p-values and Discovery

Louis Lyons

Oxford

l.lyons@physics.ox.ac.uk

KEK,

May 2007

PHYSTAT-LHC Workshop

on

Statistical Issues for LHC Physics

CERN Geneva June 27-29, 2007

Further information and registration at http://cern.ch/phystat-lhc

TOPICS

Discoveries

H0 or H0 v H1

p-values: For Gaussian, Poisson and multi-variate data

Goodness of Fit tests

Why 5σ?

Blind analyses

What is p good for?

Errors of 1st and 2nd kind

What a p-value is not

P(theory|data) ≠ P(data|theory)

THE paradox

Optimising for discovery and exclusion

Incorporating nuisance parameters

DISCOVERIES

"Recent" history:

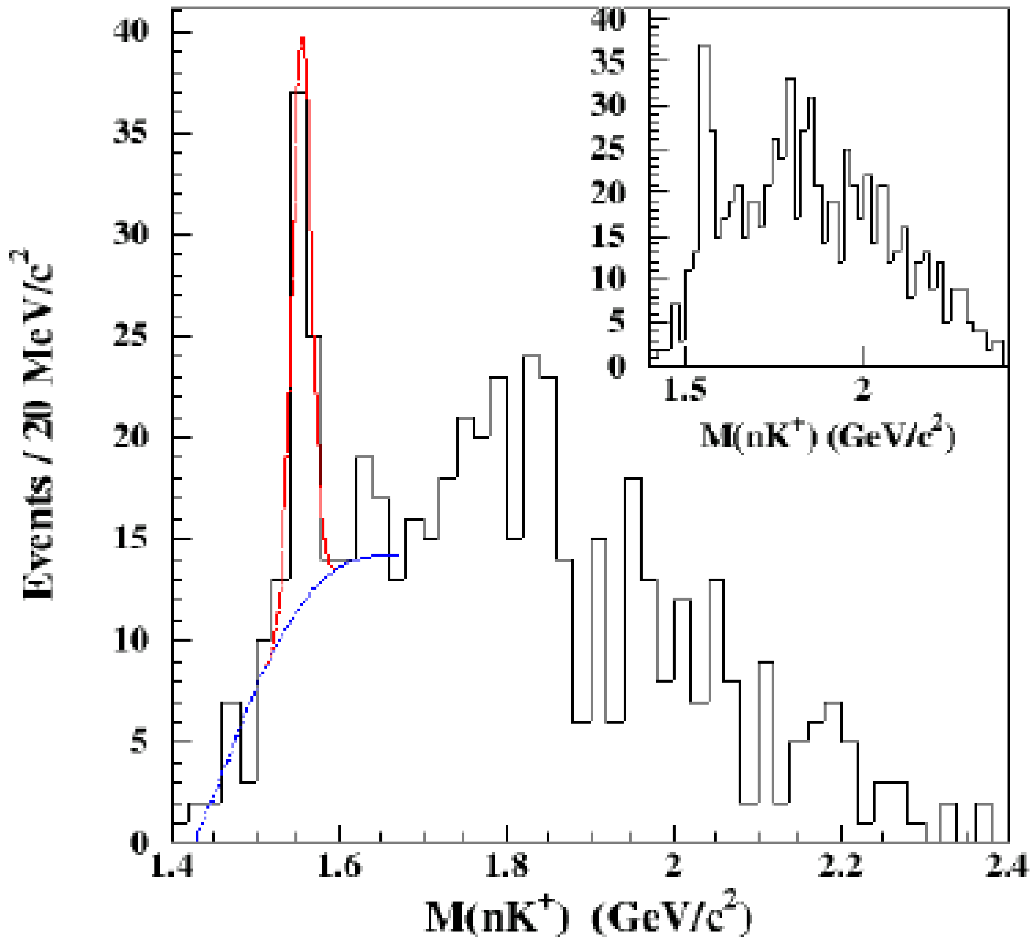
Charm	SLAC, BNL	1974
Tau lepton	SLAC	1977
Bottom	FNAL	1977
W,Z	CERN	1983
Top	FNAL	1995
{Pentaquarks	~Everywhere	2002 }
?	FNAL/CERN	2008?

? = Higgs, SUSY, q and l substructure, extra dimensions,
free q/monopoles, technicolour, 4th generation, black holes,…..

QUESTION: How to distinguish discoveries from fluctuations or goofs?

Penta-quarks?

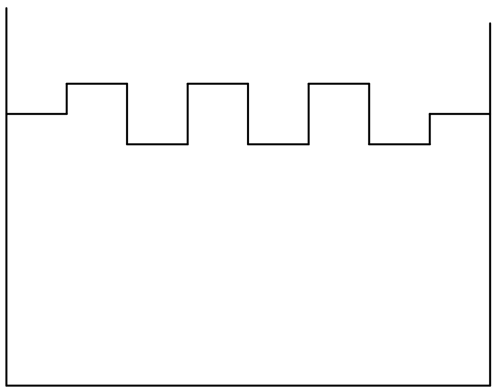
Hypothesis testing: New particle or statistical fluctuation?



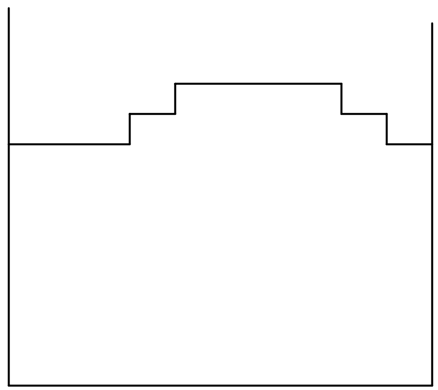
H0 or H0 versus H1 ?

H0 = null hypothesis
    e.g. Standard Model, with nothing new
H1 = specific New Physics   e.g. Higgs with M_H = 120 GeV
H0: "Goodness of Fit" e.g. χ^2 ,p-values
H0 v H1: "Hypothesis Testing" e.g. **L**-ratio
Measures how much data favours one hypothesis wrt other

H0 v H1 likely to be more sensitive



or



Testing H0:
Do we have an alternative in mind?

1) Data is number (of observed events)

"H1" usually gives larger number

(smaller number of events if looking for oscillations)

2) Data = distribution. Calculate χ^2.

Agreement between data and theory gives χ^2 ~ndf

Any deviations give large χ^2

So test is independent of alternative?

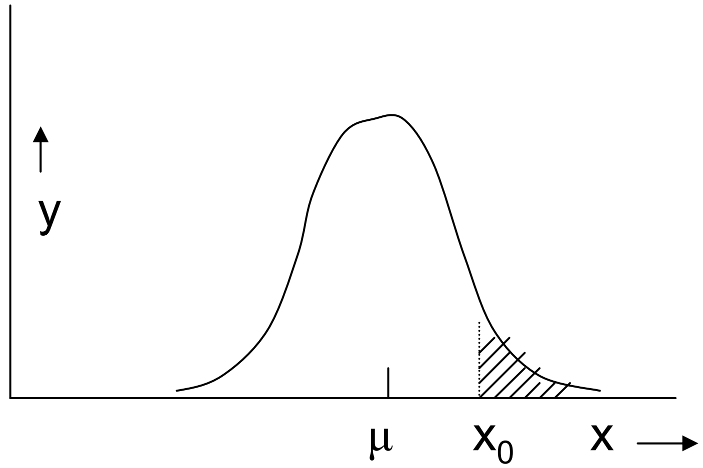
Counter-example: Cheating undergraduate

3) Data = number or distribution

Use **L**-ratio as test statistic for calculating p-value

4) H0 = Standard Model

p-values



Concept of pdf

Example: Gaussian

y = probability density for measurement x

$y = 1/(\sqrt{(2\pi)}\sigma) \exp\{-0.5*(x-\mu)^2/\sigma^2\}$

p-value: probablity that $x \geq x_0$

Gives probability of "extreme" values of data (in interesting direction)

$(x_0-\mu)/\sigma$	1	2	3	4	5
p	16%	2.3%	0.13%	0. 003%	$0.3*10^{-6}$

i.e. Small p = unexpected

p-values, contd

Assumes:

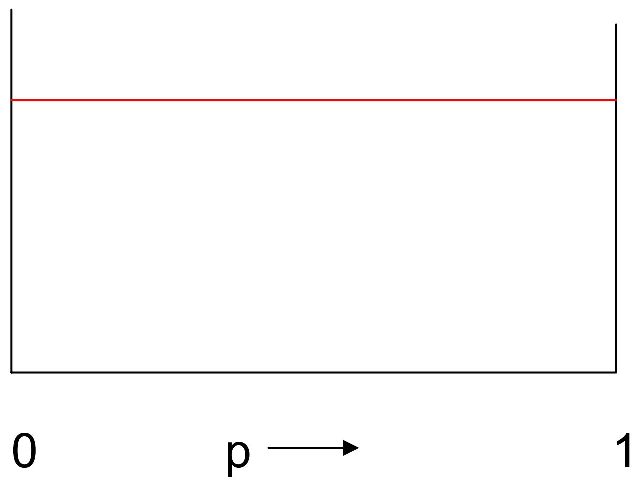
Gaussian pdf (no long tails)

Data is unbiassed

σ is correct

If so, Gaussian x ⟹ uniform p-distribution

(Events at large x give small p)

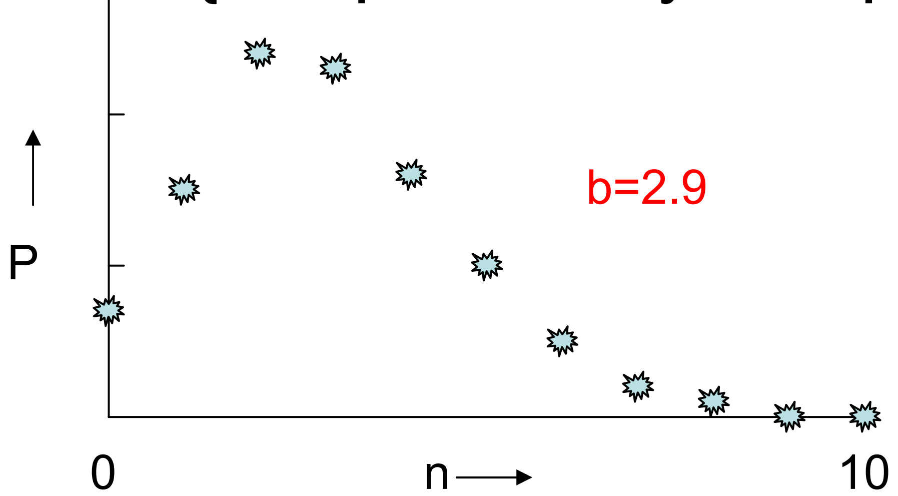


p-values for non-Gaussian distributions

e.g. Poisson counting experiment, bgd = b

$P(n) = e^{-b} * b^n/n!$

{P = probability, not prob density}



For n=7, p = Prob(at least 7 events) = P(7) + P(8) + P(9) +…….. = 0.03

Poisson p-values

n = integer, so p has discrete values

So p distribution cannot be uniform

Replace $\text{Prob}\{p \leq p_0\} = p_0$, for continuous p

by $\text{Prob}\{p \leq p_0\} \leq p_0$, for discrete p

(equality for possible p_0)

p-values often converted into equivalent Gaussian σ

e.g. $3*10^{-7}$ is "5σ" (one-sided Gaussian tail)

Significance

Significance = $S/\sqrt{B}$?

Potential Problems:

- Uncertainty in B
- Non-Gaussian behaviour of Poisson, especially in tail
- Number of bins in histogram, no. of other histograms [FDR]
- Choice of cuts (Blind analyses)
- Choice of bins (……………….)

For future experiments:

- Optimising $S/\sqrt{B}$ could give S =0.1, B = 10^{-6}

Goodness of Fit Tests

Data = individual points, histogram, multi-dimensional,
multi-channel

χ^2 and number of degrees of freedom

$\Delta\chi^2$ (or *ln*L-ratio): Looking for a peak

Unbinned L_{max}?

Kolmogorov-Smirnov

Zech energy test

Combining p-values

Lots of different methods. Software available from:

http://www.ge.infn.it/statisticaltoolkit

χ^2 with ν degrees of freedom?

1) ν = data – free parameters ?

Why asymptotic (apart from Poisson → Gaussian) ?

a) Fit flatish histogram with

$y = N\ \{1 + 10^{-6} \cos(x-x_0)\}$ x_0 = free param

b) Neutrino oscillations: almost degenerate parameters

$y \sim 1 - A \sin^2(1.27\ \Delta m^2\ L/E)$ 2 parameters

$\xrightarrow[\text{Small } \Delta m^2]{} 1 - A\ (1.27\ \Delta m^2\ L/E)^2$ 1 parameter

χ^2 with ν degrees of freedom?

2) Is difference in χ^2 distributed as χ^2 ?

H0 is true.

Also fit with H1 with k extra params

e. g. Look for Gaussian peak on top of smooth background

$$y = C(x) + A \exp\{-0.5 ((x-x_0)/\sigma)^2\}$$

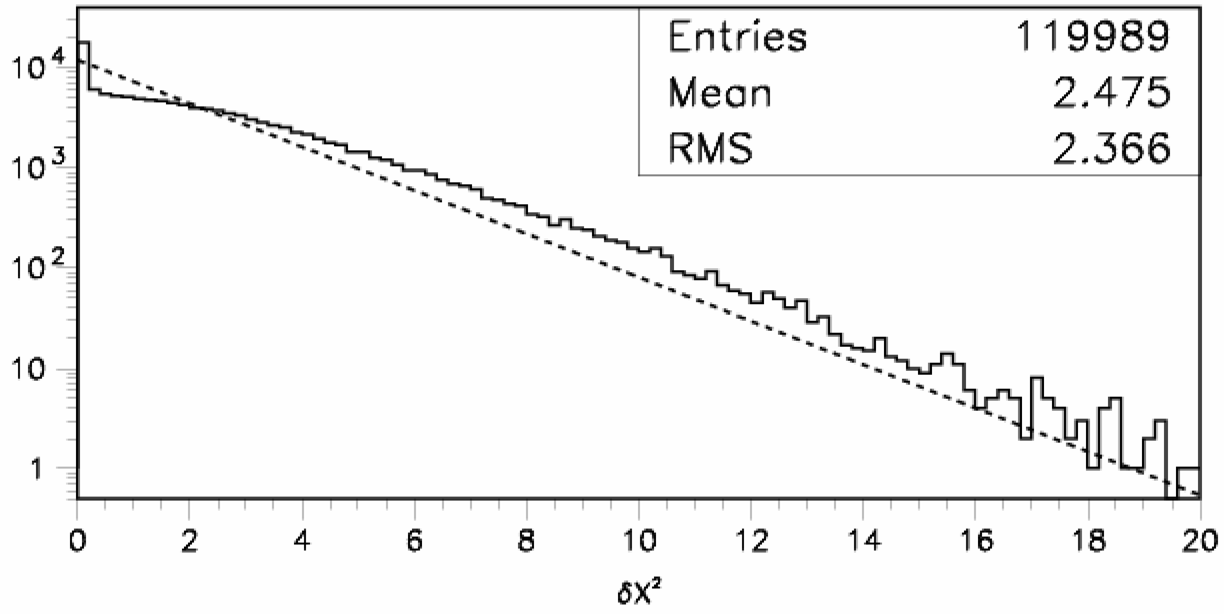
Is $\chi^2_{H0} - \chi^2_{H1}$ distributed as χ^2 with $\nu = k = 3$?

Relevant for assessing whether enhancement in data is just a statistical fluctuation, or something more interesting

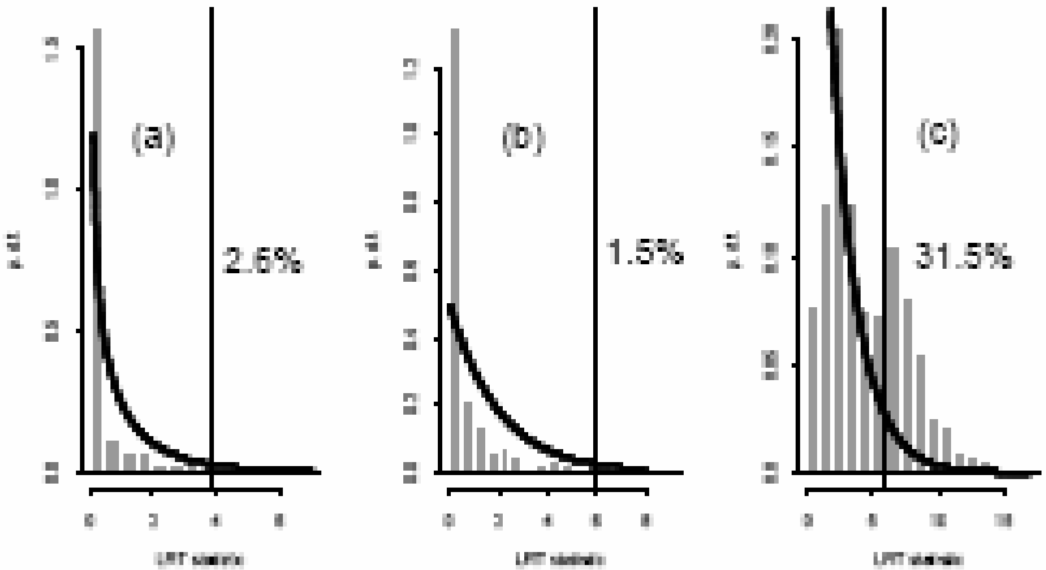
N.B. Under H0 ($y = C(x)$) : $A=0$ (boundary of physical region)

x_0 and σ undefined

Is difference in χ^2 distributed as χ^2 ?



Demortier:
H0 = quadratic bgd
H1 = ……………… +
Gaussian of fixed width,
variable location & ampl



Protassov, van Dyk, Connors, ….
H0 = continuum
(a) H1 = narrow emission line
(b) H1 = wider emission line
(c) H1 = absorption line

Nominal significance level = 5%

Is difference in χ^2 distributed as χ^2 ?, contd.

So need to determine the $\Delta\chi^2$ distribution by Monte Carlo

N.B.

1) Determining $\Delta\chi^2$ for hypothesis H1 when data is generated according to H0 is not trivial, because there will be lots of local minima

2) If we are interested in 5σ significance level, needs lots of MC simulations (or intelligent MC generation)

Unbinned L_{max} and Goodness of Fit?

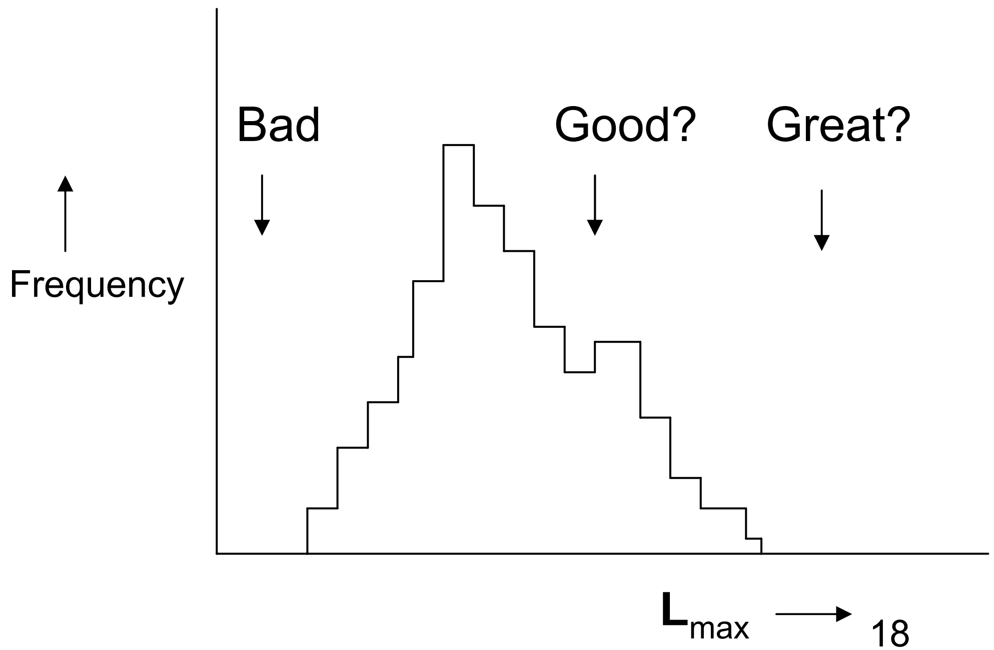
Find params by maximising L

So larger L better than smaller L

So L_{max} gives Goodness of Fit ??

Monte Carlo distribution

of unbinned L_{max}



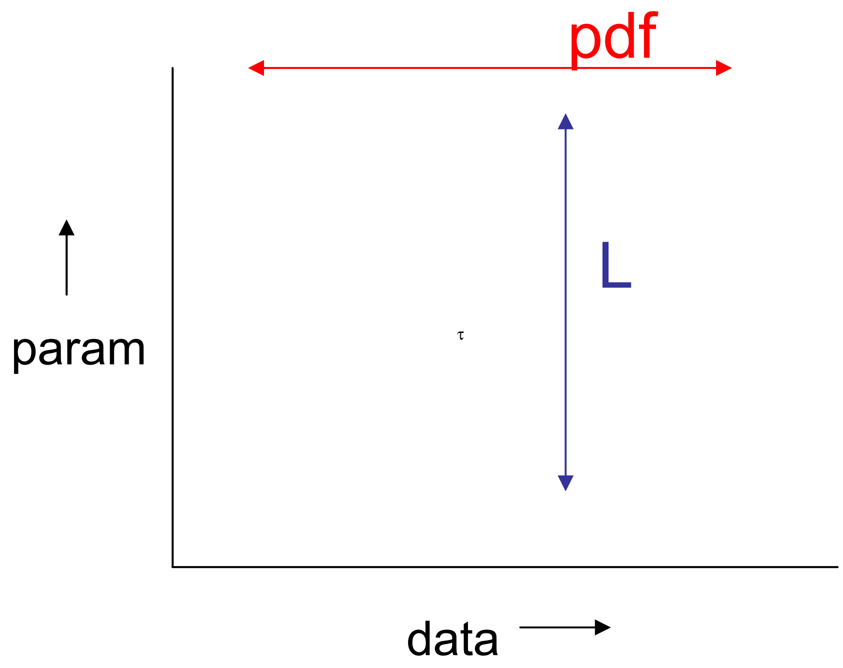
Not necessarily:

L(data,params)

↑ fixed ↑ vary

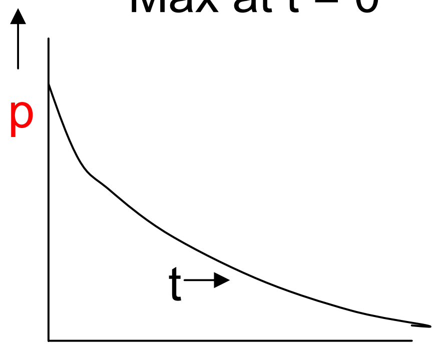
Contrast pdf(data,params)

↑ vary ↑ fixed



e.g. $p(t,\lambda) = \lambda * \exp(-\lambda t)$

Max at t = 0



Max at $\lambda = 1/t$