Tutorial Statistics Limits Part II

M. Herbst

influenced by many other unknowing contributors,
mentioned where possible

yesterday

Statistics/ Probability

Frequentist/ Bayesian

Probability Density Function

Confidence Level/ p-Value

Confidence Intervals

Exercises

today

Hypothesis Testing

Error Classification

Size/ Power of Test

Test Statistics/ Chisquare Dist.

NP Lemma/ Wilks' theorem

Likelihood Function

Systematics

POI, Nuissance Parameters

Profile Likelihood Ratio

Coverage/ Flip-Flopping/ Asymptotic Limit/ Look-Elsewhere

current ATLAS discussion: Power Constraint Limits

Bayesian Statistics, follow up

$P(B)$ is called the *marginal probability* of B: the *a priori* probability of witnessing the new evidence B under all possible hypotheses. It can be calculated as the sum of the product of all probabilities of any complete set of mutually exclusive hypotheses and corresponding conditional probabilities:

http://en.wikipedia.org/wiki/Bayesian_inference

Bayes' Law

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

$$P(B) = \quad P(E) = P(E|H)\,P(H) + P(E|\neg H)\,P(\neg H)$$

http://en.wikipedia.org/wiki/Bayesian_inference

Bayesian Statistics

π (H) does not assume anything about x

$$P(H|x) = \frac{P(x|H)\ \pi(H)}{\int P(x|H)\pi(H)dH}$$

posterior probability after seeing the data

"normalisation" sum over all hypothesis

done your Homework?

3.1 - Particle Production

Solution:

From the figure:

$p(\mu = 3, 2) = 0.4,$
$p(\mu = 4, 2) = 0.22,$
$p(\mu = 5, 2) = 0.12,$
$p(\mu = 6, 2) = 0.06$
$\rightarrow \mu \sim 5.3$
(exact solution, see e.g. PDG: $\mu = 5.32$)

3.2 - Particle Production small background

Solution:

a) $\mu_{bgr} = 0,\ N_{obs} = 2$:
See previous exercise, $\mu_{sig} = 5.3$

b) $\mu_{bgr} = 1,\ N_{obs} = 2$:
$\mu_{sig} = 5.3 - \mu_{bgr} = 4.3$

c) $\mu_{bgr} = 3,\ N_{obs} = 0$: $p = e^{-(\mu_{sig}+3)} = 0.1$
$\rightarrow \mu_{sig}$ ought to be smaller than zero $\rightarrow \mu_{sig} = 0.$

thanks to

O. Behnke, C. Kleinwort, S. Schmitt (DESY),

from Terascale Statistics School 2008 exercises

done your Homework?

3.3 - Particle Production modified frequentist

Solution:

$CL_s = CL(S+B)/CL(B) = e^{-(\mu_{sig}+\mu_{bgr})}/e^{-\mu_{bgr}} = e^{-\mu_{sig}} = 0.1 \rightarrow \mu_{sig} = -ln(0.1) = 2.3$

... as if there were no background!

(Reference: A.L. Read, (Oslo) CERN-OPEN-2000-205, Aug 2000.)

3.4 - Particle Production frequentist vs. bayesian

Solution:

a) Frequentist: from the CL curve:

$CL = 0.1 \leftrightarrow 1.28\sigma$

$\rightarrow \mu_{lim} = -2 + 1.28 = -0.72$

b) Bayesian:

Renormalised total integral in physical area:

$\int_{0.}^{\infty} dx' \frac{1}{\sqrt{2\pi}} e^{\frac{-(x'+2)^2}{2}} = CL(2) = 0.028$

Integral above limit:

$\rightarrow \int_{\mu_{lim}}^{\infty} dx' \frac{1}{\sqrt{2\pi}} e^{\frac{-(x'+2)^2}{2}} = 0.1 \cdot 0.028 = 0.0028$

$CL = 0.0028 \leftrightarrow 2.75\sigma$

$\rightarrow \mu_{lim} = -2 + 2.75 = 0.75$

thanks to

O. Behnke, C. Kleinwort, S. Schmitt (DESY),

from Terascale Statistics School 2008 exercises

Observed vs. Expected Limits

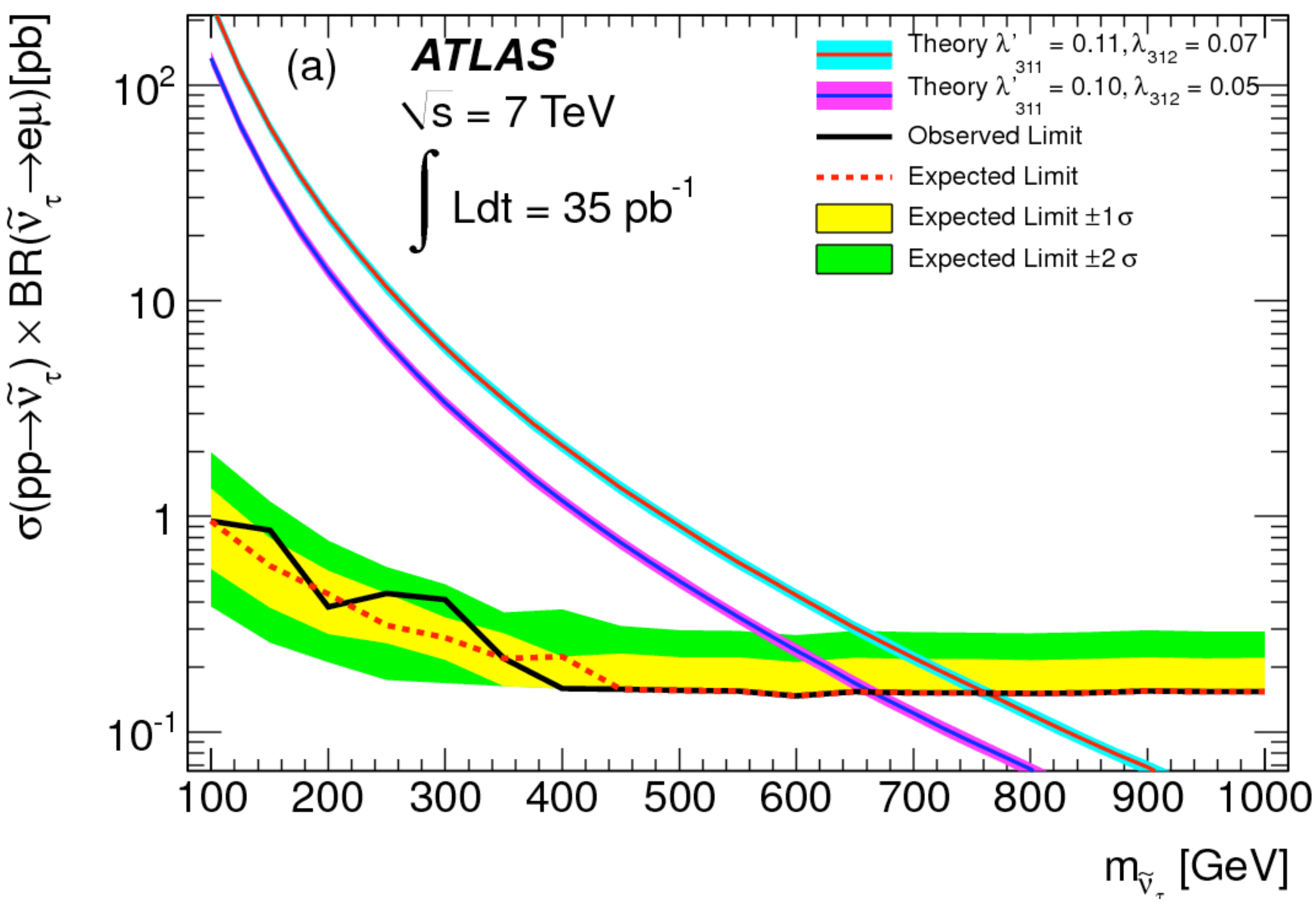
understand the jargon

Expected Limit:

calculated from background prediction only
(as if data/MC agree exactly, i.e. there is no deviation)

Observed Limit:

data is compared
to MC background
prediction,
observed
limit should wiggle
around expected!



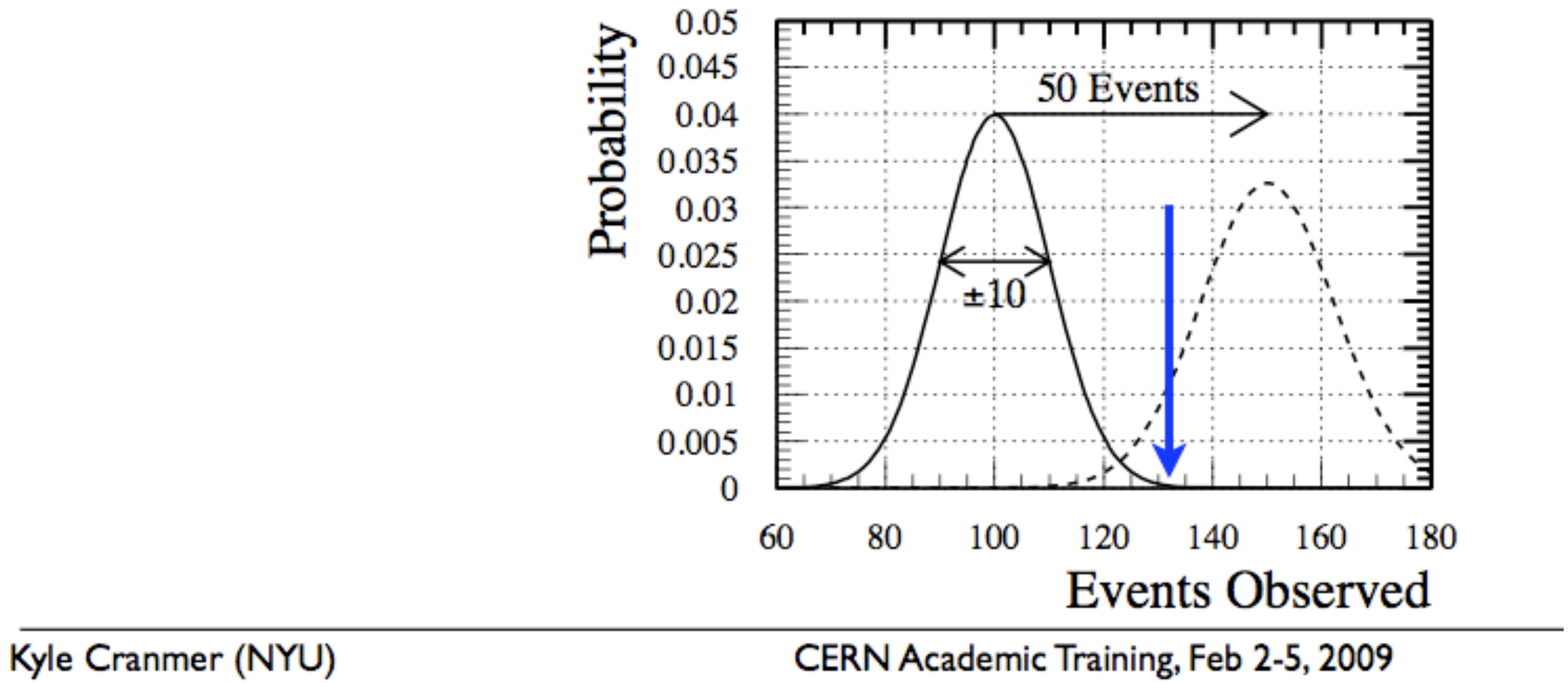
Hypothesis Testing

consider data under two Hypothesis:

H_0 Null-Hypothesis: background - only

H_1 Alternate Hypothesis: background + signal

decide whether to accept/ reject H_0



inspired by lectures of Kyle Cranmer at CERN (ATLAS, NYU)

Error Classification

can never be sure
it is the right decision!

		TRUE condition	
		guilty	*not guilty*
OUR decision	*sentenced guilty*	**TRUE POSITIVE**	**Type I Error false positive**
	not sentenced guitly	**Type II Error false negative**	**TRUE NEGATIVE**

call rate of Type I Error: **α**

call rate of Type II Error: **β**

Size and Power

call rate of Type I Error: α

treat Hypotheses asymmetrically

Null-Hypothesis is special!

Fix rate of α, call it **"Size of the Test"**

call rate of Type II Error: β

call (1 - β) the **"Power of the Test"**

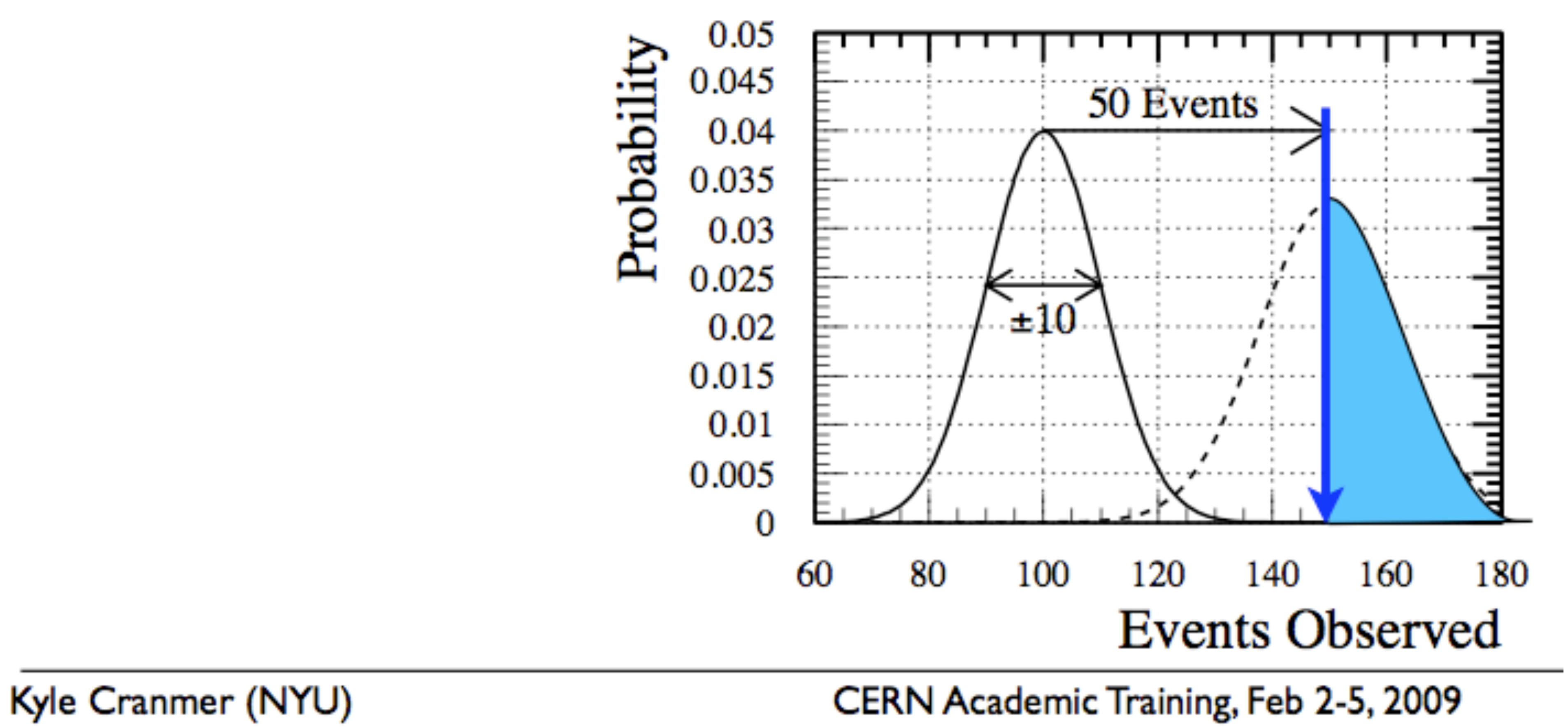
now can define a Goal:

Maximise Power for a fixed Size of the Test

Hypothesis Testing: Size and Power

think of 5σ discovery in particle physics: $5\sigma \Leftrightarrow \alpha = 2.87 \cdot 10^{-7}$

very small chance to reject the Standard Model



in general: Size is arbitrary: choose depend on ***Utility*** or ***Risk*** ...

Neyman-Pearson Lemma (1928-1938)

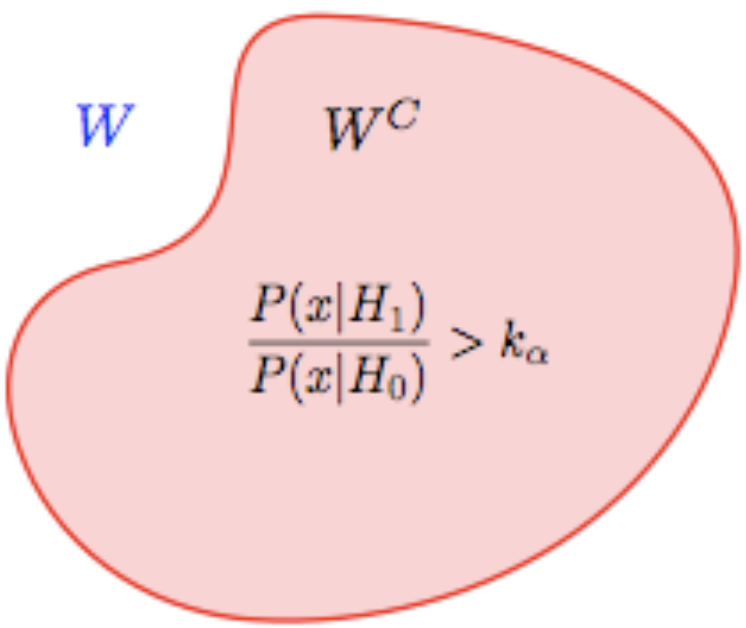
given the probability to wrongly reject Null-Hypothesis

$$\alpha = P(x \notin W \mid H_0)$$

(if data falls in W we accept H_0)

find region **W** that minimizes the probability of wrongly accepting H_0 (when H_1 is true)

$$\beta = \quad (\square \in \square \mid \quad {}_1)$$



<u>NP Lemma:</u>

region **W** is a contour of the Likelihood Ratio!

it can be shown (proof):

any other contour (same size) has less power!

Test Statistic

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

Likelihood Ratio

is an example of a Test Statistic

(real valued function, summarizing

the data in a way relevant to the Hypo-Test)

- Common test statistics
  - simple likelihood ratio (LEP) $Q_{LEP} = L_{s+b}(\mu=1)/L_b(\mu=0)$
  - ratio of profiled likelihoods (Tevatron) $Q_{TEV} = L_{s+b}(\mu=1,\hat{\hat{\nu}})/L_b(\mu=0,\hat{\hat{\nu}}')$
  - profile likelihood ratio (LHC) $\lambda(\mu) = L_{s+b}(\mu,\hat{\hat{\nu}})/L_{s+b}(\hat{\mu},\hat{\nu})$

(taken from Kyle Cranmer's talk)

ν's are nuisance parameters (shape)

Simple Hypothesis Testing

an Hypothesis is simple, if it has no free parameters
NP Lemma is the answer!

$f(x \mid H_0)$ vs. $f(x \mid H_1)$

if there are free parameters
Hypothesis is composite!

$f(x \mid H_0)$ vs. $f(x \mid H_1, m_{Higgs})$

typically pdf can be parametrized: $\mathbf{f(x \mid \theta)}$

for fixed θ it is a pdf for x,
as a function of θ call it "Likelihood function"
(not a pdf!)

divide θ into parameters of interest, nuisance parameters

LEP vs. LHC Likelihood Ratio

Simple Likelihood Ratio (LEP)

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

Profile Likelihood Ratio (LHC)

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

sophisticated ansatz:

- where $\hat{\hat{\nu}}$ is best fit with μ fixed to 0
- and $\hat{\nu}$ is best fit with μ left floating

Hypothesis Testing vs. Interval Construction

Interval Construction is "inverted" Hypothesis Test

Property of Test	*Property of Intervall*
test size α	*confidence level* α
probability of rejecting a false value of θ *power = 1 -* β	*probability of not covering a false value of* θ *1 -* β
most powerful	*uniformly most accurate*

Wilk’s Theorem

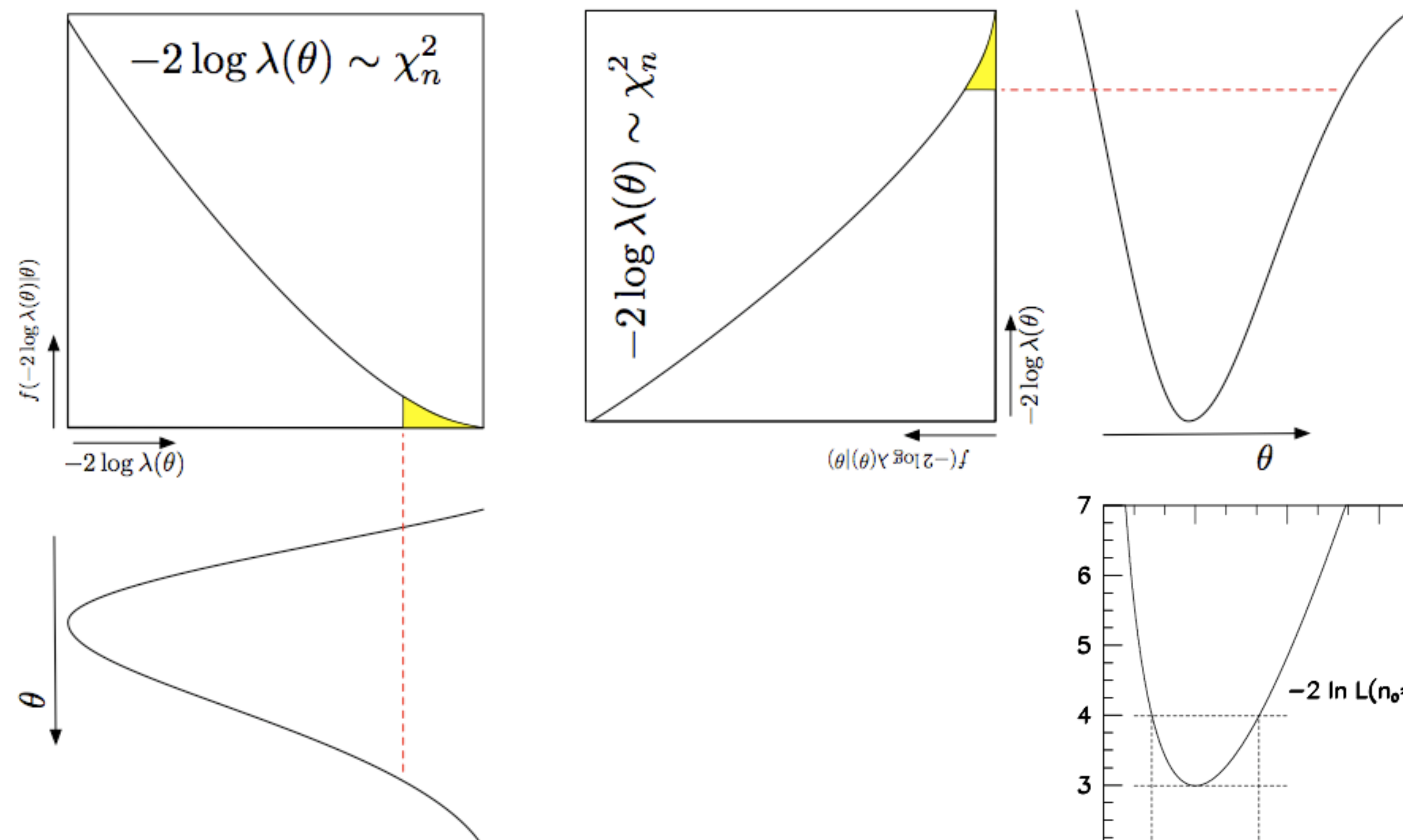
$$-2\log\lambda(\theta_0) = -2\log\frac{f(x|\theta_0)}{f(x|\theta_{best}(x))}$$

negative logarithm of test statistic approaches χ^2-distribution
in the asymptotic limit (central limit theorem)
with n degrees of freedom equal to parameters of interest!

$$-2\log\lambda(\theta) = \chi^2_n$$

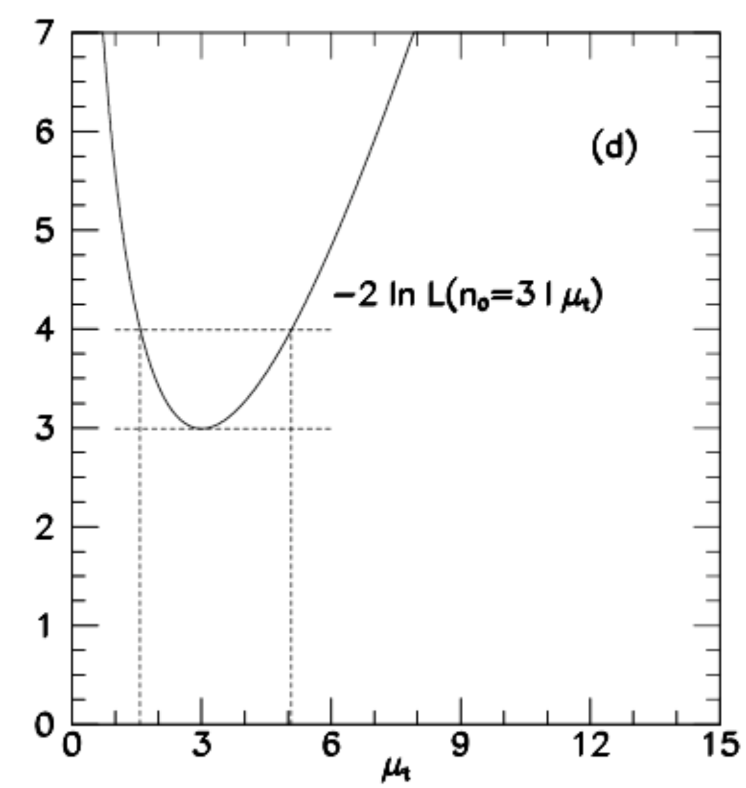
Wilk's Theorem

$$-2\log\lambda(\theta) = \chi^2_n$$



Figures from
Kyle Cranmer

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



p-Value Correspondence for χ^2_n

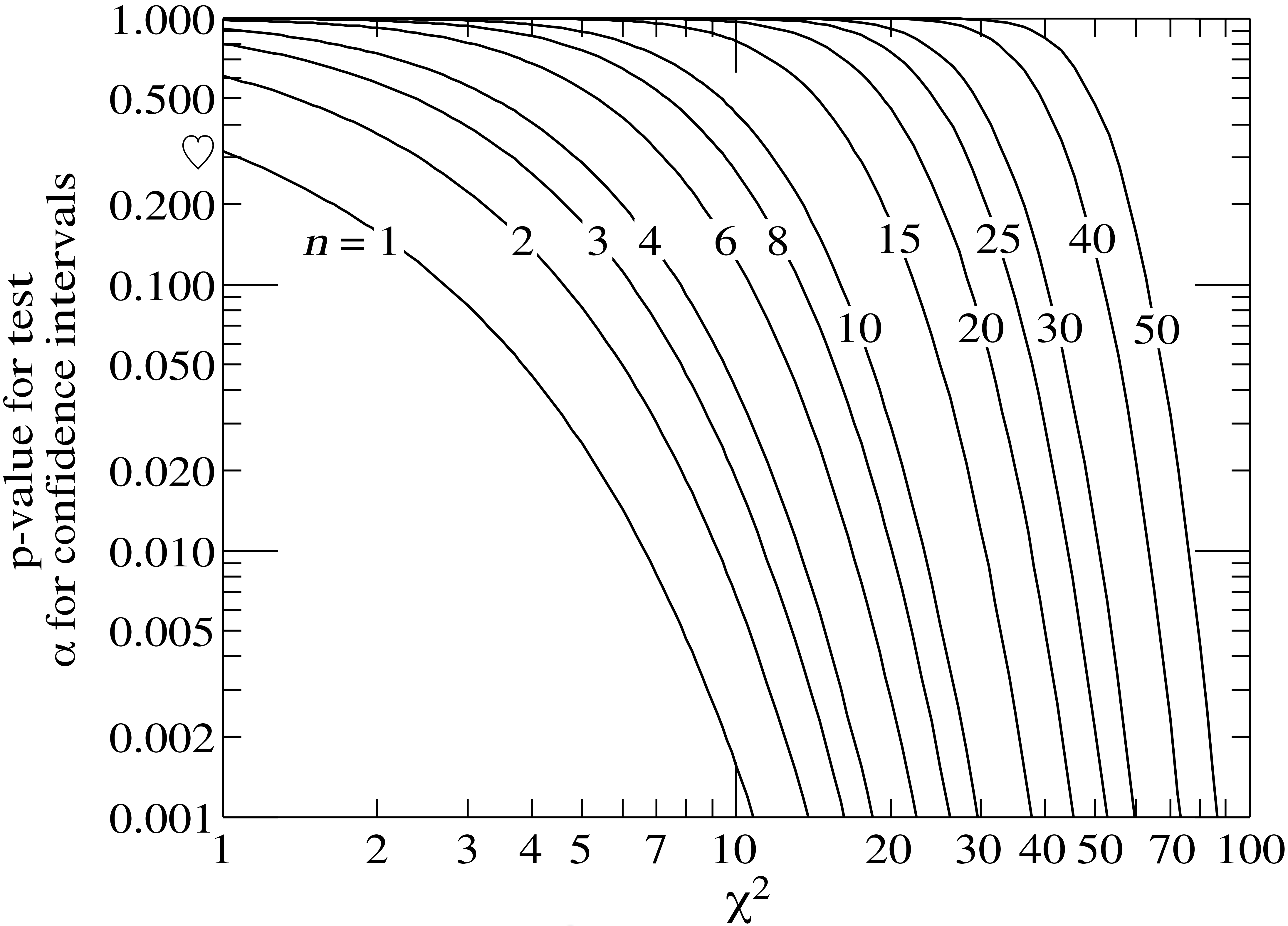
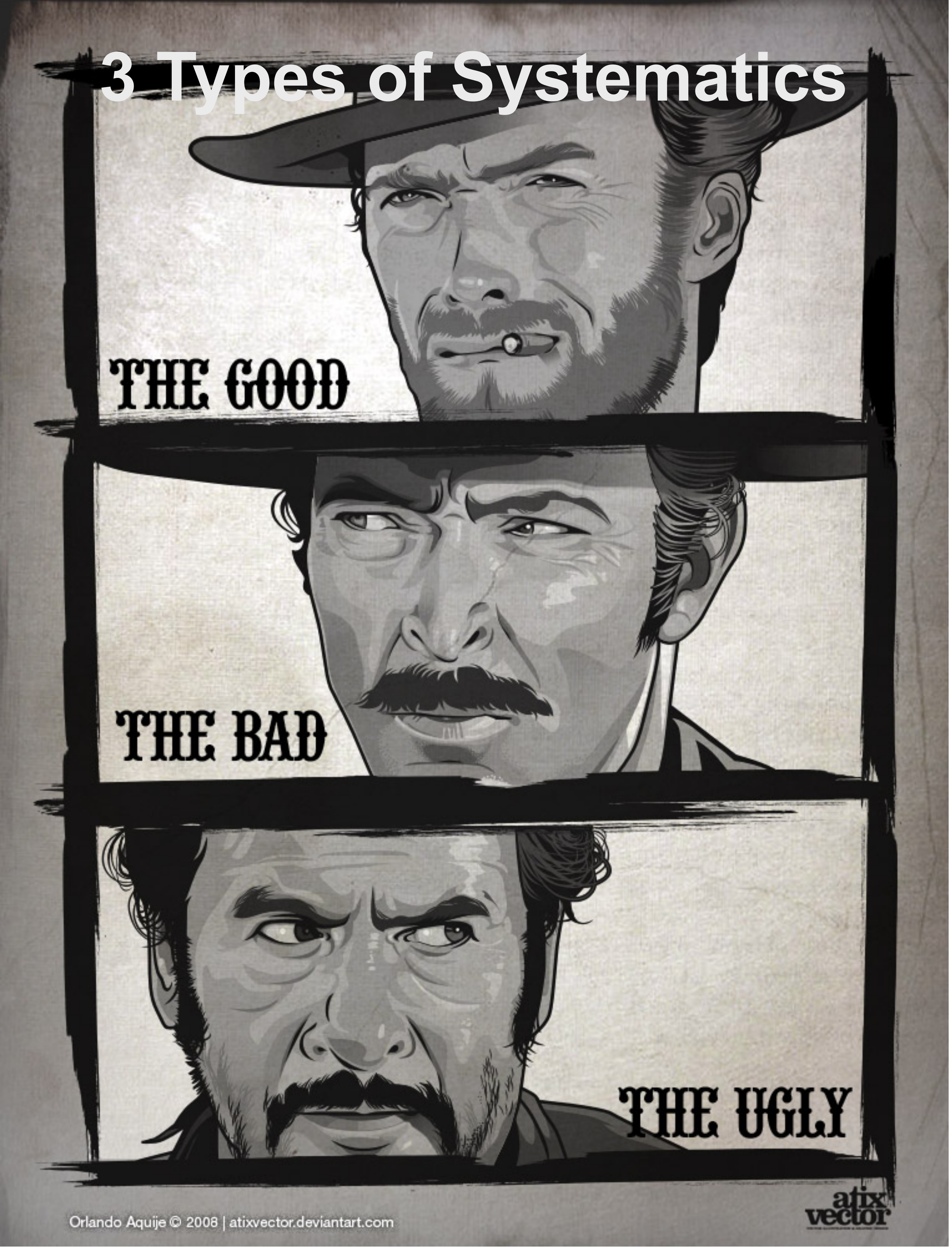


Figure 2: Probabilities to observe a χ^2 equal or larger than the given one for different degrees of freedom n (from the PDG).

3 Types of Systematics



constrain via sideband/
control region measurement

statistical uncertainty
scale with lumi

from model assumptions/
poorly understood features

shape systematics
don't scale with lumi

from underlying paradigm

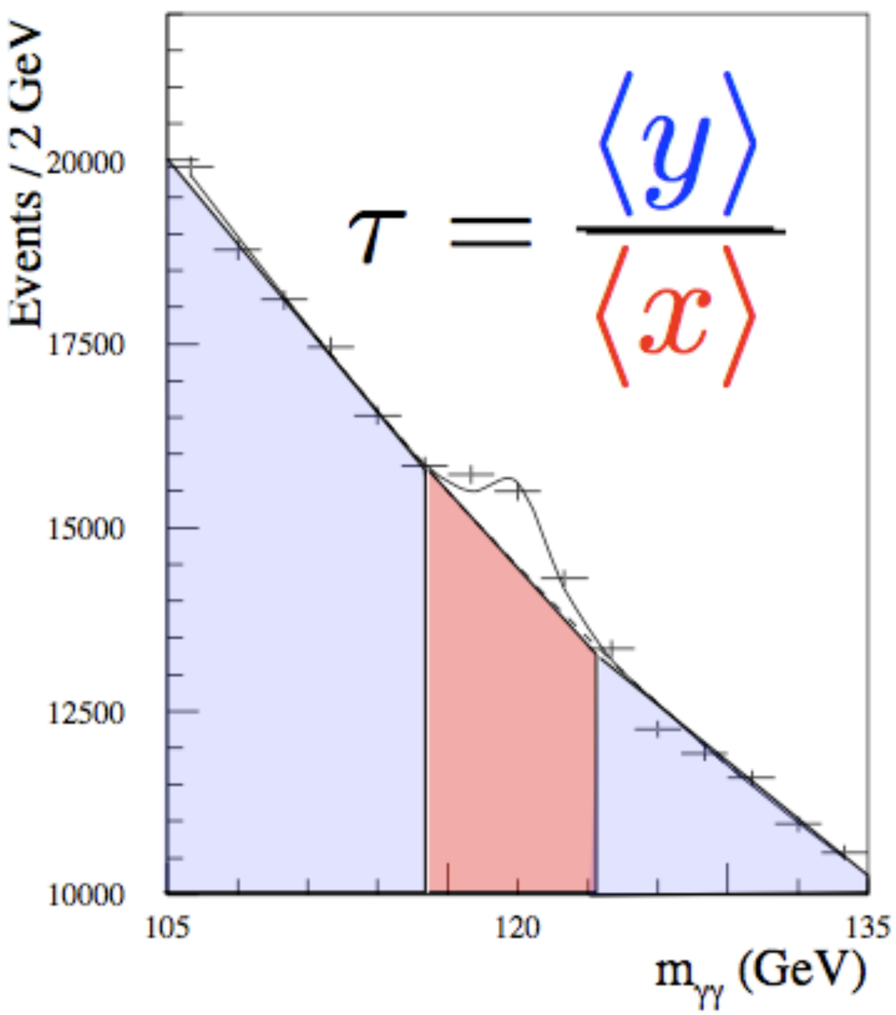
philosophical issue

Constrain Systematics

Typically, we consider an auxiliary measurement y used to estimate background (Type I systematic)

- eg: a sideband counting experiment where background in sideband is a factor τ bigger than in signal region

$$L_P(x, y|\mu, b) = Pois(x|\mu + b) \cdot Pois(y|\tau b).$$



$$\tau = \frac{\langle y \rangle}{\langle x \rangle}$$

(taken from Kyle Cranmer)

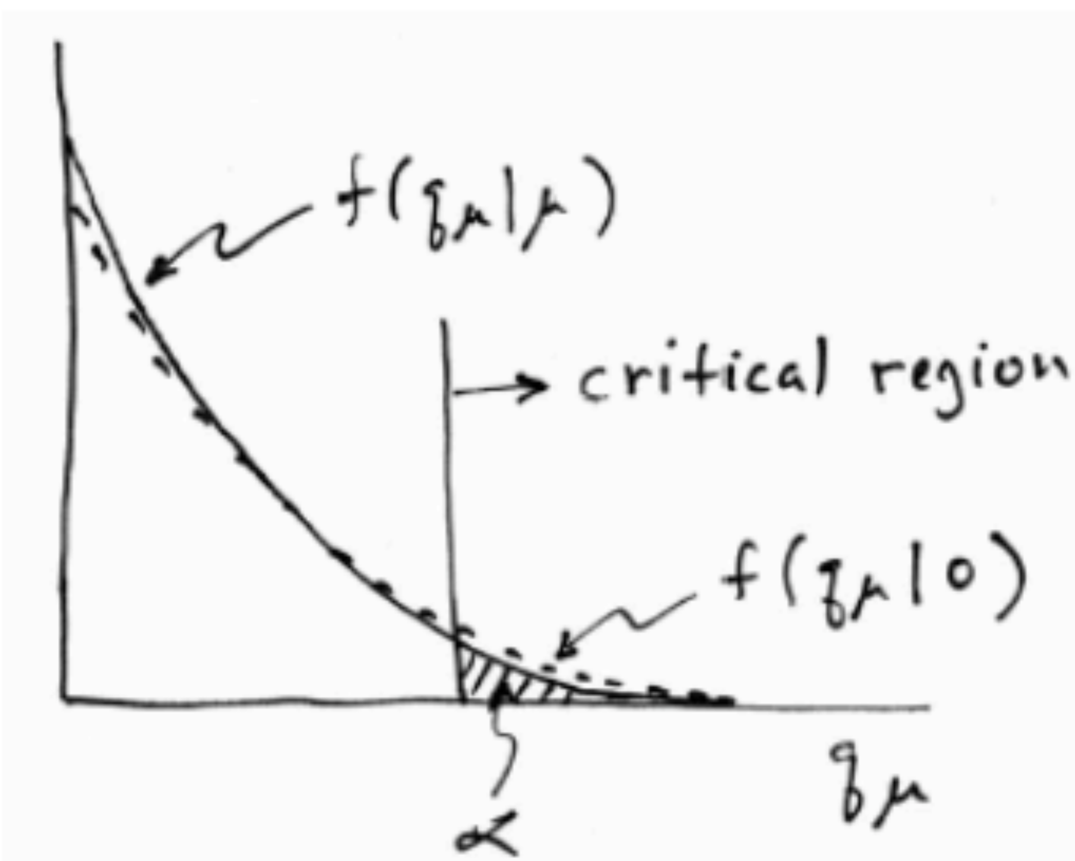
can convert systematic error into statistical one turn "The Bad" into "The Good"

Few Words on Sensitivity Issue

Spurious exclusion

Consider again the case of low sensitivity. By construction the probability to reject μ if μ is true is α (e.g., 5%).

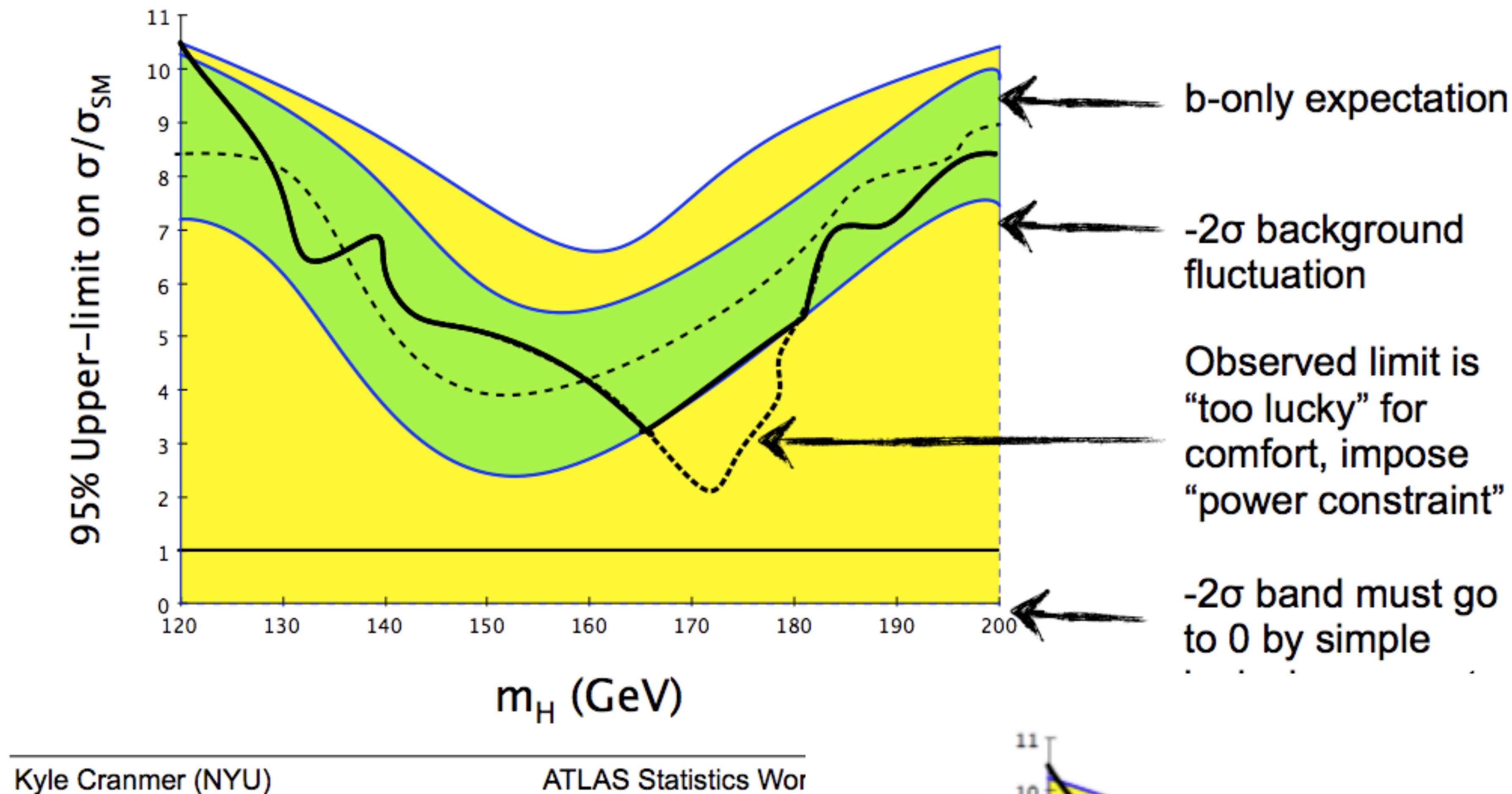
And the probability to reject μ if $\mu = 0$ (the power) is only slightly greater than α.



This means that with probability of around $\alpha = 5\%$ (slightly higher), one excludes hypotheses to which one has essentially no sensitivity (e.g., $m_H = 1000$ TeV).

"Spurious exclusion"

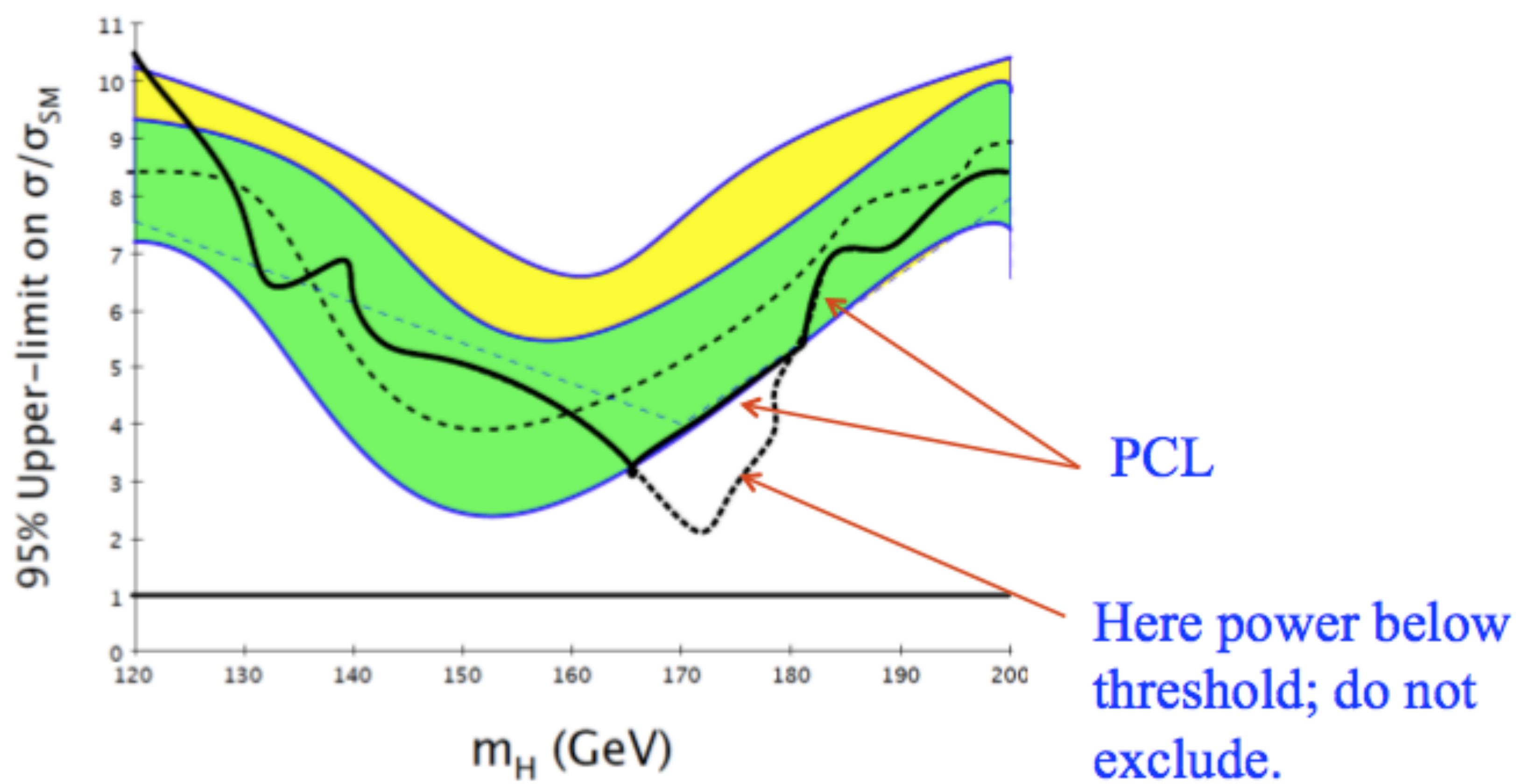
Recommendation to use PCL



from Kyle Cranmer

from Glen Cowan

(ATLAS Stat. Workshop 15.04.11)



what i couldn’t talk about

(over-/ under-) Coverage

Flip-Flopping

Look-Elsewhere Effect

Power Constraint Limits

an much more stuff that can be said about limits ....

conclusion

Hypothesis Testing

Error Classification

Power and Size of Test

Neyman-Pearson Lemma

Test Statistics

Wilk's Theorem

Systematics

now

Hands-On part

have a look at my wiki page

commonly used limit implementation

fortran routines for CL_s
written by Tom Junk
combination of search channels (eclsyst.f)

Nuclear Instruments and Methods in Physics Research A 434 (1999) 435–443

almost used everywhere

mkdir tutorialtestsite; cd tutorialtestsite;

cp -r ~mherbst/testarea/junklimit .

try it out: ./junklimit/testeclsyst

cp -r ~mherbst/testarea/cernlib .

change junklimit/testeclsyst.f and compile

also some bayesian codes (don't know myself)

RooFit/ RooStats

RooStats: Framework for the Collection of Statistical Methods

RooFit: Complex fit-machinery, maybe used in any aspect of hep

RooFit + RooStat:
 unified framework for users (coherence)
 also addresses publishing of Statistical Results

RooFit/ RooStats

The Prototype Problem in RooFit/RooStats

Early in the RooStats project, we considered this prototype problem

$$L_P(x, y|\mu, b) = Pois(x|\mu + b) \cdot Pois(y|\tau b).$$

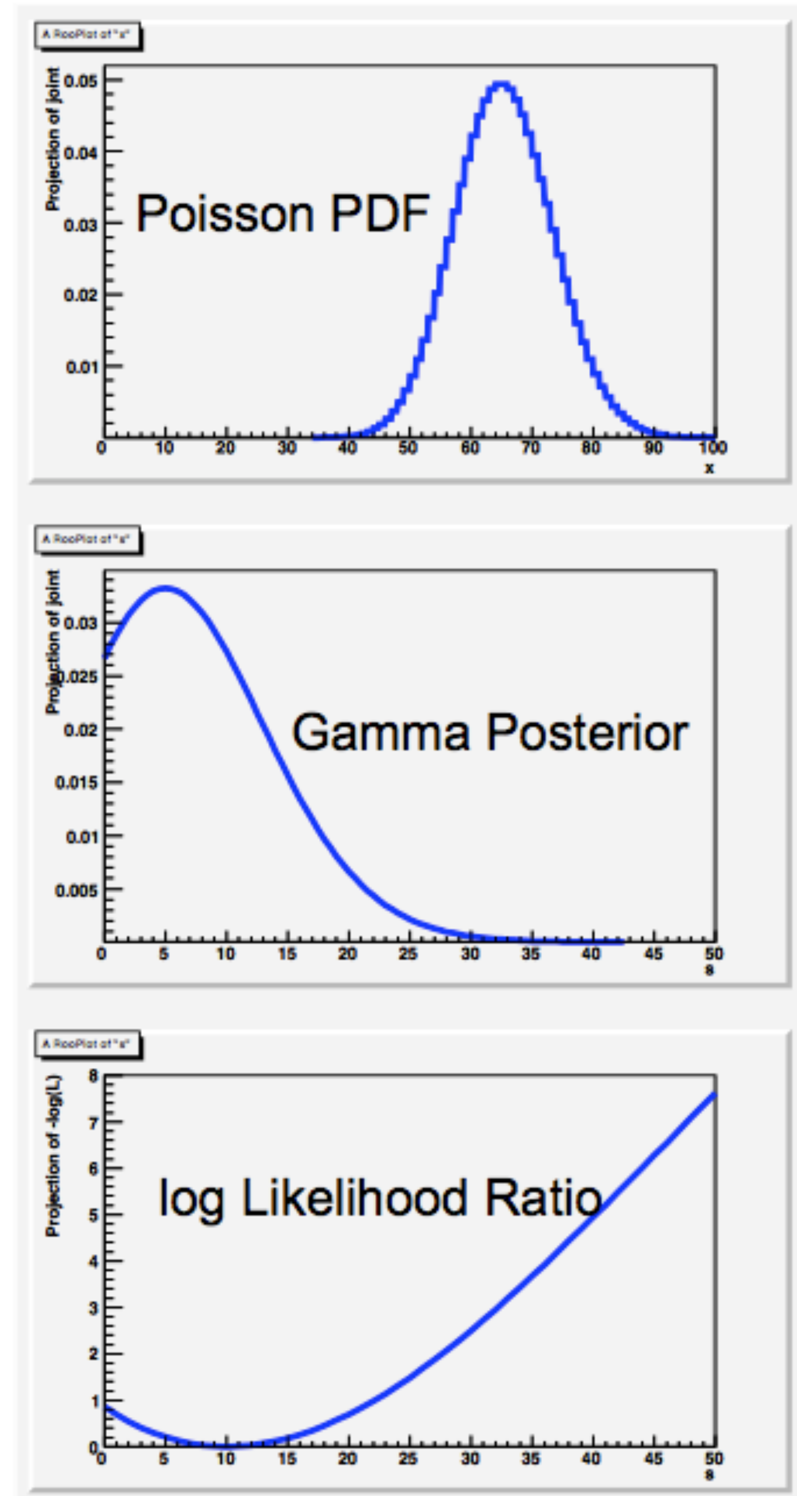
Easy to code up using RooFit:

```
RooRealVar   s("s","s",_s, 0., 100.);
RooRealVar   b("b","b",_b,  0., 200.);
RooRealVar  tau("tau","tau",_tau,0,2);
tau.setConstant(kTRUE);
RooFormulaVar  splusb("splusb","s+b",RooArgSet(s,b));
RooProduct   bTau("bTau","b*tau",RooArgSet(b, tau));
RooRealVar   x("x","x",_s+_b, 0., 200.);
RooRealVar   y("y","y",_b*_tau,  0., 200.);

RooPoisson sigRegion("sigRegion","sigRegion",x,splusb);
RooPoisson sideband("sideband","sideband",y,bTau);

RooProdPdf joint("joint","joint", RooArgSet(sigRegion,sideband) );
```

Easy to obtain relevant plots in three different approaches



Preparations to use ATLAS combination package

sorry only for ATLAS users

check out ATLAS Combination repository

svn co svn+ssh://svn.cern.ch/reps/atlasgrp/Physics/
SUSY/Analyses/Combination/trunk

init environment for ATHENA and root

source $AtlasSetup/scripts/asetup.sh 16.5.0

make library

cd trunk/Tools; make;

load Library in Macro (if you can't have your own...)

gSystem->Load("~mherbst/testarea/tutorial/trunk/lib/
libCombinationTools.so");

Making the Workspace

from Combination svn: trunk/Tools/MakeWorkSpaceOneChannel.cxx

MakeWorkSpaceOneChannel (
filename , suffix ,

data, // observartion in signal region
back_exp, // background expectaion in signal region
b_exp_gauss_sigma, // Absolute uncertainty on SM background only (without JES etc)
ds_JES_numb, // Rel. effect of 1 sigma variation from JES : for signal in signal region
db_JES_numb, // Rel. effect of 1 sigma variation from JES : for SM background in signal region
ds_lumi_numb, // Rel. effect of 1 sigma variation from lumi : for signal in signal region
db_lumi_numb, // Rel. effect of 1 sigma variation from lumi : for SM expectation in signal region
sig_exp, // signal expectation in signal region
sig_eff, // Rel. effect on 1 sigma variation from eg theory uncertainty: On signal in signal region

copy makeWorkspace.C Macro

cp ~mherbst/testarea/tutorial/makeWorkspace.C .

The Model for the PDF

```
RooFormulaVar * s= new RooFormulaVar("s","@0*(1.+@1*@2+@3*@4+@5*@6)*@7",
RooArgSet(*mu,*ds_lumi,nuis_lumi,*ds_JES,nuis_JES,*ds_sigeff,nuis_sig,*sig_exp_var));

RooFormulaVar *b = new RooFormulaVar("b","@0*@1*(1.+@2*@3+@4*@5+@6*@7)",
RooArgSet(*back_exp_w0_var,*gauss_back_mean_var,*db_lumi,nuis_lumi,
                *db_JES,nuis_JES,*gauss_back_sigma_var,*nuis_back_chan));

RooFormulaVar * s_plus_b= new RooFormulaVar("s_plus_b","@0+@1",RooArgSet(*s,*b));
```

Analysing the Workspace

cp ~mherbst/testarea/tutorial/analyseWorkspace.C .

play with RooStat tutorials:

~mherbst/testarea/tutorial/roostattuts/

or

$ROOTSYS/tutorials/roostats/