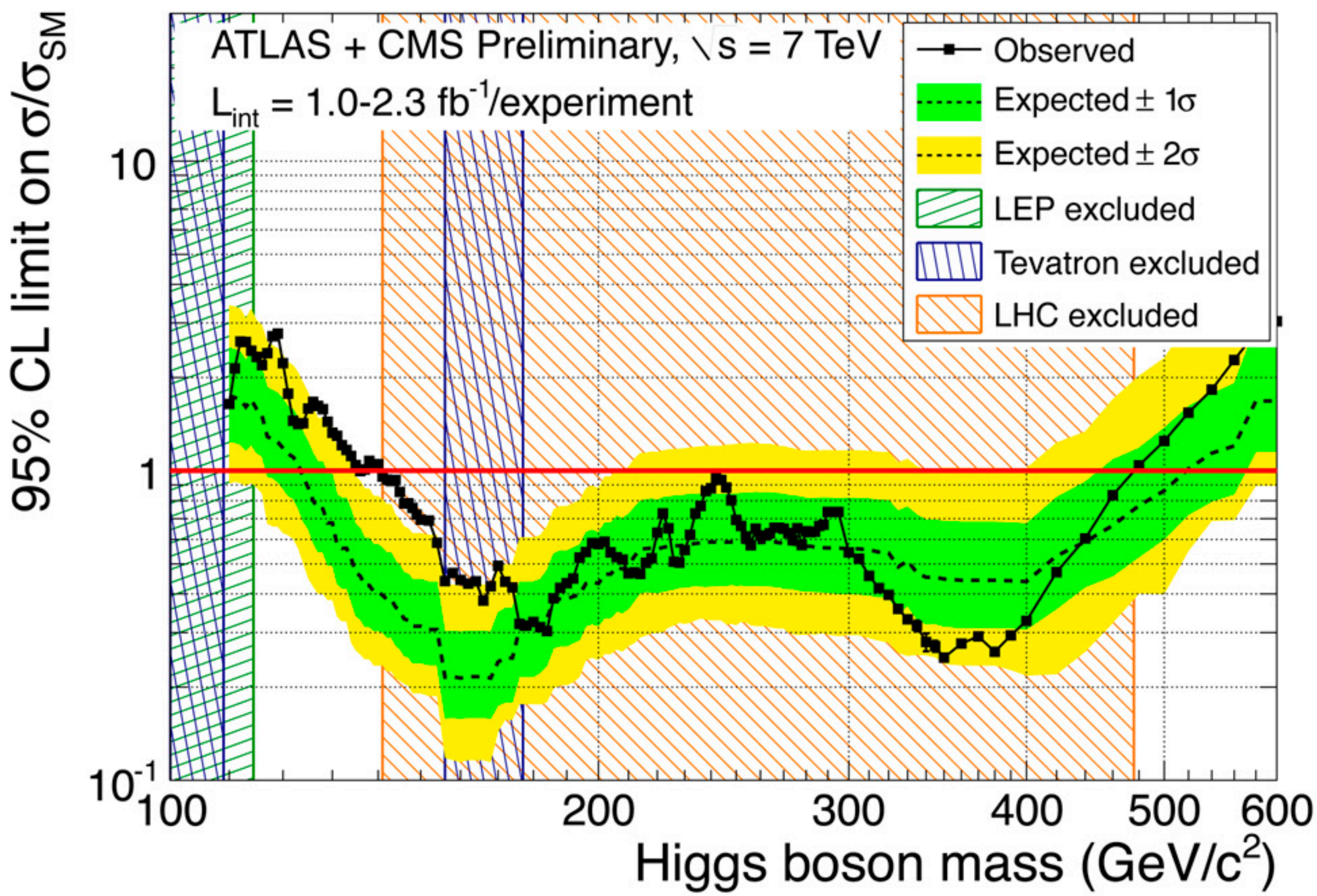


What does my interval mean?



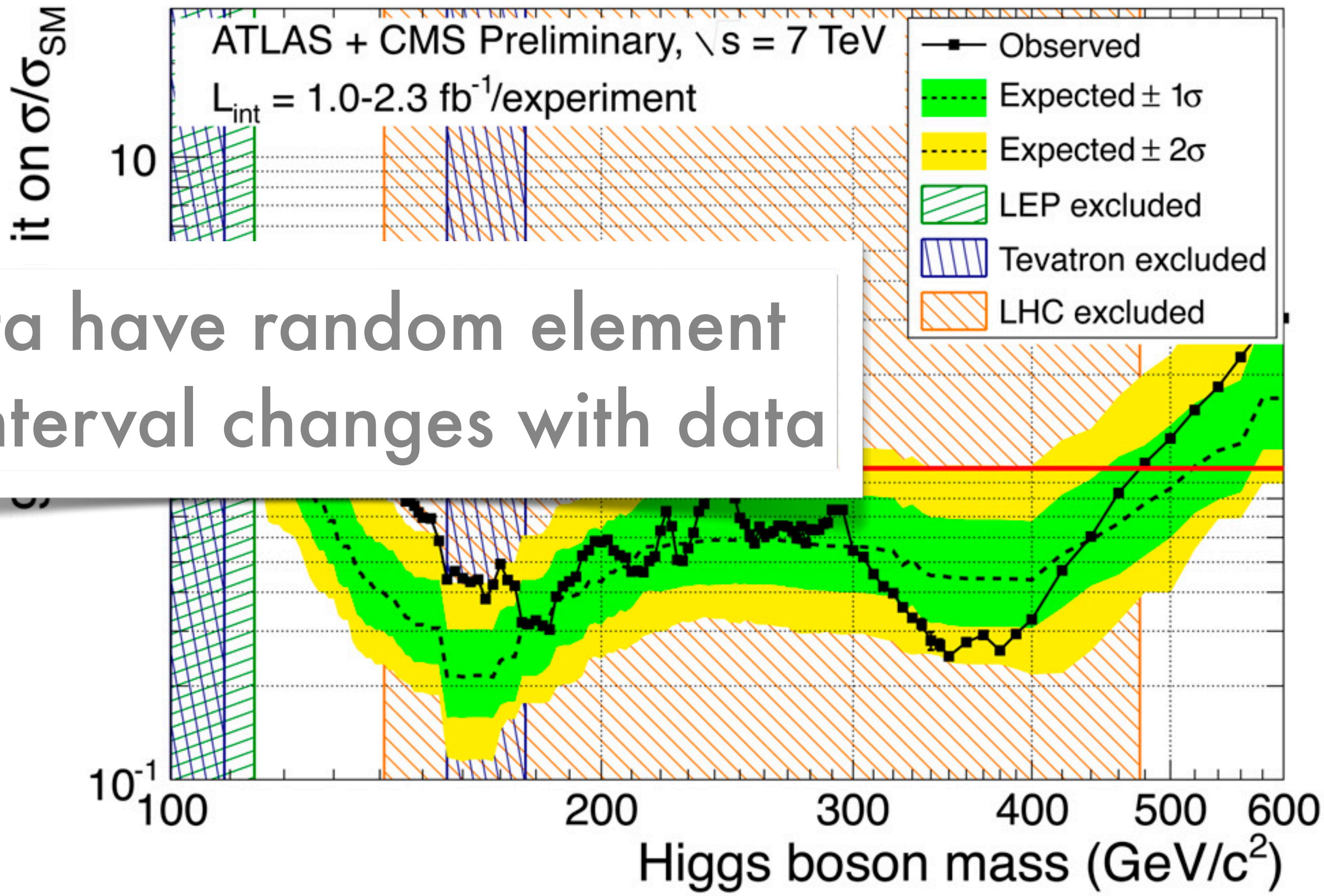
The probability that the Higgs boson mass is in this window is 5%?

What does my interval mean?



That's P(theory| data) which is Bayesian!

What does my interval mean?



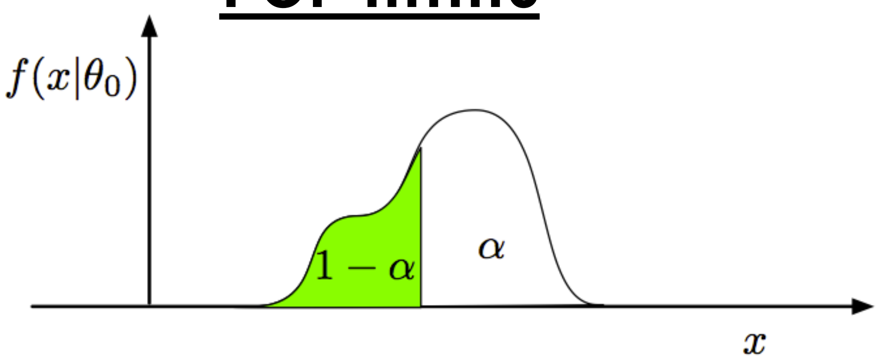
Data have random element
So interval changes with data

Frequentist: interval contains true value for 95% of experiments.

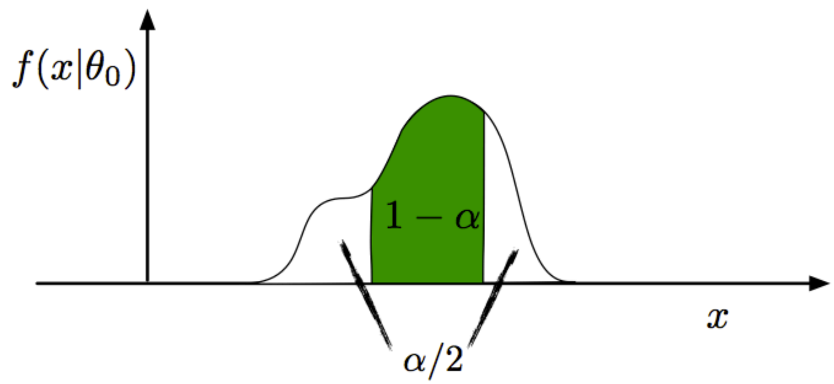
flip-flopping

You do an experiment. You don't know beforehand if you want to set an upper limit or measure a signal cross-section

For limits

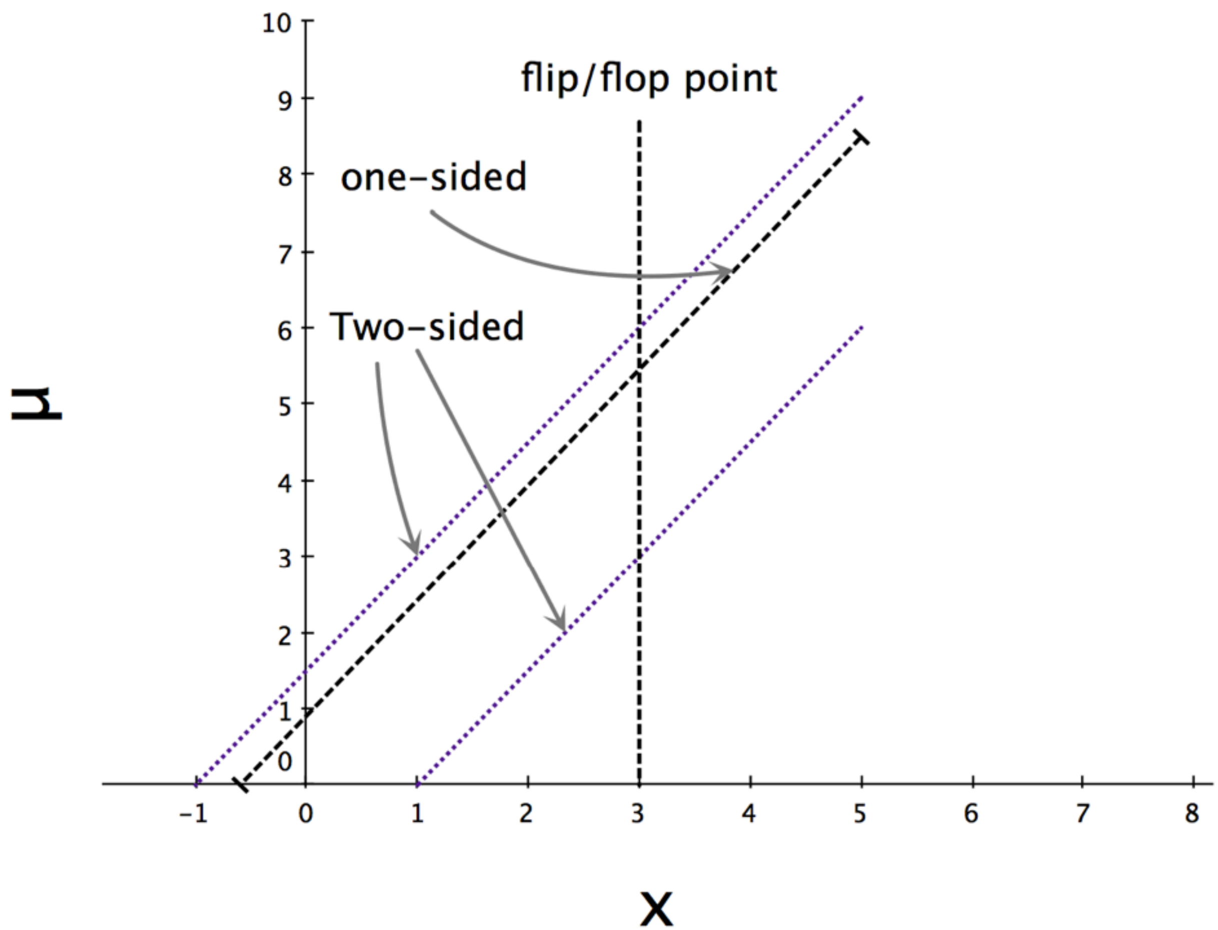


For measurements



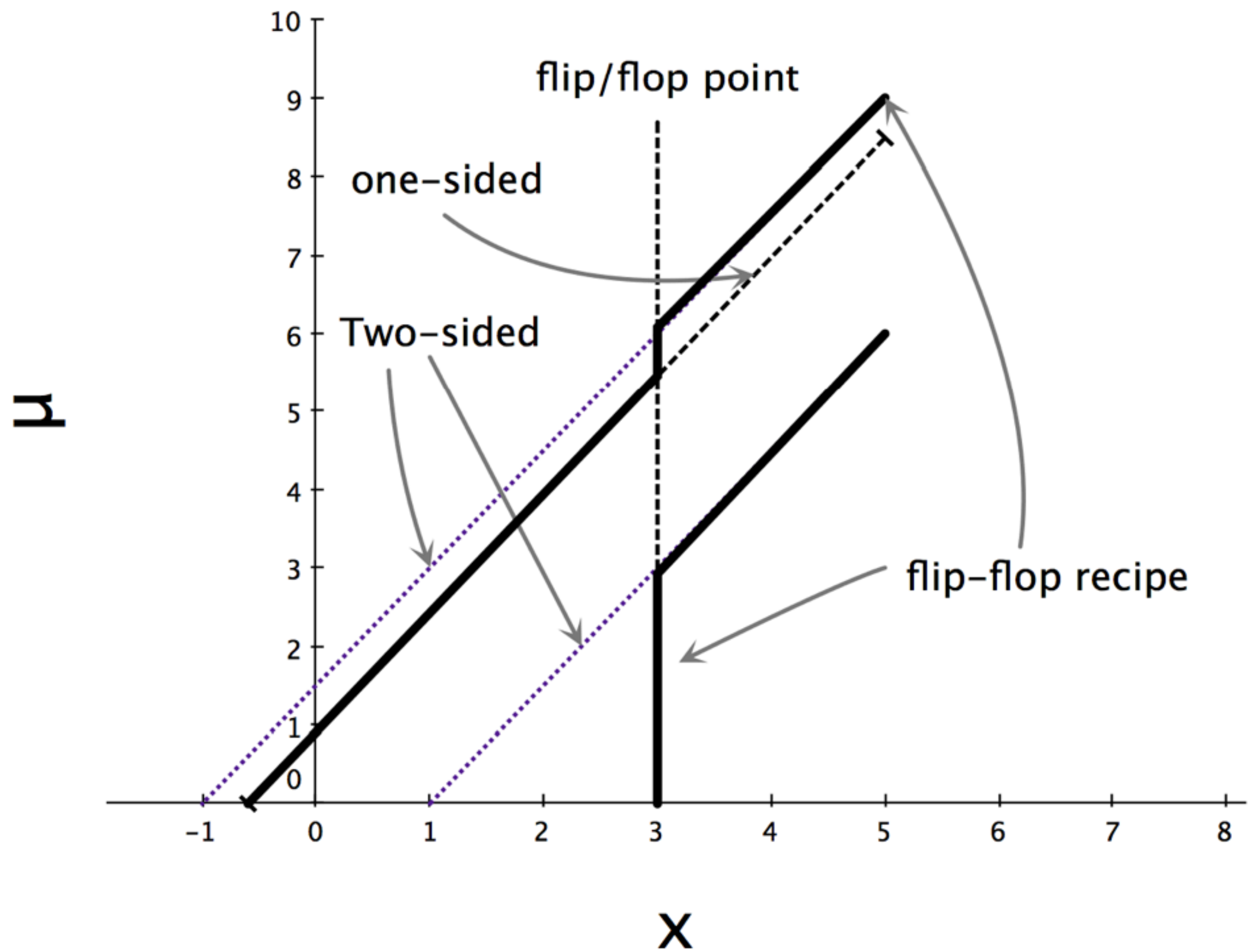
Intuitive approach:
I'll set a limit if observed < N
and make measurement if obs > N

Flip-flopping



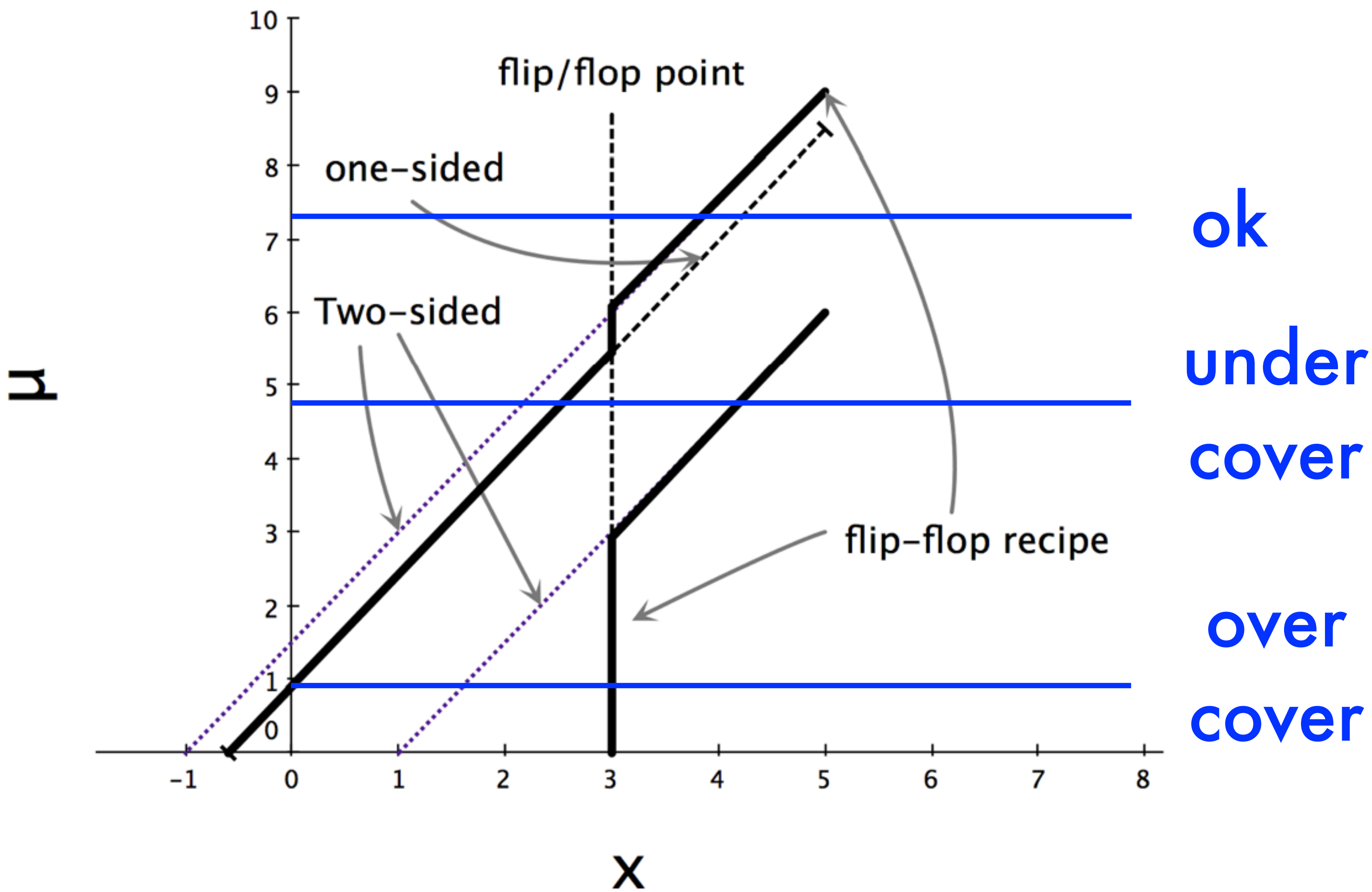
(K. Cranmer)

Flip-flop intervals



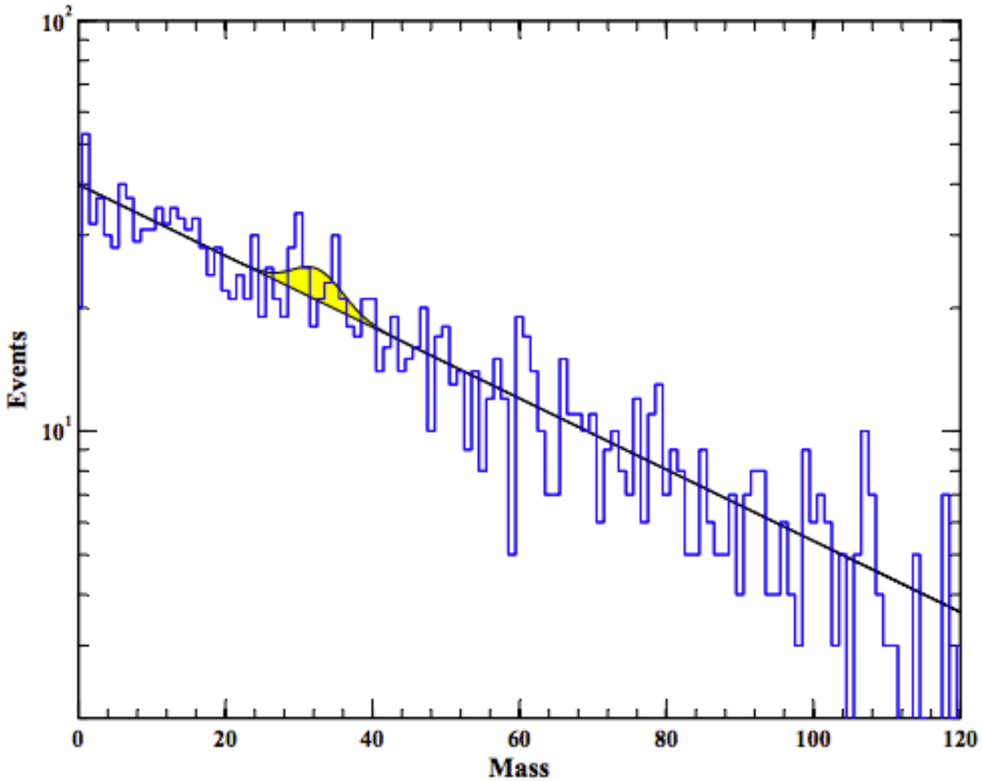
(K. Cranmer)

Coverage



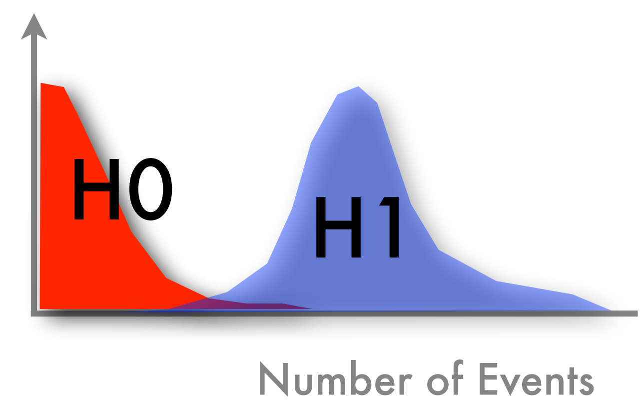
(K. Cranmer)

look elsewhere effect



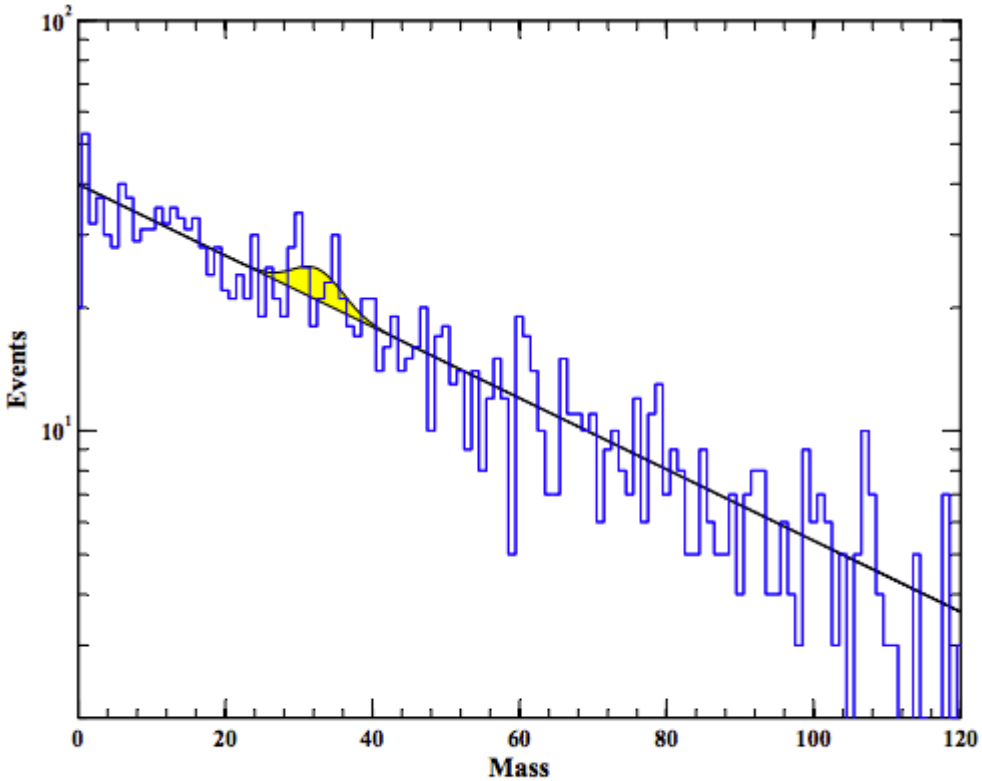
O. Vitells

For a fixed mass and width



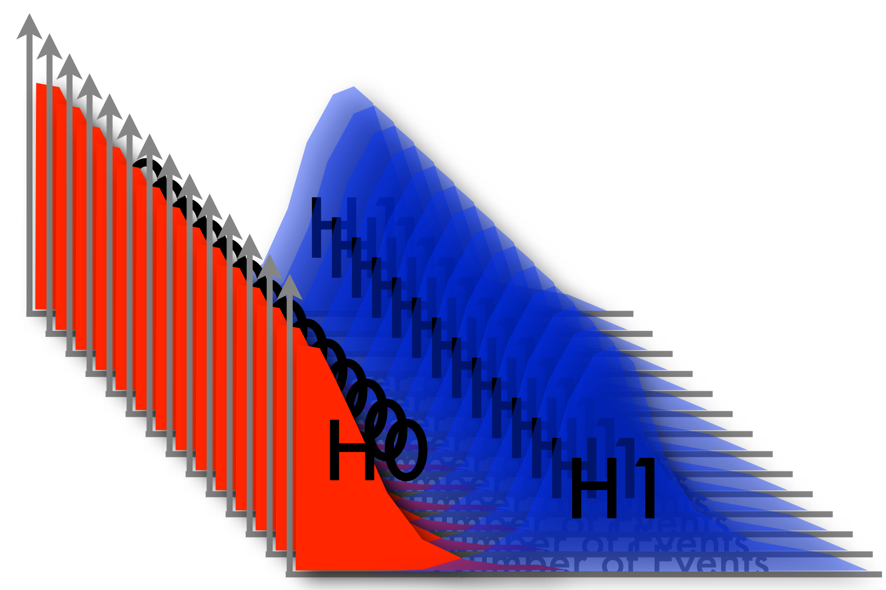
A well defined problem.

look elsewhere effect



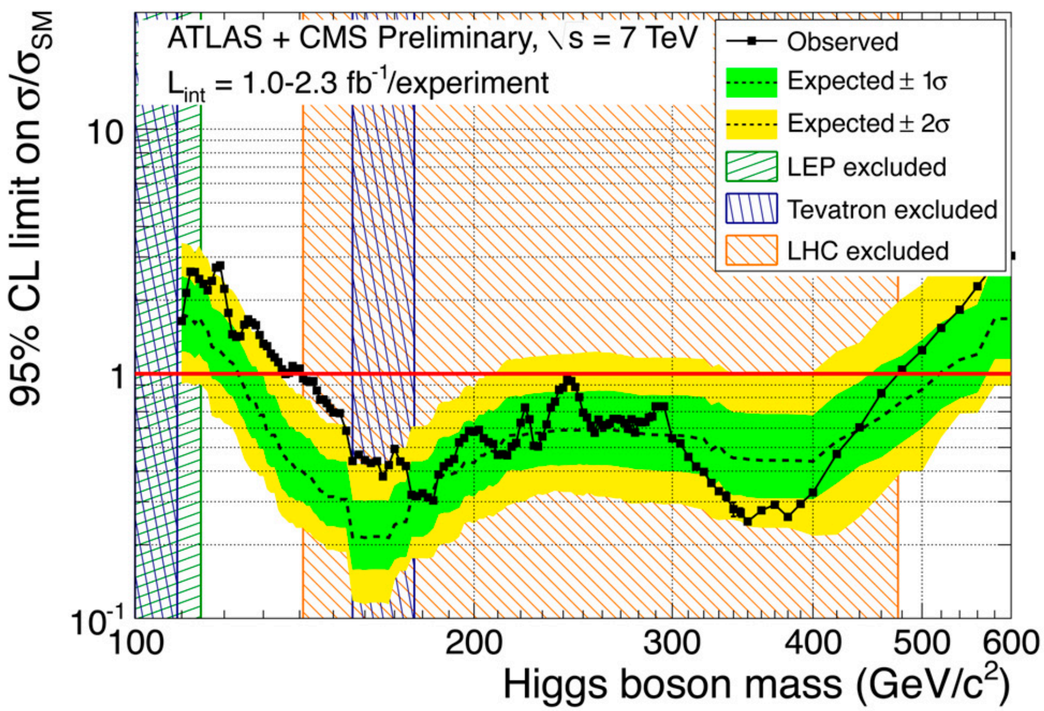
O. Vitells

For unknown mass and width



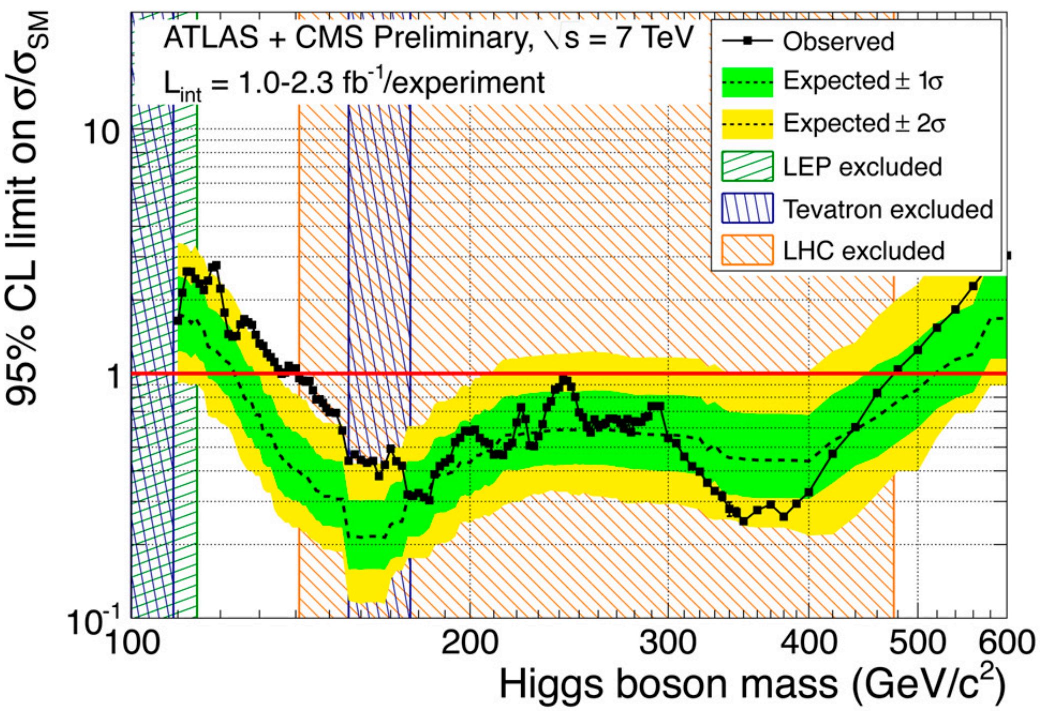
Many well defined problems.

look elsewhere effect



Can make set of well defined limits

look elsewhere effect

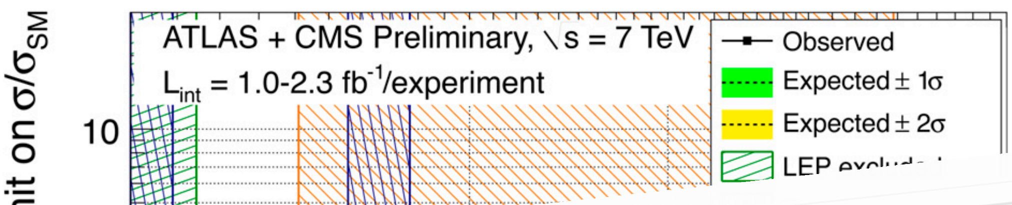


But what is significance of one-of-many results?

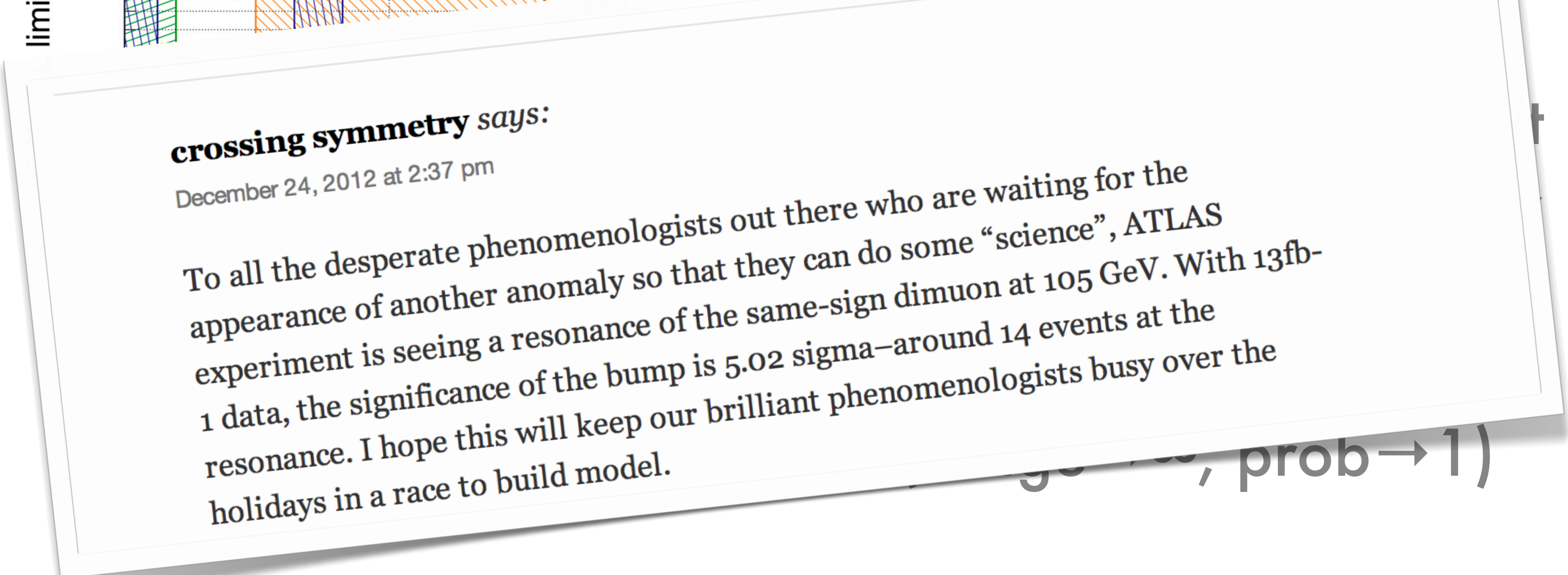
Prob to see a 5σ result depends on how many places you look!

(range→∞, prob→1)

look elsewhere effect



But what is significance



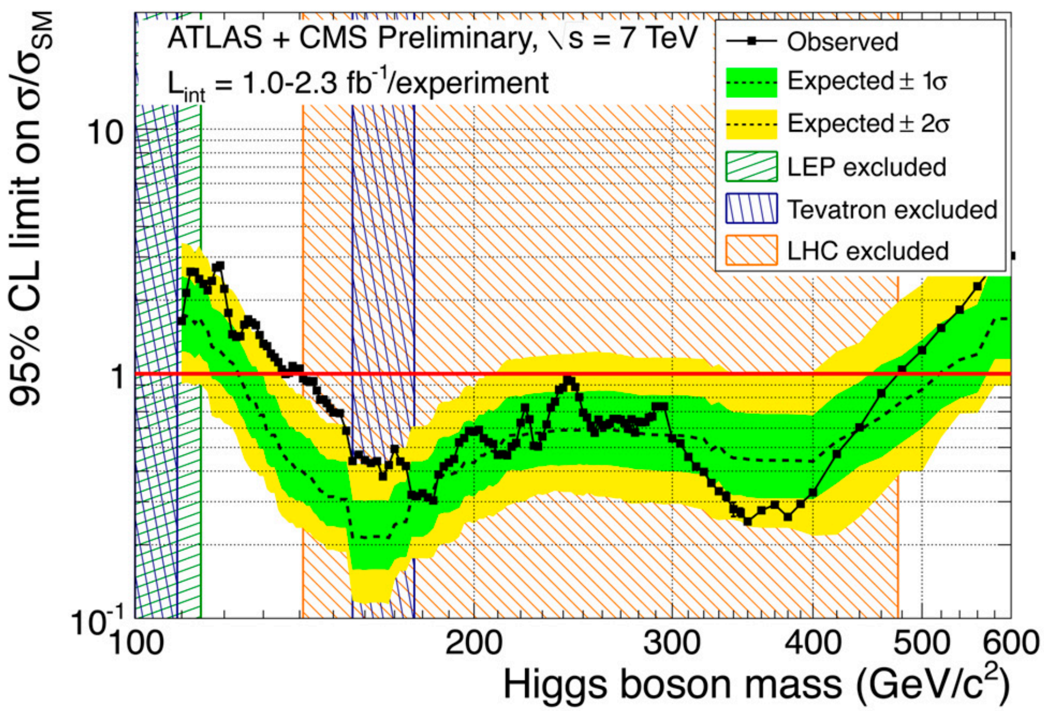
crossing symmetry *says:*

December 24, 2012 at 2:37 pm

To all the desperate phenomenologists out there who are waiting for the appearance of another anomaly so that they can do some "science", ATLAS experiment is seeing a resonance of the same-sign dimuon at 105 GeV. With 13fb-1 data, the significance of the bump is 5.02 sigma–around 14 events at the resonance. I hope this will keep our brilliant phenomenologists busy over the holidays in a race to build model.

prob→1)

look elsewhere effect



Must dilute the "local" significance by LEE.

Depends on range considered!

Philosophical: other experiment influence? Prior knowledge?

Thought experiment

What if you had 1000 graduate students

and gave them each one mass point.

Thought experiment

As the number of grad students grows

the probability that one will have a locally significant excess goes to 1

Thought experiment

As the number of grad students grows

the probability that one will have a locally significant excess goes to 1

Thought experiment

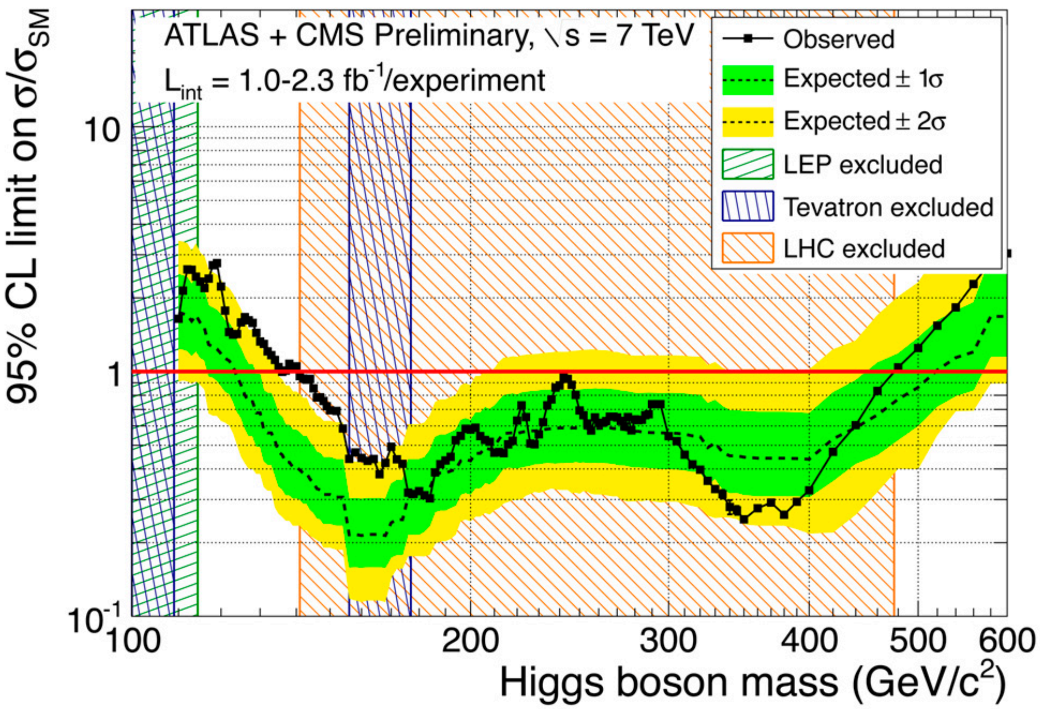
Is that student's result not valid?

Statistical fluctuations are valid results – they are expected!

LEE when you want to make statements <u>across multiple independent tests</u>

Thought experiment

Independent tests



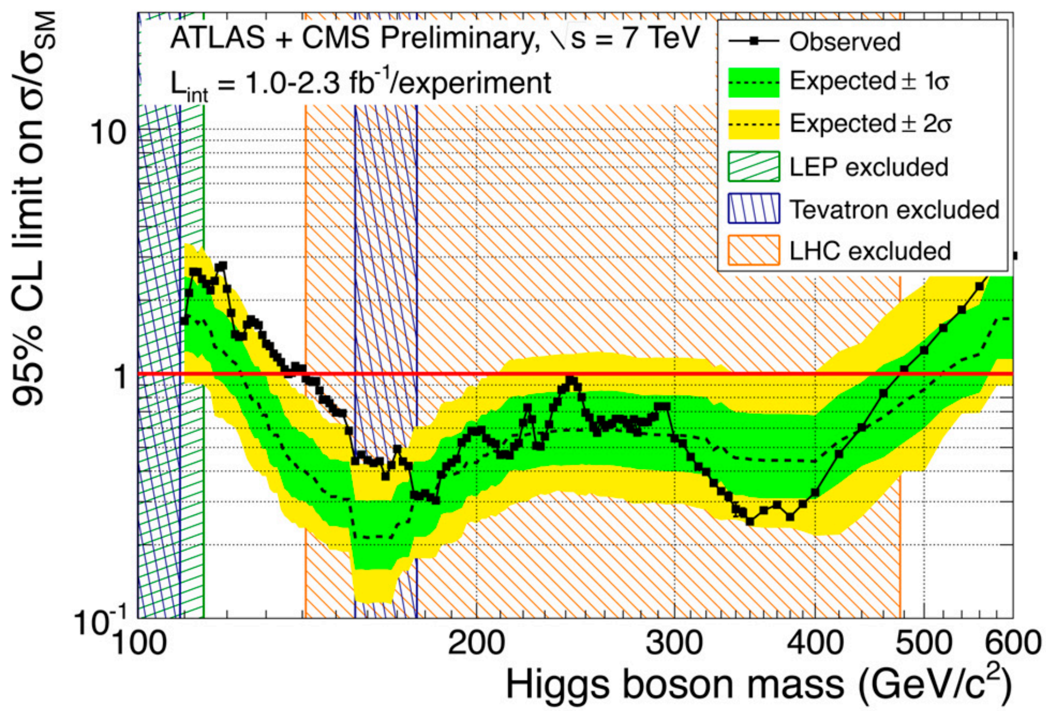
You can't make an infinite number of independent tests
because we have finite resolution.

But what about....

Does the LEE apply to a
set of papers from ATLAS?

No: if you consider each result seperately
Yes: if you take the most discrepant result from all ATLAS papers

LEE for limits?



Search done separately at each point.

Mass and width are assumed!

Fluctuations at other masses ignored

Statistical questions

- *For a given mass, what cross-sections are (in)consistent with the data? [cross-section limits]*
- *For a specific theory Z', what masses are (in)consistent with the data? [mass limits]*
- *What mass & cross-section are (in)consistent with the data? [cross-sec vs mass signifances]*

Statistical questions

- *For a given mass, what cross-sections are (in)consistent with the data? [cross-section limits]*
- *For a specific theory Z', what masses are (in)consistent with the data? [mass limits]*
- *What mass & cross-section are (in)consistent with the data? [cross-sec vs mass signifances]*

Raster Scan

For a given mass, what cross-sections are (in)consistent with the data?

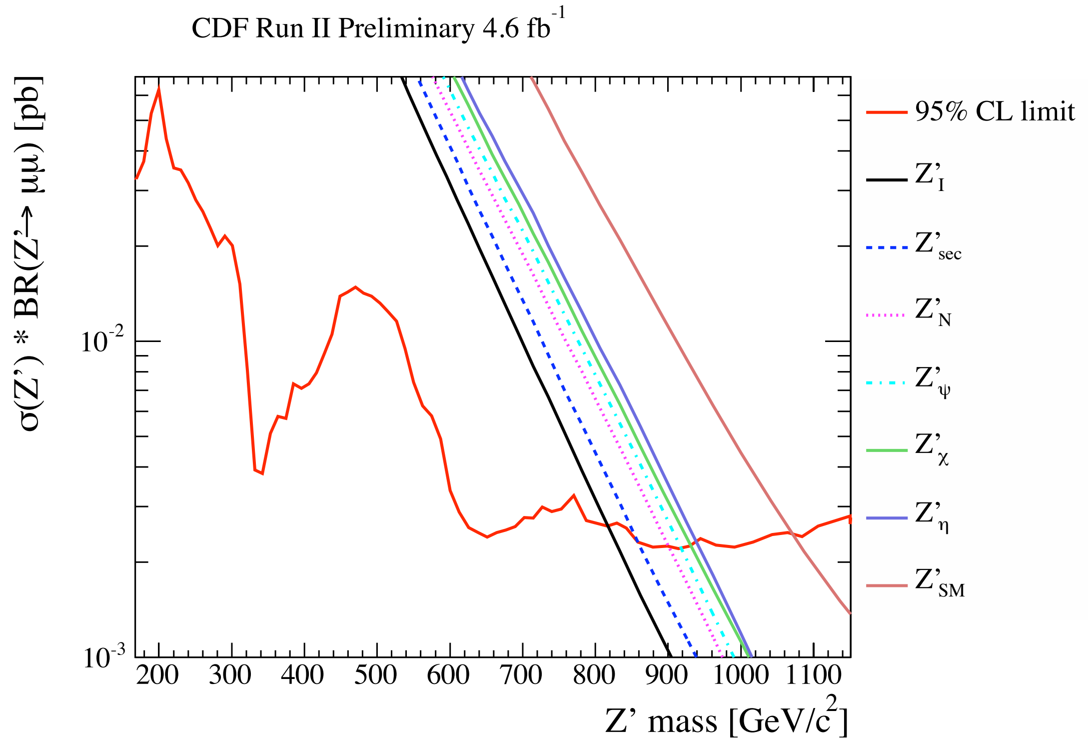
<u>Raster scan in mass</u>

At a set of masses, do a cross-section analysis.

<u>Note</u>:

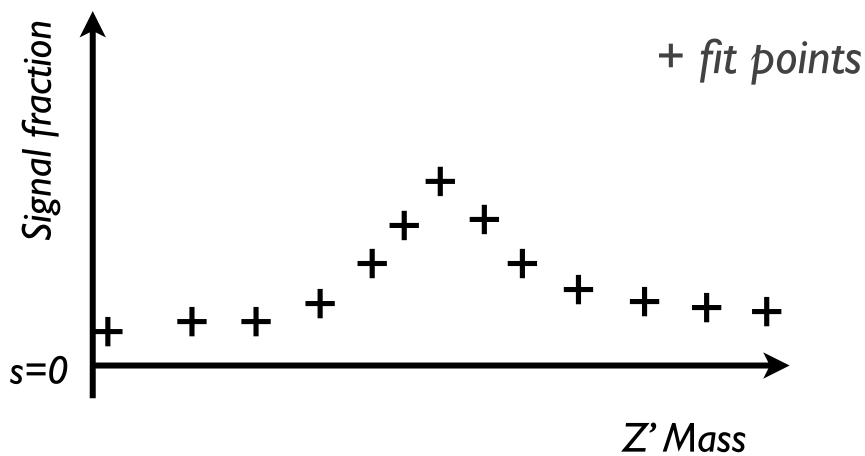
Limits are correlated in non-trivial way at different mass points

Look-elsewhere effect not accounted for



Mechanics

Raster scan

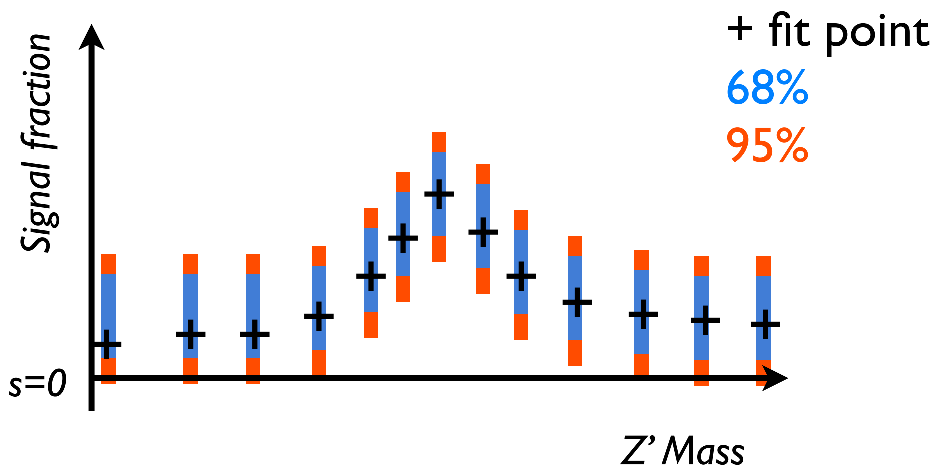


Finds set of points which maximize L(s) at each M.

The results are correlated point-to-point. By how much depends on the mass resolution and point density.

Raster scan

Raster scan



Compare *each* fit point with distribution of fit points for varying *signal at that mass*

Each specific-mass analysis interval based on comparison to fluctuations at *one* mass.

Analysis is really across mass range: you would accept bump at any mass. This requires additional dilution of claimed sensitivity here. The more places you look, the more likely to are to see a fluctuation.

"Look elsewhere effect"

(eg CDF Z' to ee bump at ~250 GeV)

Raster Scan

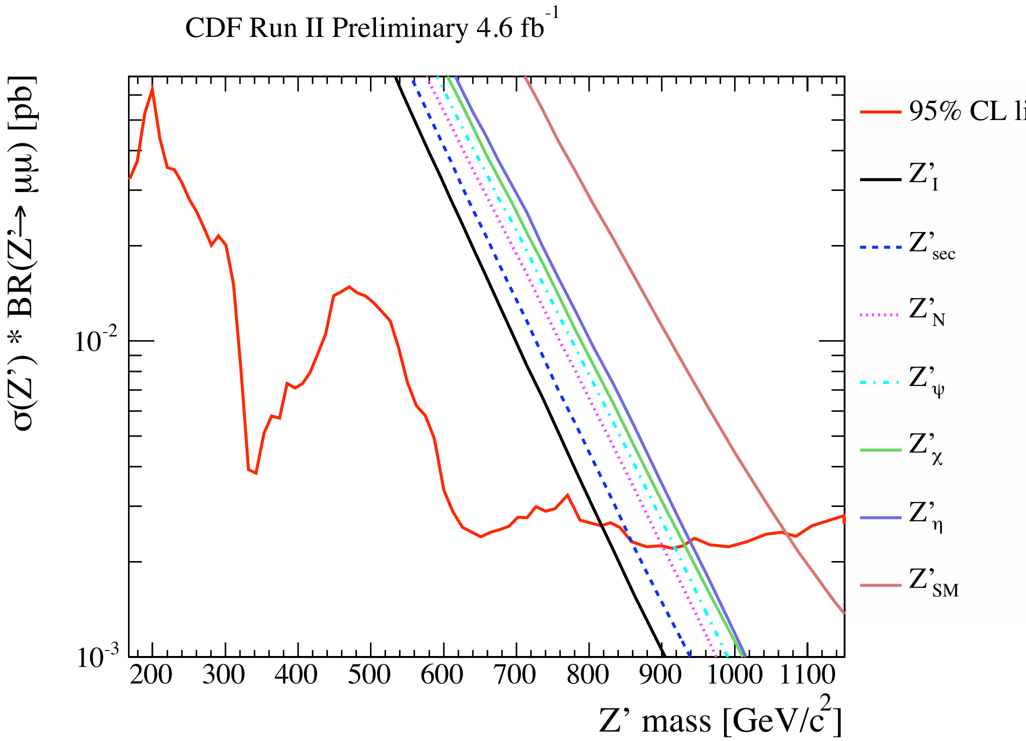
For a given mass, what cross-sections are (in)consistent with the data?

Summary

Raster scan in mass answers this question

Note:

This technique cannot be used to assess the significance of an excess or the insignificance of no-excess across masses.



Statistical questions

- *For a given mass, what cross-sections are (in)consistent with the data? [cross-section limits]*
- *For a specific theory Z', what masses are (in)consistent with the data? [mass limits]*
- *What mass & cross-section are (in)consistent with the data? [cross-sec vs mass signifances]*

Raster Scan

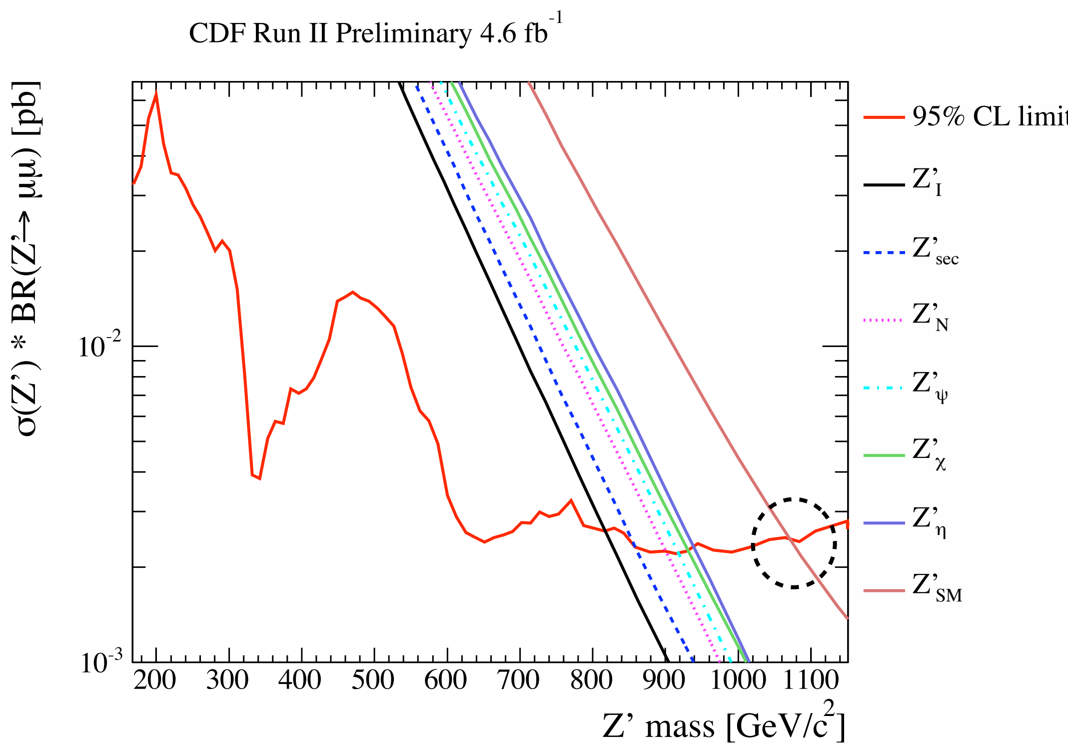
For a specific Z' theory, what masses are (in)consistent with the data?

Compare cross-section limits to theory

Find point of intersection. Quote result.

No look-elsewhere effect:
If Nature has a Z' at some mass, only need to worry about statistical correctness at that mass.

If Nature doesn't have a Z', then all exclusions statements are correct.



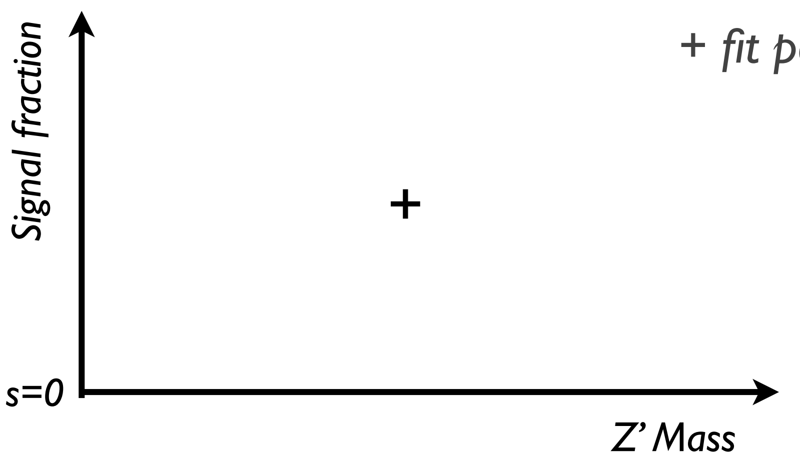
Statistical questions

- *For a given mass, what cross-sections are (in)consistent with the data? [cross-section limits]*
- *For a specific theory Z', what masses are (in)consistent with the data? [mass limits]*
- *What mass & cross-section are (in)consistent with the data? [cross-sec vs mass significances]*

Mechanics

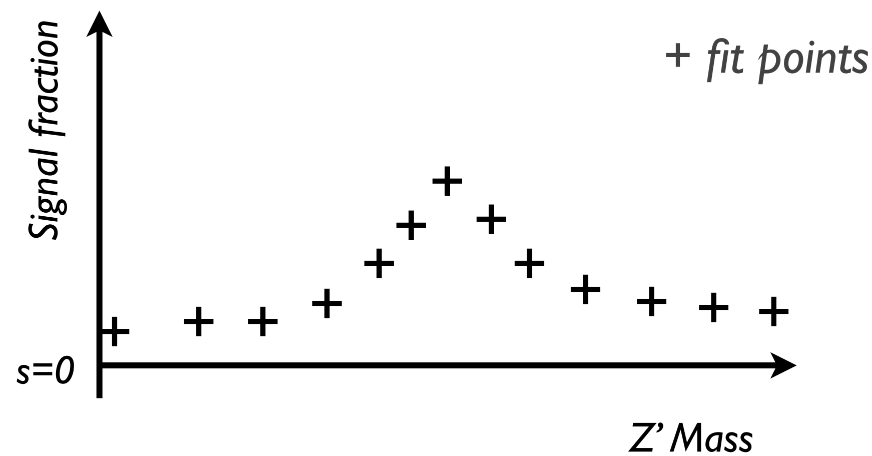
vs

Mass & rate analysis



Finds single point which maximizes L(M,s)

Raster scan

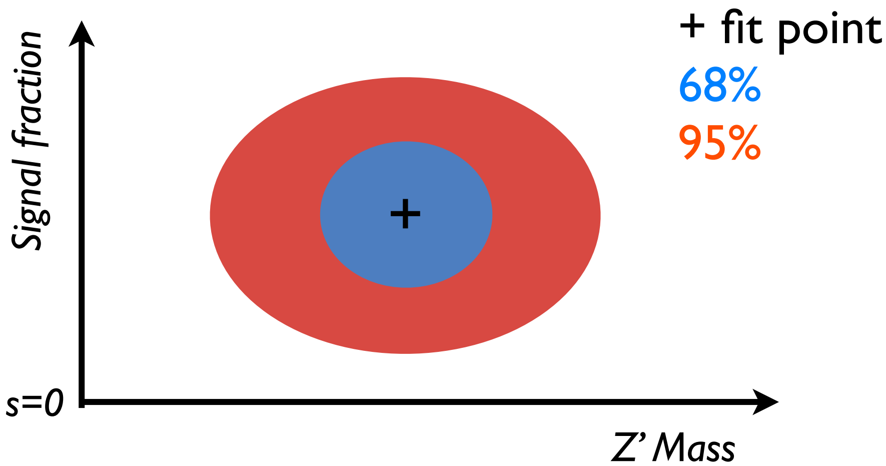


Finds set of points which maximize L(s) at each M.

What does discovery look like?

vs

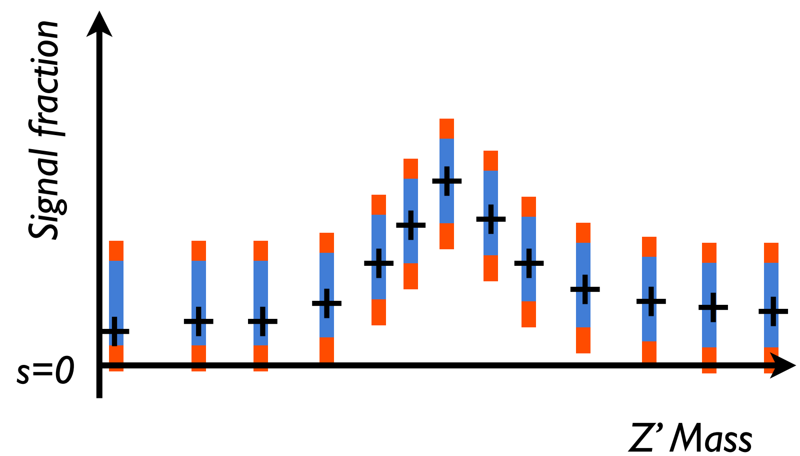
Mass & rate analysis



Compare fit point with distribution of fit points for varying *mass and signal*

Discovery if result inconsistent with s=0

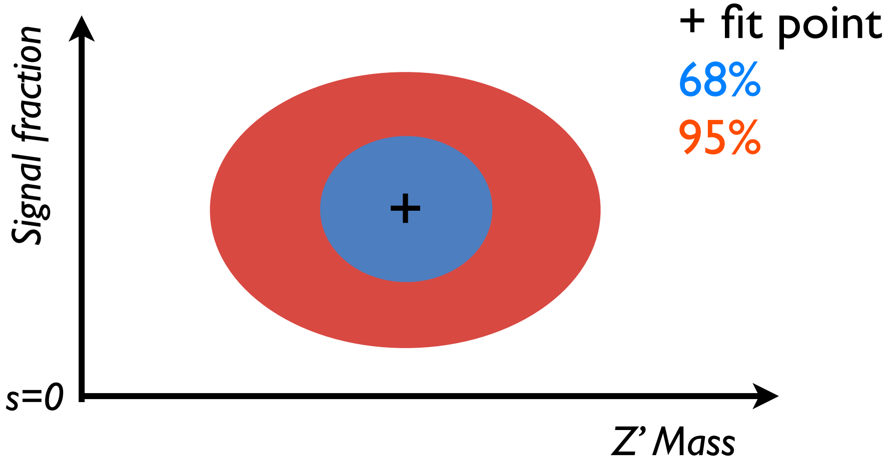
Set of rate analyses at several masses



What does discovery look like?

vs

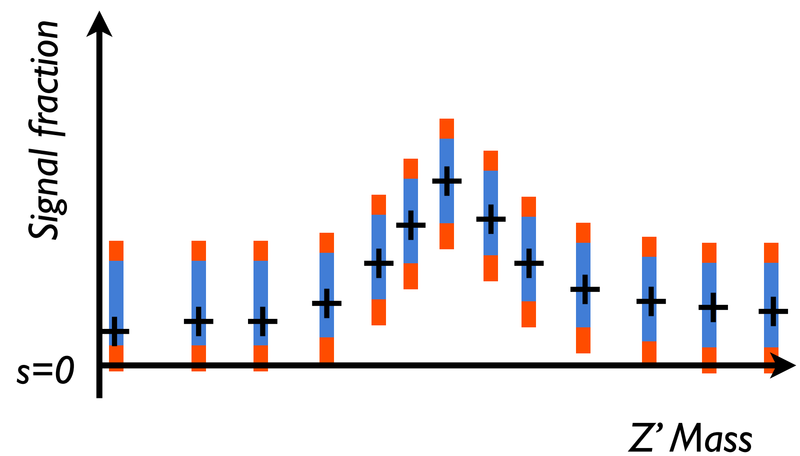
Mass & rate analysis



Intervals based on comparison to fluctuations at all masses.

Look-elsewhere effect naturally accounted for.

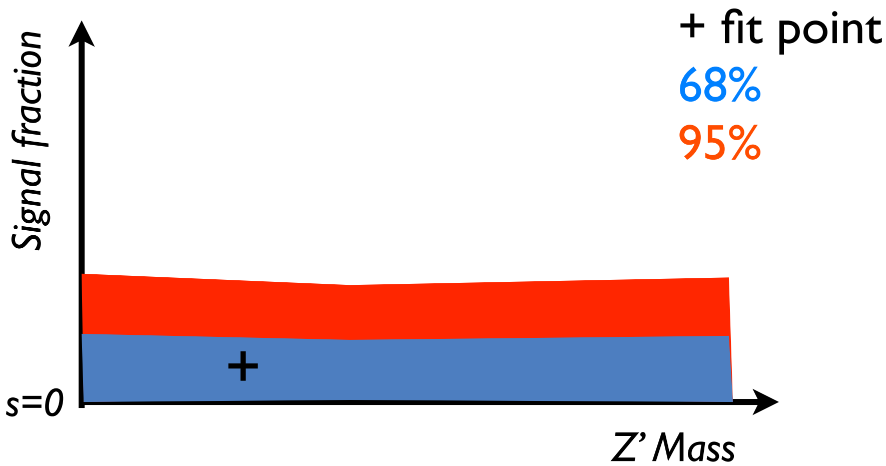
Set of rate analyses at several masses



What do limits look like?

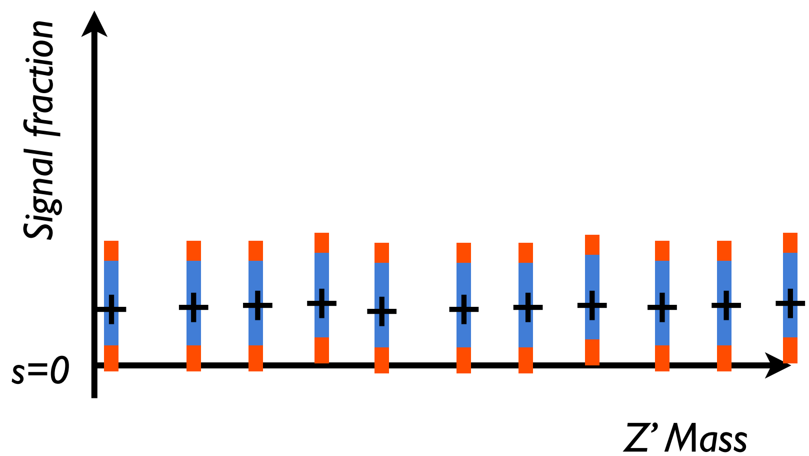
VS

Mass & rate analysis

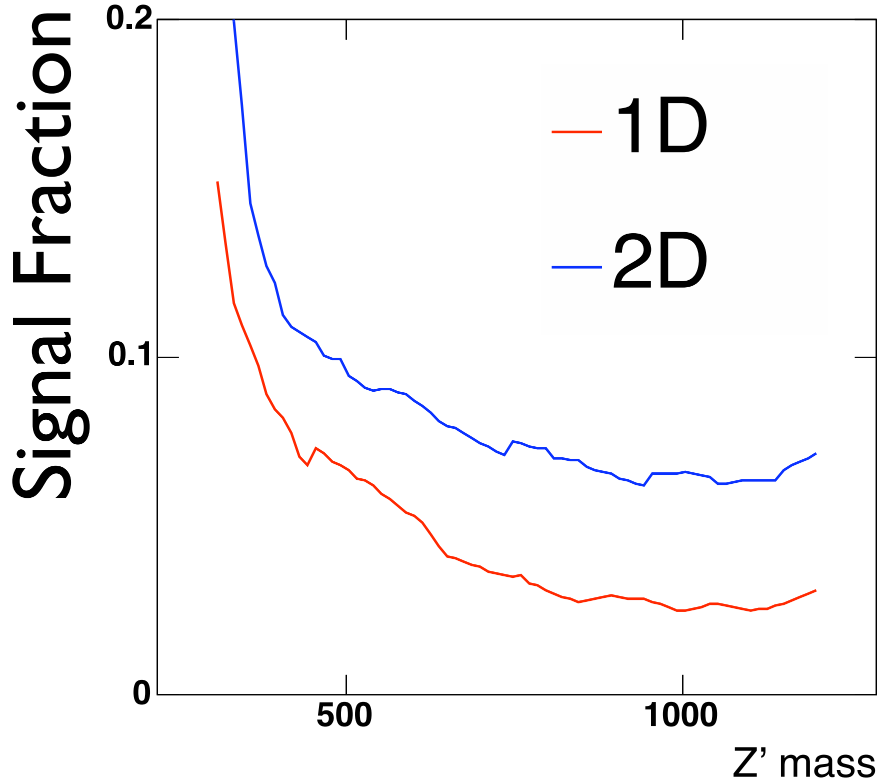


Result consistent with s=0

Set of rate analyses at several masses



Exclusion with 2D?



<u>2D limits are weaker</u>

More fluctuations everywhere

Upshot

Raster scan in mass

Statistically kosher at each point. If mass is unknown can only be used for exclusion. Gains exclusion power by sacrificing discovery potential.

mass vs cross-section

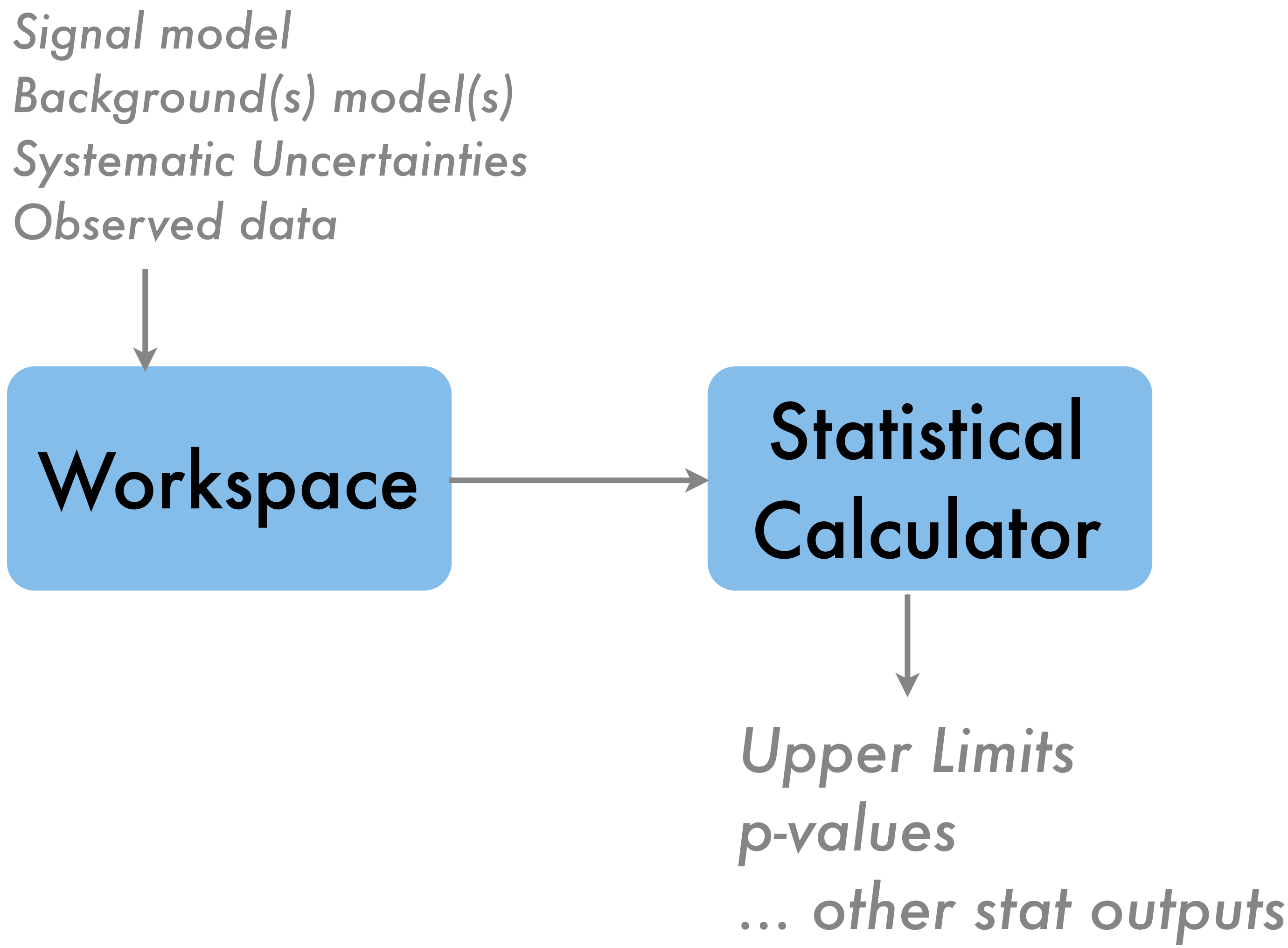
Well founded, more power for discovery, but weaker for exclusion.

Philosophy

Kosher to use both, as long as you always quote both

Tools and How-tos

RooStats basics



Tutorials

Google | roostats tutorial

Web | Videos | Shopping | News | Images | More | Search tools

About 4,740 results (0.27 seconds)

Roostats tutorials - Root
root.cern.ch/root/html534/**tutorials**/**roostats**/index.html ▾ ROOT ▾
C **tutorial** demonstrating and validates the RooJeffreysPrior class; 7. ModelInspector.C **RooStats** Model Inspector new; 8. MultivariateGaussianTest.
You've visited this page 2 times. Last visit: 8/17/14

WebHome < **RooStats** < TWiki
https://twiki.cern.ch/twiki/bin/view/**RooStats**/WebHome ▾ CERN ▾
Oct 21, 2013 - NEW Opportunity to contribute to RooFit/RooStats development ... **RooStats tutorials** at 2012 Bonn school on limit setting and global fits: slides ...
You've visited this page many times. Last visit: 4/6/14

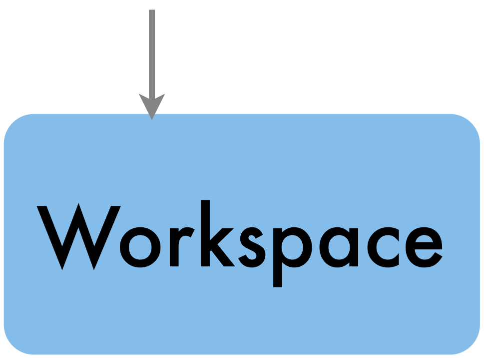
RooStatsTutorialsAugust2012 < **RooStats** < TWiki
https://twiki.cern.ch/twiki/.../**RooStats**/**RooStatsTutorials**August201... ▾ CERN ▾
Jump to Exercises for this **tutorial** (Part 1) -). Below will given code snippets showing how to build some different analysis models and then on how to run ...
Introduction to RooStats - Getting started with the software
You've visited this page 2 times. Last visit: 4/6/14

RooFit/**RooStats tutorials** for INFN School of ... - TWiki
twiki.cern.ch › TWiki › RooStats Web ▾ CERN ▾
Jun 14, 2013 - This is a set of RooFit/**RooStats tutorials** given at INFN school of statistics (3-7 June 2013) , see the school agenda. The required materials used ...

[PDF] **RooStats tutorials**
mon.iihe.ac.be/trac/t2b/.../RooFit**RooStats**_**tutorial**_summary_GVO.pdf ▾

Workspaces

Signal model
Background(s) model(s)
Systematic Uncertainties
Observed data



Example:

Number counting exp.

Nsig = 3.0

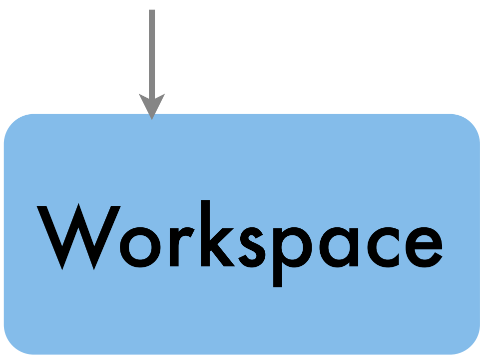
Nbg = 0.5

Nobs = 3

No systematics (other than Lumi)

Workspaces

Signal model
Background(s) model(s)
Systematic Uncertainties
Observed data



Example:

Number counting exp.

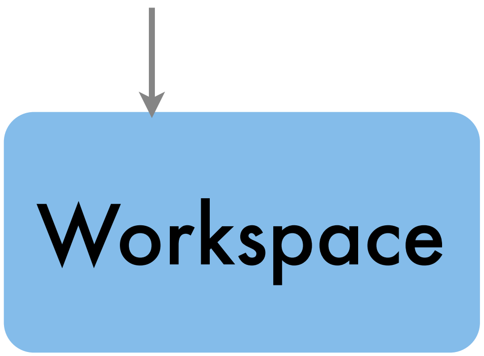
Nsig = 3.0

Nbg = 0.5 ± 0.1

Nobs = 3

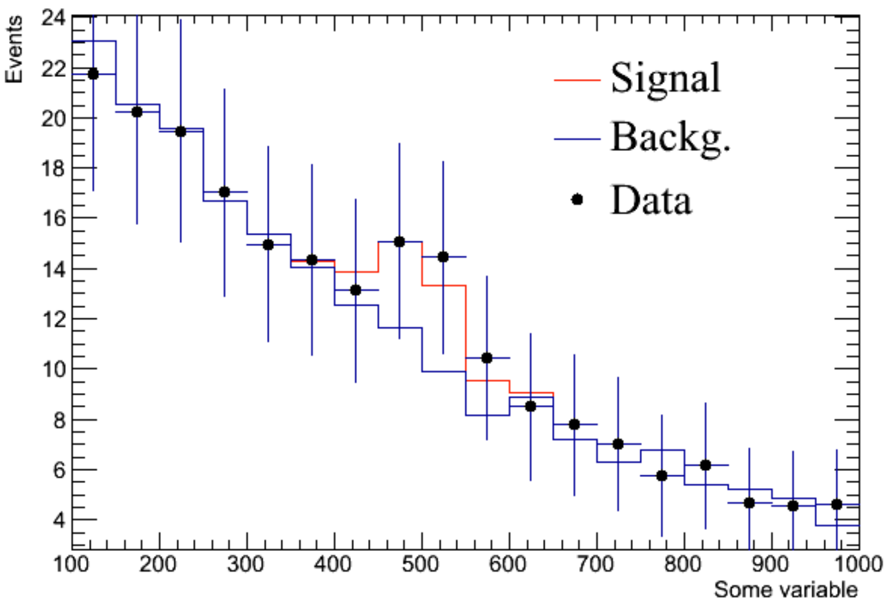
Workspaces

Signal model
Background(s) model(s)
Systematic Uncertainties
Observed data



Example:

Shape fit
Nsig = 10.0
Nbg = 200
Nobs = 210



Workspaces

Signal model
Background(s) model(s)
Systematic Uncertainties
Observed data

↓

Workspace

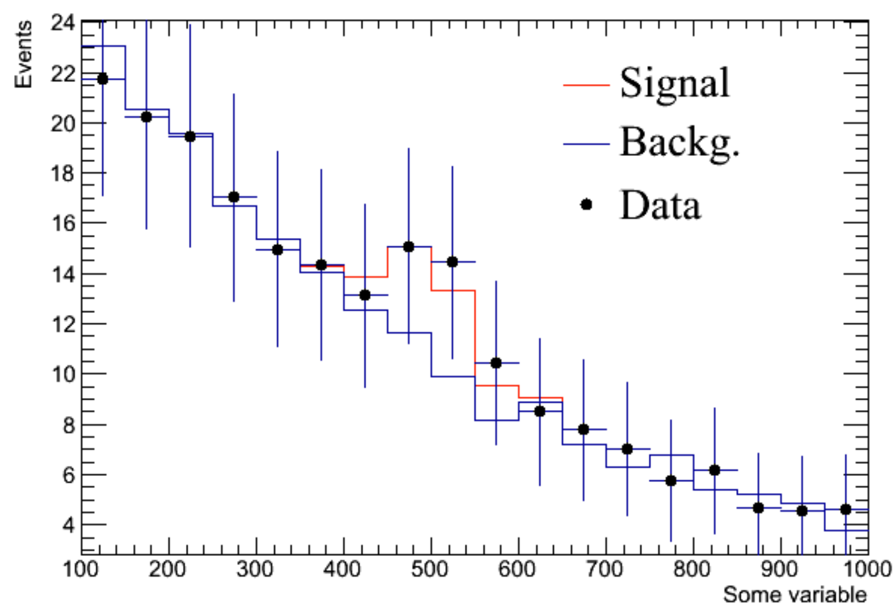
Example:

Shape fit

Nsig = 10.0

Nbg = 200 ± 20

Nobs = 210



Workspaces

Signal model
Background(s) model(s)
Systematic Uncertainties
Observed data

↓

Workspace

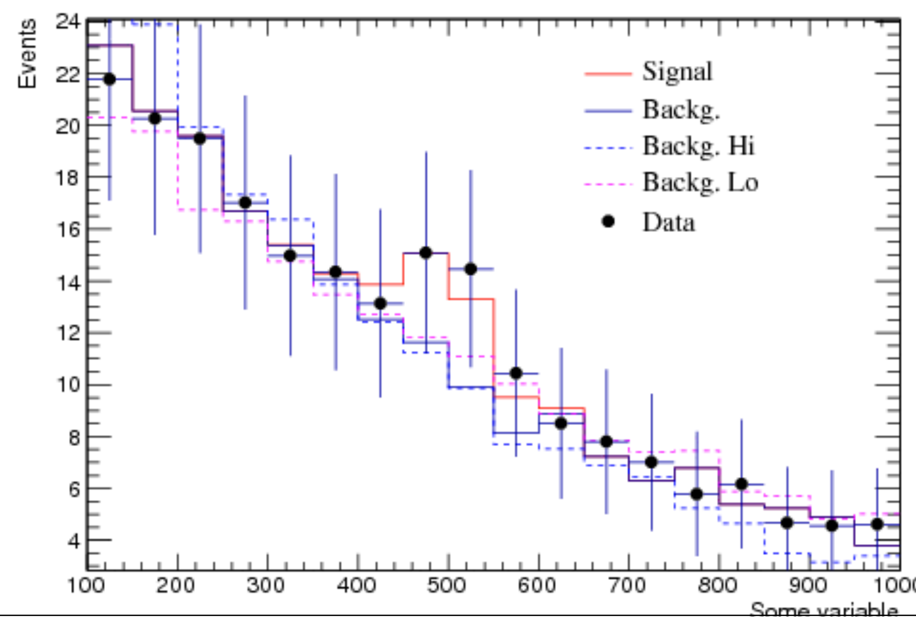
<u>Example</u>:

Shape fit

Nsig = 10.0

Nbg = 200 with shape unc.

Nobs = 210



The end!

“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