Statistical issues in Particle Physics analyses

Louis Lyons
Imperial College and Oxford

Banff **July 2010**

Particle Physics + This Week's Topics & Talks

1) What is Particle Physics?

2) Search experiments

Wilks' Theorem. Cunning MC

Look Elsewhere Effect

p-values, likelihoods

5σ for discovery?

Exclusion, Upper Limits/bounds, no discrimination, CL_s

Banff Challenge 2

Systematics. Partons

Combining p-values

3) Particle Physics and Statisticians

Particle Physics

- What it is
- Typical experiments
- Typical data
- Typical analysis

What is it?

Search for ultimate constituents of matter

Ancient tradition

What constitutes acceptable theory?

“Number of fundamental entities at most

A

F

E

W ”

Early ideas due to Greeks

What is it?

Search for ultimate constituents of matter

Ancient tradition

What constitutes acceptable theory?

"Number of fundamental entities at most

A ir

F ire

E arth

W ater "

How many?

Greeks' AFEW	4
Dalton's atoms	~20 → 120
e and p	2
and n, π, μ, γ, Ω…	>1000 "Elementary" particles
Quark model	3 → 35
Future substructure?	1?

How big?

Human ~ 1 metre

Dust ~10^{-5} metres

Atom ~ 10^{-10} metres

Nucleus ~ 10^{-15} metres

Quarks <10^{-18} metres

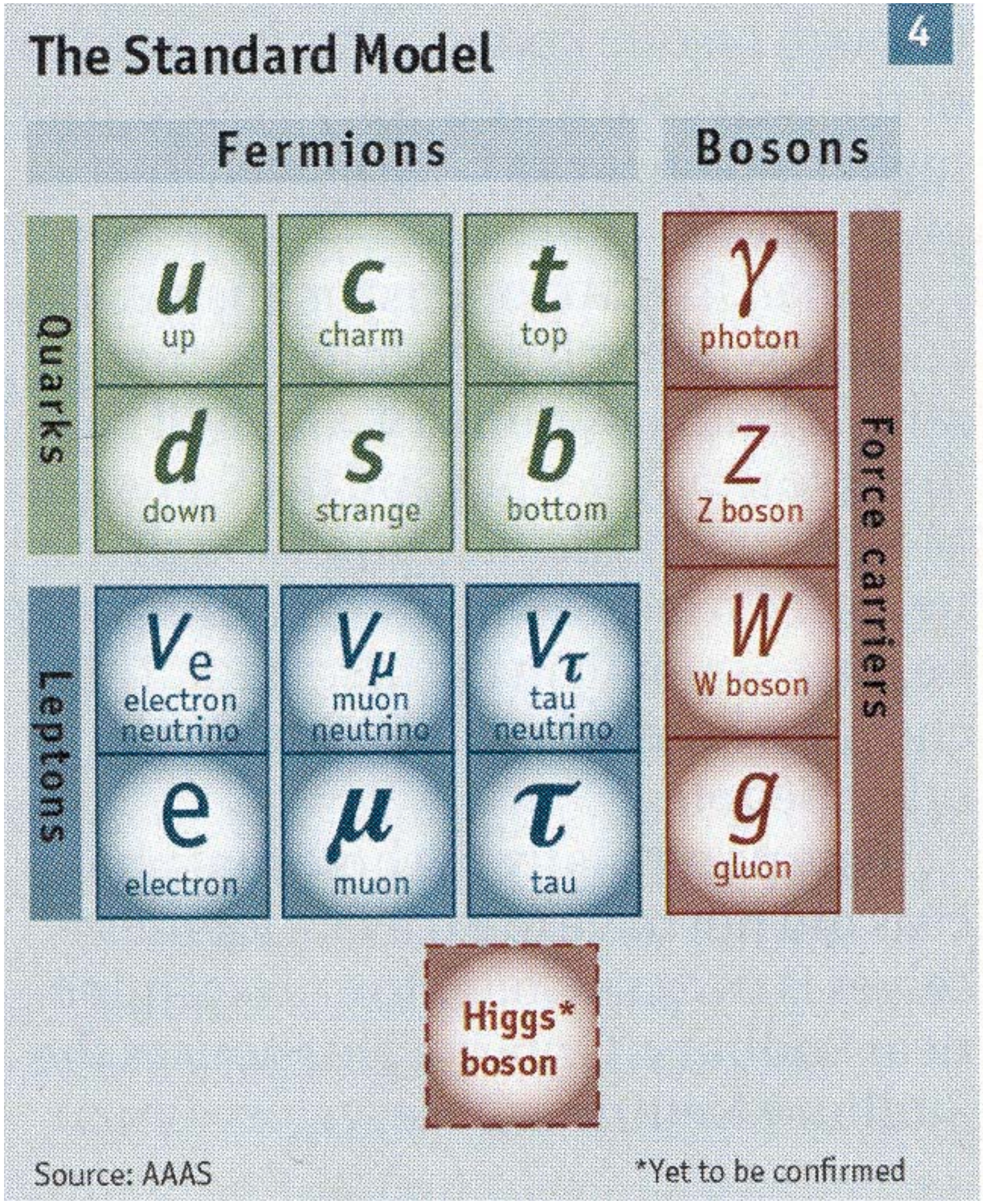
Mass of proton = 1 GeV = 10^{9} eV

neutrino < 2 eV

top = 175 GeV

Forces of nature

1) Gravity (mediated by Gravitons,   )
2) Electromagnetism (Photons, γ)
3) Nuclear strong force (Gluons, g)
4) Nuclear weak force (Int Vect Bosons, W Z)


The Standard Model
4
Fermions
Bosons
Quarks
u
up
c
charm
t
top
d
down
s
strange
b
bottom
Leptons
νe
electron neutrino
νμ
muon neutrino
ντ
tau neutrino
e
electron
μ
muon
τ
tau
γ
photon
Z
Z boson
W
W boson
g
gluon
Force carriers
Higgs* boson
Source: AAAS
*Yet to be confirmed

Elementary particles:

Force carriers: photon, W and Z, graviton, gluons

Higgs for the masses

LEPTONS: e μ τ and neutrinos

Composite particles:

HADRONS:

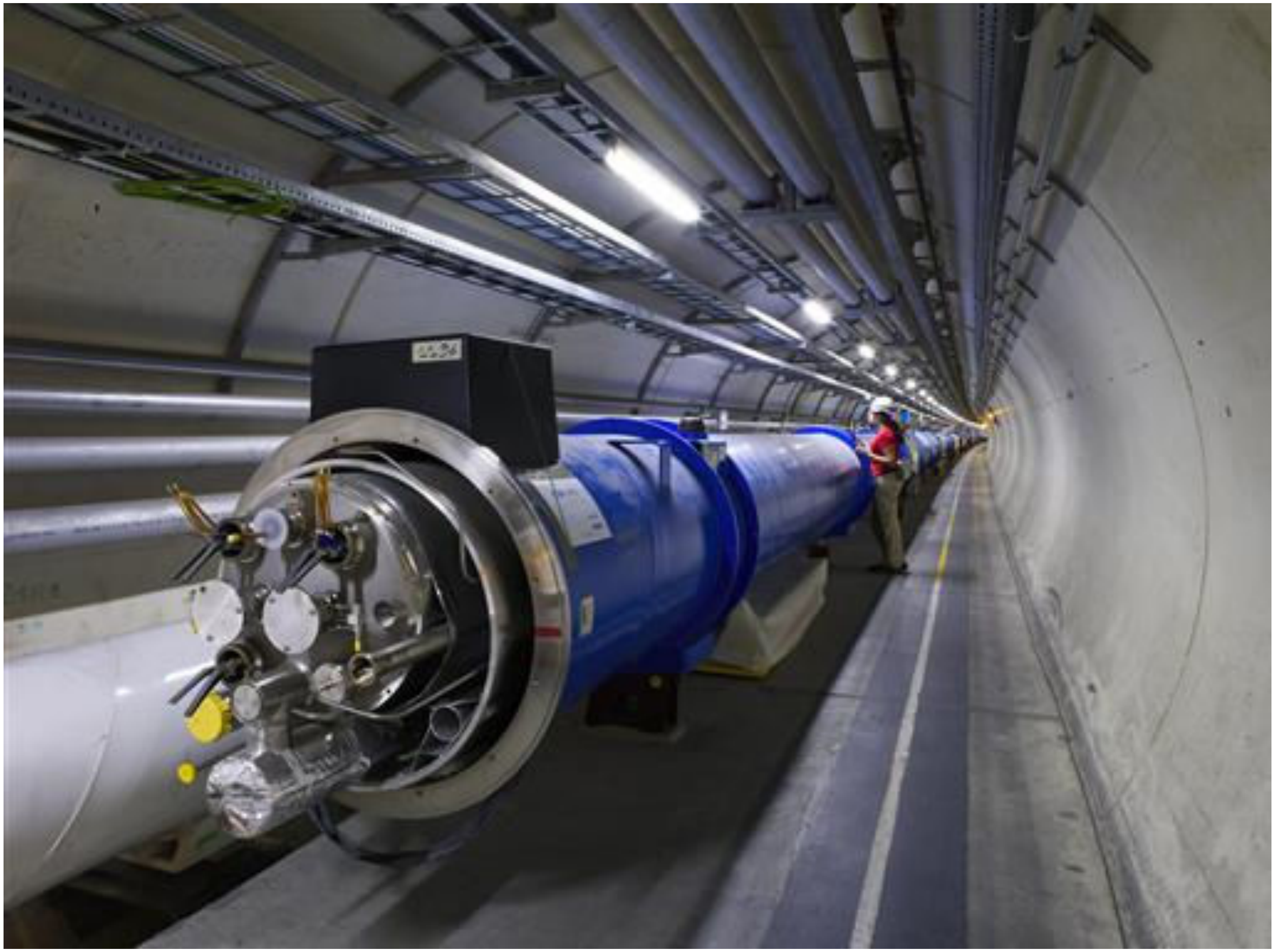
Made from qqq or q $\bar{q}$

e.g. proton = uud, $\pi^+ = u\,\bar{d}$

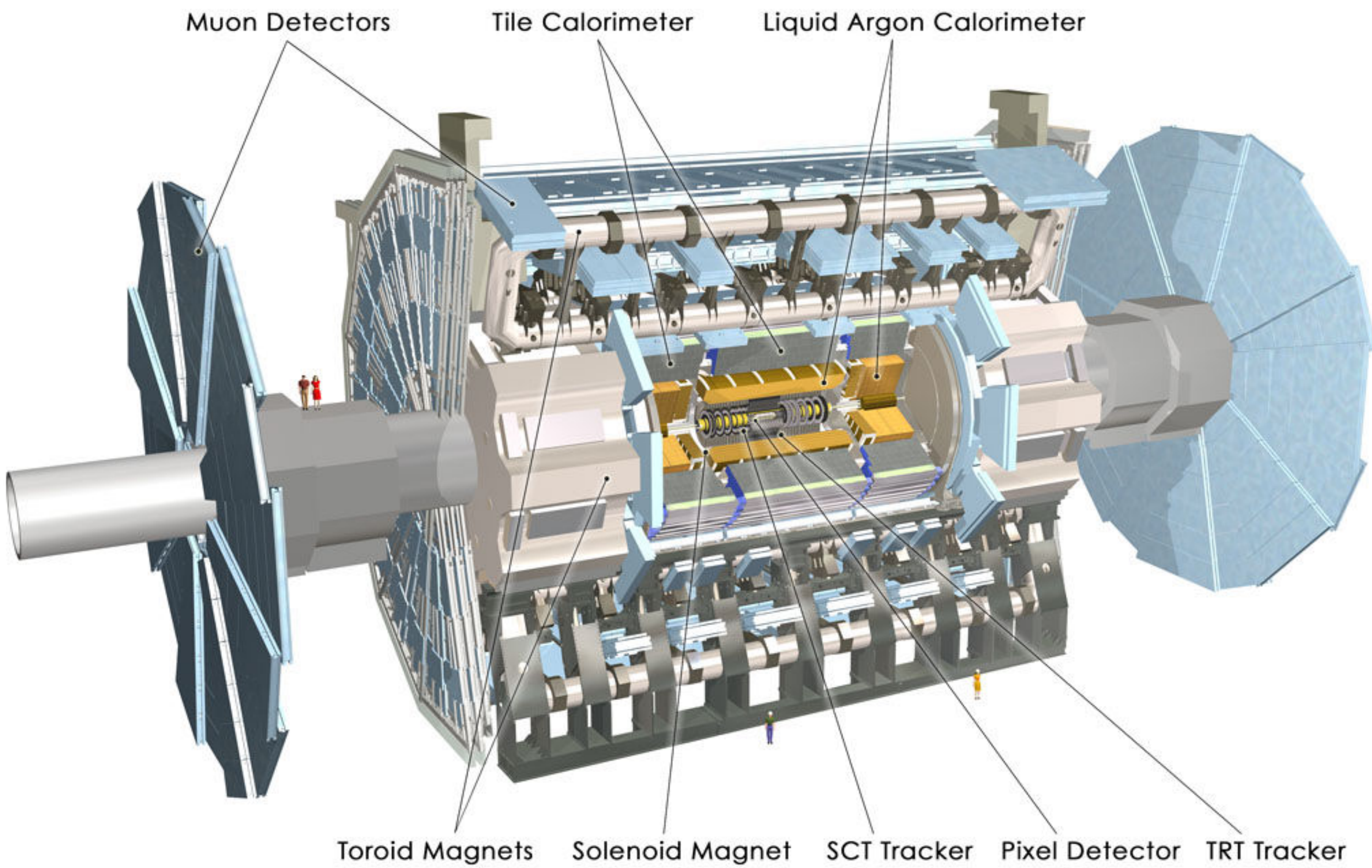
Typical Experiments

Experiment	Energy	Beams	# events	Result
LEP	200 GeV	e+ e-	10^7 Z	$N_\gamma = 2.987 \pm 0.008$
BaBar/Belle	10 GeV	e+ e-	10^8 B anti-B	CP-violation
Tevatron	2000 GeV	p anti-p	"10^{14}"	SUSY?
LHC	14000 GeV	p p	100/sec	Higgs?
K→K	~3 GeV	ν_μ	100	ν oscillations

In the LHC tunnel

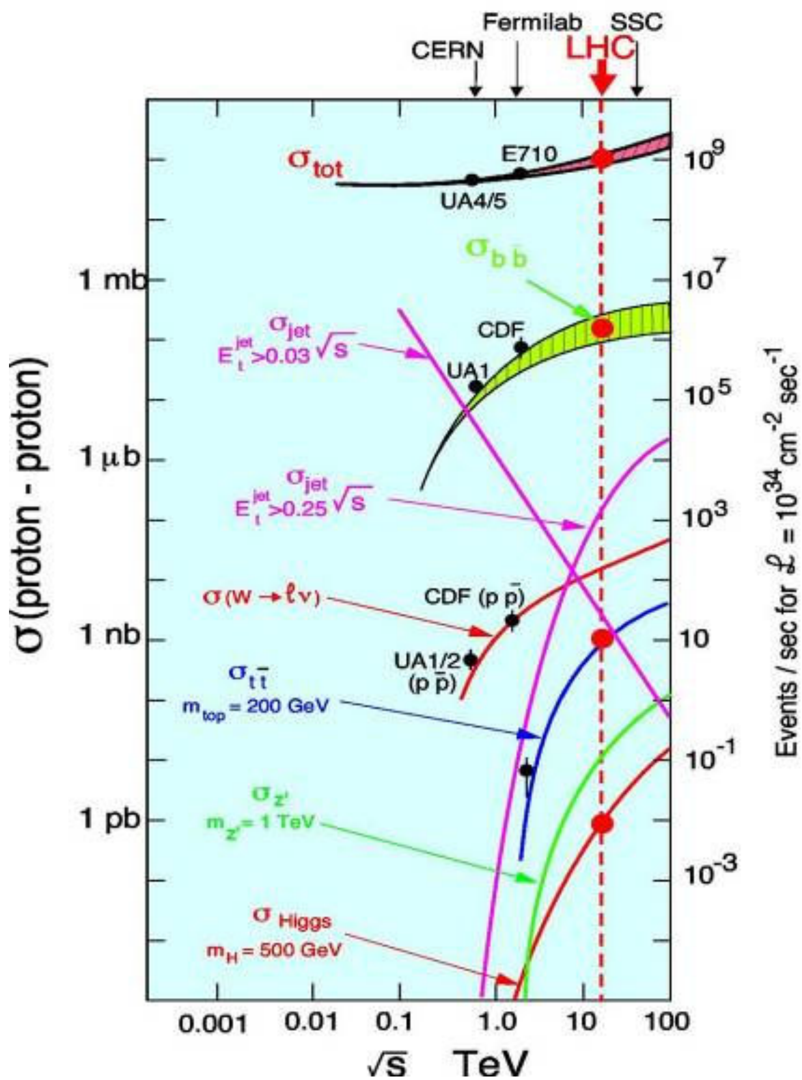


ATLAS Detector at LHC



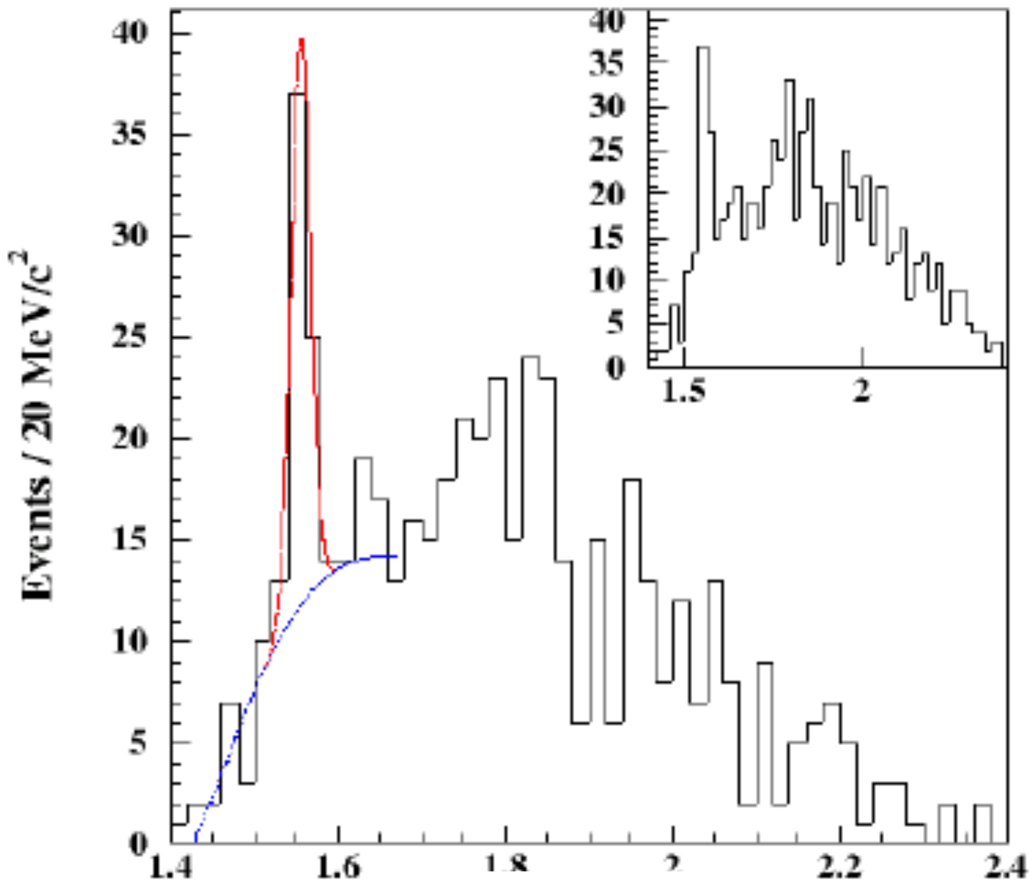
45 metres long, 25 m high. 3000 physicists from 40 countries

Interesting Physics is rare



Typical Analysis

Hypothesis testing: Peak or statistical fluctuation?



Read more

LL, "Open statistical issues in Particle Physics", Annals of Applied Statistics **2** (2008) 887

PHYSTAT-LHC and earlier workshops

CDF Statistics Committee web-page

Next Workshop
CERN, Jan 2011

Topics de-emphasised at Banff

Separating wanted events from background

1) On-line trigger
2) Reduce sample size by ‘cuts’
3) Apply machine learning technique

GoF for sparse (unbinned) multi-dim data wanted

Bayes or Frequentist (or other)?

"Particle Physicists are last living Frequentist fossils" Michael Goldstein, PHYSTAT 2002

Actually tend to be pragmatic, using one or another as convenient, or even both at the same time {Bayesian treatment of systematics in frequentist analysis}

"It doesn't matter whether method's motivation was B or F, just study its performance" Peter Clifford

We (except for D+J+Prosper) tend to use priors that are constant in the first variable we can think of.

Monte Carlo simulations in HEP

Simulate almost every aspect of experiment, detector, analysis

e.g. Tracking

Overlying events

Machine learning (separate signal and bgd)

Analysis technique

Statistical analysis (Michael Woodroofe)

Search Experiments

Look for New Physics

a) Measured parameter different from expected

e.g. speed of earth through "aether"; or

observed number of events

b) Observed distribution unexpected

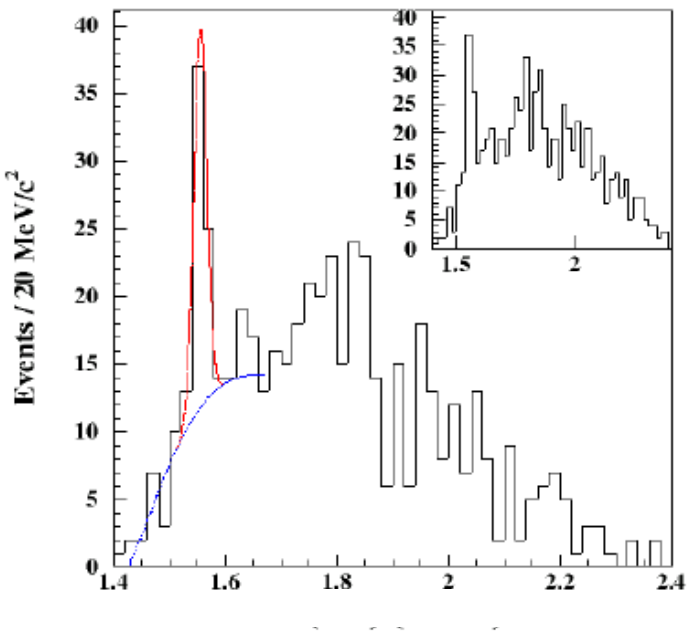
e.g. peak, rather than smooth distribution; or

enhancement; or

oscillatory behaviour; etc.

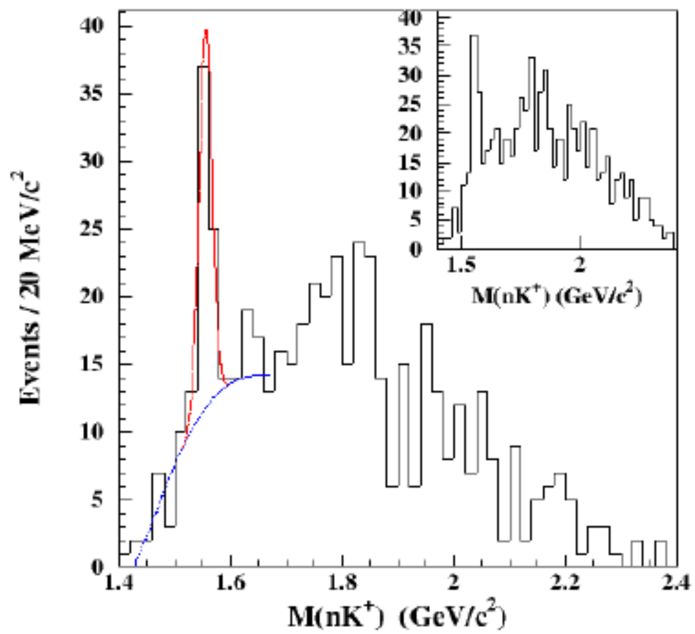
Use statistic (e.g. a likelihood ratio)

Typical Analysis



Is there evidence for a peak in this data, or is it a statistical fluctuation?

Is there evidence for a peak in this data?



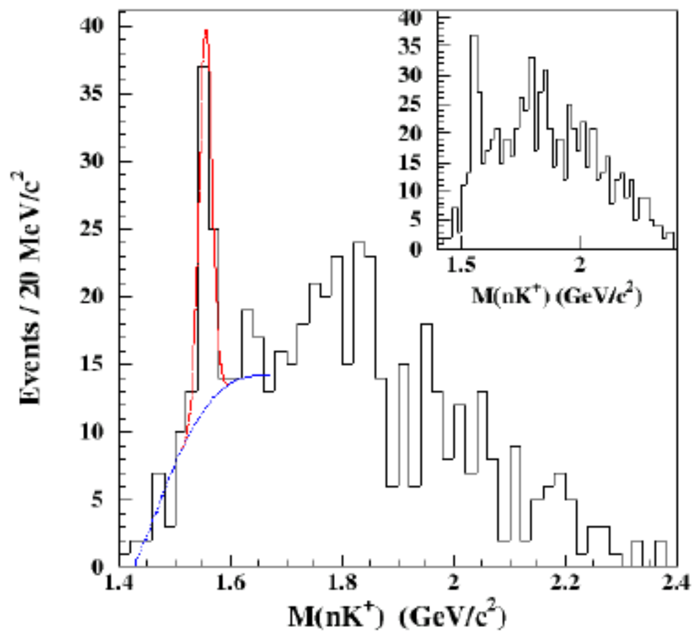
"Observation of an Exotic S=+1

Baryon in Exclusive Photoproduction from the Deuteron"

S. Stepanyan et al, CLAS Collab, Phys.Rev.Lett. 91 (2003) 252001

"The statistical significance of the peak is 5.2 ± 0.6 σ"

Is there evidence for a peak in this data?



"Observation of an Exotic S=+1 Baryon in Exclusive Photoproduction from the Deuteron"

S. Stepanyan et al, CLAS Collab, Phys.Rev.Lett. 91 (2003) 252001

"The statistical significance of the peak is $5.2 \pm 0.6\ \sigma$"

"A Bayesian analysis of pentaquark signals from CLAS data"

D. G. Ireland et al, CLAS Collab, Phys. Rev. Lett. 100, 052001 (2008)

"The ln(RE) value for g2a (-0.408) indicates weak evidence in favour of the data model without a peak in the spectrum."

Comment on "Bayesian Analysis of Pentaquark Signals from CLAS Data" Bob Cousins, http://arxiv.org/abs/0807.1330

Significance

In counting expts, use Poisson p-value for $n \geq n_{obs}$
Approximations like S/√B not good, especially for optimising analysis

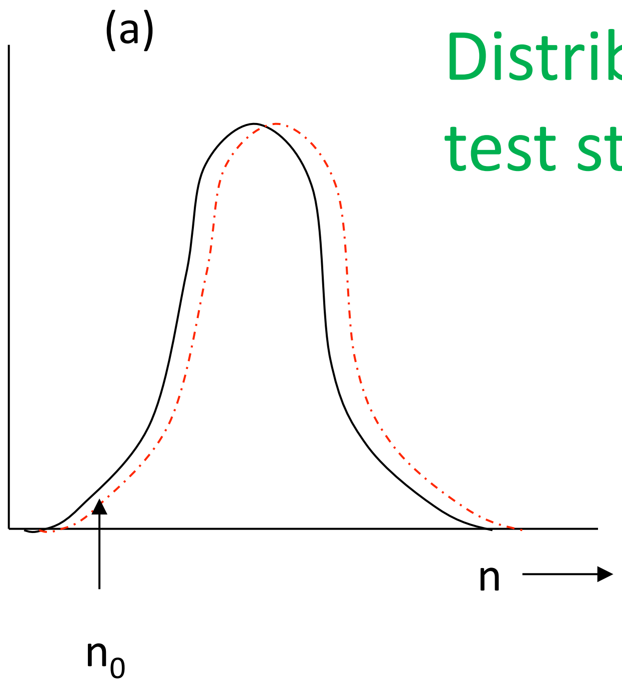
For distributions, use pdf of statistic according to $H_0 \rightarrow p_0$

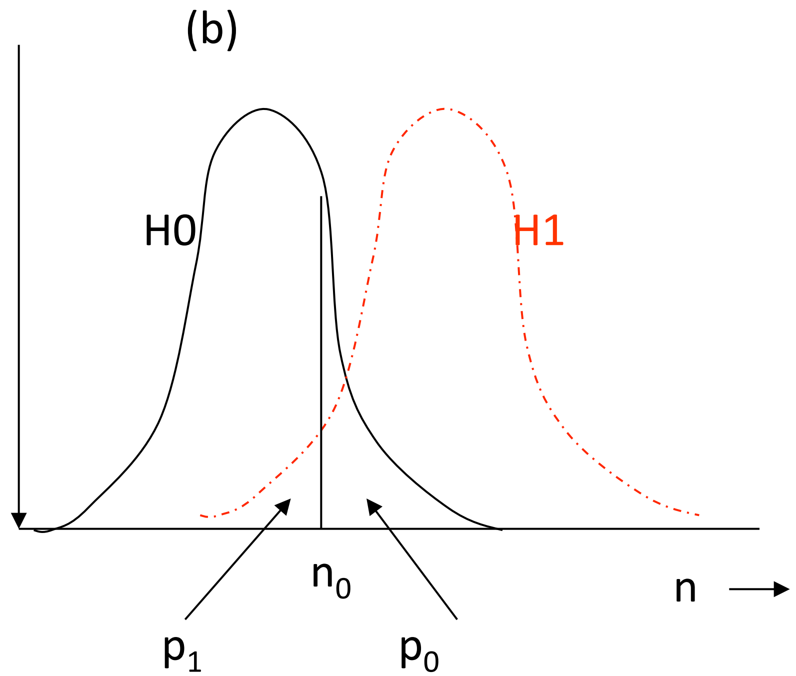
Linnemann

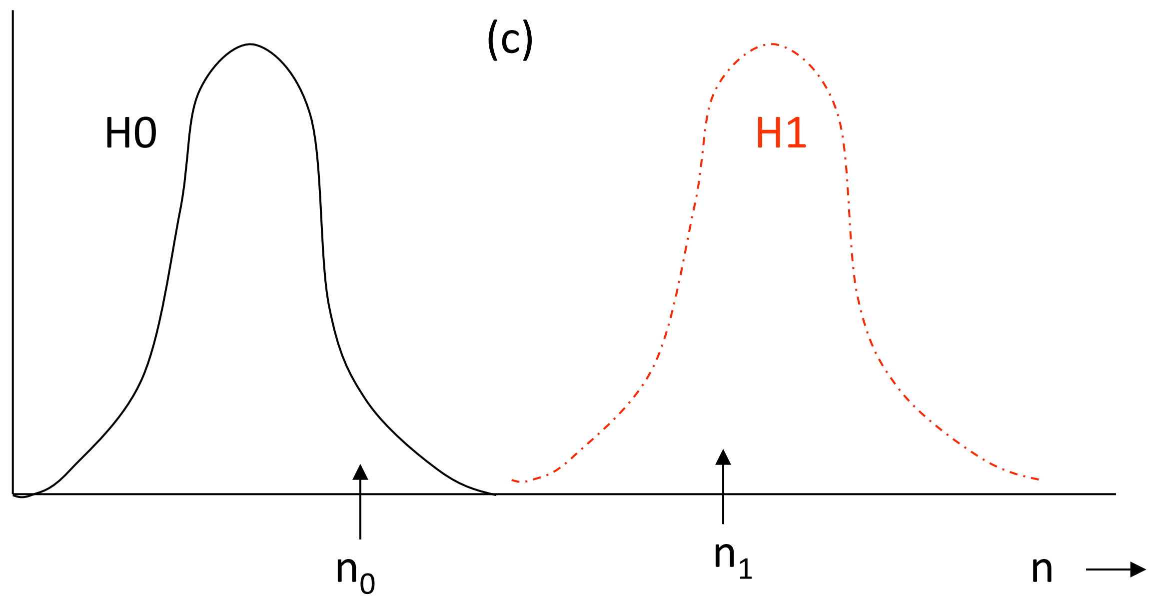
Convert p to σ's according to one-sided Gaussian tail (just convention)

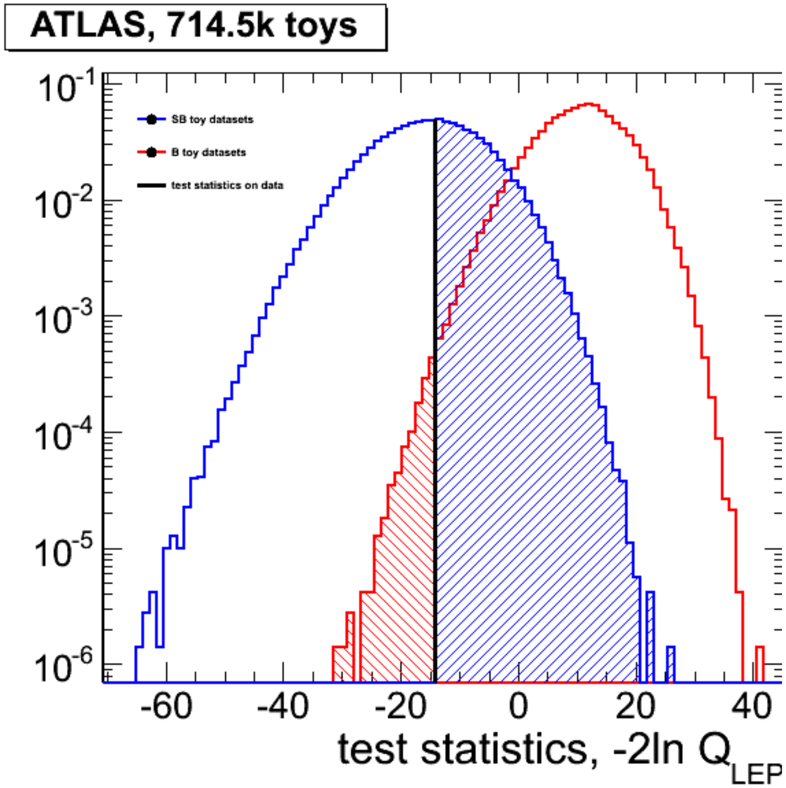
e.g. $p = 3 * 10^{-7} \rightarrow 5.0\ \sigma$

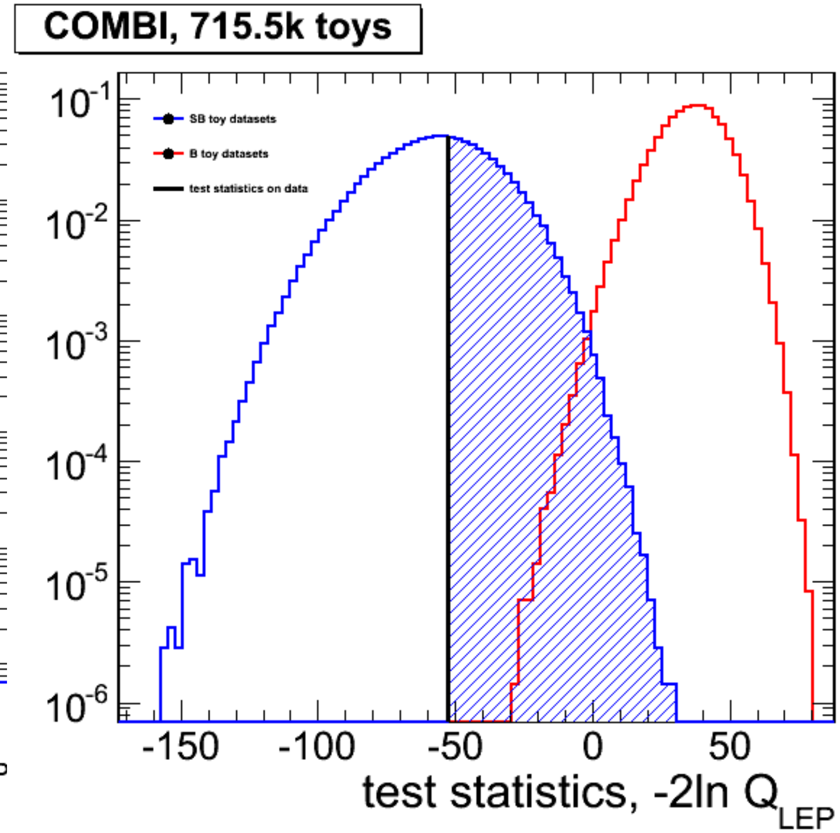
Distribution of test statistic

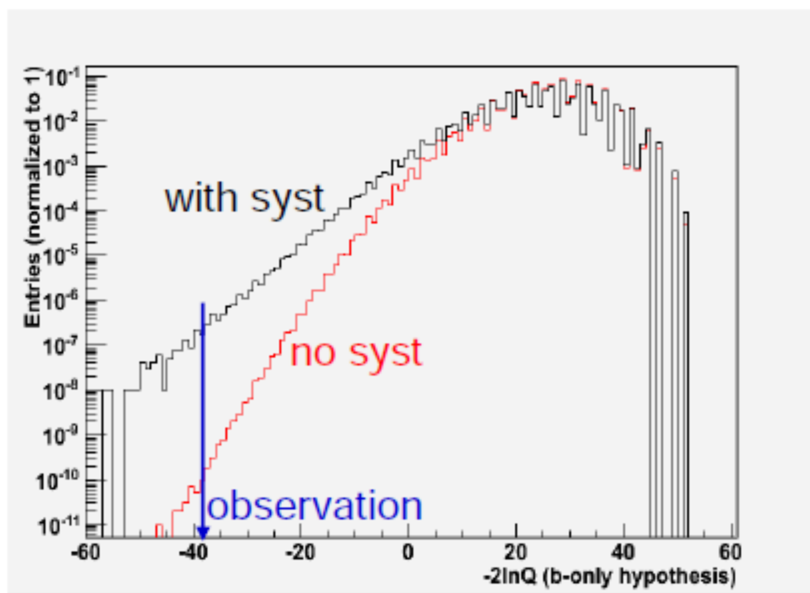







ATLAS, 714.5k toys
SB toy datasets
B toy datasets
test statistics on data
10-1
10-2
10-3
10-4
10-5
10-6
-60
-40
-20
0
20
40
test statistics, -2ln Q_LEP


COMBI, 715.5k toys
SB toy datasets
B toy datasets
test statistics on data
10-1
10-2
10-3
10-4
10-5
10-6
-150
-100
-50
0
50
test statistics, -2ln Q_LEP


Entries (normalized to 1)
with syst
no syst
observation
-60
-40
-20
0
20
40
60
-2lnQ (b-only hypothesis)

Which Statistic?

For counting expt, statistic = n

For distribution, use likelihood ratio

BEWARE: Likelihood ratio is ambiguous

$p(x;\ \mu=1,\ \nu)\ /\ p(x;\ \mu=0,\ \nu)$ ν fixed

$p(x;\ \mu=1,\ \nu'_1)\ /\ p(x;\ \mu=0,\ \nu'_0)$ Profile L (Cowan)

$P(x;\ \mu',\ \nu')/\ p(x;\ \mu=0,\ \nu)$ For Wilks?

Histogram: $p(n_i;\ \mu_i)/p(n_i;\ \mu_i=n_i) \sim \chi^2$ for one hyp.

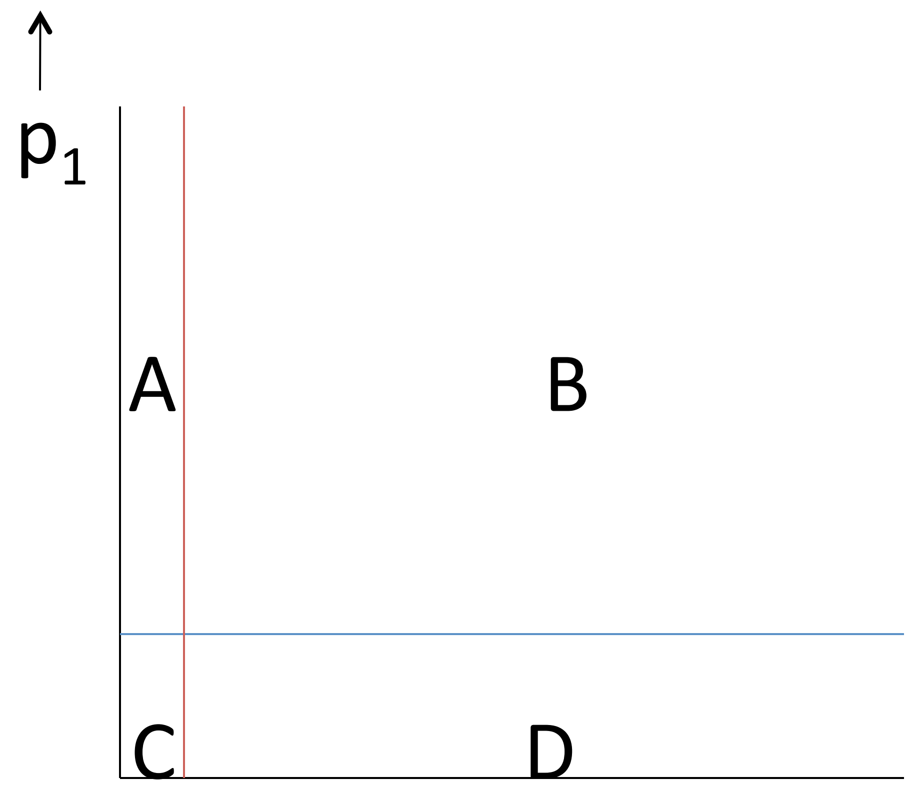
Possible outcomes

A = Reject H_0

B = Make no choice

C = ?

D = Exclude H_1



N.B. Reject/exclude levels really tighter $p_0 \rightarrow$

If H_1 true: D = false exclusion, B+D = Error of 2^{nd} kind (for H_0)

Why 5σ?

- Past experience with 3σ, 4σ,… signals
- Look elsewhere effect:

  Different cuts to produce data

  Different bins (and binning) of this histogram

  Different distributions Collaboration did/could look at

  Defined in SLEUTH
- Bayesian priors:

$$\frac{P(H_0|data)}{P(H_1|data)} = \frac{P(data|H_0) * P(H_0)}{P(data|H_1) * P(H_1)}$$

Bayes posteriors — Likelihoods — Priors

Prior for $\{H_0 = S.M.\}$ >>> Prior for $\{H_1 = \text{New Physics}\}$

Statistician's comment:

"No distribution is valid at 5σ tail"

We believe our distributions are (essentially) Poisson

BUT we always have systematics (nuisance parameters), and these are usually less precisely determined than statistical effects.

CONCLUSION:

It is hard to convince people of 5σ discovery in systematics dominated analysis.

Wilks' Theorem

My understanding:

H_0 and H_1 are nested hypotheses. H_1 has k extra parameters.

If H_0 true, $\chi^2{}_0 - \chi^2{}_1$ distributed like χ^2 with NDF=k provided

a) Asymptotic

b) Extra params all defined for $H_1 \rightarrow H_0$

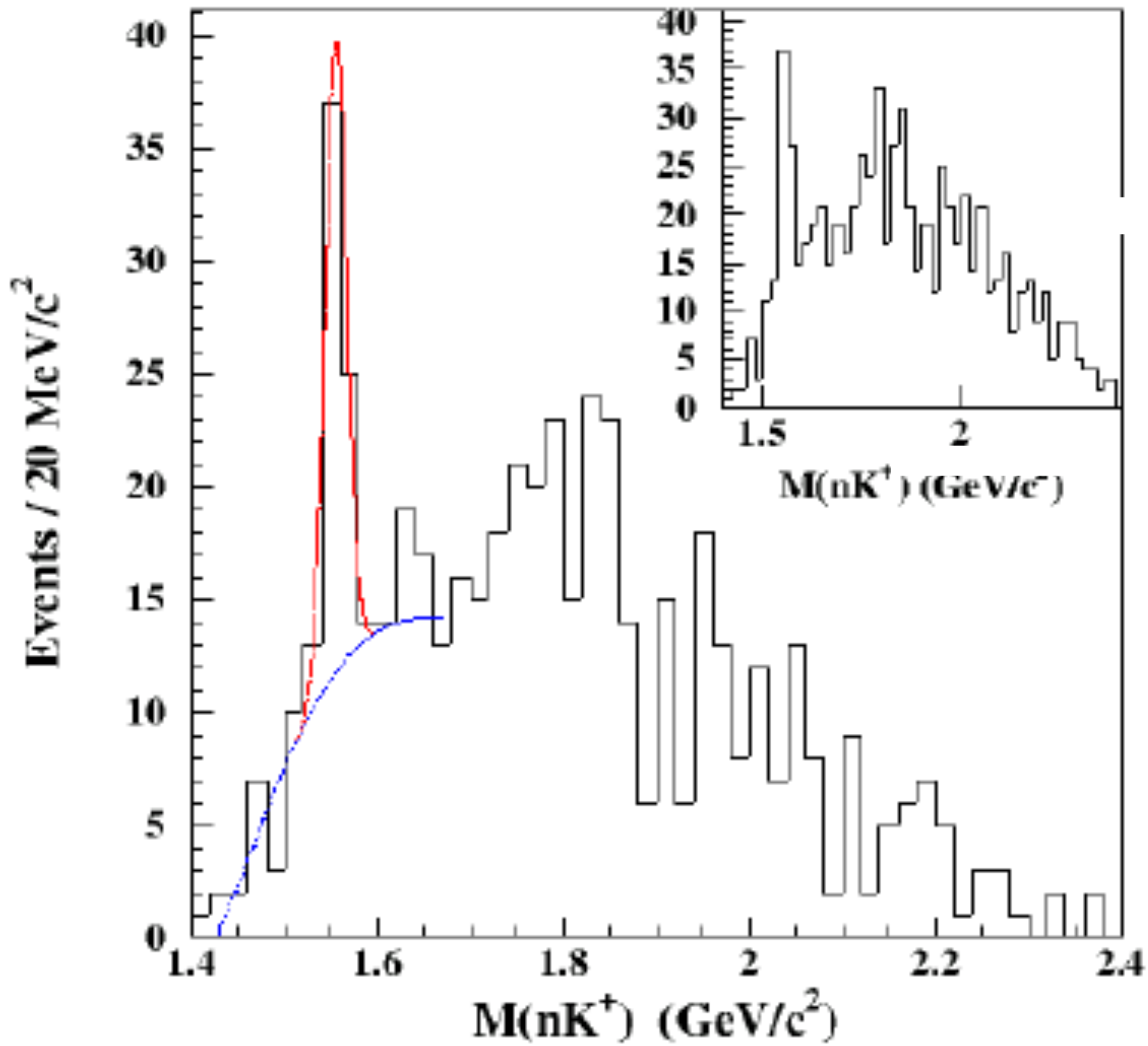
c) Params not on boundary of physical region

Our case: H_0 = smooth background

H_1 = smooth bgd + peak, e.g. $A*\exp(-0.5(x-x_0)^2/\sigma^2)$

Problem b) For A=0, x_0 and σ irrelevant

Problem c) If A forced to be non-negative, A=0 is on boundary



Events / 20 MeV/c^2
M(nK+) (GeV/c^2)
0
5
10
15
20
25
30
35
40
1.4
1.6
1.8
2
2.2
2.4
1.5
2

N.B. If Wilks' Th not true, need lots of Monte Carlo, to assess signif of χ^2 difference

Needs cunning Monte Carlo (Michael Woodroofe)

{Monte Carlos show differing agreement with χ^2. Maybe depends on ....} Elliott Bloom

Cousins: Dorigo's comment that $A_{fit} = 0$ not being very likely

Cowan/Gross/Vitells have insights

Subsidiary question

What is procedure for calculating significance of possible second peak?

Cf: From data on position wobble of star, what is significance for existence of second planet?

Is it just comparing χ^2 for 1 peak with χ^2 for 2 peaks? (Need MC to assess significance)

A and x_0

H_1 = smooth bgd + $A*\exp(-0.5(x-x_0)^2/\sigma^2)$

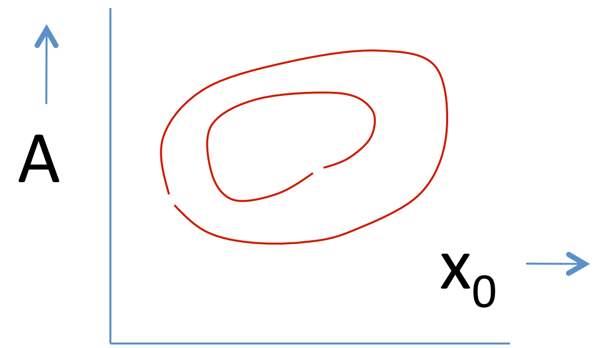
For measurement, vary A and x_0 together

For exclusion, vary A at a series of fixed x_0

For discovery, do either

Signif is chance of obtaining observed effect (or more extreme) at any relevant x_0, assuming bgd only

Likelihood contours



"Look Elsewhere Effect"

Chance of statistical fluctuation giving bump **anywhere**

Where is "anywhere"?

1) Any mass in plot (but some parts may be irrelevant)
2) In other plots considered in analysis
3) Elsewhere in experiment
4) Etc. ?

LHC suggestion:

Local signif + 1) (main number to quote) + 2) (if poss)

LEE grows with number of σ

Gross and Vitells: Perhaps this is plausible.

Look ElseWHEN Effect?

Not necessary to consider, as relevant data is (essentially) all data up to present.

(Might involve ‘undiscovering’)

LEE for exclusion?

Currently quote at each mass separately

Interesting if **whole** mass range is excluded

e.g. 114 -1000 GeV for S.M. Higgs

What confidence level to attach to whole exclusion?

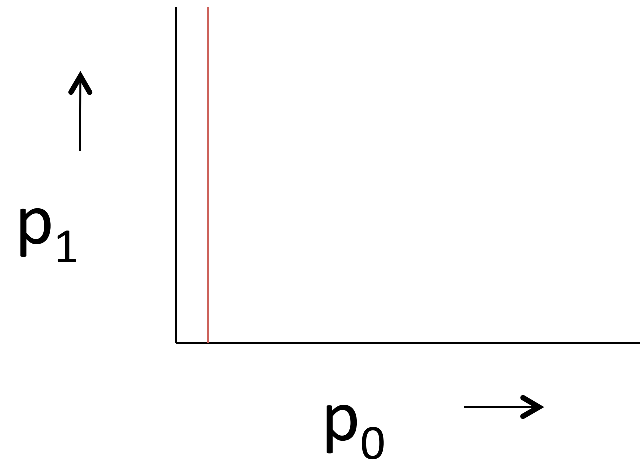
N.B. False exclusions can be

a) When excluded particle actually exists

b) When particle does not exist, but really expected to be in 'no decision' situation

A LEE applies to b)

Exclusion



Standard statistical test of H_0:

Either reject H_0, or don't reject H_0

Particle Physics:

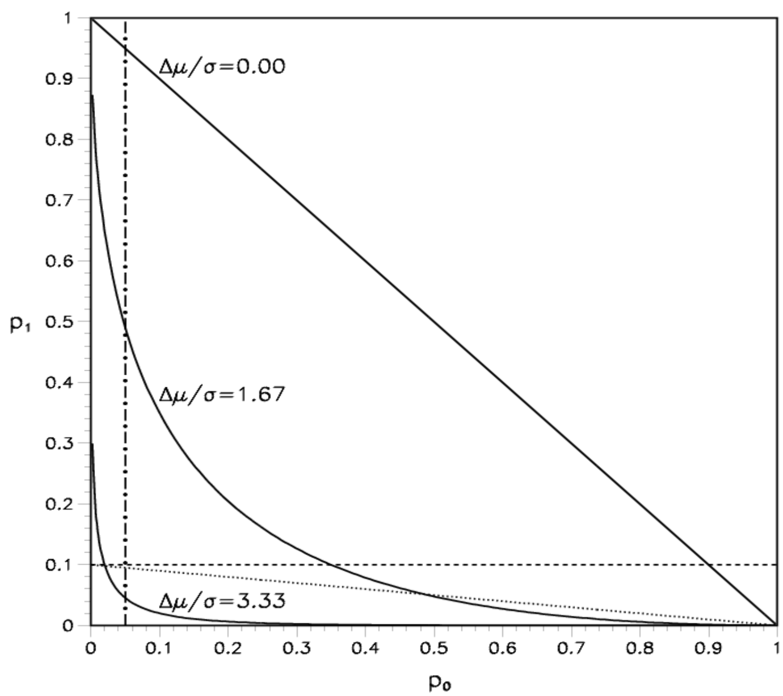
When don't reject H_0, is H_1 rejected ("EXCLUSION")?

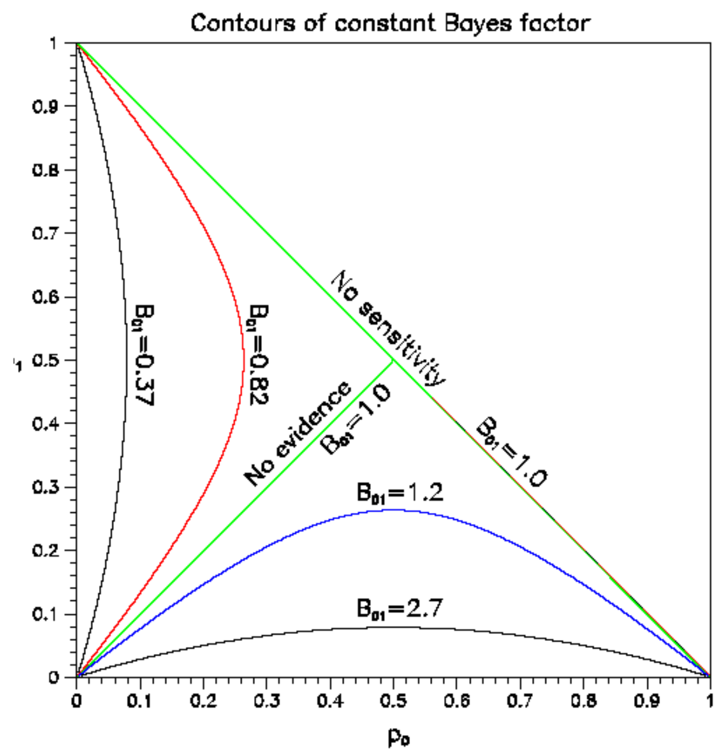
Big industry: e.g. Aether,   Higgs below 114 GeV

N.B. False exclusion ≠ Error of 2nd kind

p_0 v p_1 plot. "No decision" region (and double decision region)

Cuts at 5σ and 95%





p_0 versus p_1 plots

Look for something (e.g Higgs, SUSY,…) and don't see it: Set limit

(With pred for A(m), limit on A(m)→ excluded m's)

N.B. Language problem:

Van Dyk et al: Upper limit = expected. Upper bound from data

HEP: Sensitivity = expected Upper limit from data

(Expected = mean, median, Asimov)

Subject of first 2 PHYSTAT Workshops (2000)

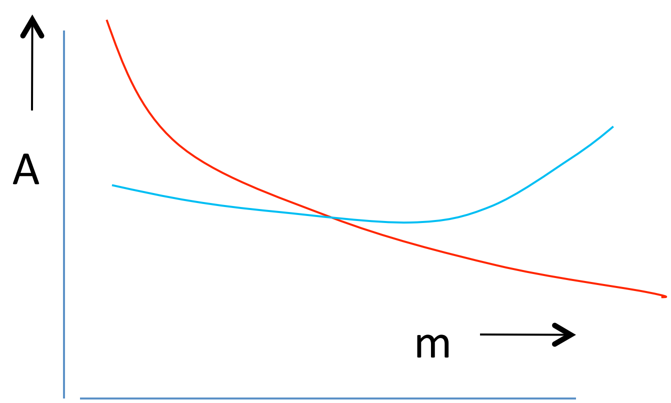
Google: CERN CLW and FNAL CLW

Prototype problem: n = Poisson ($\mu = \varepsilon s + b$)

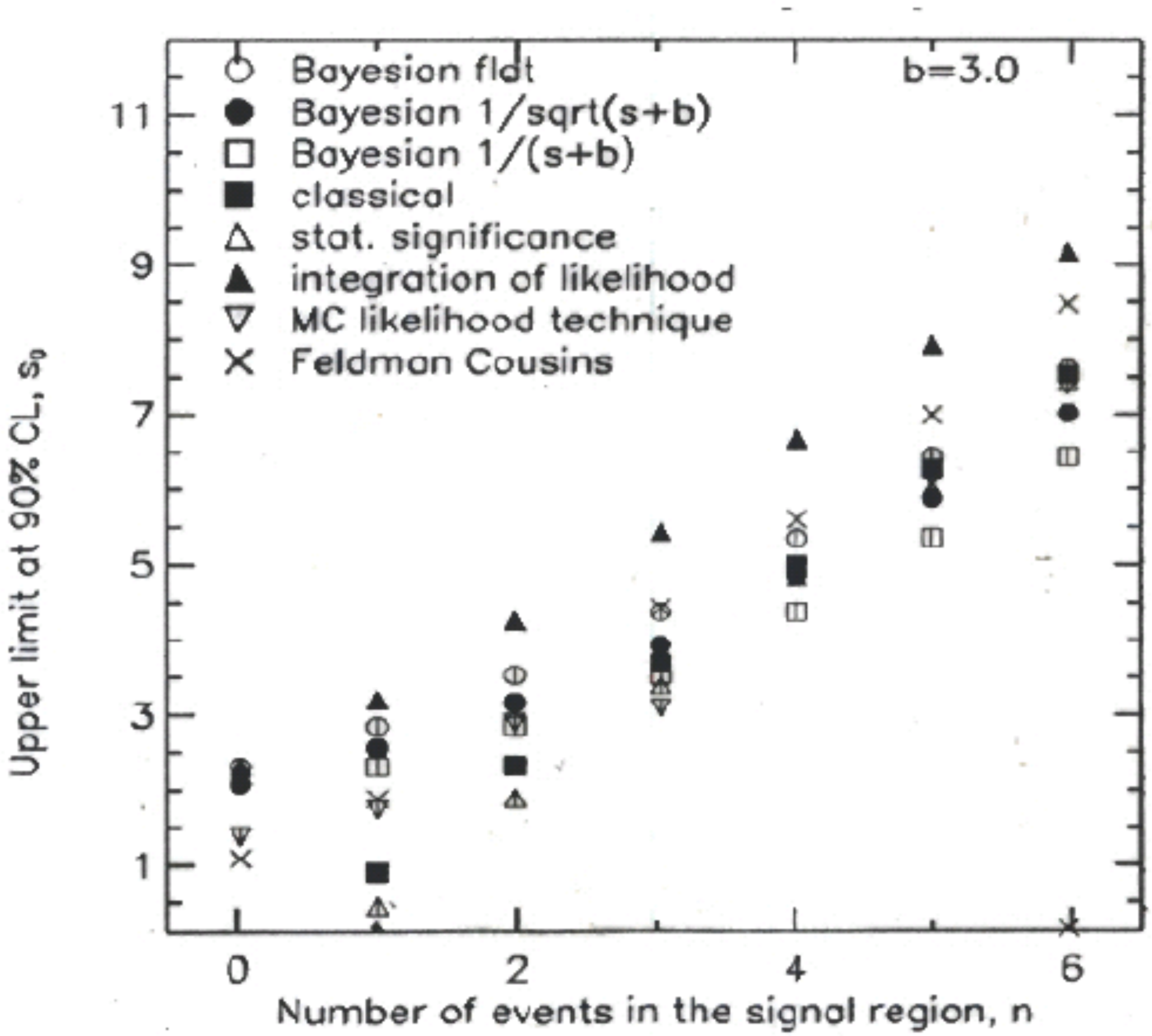
Lots of different methods:

Likelihood;
Frequentist with different ordering rules (e.g. Feldman-Cousins);
Bayes with different priors

Predicted A(m)
Exptl upper limit



Ilya Narsky, FNAL CLW 2000



Upper Limits: Desirable properties

(More difficult for 2-sided intervals)

Coverage: Does it need to be strict?

Bayesian credibility: Short but not too short

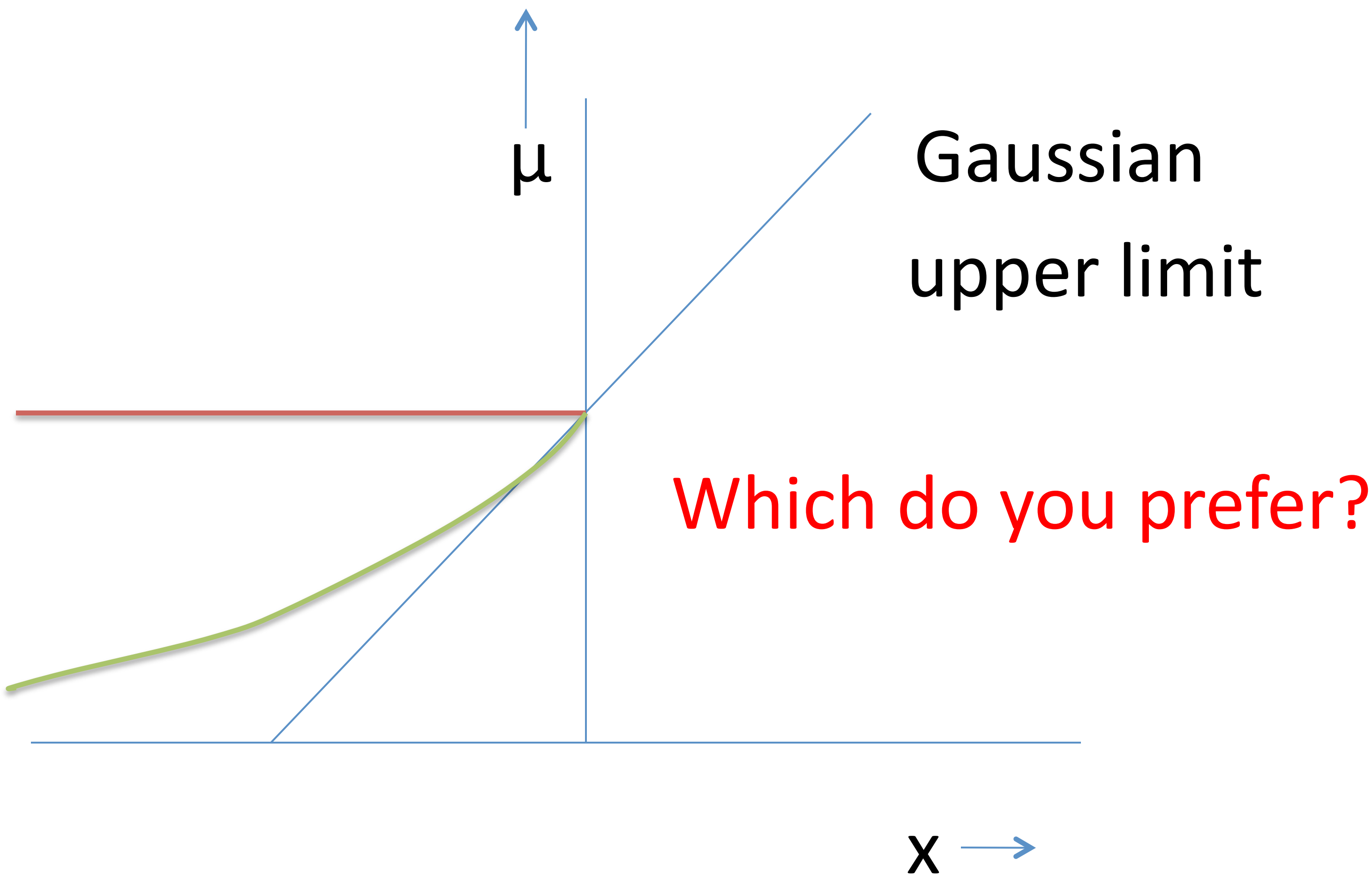
Behaviour when n_{obs} less than expected bgd

Behaviour wrt bgd

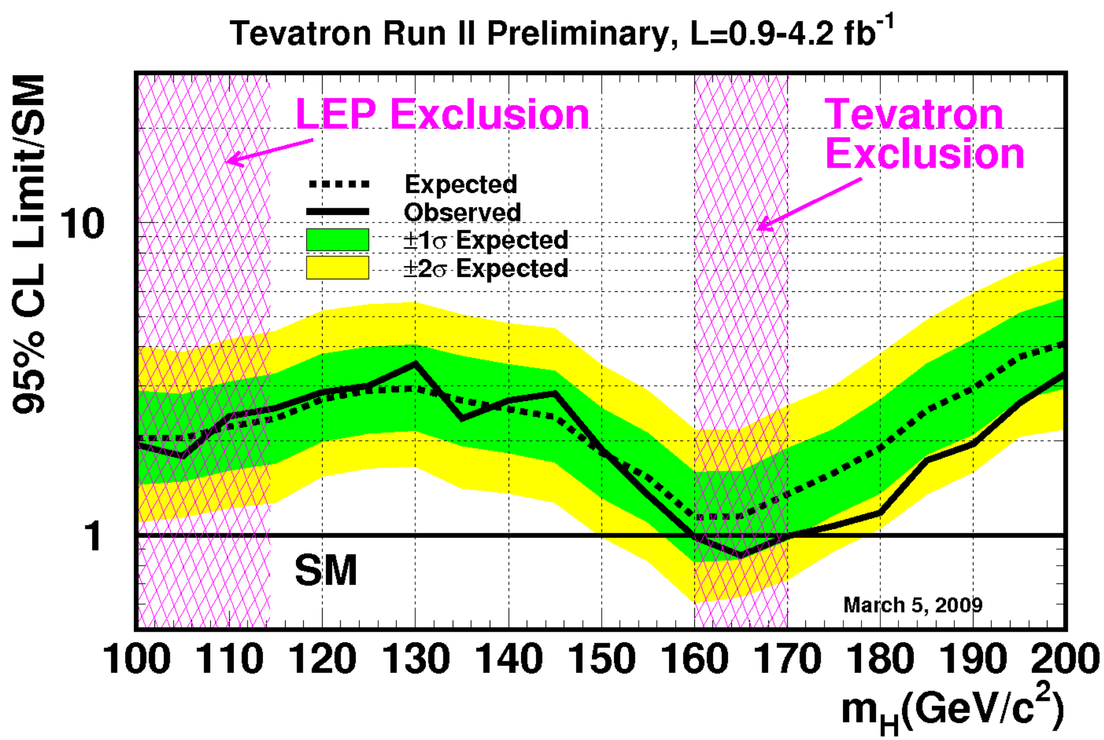
Inclusion of nuisance parameters

David van Dyk et al: preprint on reading list

Behaviour when $x < \mu$ (or $n_{obs} < b$)



Higgs exclusion at Tevatron



On-Off problem

Estimate signif of “on” counts, but with bgd estimated from “off” counts

n_{on} = Poisson($\varepsilon s + b$)

n_{off} =Poisson(τb)

Linnemann; Cousins, Linnemann and Tucker

Sometimes τ has uncertainty. How to incorporate this?

Banff Challenge

Banff 2006:

Upper limit calculation , given signal and background counts, for 1 or several channels

See: Joel Heinrich in PHYSTAT-LHC

("On-off" as above, but for signif.)

Banff 2010:

Background, signal and 'data' distributions

Does data contain signal?

Organised by Wade Fisher, Tom Junk and Jim Linnemann.

Taken up by Scargle, Vitells, Schafer,………

CL_s

Danger of exclusion (5%) when no sensitivity

Consensus in HEP community: Protect against this

One way: Select according to

$$CL_s = p_1/(1-p_0) = p_{bgd+sig}/\{1-p_{bgd}\}$$

Statisticians don't like ratio of p-values, but it is 'conservative Frequentist' (Bill Murray)

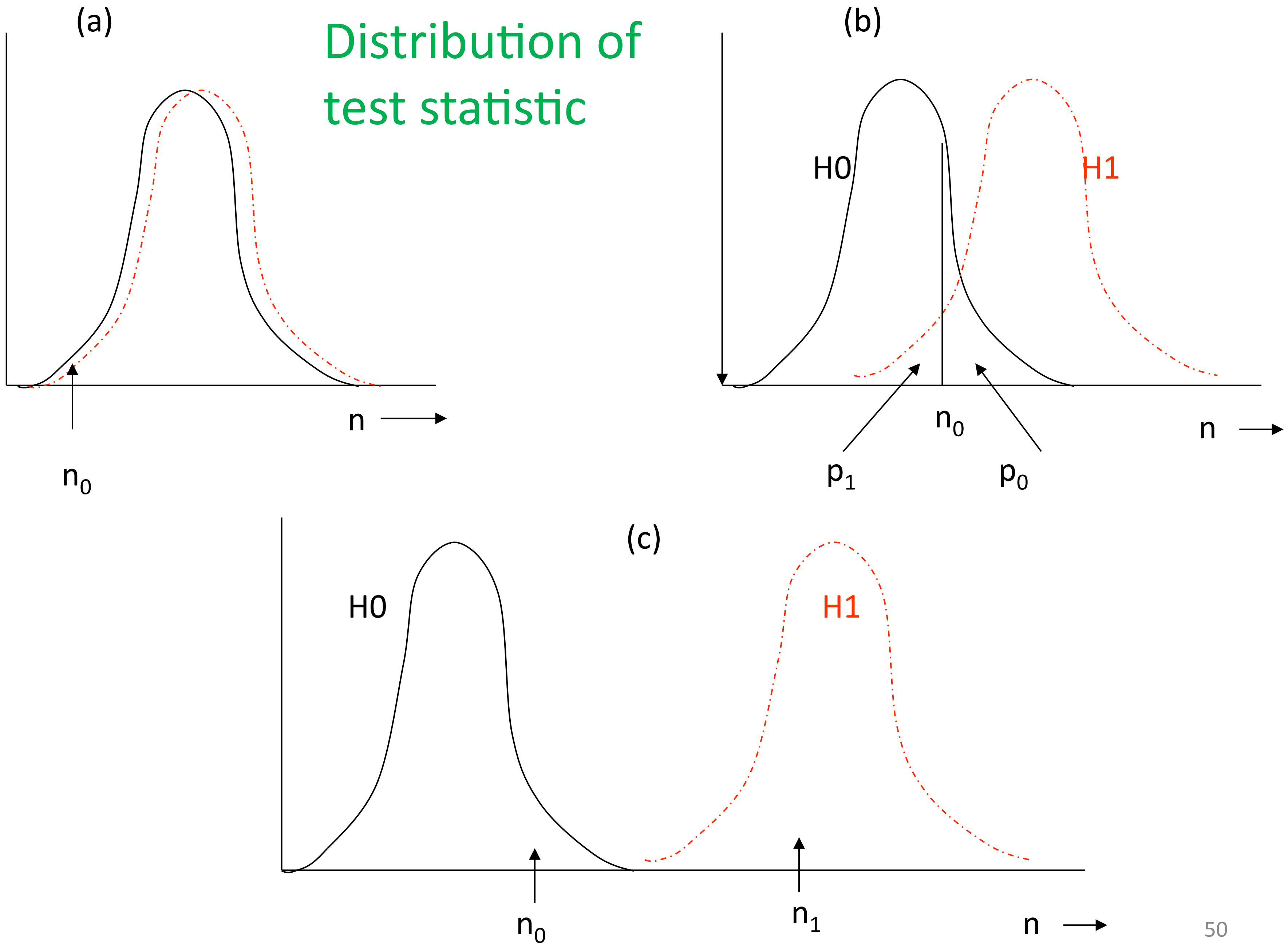
Changes exclusion region on p_0 v p_1 plot

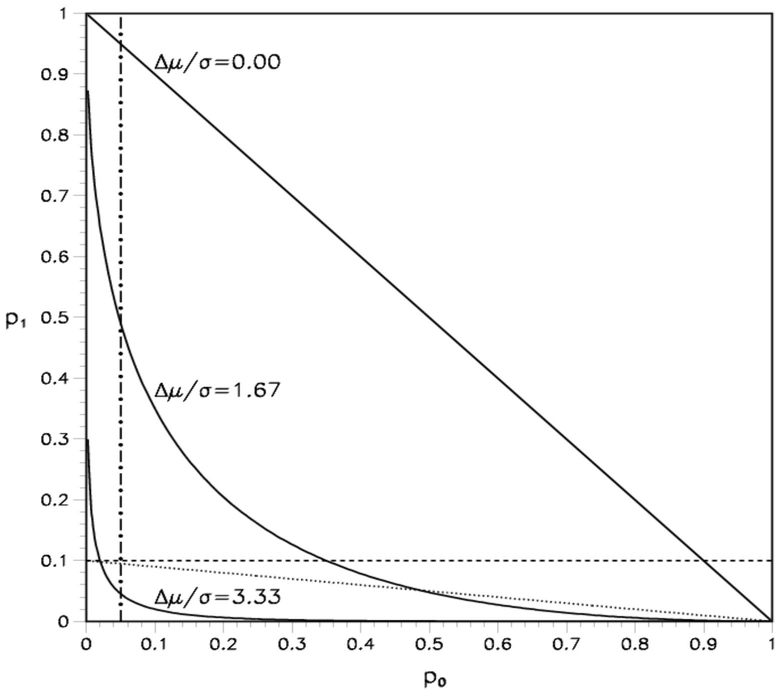
Alternatives possible:

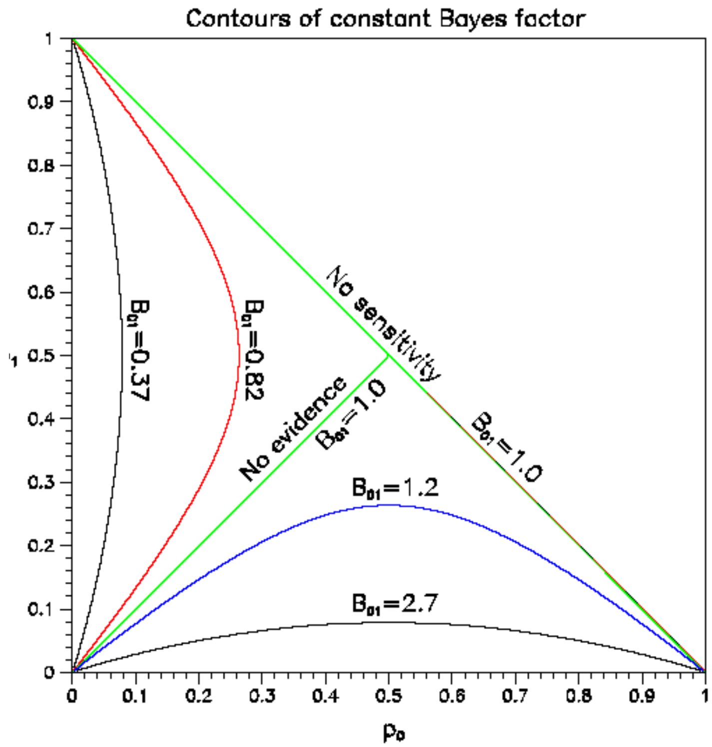
e.g. Limit never tighter than some sensitivity criterion

Van Dyk: Quote both 'upper limit' and 'upper bound'

Distribution of test statistic







p_0 versus p_1 plots

p_0 v p_1 plots

- Exclusion/discovery/no-decision regions
- Contours for fixed separations
- CL_s exclusion
- Punzi sensitivity definition
- Distribution of p_1 for H_0, or p_0 for H_1
- Varying likelihood ratio at fixed p_0

Choosing between Hypotheses

H_0 = No New Physics

H_1 = Specific New Physics

Frequentist approaches = p-values of some statistic

Likelihood ratio , or difference in χ^2

Bayesian approaches (Jim Berger):

- Posterior odds (problem of priors)
- Bayes factor
- BIC, AIC

Why p $\neq$ Bayes factor

Measure different things:

p_0 refers just to H_0; B_{01} compares H_0 and H_1

Depends on amount of data:

e.g. Poisson counting experiment little data:

For H_0, μ_0 = 1.0. For H_1, μ_1 =10.0

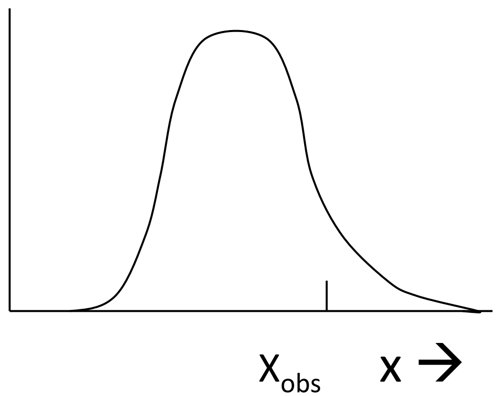
Observe n = 10 $p_0 \sim 10^{-7}$ $B_{01} \sim 10^{-5}$

Now with 100 times as much data, μ_0 = 100.0 μ_1 =1000.0

Observe n = 160 $p_0 \sim 10^{-7}$ $B_{01} \sim 10^{+14}$

Version of Lindley's paradox (Cousins)

p-values or Likelihood ratio?



L = height of curve

p = tail area

Different for distributions that

a) have dip in middle

b) are flat over range

Likelihood ratio favoured by Neyman-Pearson lemma (for simple H_0, H_1)

Use L-ratio as statistic, and use p-values for its distributions for H_0 and H_1

Think of this as either

i) p-value method, with L-ratio as statistic; or

ii) L-ratio method, with p-values as method to assess value of L-ratio

Incorporating systematics in p-values

Simplest version:

Observe n events

Poisson expectation for background only is b ± σ_b

σ_b may come from:

acceptance problems

jet energy scale

detector alignment

limited MC or data statistics for backgrounds

theoretical uncertainties

Include systematics

H_0 = No New Physics

H_1 = Specific New Physics (Usually Physics params)

Frequentist approaches = p-values of some statistic

Hybrid method

Likelihood ratio Profile

Bayesian approaches Bayesian

Ways to incorporate nuisance params in p-values

- Supremum — Maximise p over all ν. Very conservative
- Conditioning — Good, if applicable
- Prior Predictive — Box. Most common in HEP

  $p = \int p(\nu)\, \pi(\nu)\, d\nu$
- Posterior predictive — Averages p over posterior
- Plug-in — Uses best estimate of ν, without error
- L-ratio
- Confidence interval — Berger and Boos.

  $p = \mathrm{Sup}\{p(\nu)\} + \beta$, where $1-\beta$ Conf Int for ν
- Generalised frequentist — Generalised test statistic

Performances compared by Demortier

Profile Likelihood

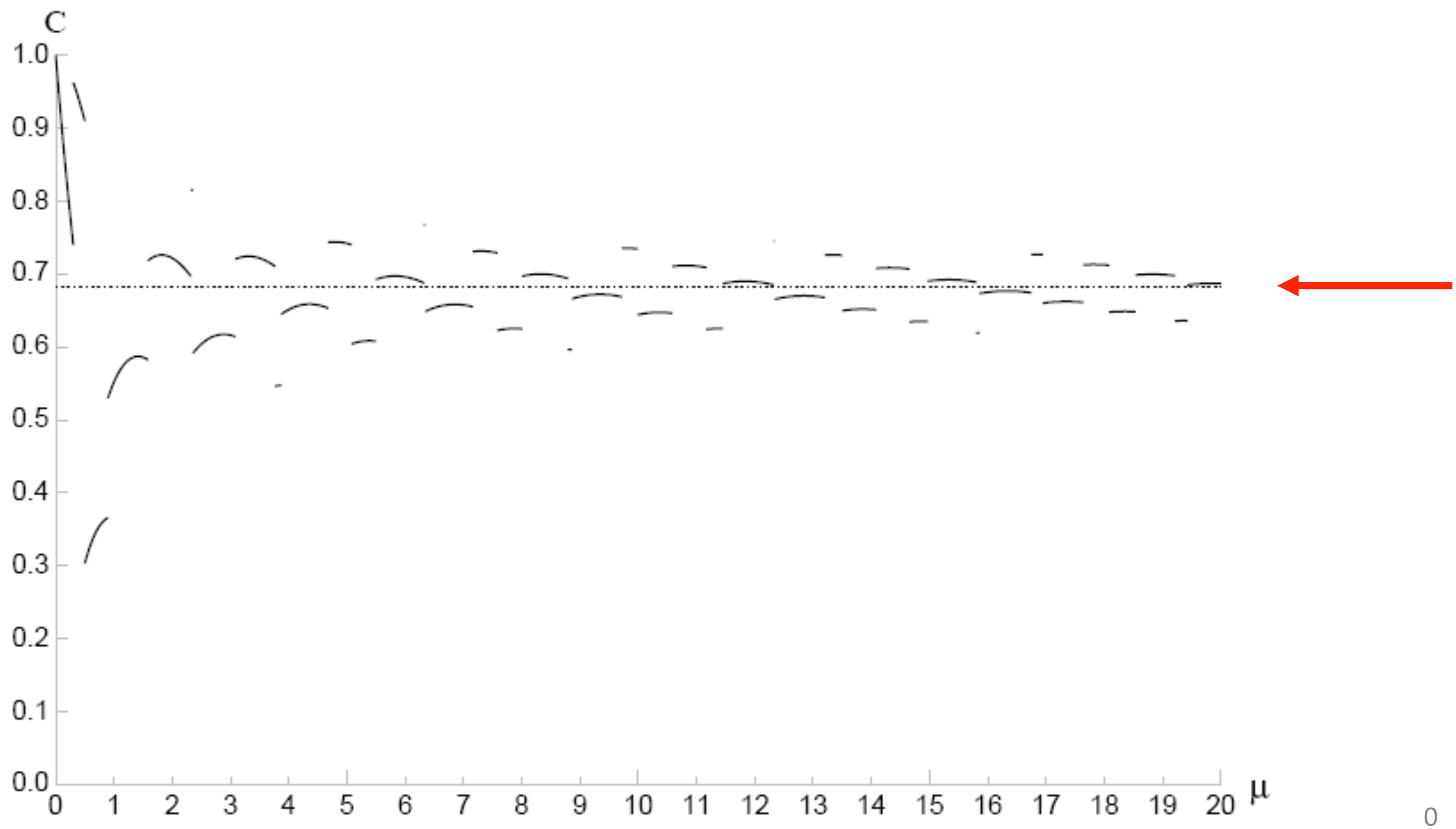
- Coverage studies almost all show overcoverage
- Ordinary likelihood can significantly undercover

Is the profiling somehow averaging over over- and under-coverage?

Coverage : L approach (Not frequentist)

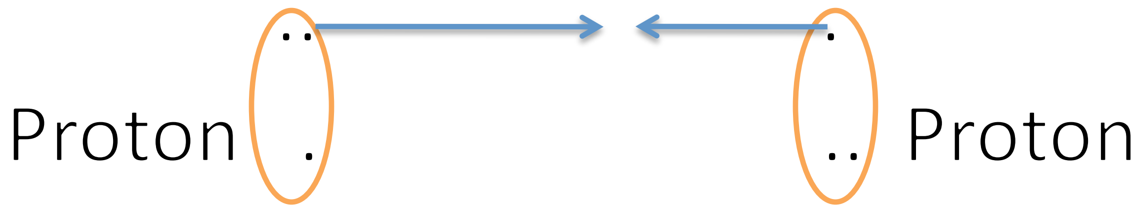
$P(n,\mu) = e^{-\mu}\mu^n/n!$ (Joel Heinrich CDF note 6438)

$-2\ \ln\lambda < 1$ $\lambda = P(n,\mu)/P(n,\mu_{best})$ UNDERCOVERS



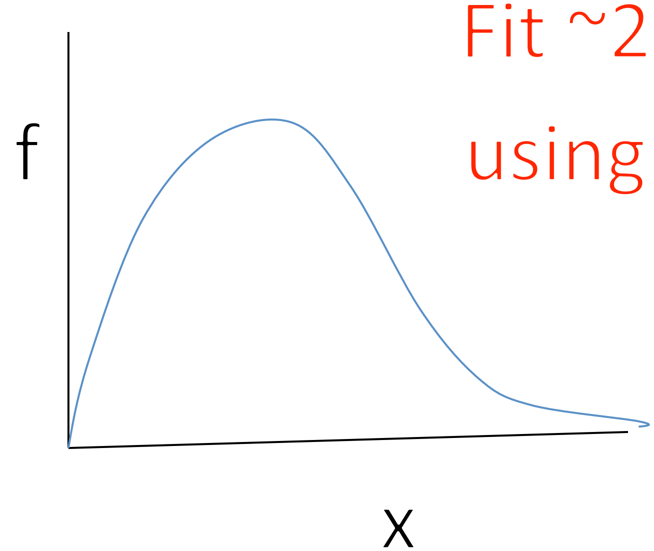
Coverage (C) vs μ: $-2\ln\lambda < 1$ ($C \to 0.6827$ as $\mu \to \infty$)

Partons



$f(x,Q^2)$ for each parton type $(u, d, s, \bar{u}, \bar{d}, \bar{s}, g)$

Parametrise e.g.



Fit ~2000 data points

using ~20 params

Partons important for predicting particle production.

Uncertainties important for systematics.

Fits have χ^2/NDF ~ 1 (within ~20%)

But param uncertainties from $\Delta\chi^2$ ~ 50 (or equiv)

For analyses where this systematic is important, this is worrying

Various suggestions about why $\Delta\chi^2$ ~ 50 is needed

e.g. parametrisation inadequate

χ^2/NDF and $\Delta\chi^2$ can be independent

Talks by Jon Pumplin and Robert Thorne

http://www.hep.ucl.ac.uk/pdf4lhc/PDF4LHCrecom.pdf

for advice about how to incorporate uncertainties

Combining p-values

Q: Given independent analyses giving uniform p-value distributions under H_0, how to combine results? (Better to combine data)

A: Prescription is not unique

e.g. Analysis 1 gives χ^2 = 90 for NDF=100

Analysis 2 gives χ^2 = 25 for NDF=1

Does analysis 2's low p-value mean H_0 is excluded?

Or does χ^2=115 for NDF=101 mean H_0 is fine?

Bob Cousins: "Variables p_1, p_2.... have uniform distribution over hypercube. What $f(p_1,p_2,..)$ would result in uniform distribution in f?"

Common options are:

Calculate probability that product of p-values is smaller than observed; or

Calculate prob that smallest p is smaller than observed

Slightly disconcerting that not associative

Combining very correlated results

e.g. Different analyses on same data.

Best estimate can lie outside range ; and

Its uncertainty tends to zero, as $\rho \rightarrow 1$.

Rather than combine, choose 'better' analysis

Q: 'Better' = smaller observed or expected uncertainty?

- Observed uncertainty can depend on result → bias
- Cox's measuring instruments (or ALEPH's measurement of mass of ν_τ)

Particle Physicists and Statisticians

Several active collaborations between statisticians and astrophysicists/cosmologists

Particle Physicists have 'interesting' problems too

HEP less happy about sharing data with others (especially other HEP Collaborations), but gradually changing (CDF agreed to have Statisticians as 'associates')

Welcome to anyone who wants to be active