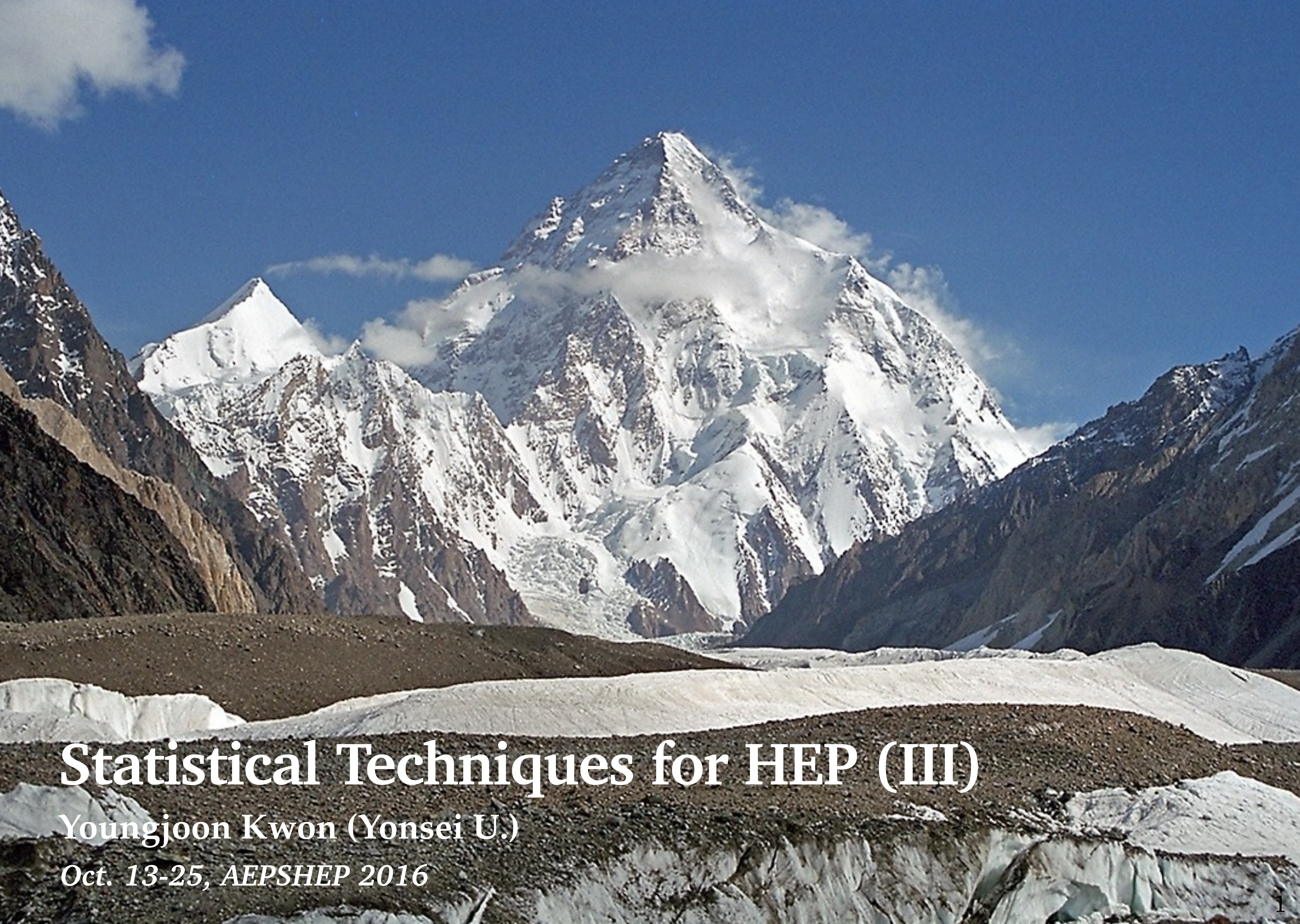


Statistical Techniques for HEP (III)

Youngjoon Kwon (Yonsei U.)

Oct. 13-25, AEPSHEP 2016

Parameter Estimation

Likelihood function

- Suppose the entire result of an experiment (*set of measurements*) is a collection of numbers $\vec{x}$, and suppose the joint PDF for the data $\vec{x}$ is a function depending on a set of parameters $\vec{\theta}$: $f(\vec{x};\vec{\theta})$
- Evaluate this function with the measured data $\vec{x}$, regarding this as a function of $\vec{\theta}$ only. This is the **likelihood function**.

$$L(\vec{\theta}) = f(\vec{x};\vec{\theta})\ \ (\vec{x}, \text{fixed})$$

The likelihood function for i.i.d. data

i.i.d. = *independent and identically distributed*

- Consider n independent observations of $\{x : x_1, \cdots, x_n\}$, where x follows $f(x, \theta)$.
  The joint PDF for the whole data sample is:

$$f(x_1, \cdots, x_n; \vec{\theta}) = \prod_{i=1}^{n} f(x_i; \vec{\theta})$$

- In this case, the likelihood function is

$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i; \vec{\theta}) \quad (x_i \ \text{constant})$$

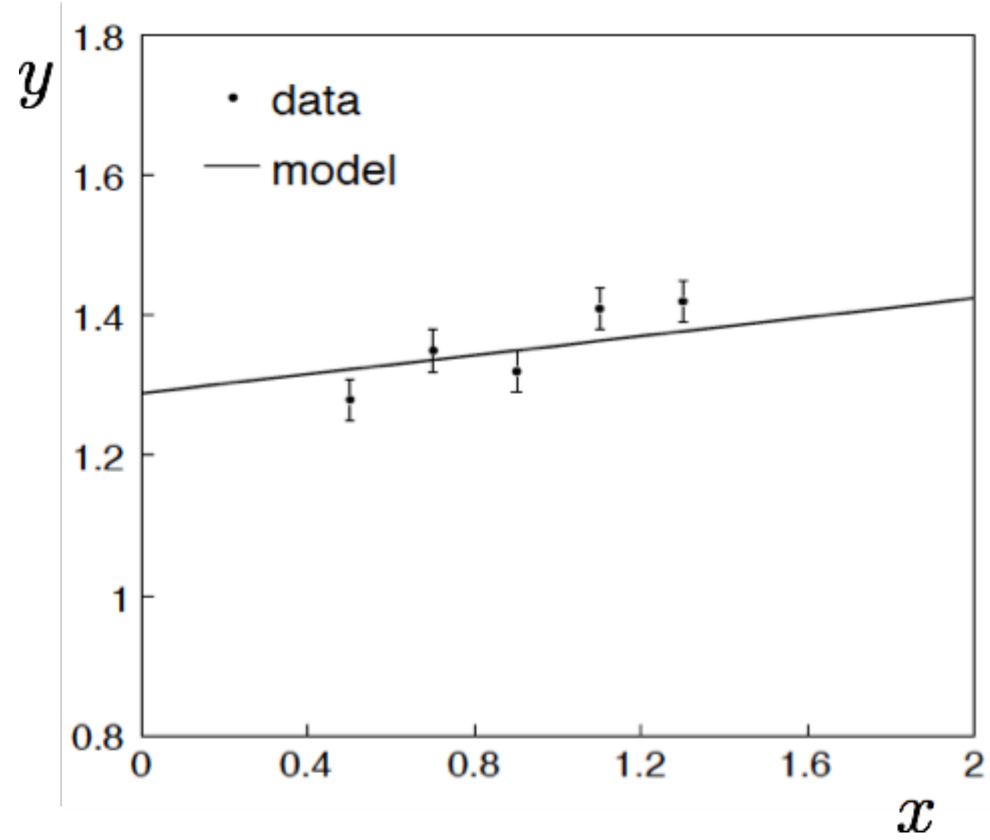
So we define the max. likelihood (ML) estimator(s) to be the parameter value(s) for which the L becomes maximum.

ML estimator example: fitting to a straight line

- Suppose we have a set of data: $(x_i, y_i, \sigma_i),\ i = 1, \cdots, n.$
- Modeling: y_i are independent and follow $y_i \sim G(\mu(x_i), \sigma_i)$ (G: Gaussian) where $\mu(x_i)$ are modelled as $\mu(x; \theta_0, \theta_1) = \theta_0 + \theta_1 x$

  Assume x_i and σ_i are known.
- Goal: to estimate θ_0

  Here, let's suppose we don't care about θ_1 (an example of a *nuisance parameter*)



ML fit with Gaussian data

- In this example, the y_i are assumed independent, so that likelihood function is a product of Gaussians:

$$L(\theta_0, \theta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\frac{(y_i - \mu(x_i;\theta_0,\theta_1))^2}{\sigma_i^2}\right]$$

- Then maximizing L is equivalent to minimizing

$$\chi^2(\theta_0, \theta_1) = -2\ln L(\theta_0, \theta_1) + C = \sum_{i=1}^{n} \frac{(y_i - \mu(x_i;\theta_0,\theta_1))^2}{\sigma_i^2}$$

i.e., for Gaussian data, ML fitting is the same as the method of least squares

Wilks' theorem

the Wilks' theorem

THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

BY S. S. WILKS

By applying the principle of maximum likelihood, J. Neyman and E. S. Pearson[2] have suggested a method for obtaining functions of observations for testing what are called *composite statistical hypotheses*, or simply *composite*

...

[1] Presented to the American Mathmatical Society, March 26, 1937.

...

We can summarize in the

Theorem: If a population with a variate x is distributed according to the probability function $f(x, \theta_1, \theta_2 \cdots \theta_h)$, such that optimum estimates $\bar{\theta}_i$ of the θ_i exist which are distributed in large samples according to (3), *then when the hypothesis H is true that $\theta_i = \theta_{0i}$, $i = m+1, m+2, \cdots h$, the distribution of $-2 \log \lambda$, where λ is given by* (2) *is, except for terms of order $1/\sqrt{n}$, distributed like χ^2 with $h - m$ degrees of freedom.*

the Wilks' theorem

http://wwwusers.ts.infn.it/~milotti/Didattica/StatisticaAvanzata/Cowan_2013.pdf

Suppose we model the data $\vec{X}$ with a likelihood $L(\vec{\mu})$ that depends on a set of N parameters $\vec{\mu} = (\mu_1, \cdots, \mu_N)$. (For simplicity, let's just consider a single parameter μ.)

- Define the statistic $t_\mu = -2\ln[L(\mu)/L(\hat{\mu})]$, where $\hat{\mu}$ is the ML estimator.
- The value of t_μ is a measure of how well the hypothesized parameter μ stand in agreement with the observed data.
- Larger values of t_μ indicate increasing incompatibility between the data and the hypothesized μ.
- According to Wilks' theorem, if the parameter value μ is true, then in the asymptotic limit of a large data sample, the PDF of t_μ is a χ^2 distribution for N d.o.f.

$$f(t_\mu|\mu) \sim \chi^2_N$$

ML fit or Least-square fit?

- **Consider we have a random variable $x \in [0, 3]$, and a distribution $f(x)$.**
- **In a series of measurements, we obtained**
  - 9 events in [0,1), 10 events in [1,2), and 8 events in [2,3]
  - We have a model of uniform $f(x)$, and would like to estimate the mean value of $\int f(x)\,dx$ for each histogram bin.
- **Run a thought-experiment, comparing**
  - maximum likelihood method, and least-square method
  - *Do they give the same result?*

Bayesian likelihood function

- Suppose our L-function contains two parameters θ_0 and θ_1, where we have some knoweldege about the prior probability on θ_1 from previous measurements:

$$\pi(\theta_0, \theta_1) = \pi_0(\theta_0)\pi_1(\theta_1)$$
$$\pi_0(\theta_0) = \text{const.}$$
$$\pi_1(\theta_1) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-(\theta_1-\theta_p)^2/2\sigma_p^2}$$

- Putting this into the Bayes' theorem gives the posterior probability:

$$p(\theta_0, \theta_1|\vec{x}) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu(x_i;\theta_0,\theta_1))^2/2\sigma_i^2} \; \pi_0 \; \frac{1}{\sqrt{2\pi}\sigma_p} e^{-(\theta_1-\theta_p)^2/2\sigma_p^2}$$

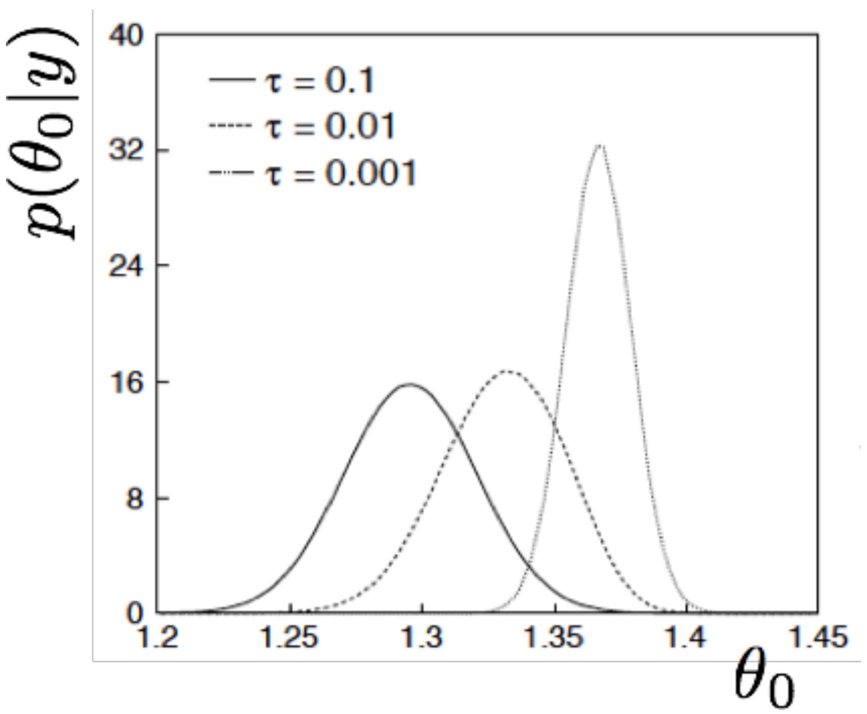
- Then, $p(\theta_0|\vec{x}) = \int p(\theta_0, \theta_1|\vec{x})\, d\theta_1$

with alternative priors

- Suppose we don't have a previous measurement of θ_1 but rather a theorist saying that θ_1 should be > 0 and not too much greater than, say, 0.1 or so. In that case, we may try modeling the prior for θ_1 as something like

$$\pi_1(\theta_1) = \frac{1}{\tau} e^{-\theta_1/\tau}, \ \theta_1 \geq 0, \ \tau = 0.1$$

- From this we obtain (numerically) the posterior PDF for θ_0



- This plot summarizes all knowledge about θ_0.

other advanced topics

- nuisance parameters & systematic uncertainties
- spurious exclusion ➔ the CL_s procedure
- look-elsewhere effect
- blind analysis

What can go wrong in a measurement?

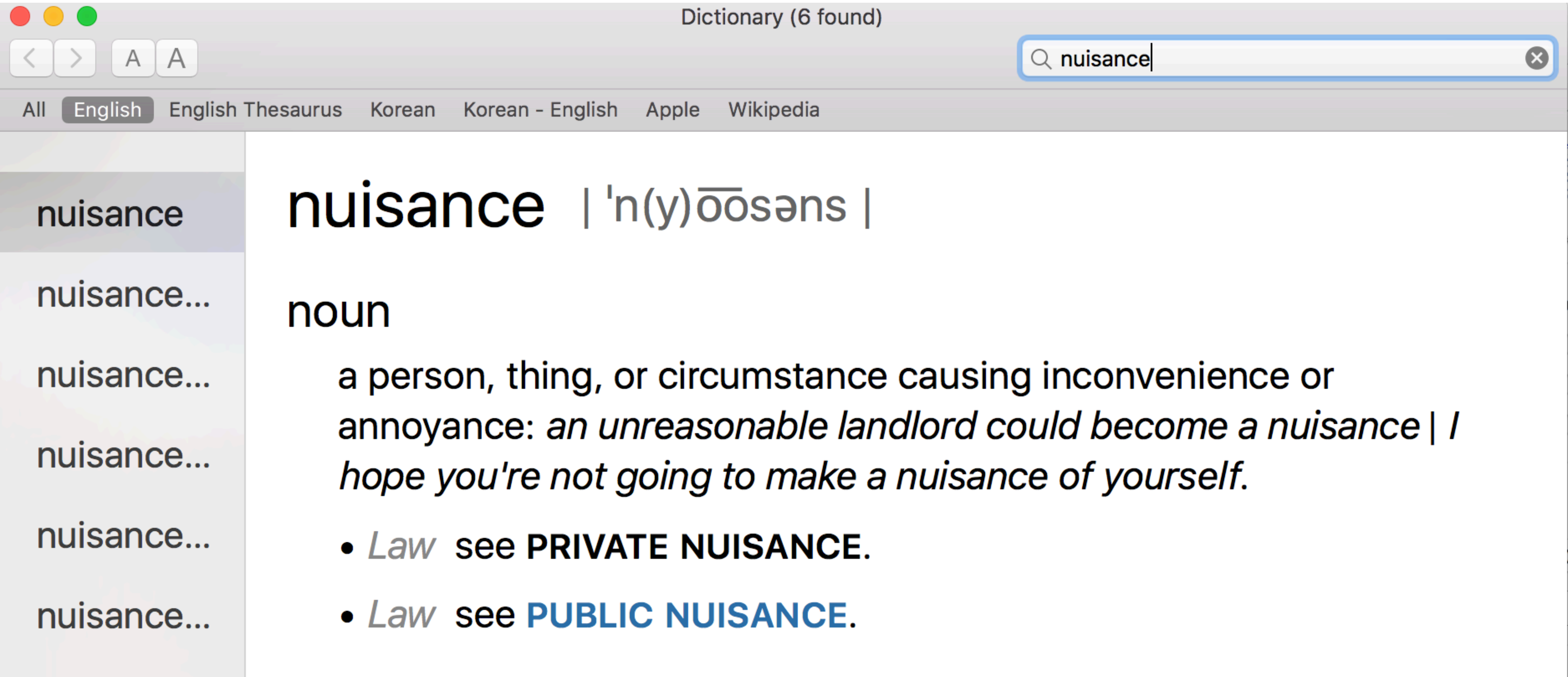
Consider a typical branching fraction measurement

$$\mathcal{B} = \frac{(N_{\rm obs} - N_{\rm bkg})}{N_{\rm norm}} \frac{\epsilon_{\rm norm}}{\epsilon_{\rm sig}}$$

- Determination of any elements in the final number can be wrong due to incomplete knowledge about the experimental apparatuses, background contaminations, etc.
- All such sources shall be studied and corrected for. Any uncertainties in these shall be included in the systematic uncertainty.

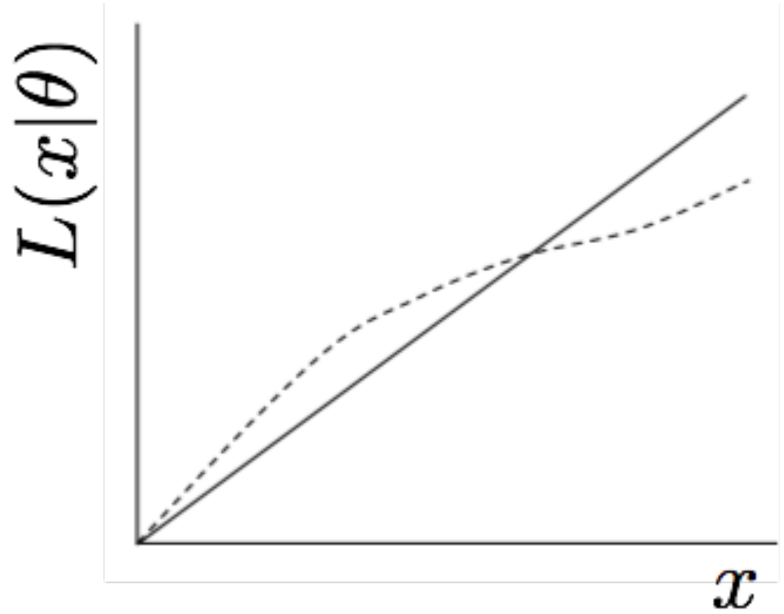
Systematic uncertainties?

In statistics, they call it the "nuisance parameter"



Nuisance parameters

- In general our model of the data is *not perfect*



model: $L(x|\theta) = \theta x$

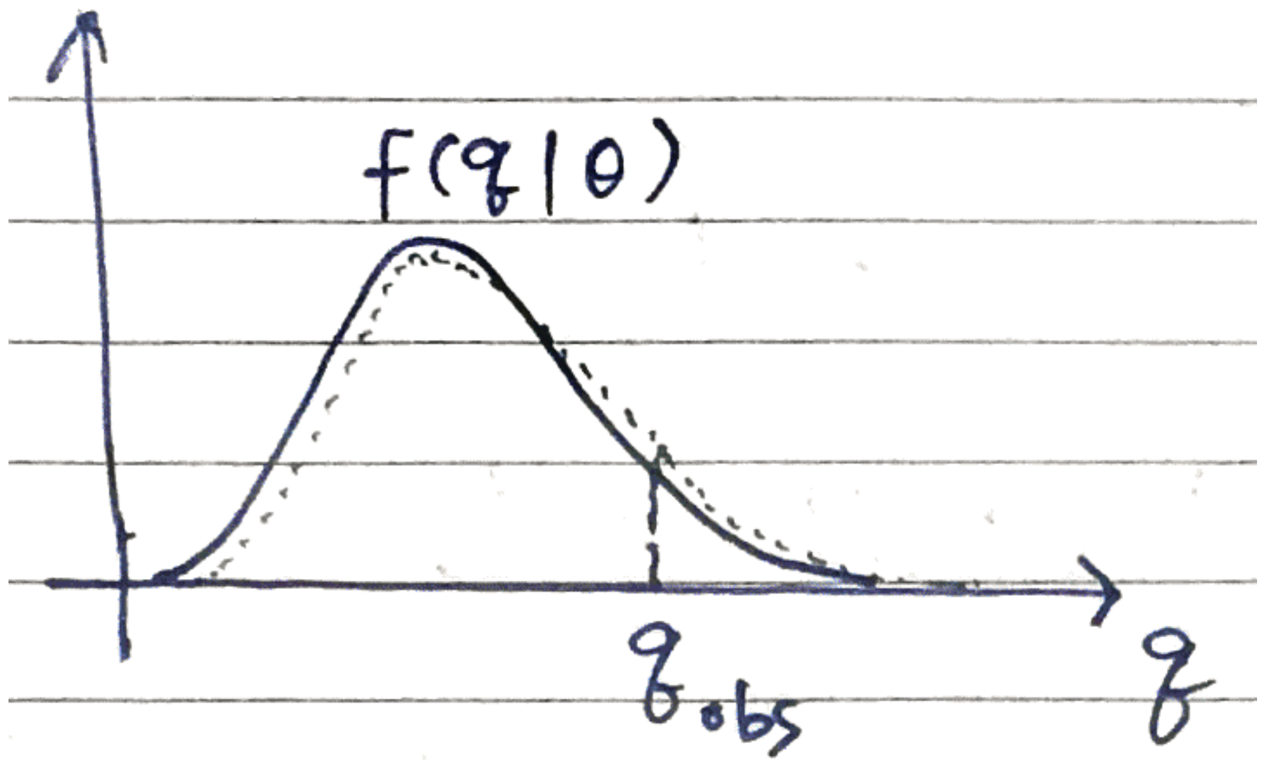
truth: $L(x|\theta) = \theta x + \alpha x^2 + \beta x^3 + \cdots$

- can improve model by including additional adjustable parameters: $L(x|\theta) \to L(x|\theta, \nu)$
- Nuisance parameter $\leftrightarrow$ systematic uncertainty
  Some point in the parameter space of the enlarged model must be "true"
- Presence of nuisance parameter(s) decreases sensitivity of analysis to the parameter of interest (e.g. larger variance of estimate).

- Suppose we have a statistic q to test a hypothesized value of a parameter θ, such that the p-value of θ is

$$p_\theta = \int_{q_\theta,\mathrm{obs}}^{\infty} f(q_\theta|\theta,\nu)\, dq_\theta$$

- But what value of ν should we use for $f(q_\theta|\theta,\nu)$?



Profile likelihood ratio

- Base significance test on the profile likelihood ratio

profile likelihood

maximizes L for Specified μ

$$\lambda(\mu) = \frac{L_{\rm p}(\mu)}{L_{\max}} = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

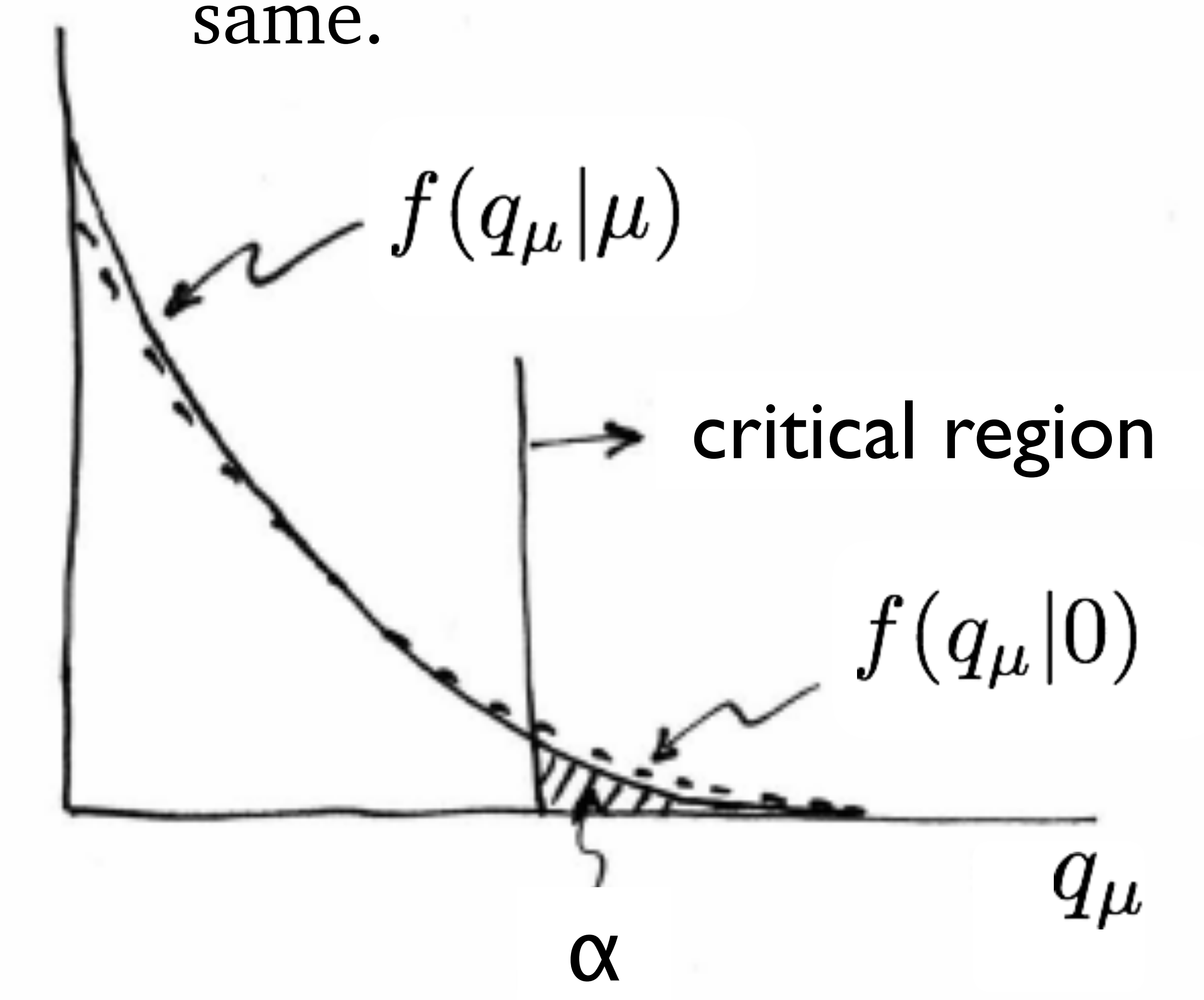
maximize L

  - the likelihood ratio of point hypotheses gives optimal test (by Neyman-Pearson lemma)
  - the statistic above is nearly optimal

- Advantage of $\lambda(\mu)$ – in large sample limit, $f(-2\ln\lambda(\mu)\,|\mu\,)$ approaches a χ^2 pdf for $n = 1$ (by *Wilks' theorem*)

Spurious exclusion

- **Sometimes, the effect of a given hypothesized μ is very small relative to the null (μ =0) prediction**
  - In that case, the distributions $f\,(q_\mu|\mu)$ and $f\,(q_\mu|0)$ will be almost the same.

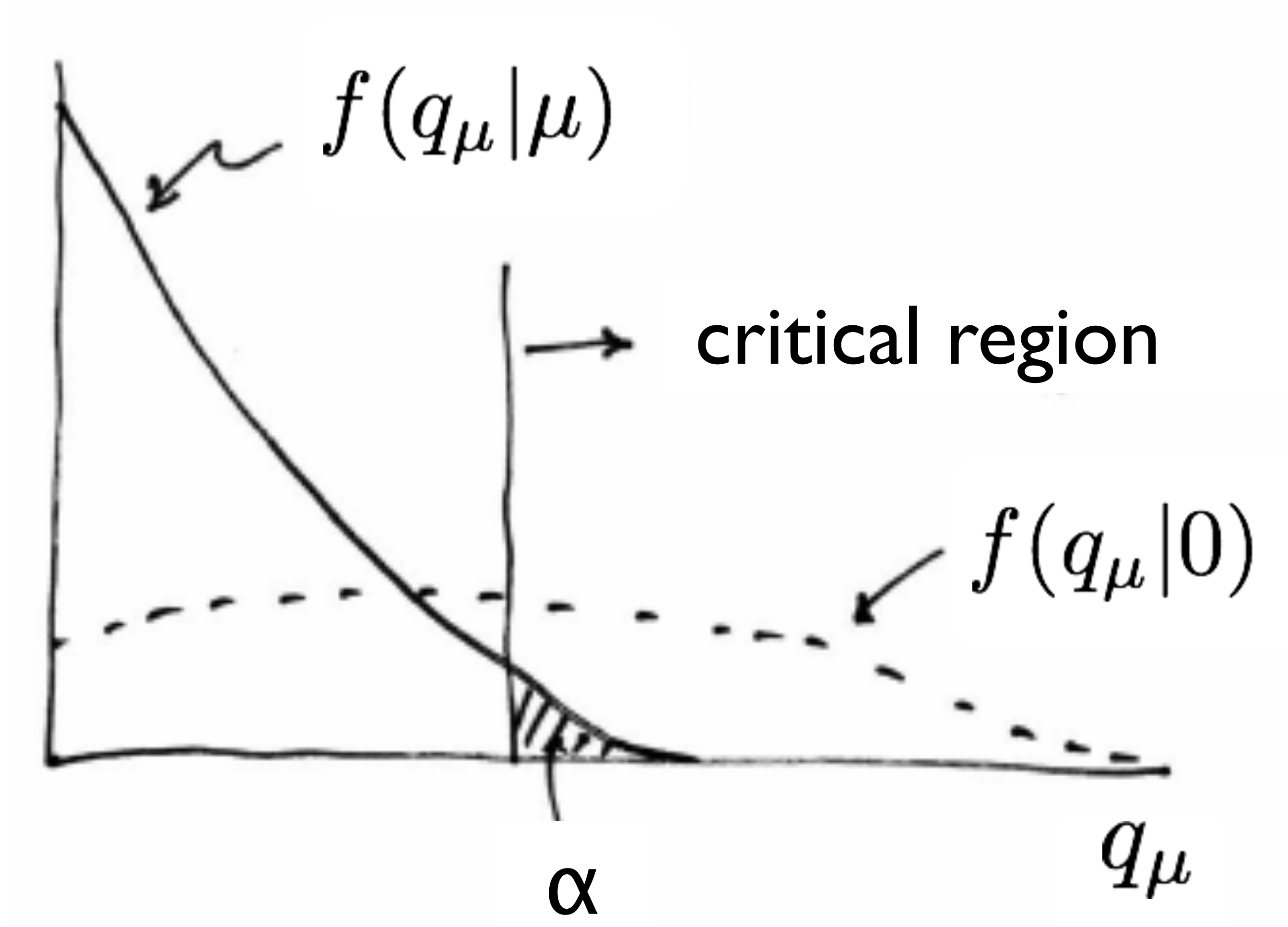


- **This means that one excludes hypotheses to which one has essentially no sensitivity (e.g. m_H = 1000 TeV)**
- **It is called the "spurious exclusion"**

"spurious" = not being what it claims to be

Spurious exclusion

- **In contrast, for a high-sensitivity test, the two pdf's --** $f\,(q_\mu|\mu)$ and $f\,(q_\mu|0)$ **-- are well separated**

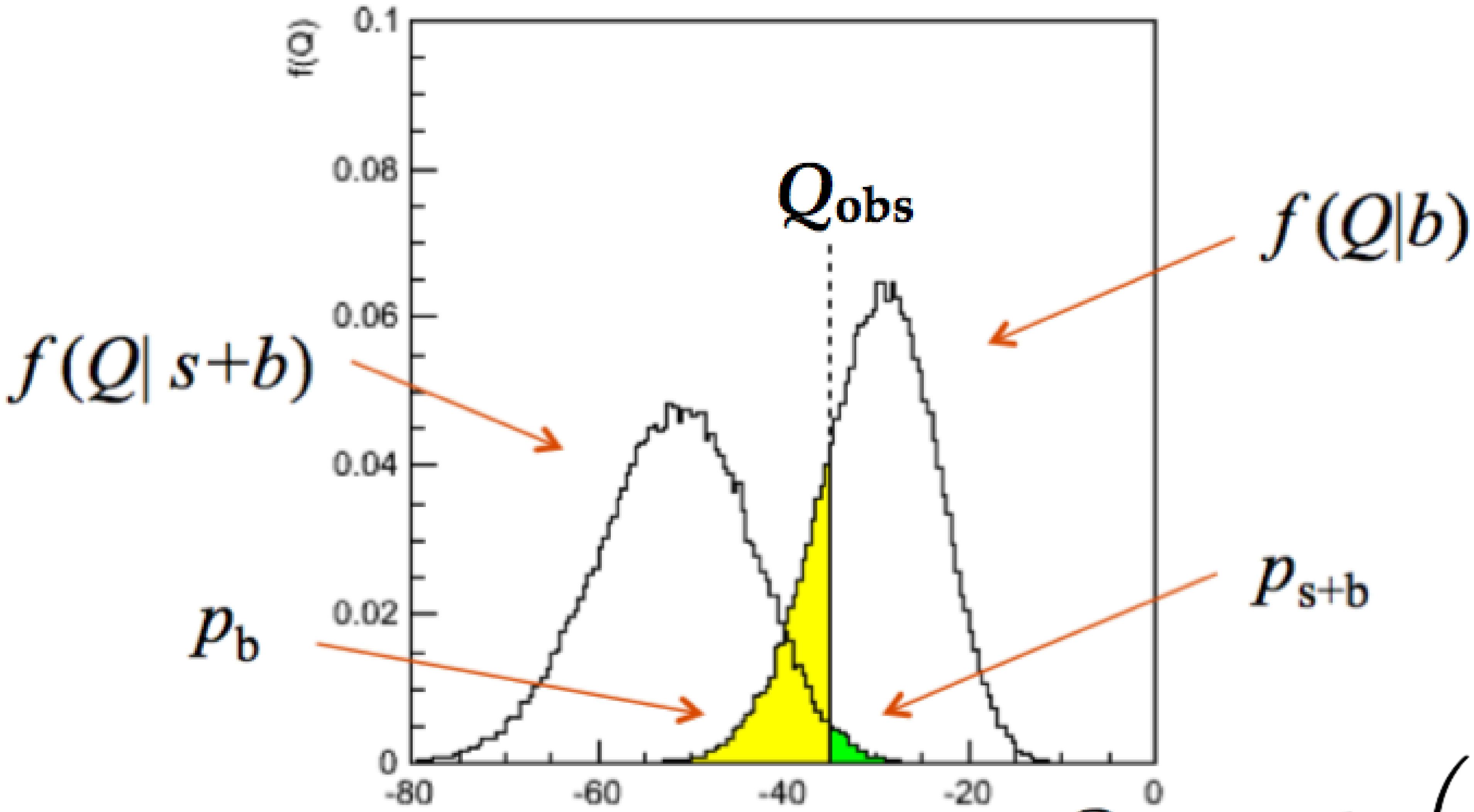


In this case, the power is substantially higher than 1−α.
Use this 'power' as a measure of the sensitivity.

- The problem of excluding values to which one has no sensitivity is known for a long time
- In the 1990s this problem was re-examined for the LEP Higgs search, e.g.
  T. Junk, NIM A 434, 435 (1999); A.L. Read, J. Phys. G 28, 2693 (2002).
  and led to the "CL_s" procedure for upper limits

The CL$_s$ procedure

- In the CL$_s$ formulation, one tests both the $\mu = 0$ (b) and $\mu > 0$ $(s + b)$ hypotheses with the same statistic $Q = -2\ln L_{s+b}/L_b$

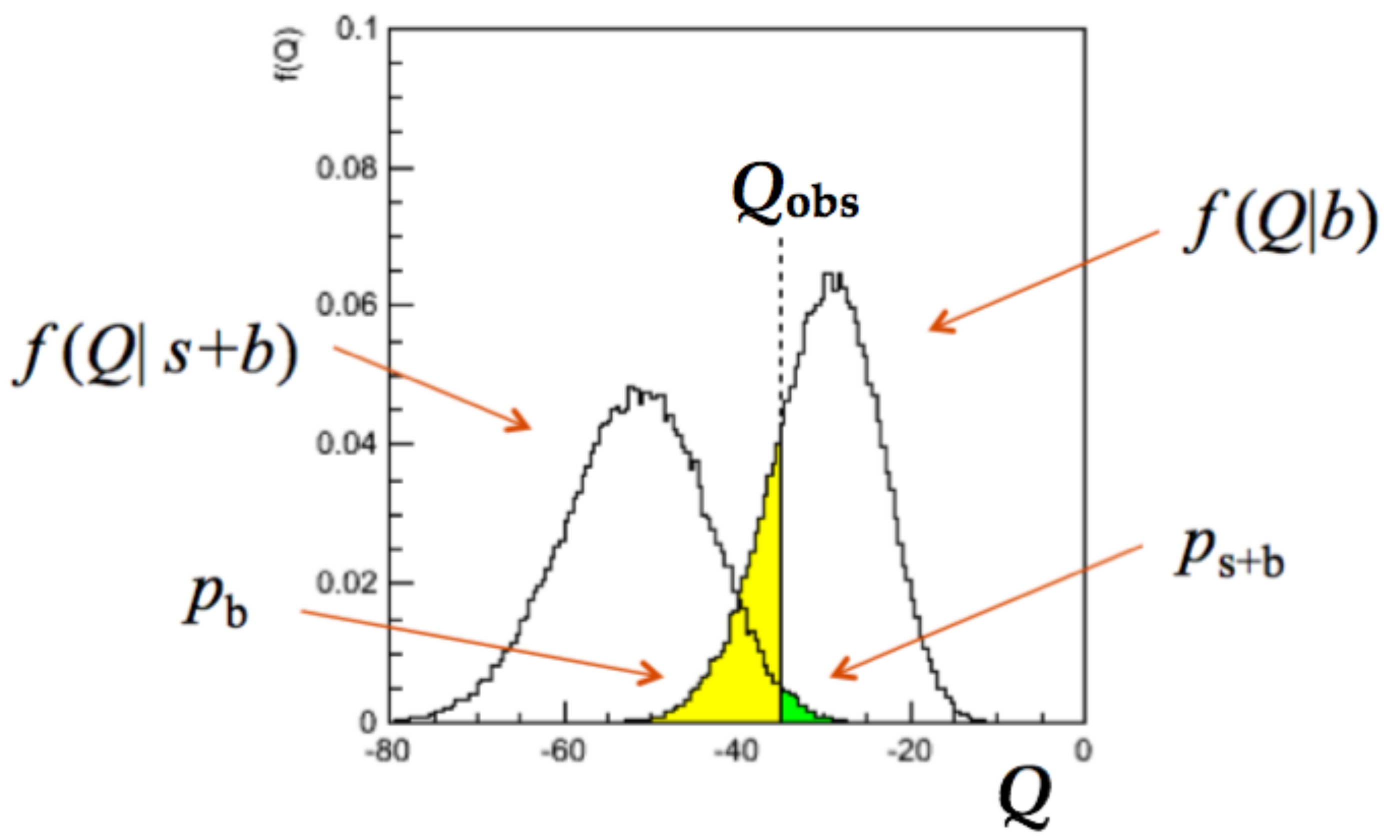


$$Q = -2\ln\left(\frac{L_{s+b}}{L_b}\right)$$

The CL$_s$ procedure

- The CL$_s$ prescription is to base the test on the usual p-value (CL$_{s+b}$), but rather to divide this by CL$_b$$(= 1 - p_b)$

$$\mathrm{CL}_s \equiv \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{p_{s+b}}{1 - p_b}$$

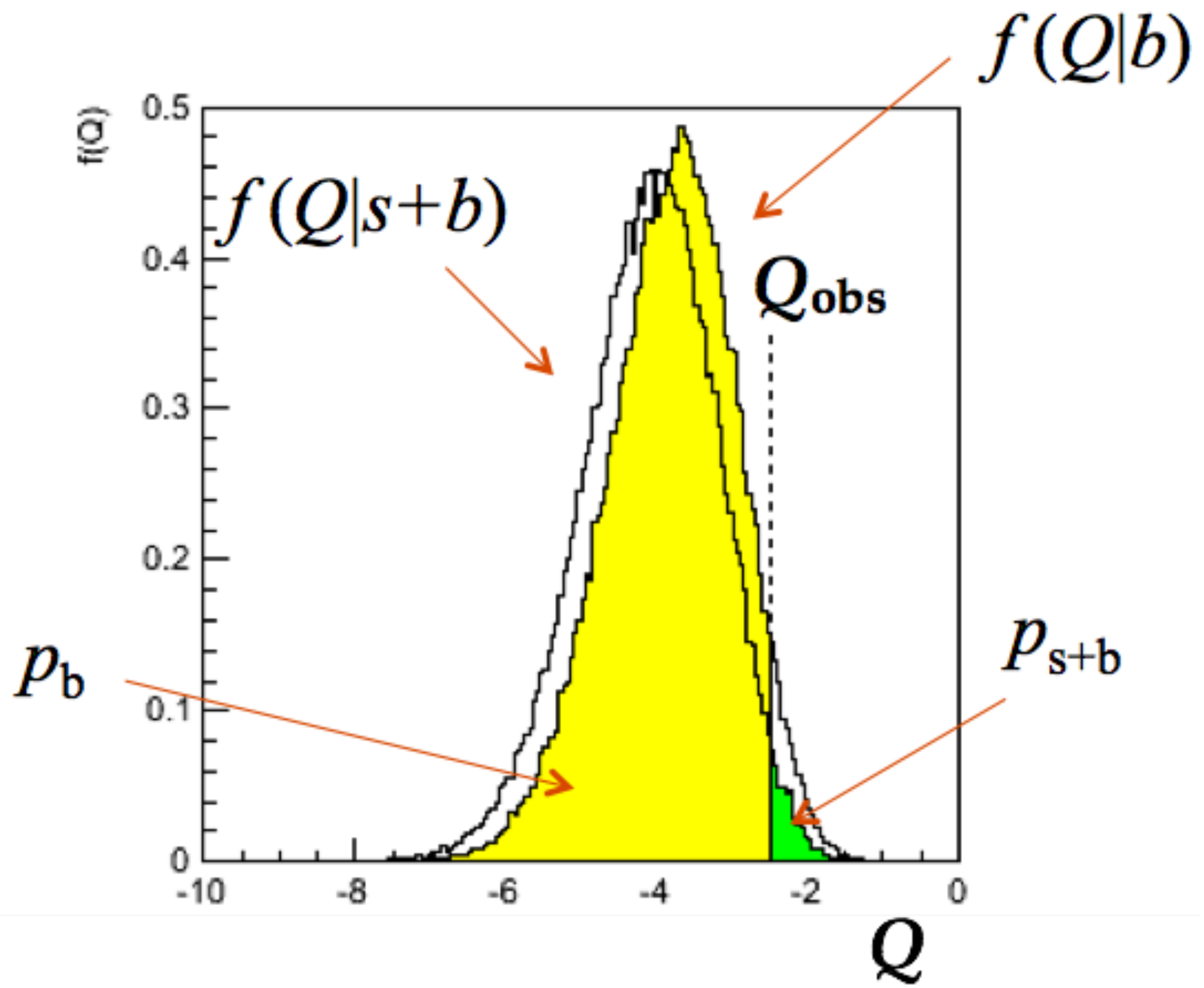


- Reject $s + b$ hypothesis if CL$_s < \alpha$
- Makes "effective" p-value bigger when the two distributions become close, thus preventing exclusion if sensitivity is low

The CL_s procedure

$$\mathrm{CL}_s \equiv \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{p_{s+b}}{1 - p_b}$$

- Reject $s + b$ hypothesis if $\mathrm{CL}_s < \alpha$
- Makes “effective” p-value bigger when the two distributions become close, thus preventing exclusion if sensitivity is low

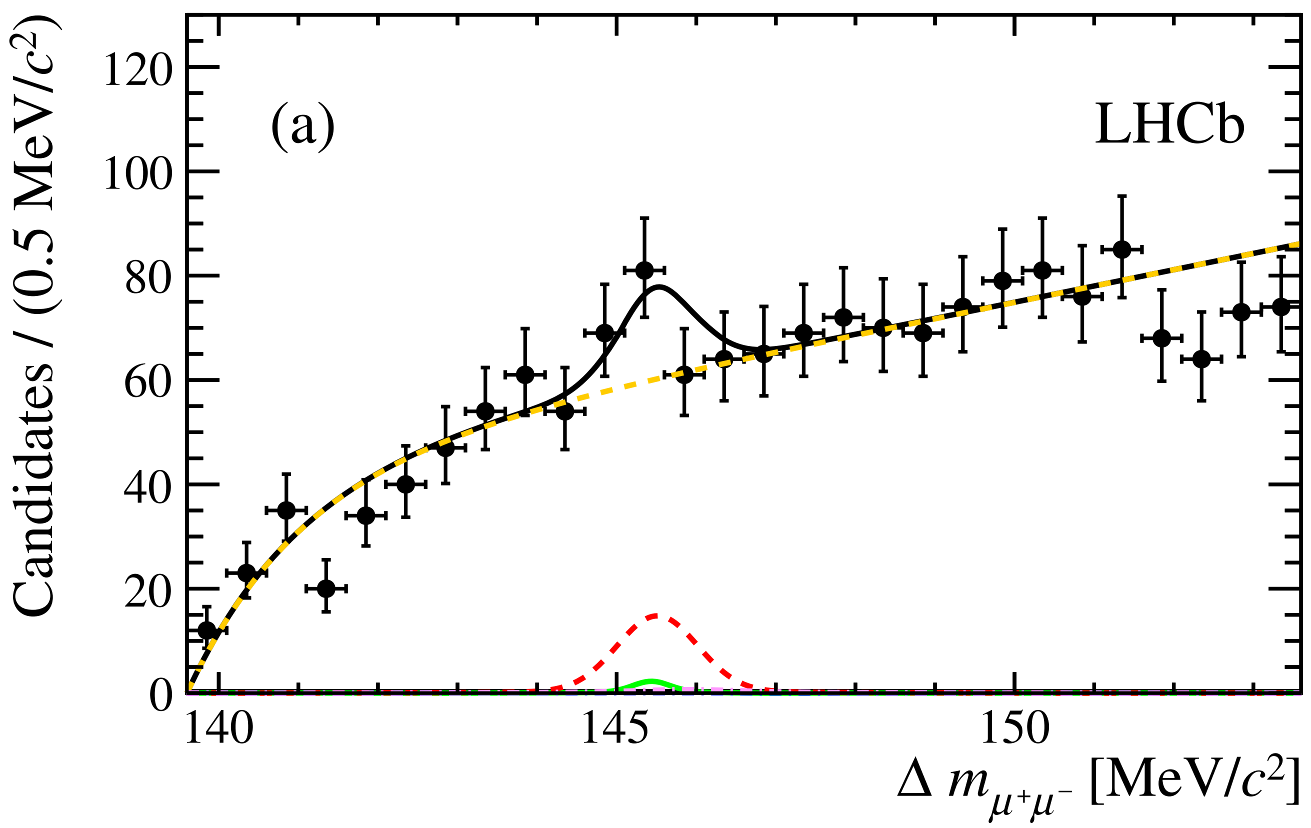


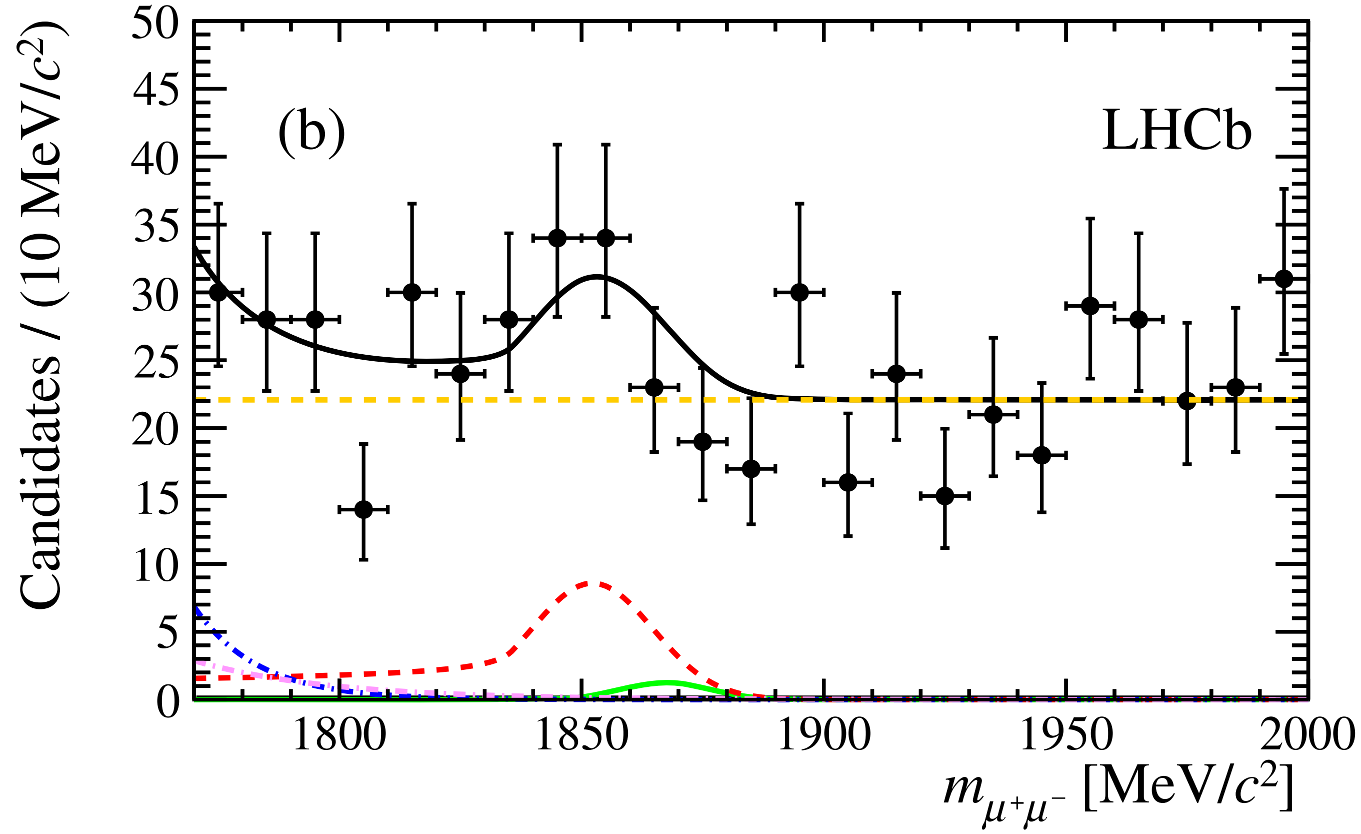
The CL$_s$ procedure - an example

Physics Letters B 725 (2013) 15–24

Search for the rare decay $D^0 \to \mu^+\mu^-$

LHCb Collaboration

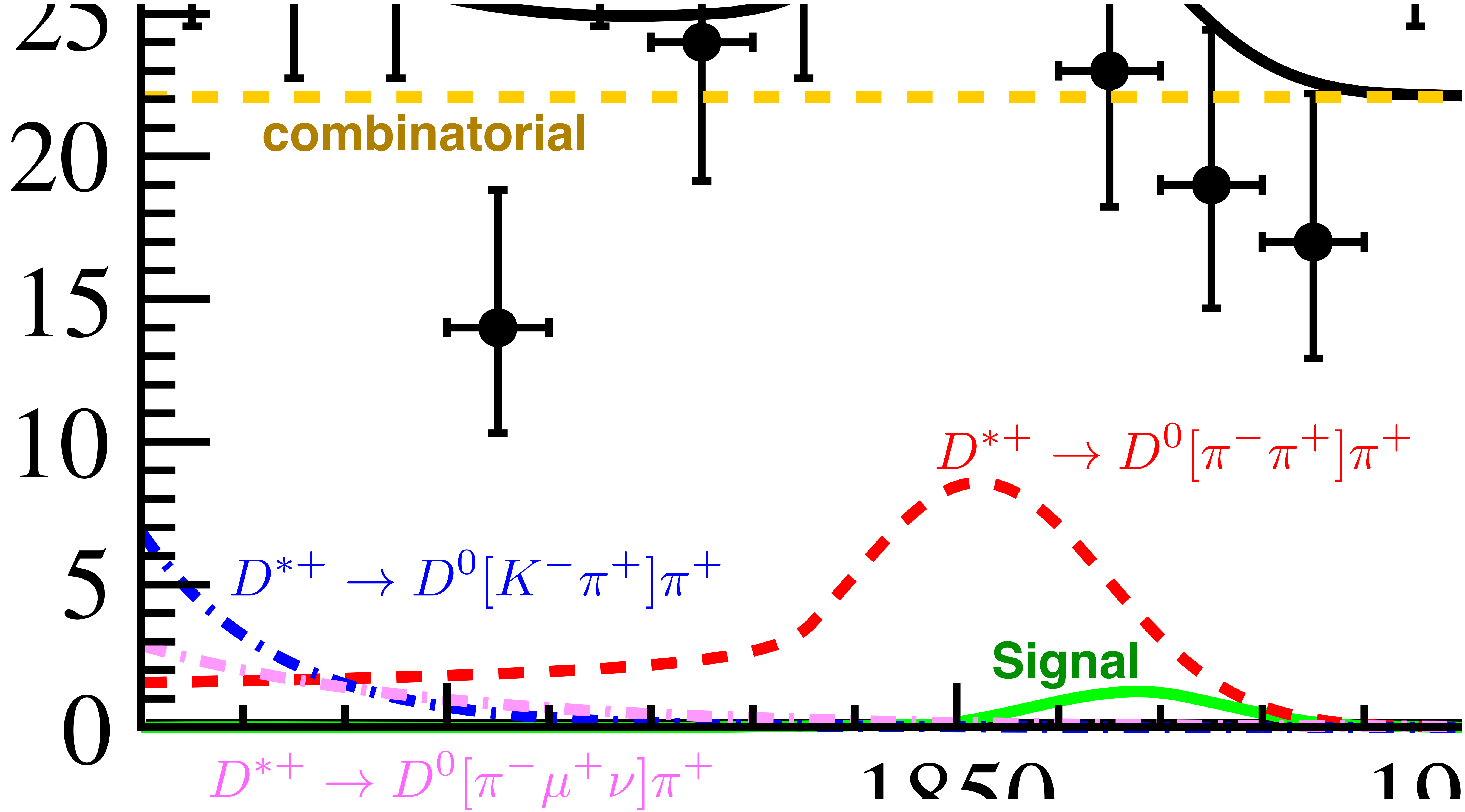




Physics Letters B 725 (2013) 15–24

Search for the rare decay $D^0 \to \mu^+\mu^-$

LHCb Collaboration



Physics Letters B 725 (2013) 15–24

Search for the rare decay $D^0 \to \mu^+\mu^-$

LHCb Collaboration

The $D^0 \to \mu^+\mu^-$ branching fraction is obtained from

$$\mathcal{B}(D^0 \to \mu^+\mu^-) = \frac{N_{\mu^+\mu^-}}{N_{\pi^+\pi^-}} \times \frac{\varepsilon_{\pi\pi}}{\varepsilon_{\mu\mu}} \times \mathcal{B}(D^0 \to \pi^+\pi^-)$$
$$= \alpha \times N_{\mu^+\mu^-} \qquad (1)$$

using the decay $D^{*+} \to D^0(\pi^+\pi^-)\pi^+$ as a normalisation mode, where α is the single event sensitivity, $N_{\pi^+\pi^-(\mu^+\mu^-)}$ are the yields and $\varepsilon_{\pi\pi(\mu\mu)}$ the total efficiencies for $D^{*+} \to D^0(\pi^+\pi^-)\pi^+$ $(D^{*+} \to D^0(\mu^+\mu^-)\pi^+)$ decays. In this section the various contri-

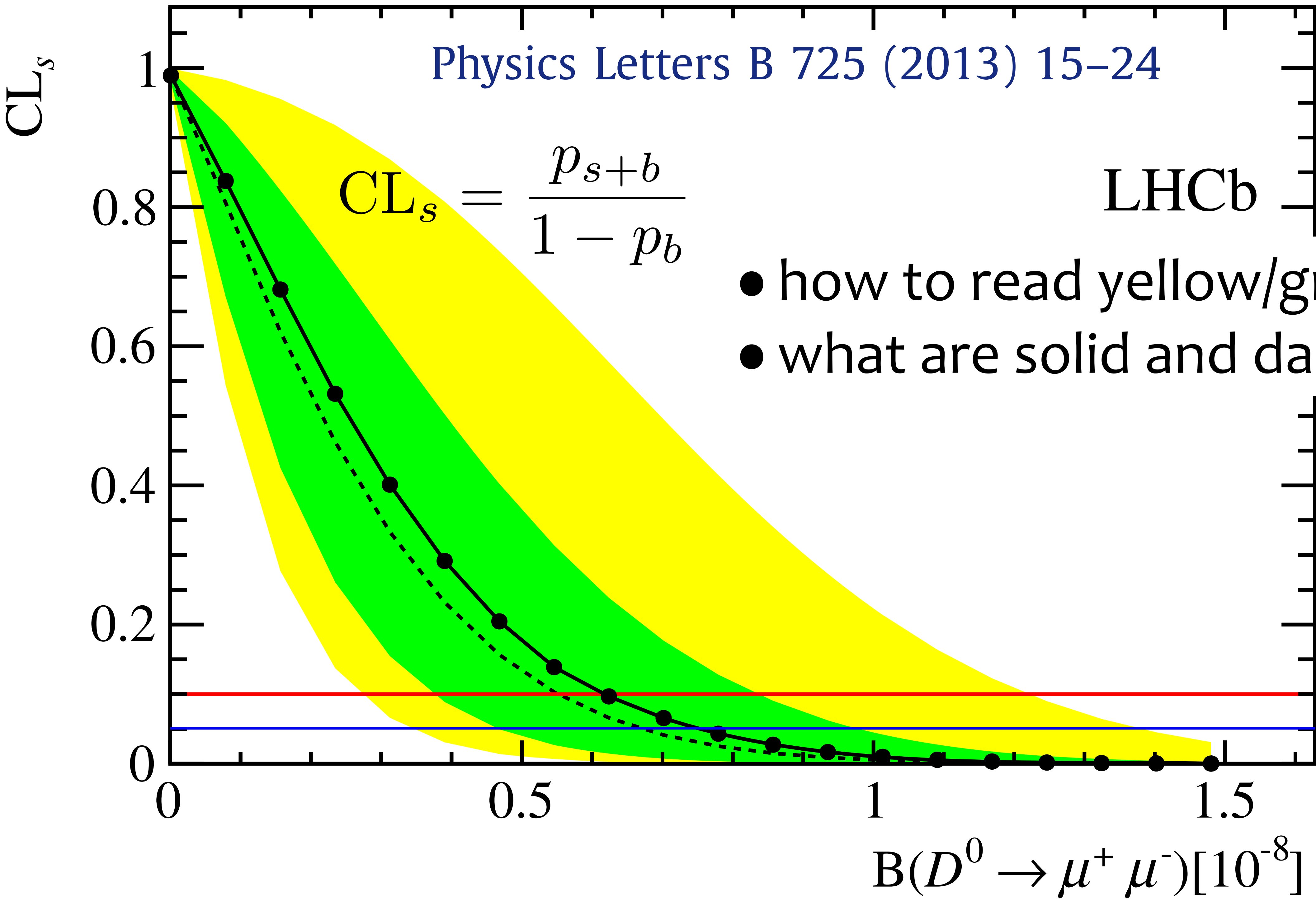
Physics Letters B 725 (2013) 15–24

Search for the rare decay $D^0 \to \mu^+\mu^-$

LHCb Collaboration

- ... shows the Δm and $m_{\mu\mu}$ distributions, together with the one-dimensional binned projections of the two-dimensional fit overlaid. The χ^2/ndf of the fit projections are 1.0 and 1.3, corresponding to probabilities of 44% and 19%, respectively.
- The data are consistent with the expected backgrounds. In particular, a residual contribution from $D^{*+} \to D^0[\pi^-\pi^+]\pi^+$ events is visible among the peaking backgrounds.
- The value obtained for the $D^0 \to \mu^+\mu^-$ branching fraction is $(0.09 \pm 0.30) \times 10^{-8}$.
- Since no significant excess of signal is observed with respect to the expected backgrounds, an upper limit is derived.
- The limit determination is performed, ... in the **RooStats** framework, using the asymptotic CL_s method.

CL_s example

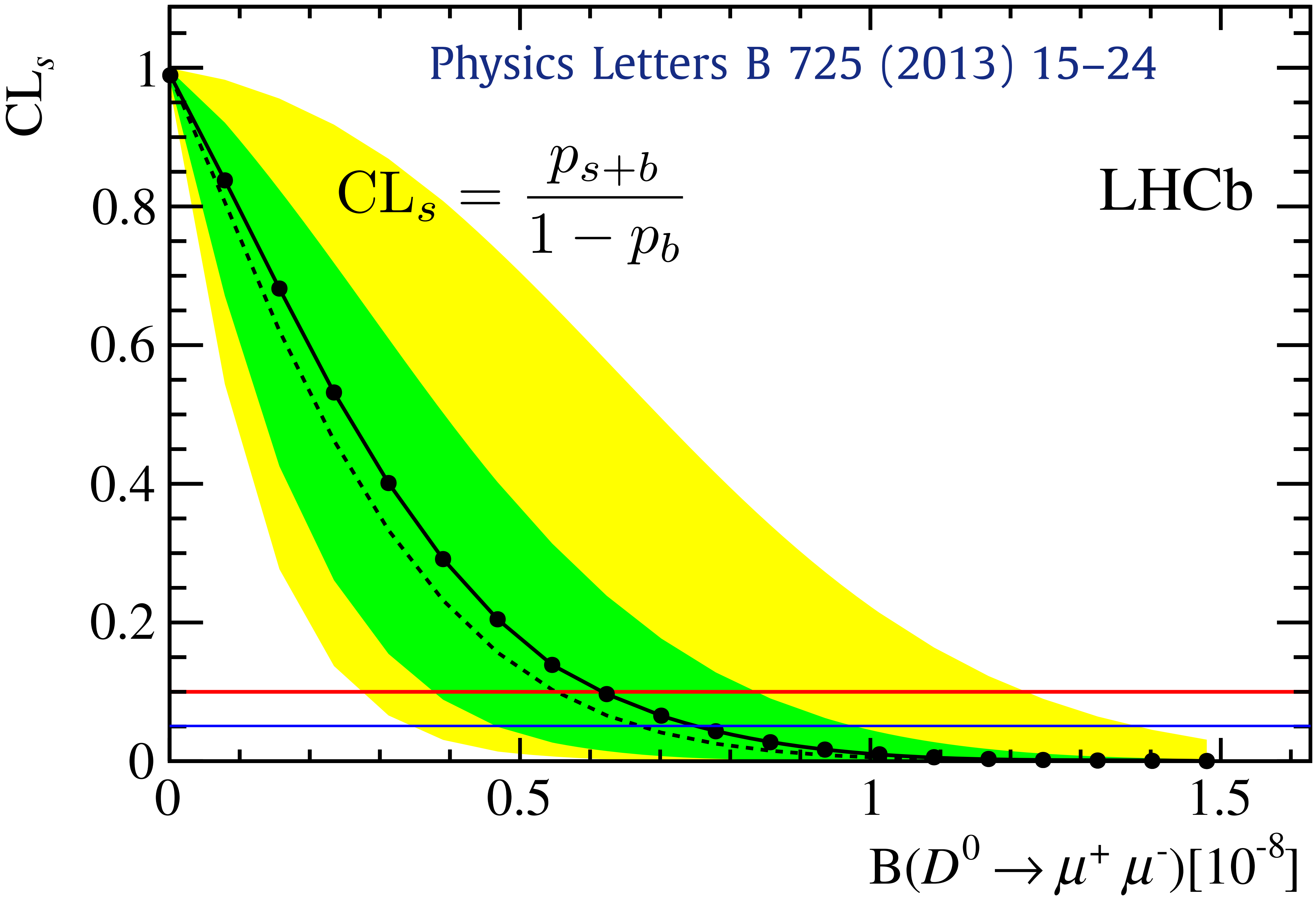


- how to read yellow/green band?
- what are solid and dashed curves?

(*x* axis) assumed branching fraction of signal decay

- —- CL_S distribution from data
- - - - simulated by bkg. only sample; median for the expected CL_S distribution, with given value of s

CL_s example



(Q1) Expected upper limit @ 90% CL, and @ 95% CL?

(Q2) What about observed upper limits?

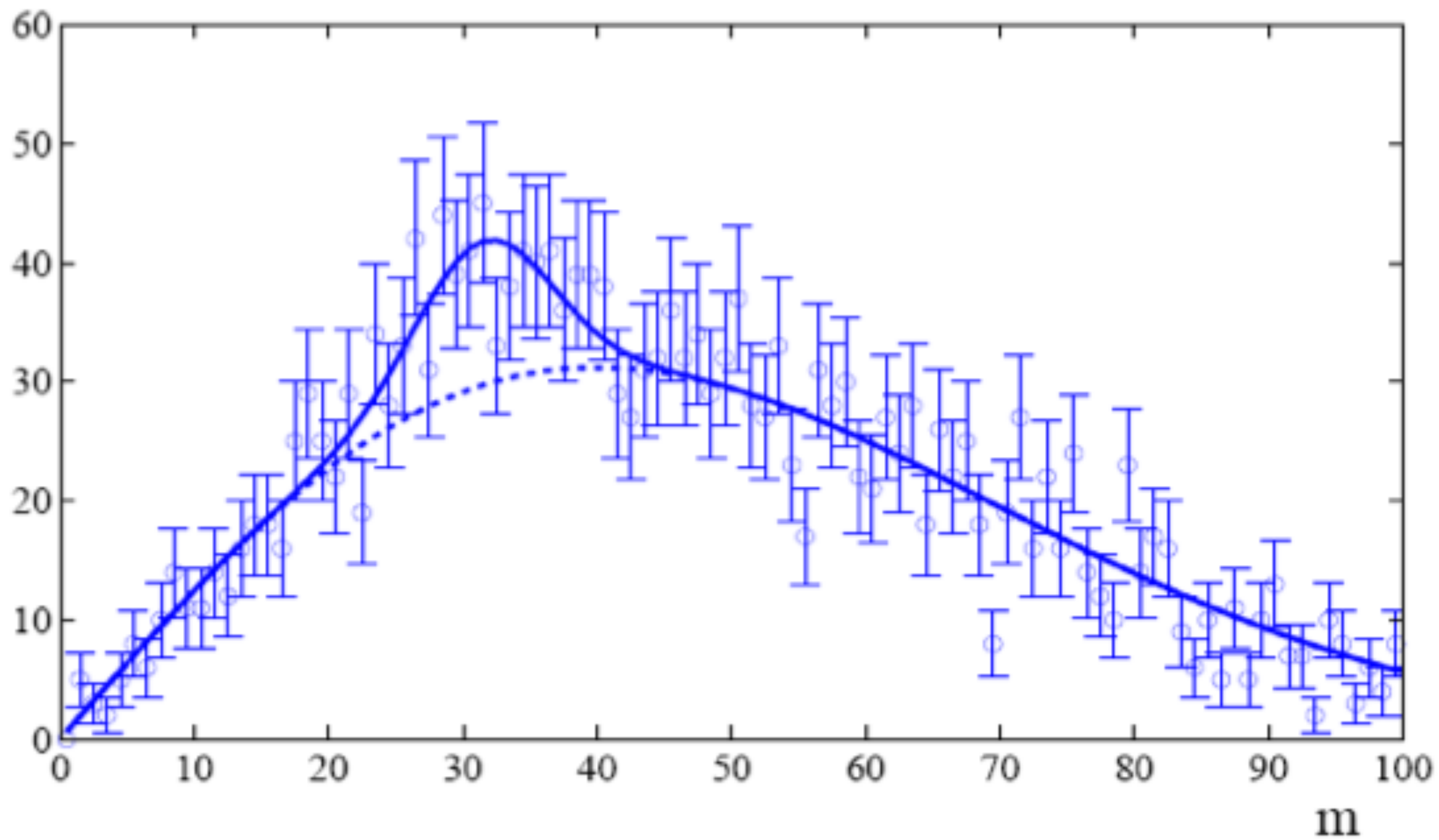
the Look Elsewhere Effect

consider…

- **Suppose you throw a coin 10 times, and you've got 10 heads, zero tails.**
  - It's very unusual.
  - Can you quantify how unusual this result is?
- **In particular, can you say the probability for this kind of peculiarity happening is $1/2^{10}$?**
  - No! Think why!
- **What must then be the correct answer?**

Look-Elsewhere Effect

- Suppose a model for a mass distribution allows for a peak at a mass m with amplitude μ
- and the data show a bump at a mass m_0



How consistent is this with the no-bump (μ =0) hypothesis?

- First, suppose that the mass peak value m_0 was known a priori.
- Test consistency of bump with the $\mu = 0$ hypothesis with e.g. L-ratio

$$t_{\text{fix}} = -2\ln\left(\frac{L(0, m_0)}{L(\mu, m_0)}\right)$$

where “fix” indicates that the mass peak value is fixed to m_0.

- The resulting p-value

$$p_{\text{local}} = \int_{t_{\text{fix,obs}}}^{\infty} f(t_{\text{fix}}|0)\, dt_{\text{fix}}$$

gives the probability to find a value of t_{fix} at least as great as the observed value at the specific mass m_0, and is called the local p-value.

- Now, suppose we did not know where to expect a peak. In other words, the signal can be found at every value of m.
- What we want is the probability to find a peak at least as significant as the one observed **anywhere** in the distribution
- For this, include the mass as an *adjustable parameter* in the fit, then test significance of peak using

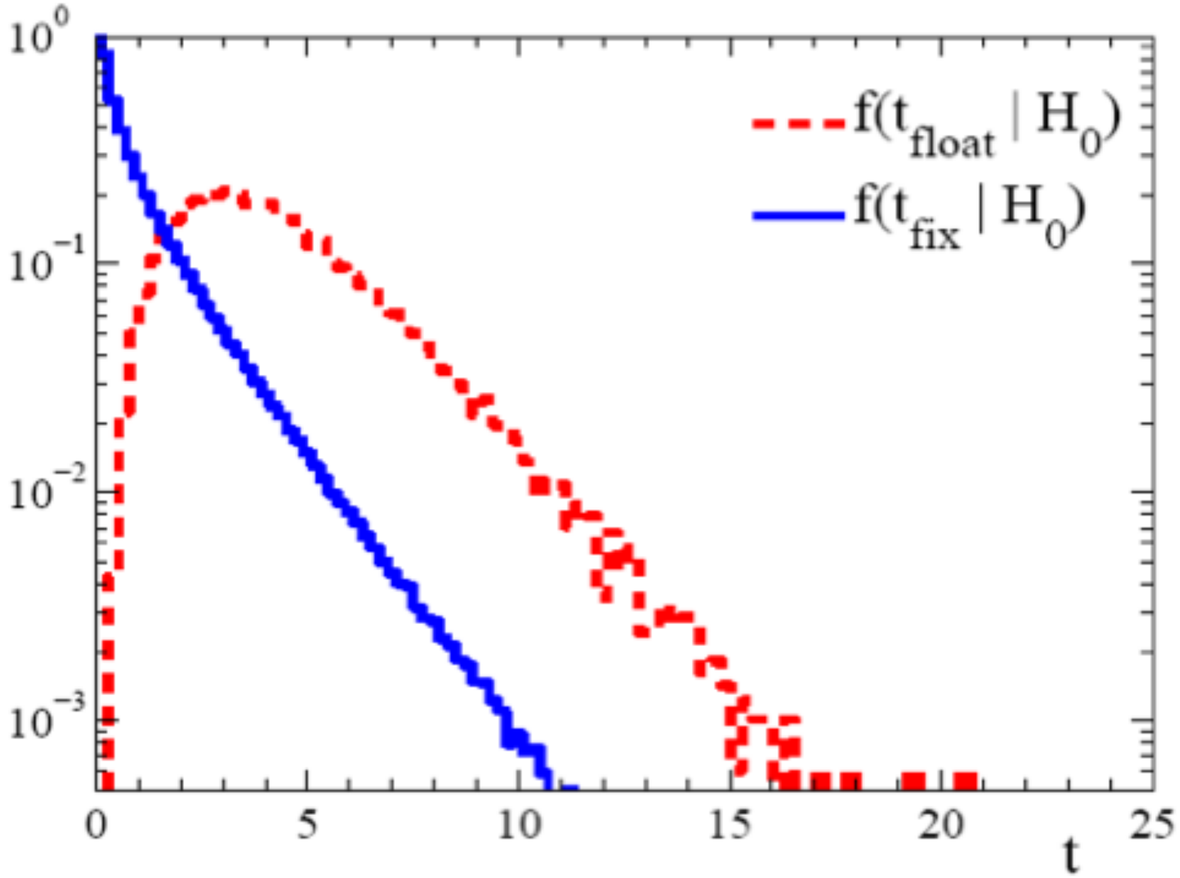
$$t_{\mathrm{float}} = -2 \ln \frac{L(0)}{L(\mu, m)}$$

Note: m does not appear in the μ=0 model

$$p_{\mathrm{global}} = \int_{t_{\mathrm{float,obs}}}^{\infty} f(t_{\mathrm{float}}|0)\, dt_{\mathrm{float}}$$

t_{fix} vs. t_{float}

- For a sufficiently large data sample, $t_{\text{fix}} \sim \chi^2$ for 1 deg. of freedom (*Wilk's theorem*)
- For t_{float} there are two adjustable parameters, μ and m, and naively Wilk's theorem says $t_{\text{float}} \sim \chi^2$ for 2 d.o.f.



But, Wilk's theorem does not hold in the floating mass case because one of the parameters (m) is not defined in the $\mu = 0$ model.

$\therefore$ getting t_{float} distribution is more difficult.

Approximate correction for LEE

- Need to relate the p-values for the fixed and floating-mass analyses (at least approximately)
- (Gross & Vitells) The p-values are approximately related by
  Gross and Vitells, EPJC 70:525-530 (2010), arXiv:1005.1891

$$p_{\text{global}} \approx p_{\text{local}} + \langle N(c) \rangle$$

where $\langle N(c) \rangle$ = mean # of *upcrossings* of $-2\ln L$ in the fit range based on a threshold

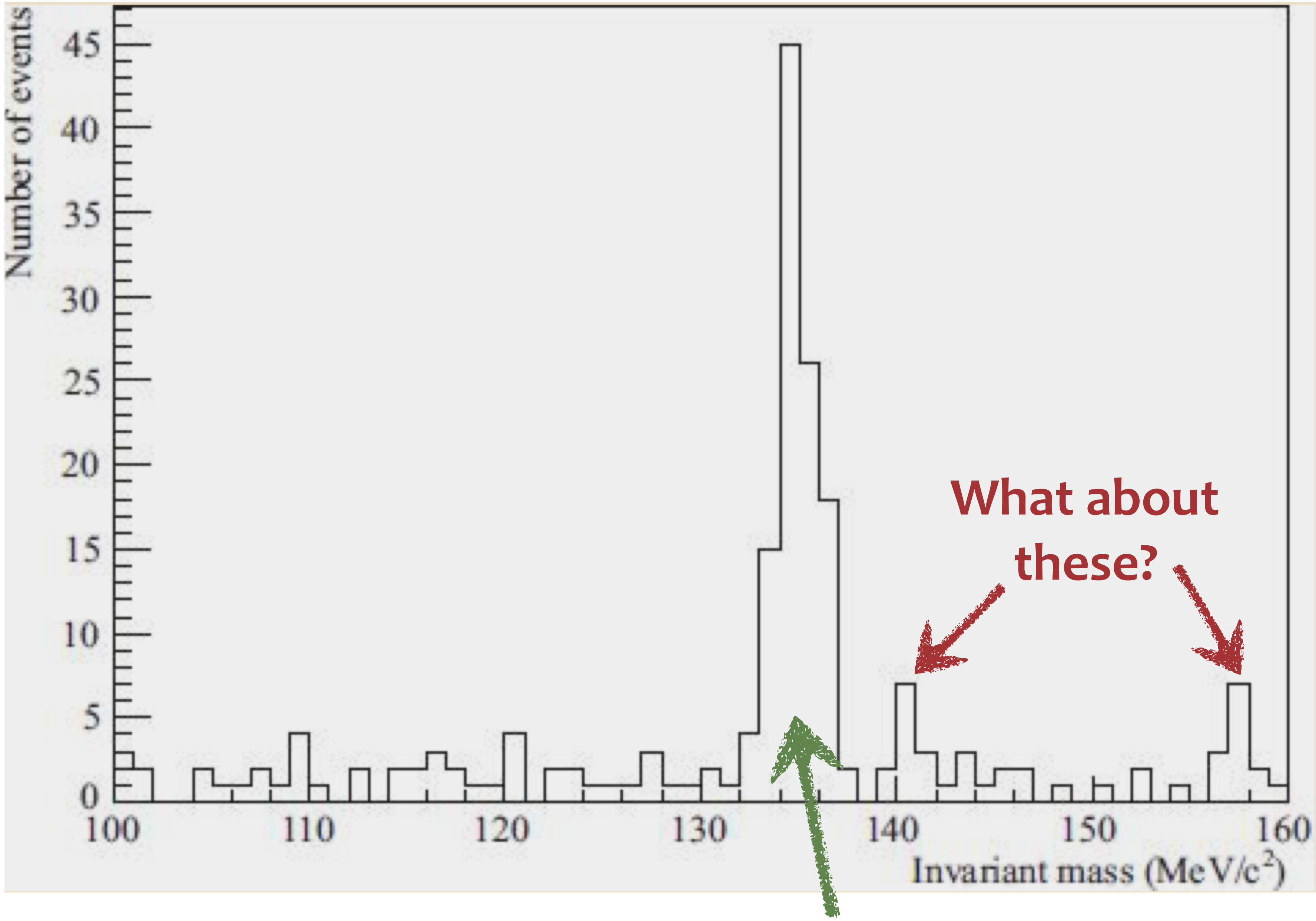
$$c = t_{\text{fix}} = Z^2_{\text{local}}$$

- We may carry out the full MC (time and CPU-consuming) or do fixed-m analysis and apply a correction factor (much faster!)

looking for $\pi^0 \to e^+ e^- \gamma$

a simulation shown in A. Bevan's book

- 100 signal& 100 background events are generated over [100 MeV, 160 MeV]
- histogram in 60 bins



OK, we have a clear peak at a known location!

bonus topic

Experimental Bias & Blind Analysis

vs.


Wilhelm van Osten
& “Clever” Hans
Ask Hans the horse to add any two numbers, and he tapped his hoof the correct number of time!

the "Clever Hans effect"

Hans answered questions correctly even when his trainer was not in the room!

- Psychologist Oskar Pfungst made a very important discovery:

  if no one in the room knew the correct answer to the question being asked of Hans, Hans didn't know the answer either!

- Apparently, Hans was picking up on subtle (conscious or unconscious) cues given by the questioners.

  Hans was indeed clever, but not in the way people thought.

- Medical applications: double-blind study of placebo effects

When will experimental results be biased?

Consider a typical branching fraction measurement

$$\mathcal{B} = \frac{(N_{\rm obs} - N_{\rm bkg})}{N_{\rm norm}} \frac{\epsilon_{\rm norm}}{\epsilon_{\rm sig}}$$

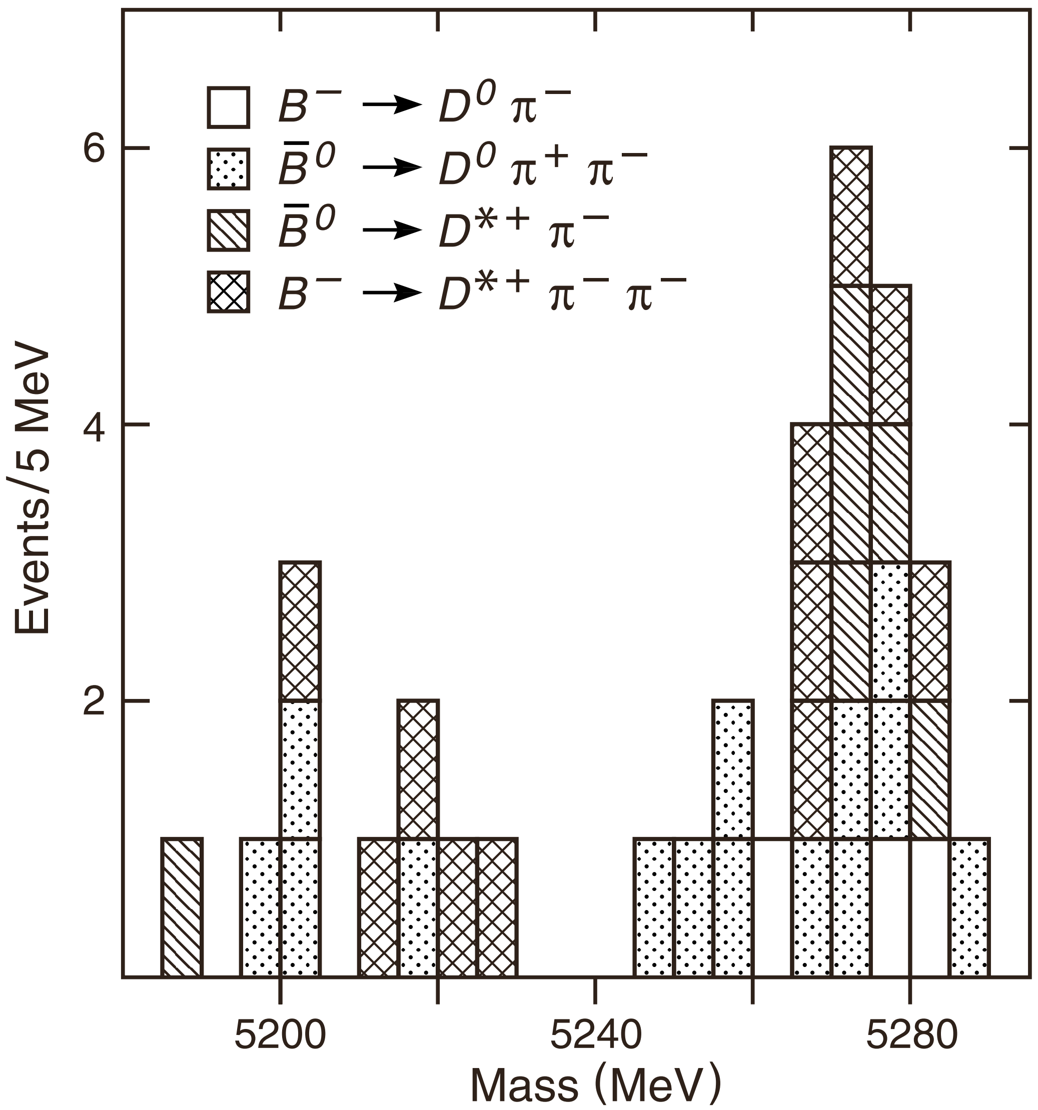
Experimental biases

- Determination of any elements in the final number can be wrong due to incomplete knowledge about the experimental apparatuses, background contaminations, etc.
- All such sources shall be studied and corrected for. Any uncertainties in these shall be included in the systematic uncertainty.

Experimenters' bias

- This is difficult (impossible?) to assess, and has to be prevented at all costs.

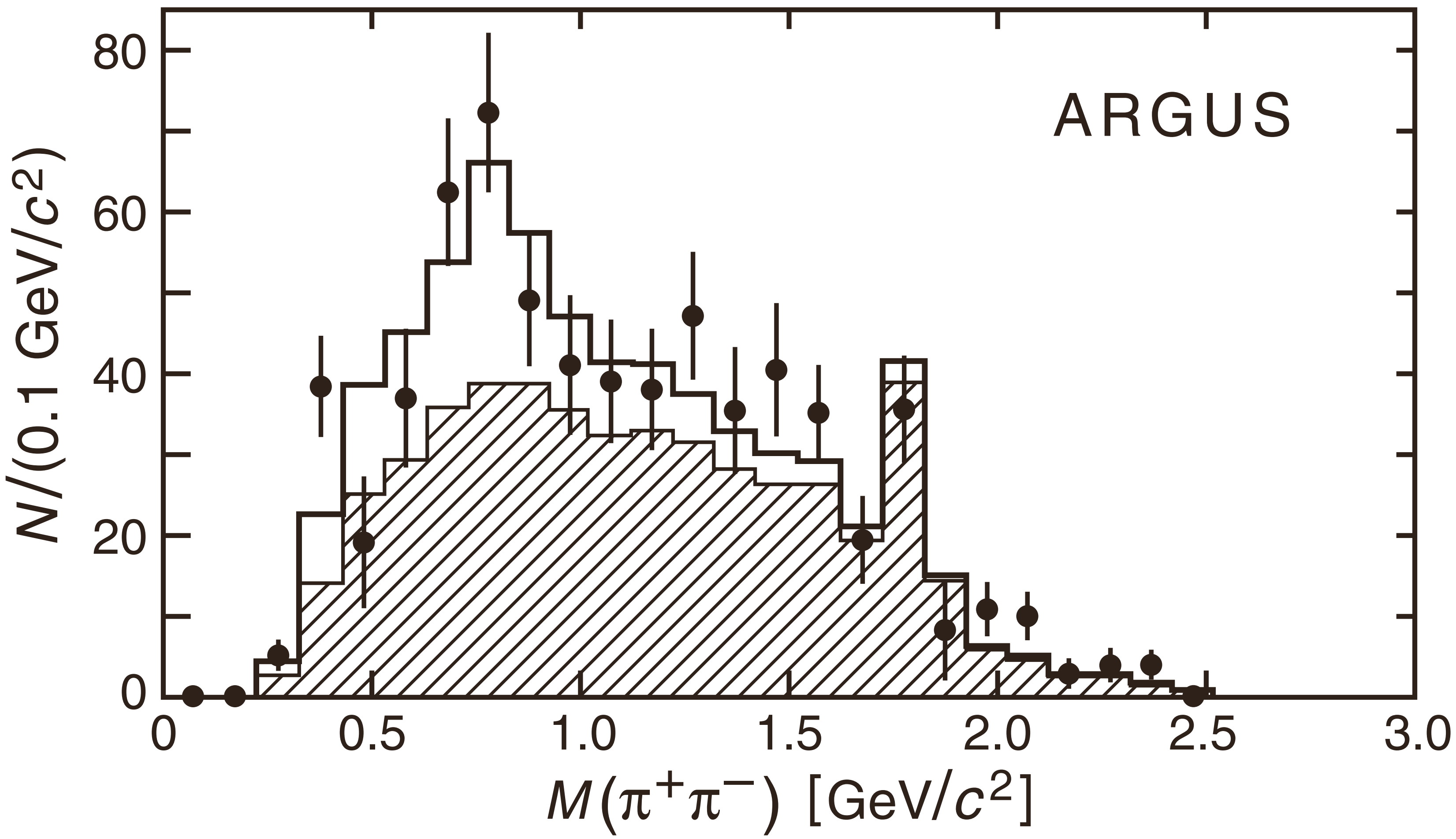
(Ex) experimental bias



First ever observation of exclusive B decays by CLEO (1983)

Mode	CLEO I branching fraction (%)		PDG 97 branching fraction (%)
$B^- \to D^0\pi^-$	4.2 ± 4.2		0.53 ± 0.05
$B^- \to D^{*+}\pi^-\pi^-$	4.8 ± 3.0	⬅ ?? ➡	0.21 ± 0.06
$\bar{B}^0 \to D^0\pi^+\pi^-$	13 ± 9		dominated by $D^{*+}\pi^-$
$\bar{B}^0 \to D^{*+}\pi^-$	2.6 ± 1.9		0.26 ± 0.04
Sum	24.6 ± 10.5		1.26 ± 0.10

(Ex) experimental bias

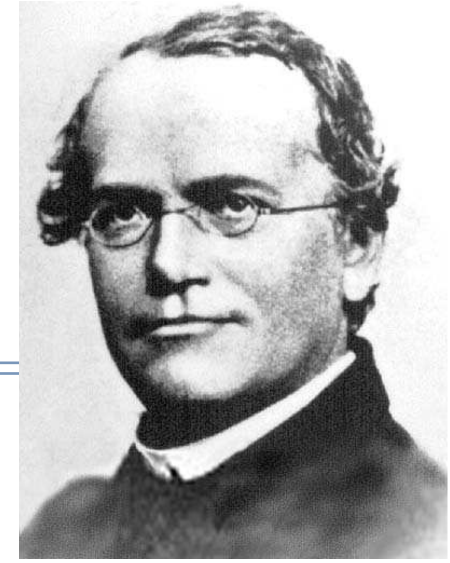


$B(B^- \to \rho^0 \ell^- \bar{\nu}) = (1.13 \pm 0.36 \pm 0.27) \times 10^{-3}$ **from the analysis**

$\mathcal{B} = (1.58 \pm 0.11) \times 10^{-4}$ from PDG

perhaps, due to incomplete determination of background shape & amount ?

Was Gregor Mendel lucky?



- **Mendel discovered the law of genetic inheritance.**
- **But his published data fits his model too well:**
- **speculations**
  - publishing only his "best data", throwing out the others, and/or
  - taking data until the results seem to agree his pre-formulated theory, then deciding to stop and publish
  - …

Experimenter's bias?

content.usatoday.com

Sep 21, 2011

File drawer effect: Science studies neglecting negative results

Comment | Recommend | Share 35 | Tweet | G+1 8

By Dan Vergano, USA TODAY

Some scientific disciplines are reporting far fewer experiments that didn't work out than they did twenty
go, suggests an analysis of the scientific literature.

n particular, economists,
usiness school
esearchers and other
ocial scientists, as well as
ome biomedical fields,
ppear increasingly
usceptible to the "file-
rawer" effect -- letting experiments that fail to prove
go unpublished -- suggests the *Scientometrics* journal
by Daniele Fanelli of Scotland's University of Edinbur

CAPTION By Gaston

"One of the most worrying ... is the loss of negative data. Results that do not confirm expectations—because they yield an effect that is either not statistically significant or just contradicts an hypothesis—are crucial to scientific progress, ... Yet, a lack of null and negative results has been noticed in innumerable fields.

Stopping bias

… how to handle some of the ways we fool ourselves. One example: Millikan measured the charge on an electron … It's a little bit off because he had the incorrect value for the viscosity of air. … look at the history of measurements of the charge of an electron, after Millikan. If you plot them as a function of time, you find that one is a little bit bigger than Millikan's, and the next one's a little bit bigger than that, and the next one's a little bit bigger than that, until finally they settle down to a number which is higher.

Why didn't they discover the new number was higher right away? It's a thing that scientists are ashamed of—this history—because it's apparent that people did things like this: When they got a number that was too high above Millikan's, they thought something must be wrong—and they would look for and find a reason why something might be wrong. When they got a number close to Millikan's value they didn't look so hard. And so they eliminated the numbers that were too far off, and did other things like that..

by R. Feynman

Some suggestions

Never determine your event-selection criteria using the data sample that you will use to measure the signal

Always check to see whether your signal is robust as you vary your cuts

Look at all the distributions you can think of for your signal and compare them with what you expect

Be careful not to underestimate the systematic errors associated with ignorance of (1) the signal efficiency, (2) background composition, and (3) background shapes

...

from Jeff Richman's lecture in Les Houches (1997)

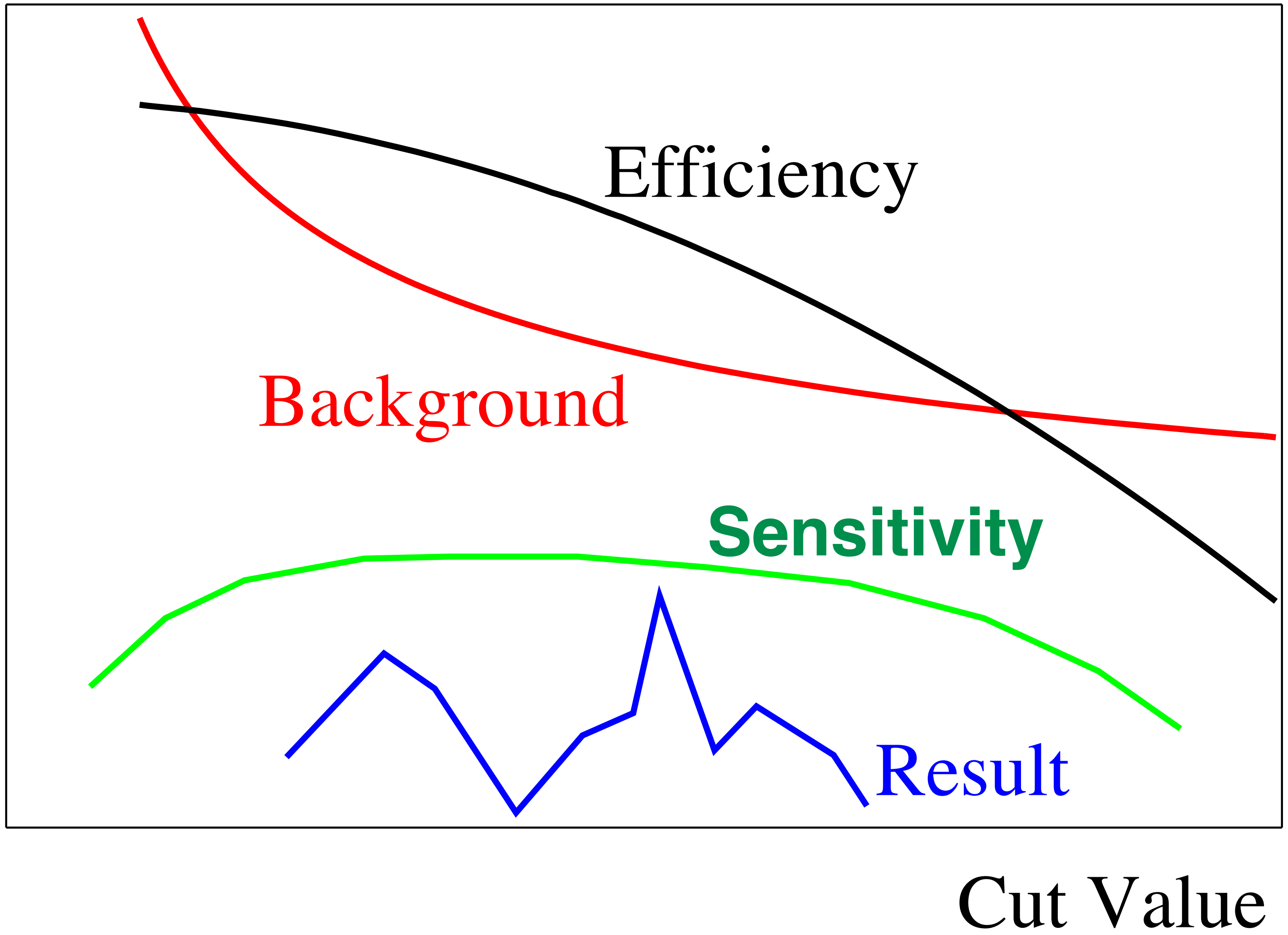
Blind analysis

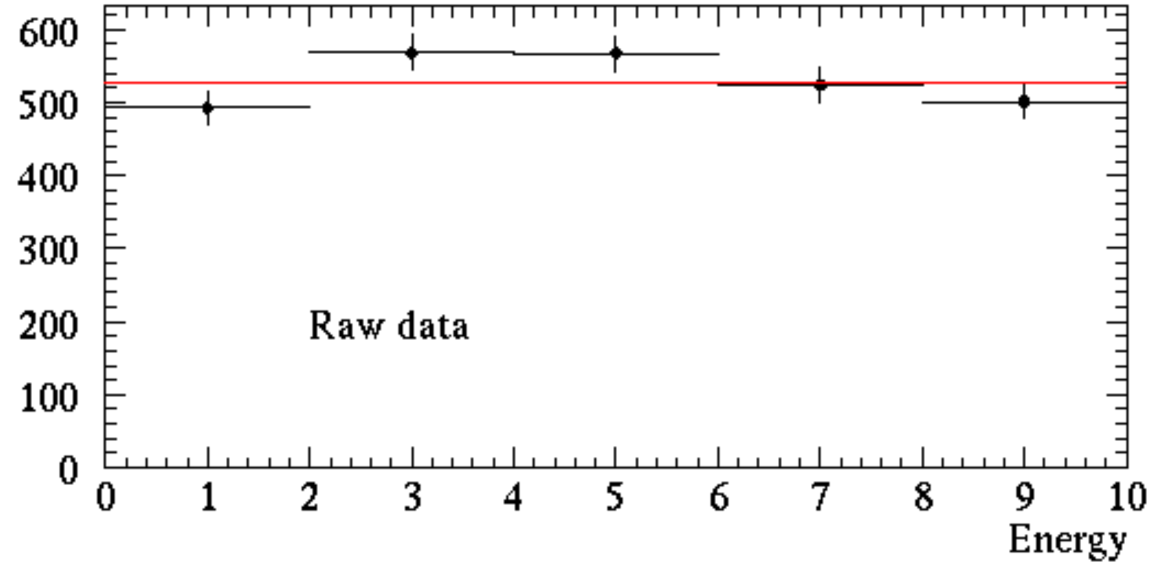
- a technique for avoiding experimenter's biases
- You commit ahead of time you will publish the result you get when you "unblind"
- Blind analysis does NOT mean
  - You never look at the data
  - You can't correct a mistake if you find one after unblinding
  - The analysis is necessarily correct — It just means that it's blind and less prone to experimenter's bias
  - A non-blind analysis is not necessarily wrong. It's only left more open to the risk of biases

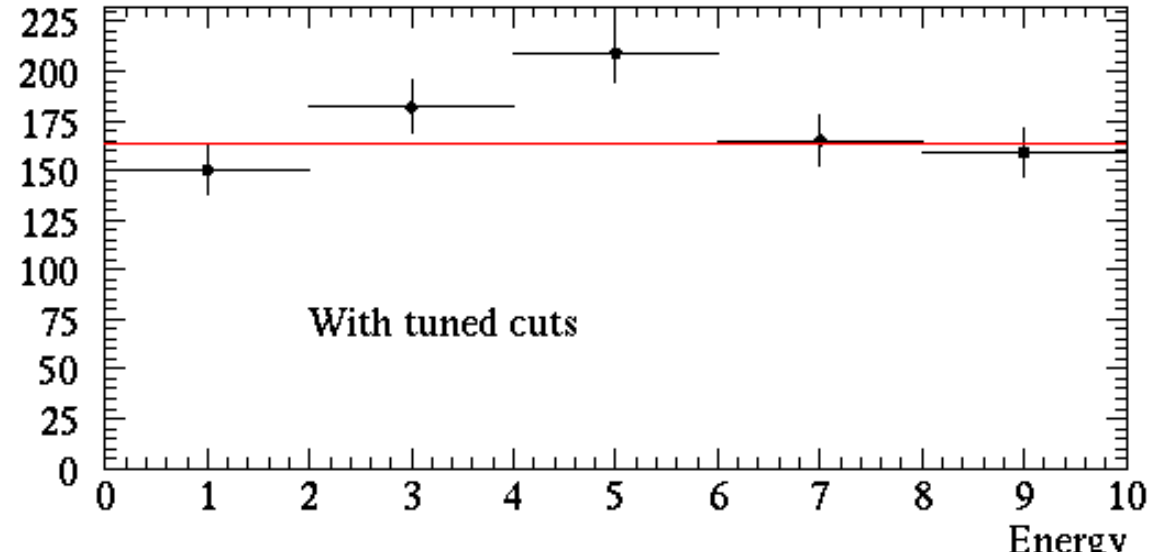
mechanisms producing human biases

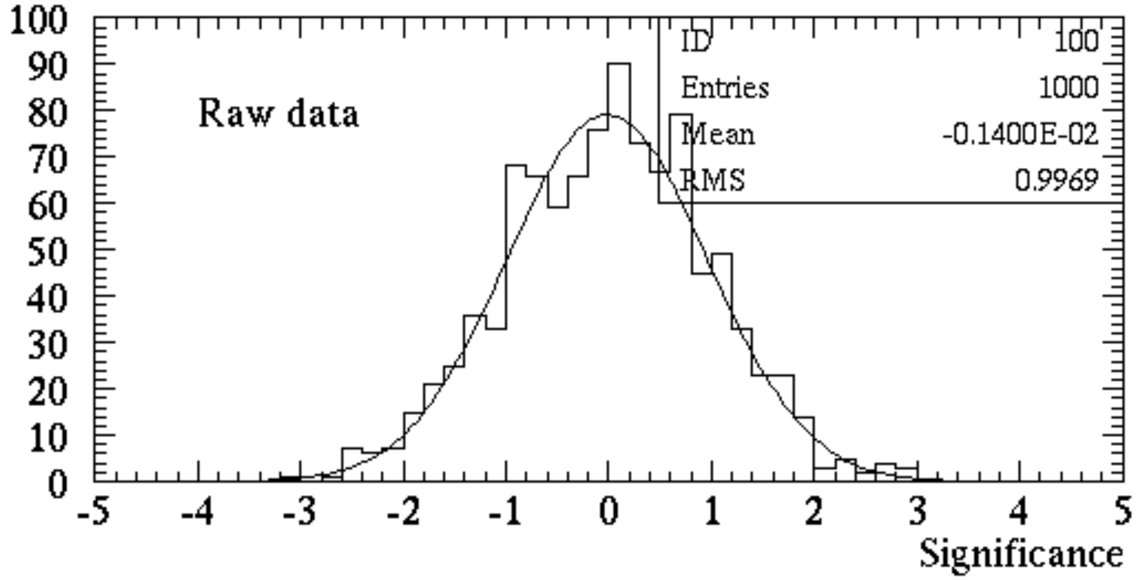
- **cut tuning on data**
- **when to stop?**
  - You do an analysis and get a very strange result
  - Spend a few days for checking, find a bug in the code, fix it.
  - Then your result is consistent with prediction.
  - You decide to stop and write a paper.

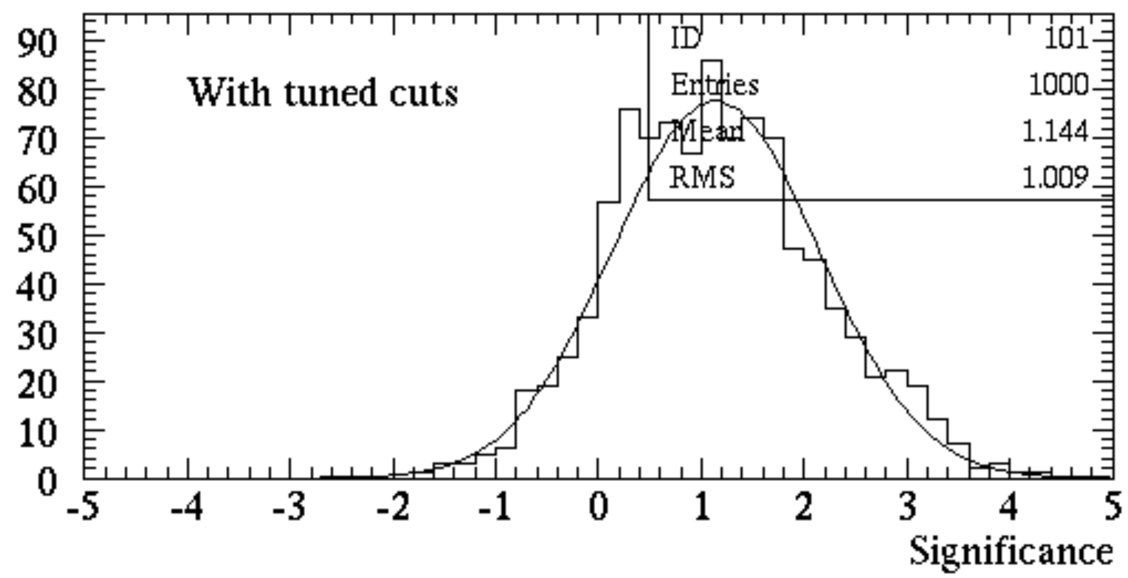
(Q) Had your initial result agreed with the prediction, would you have ever detected a bug in your code?











All the figures in this slide are just cartoon pictures, nothing to do with any real incidence.

Blind analysis — examples

VOLUME 70, NUMBER 8 PHYSICAL REVIEW LETTERS 22 FEBRUARY 1993

Improved Upper Limit on the Branching Ratio $B(K_L^0 \to \mu^\pm e^\mp)$

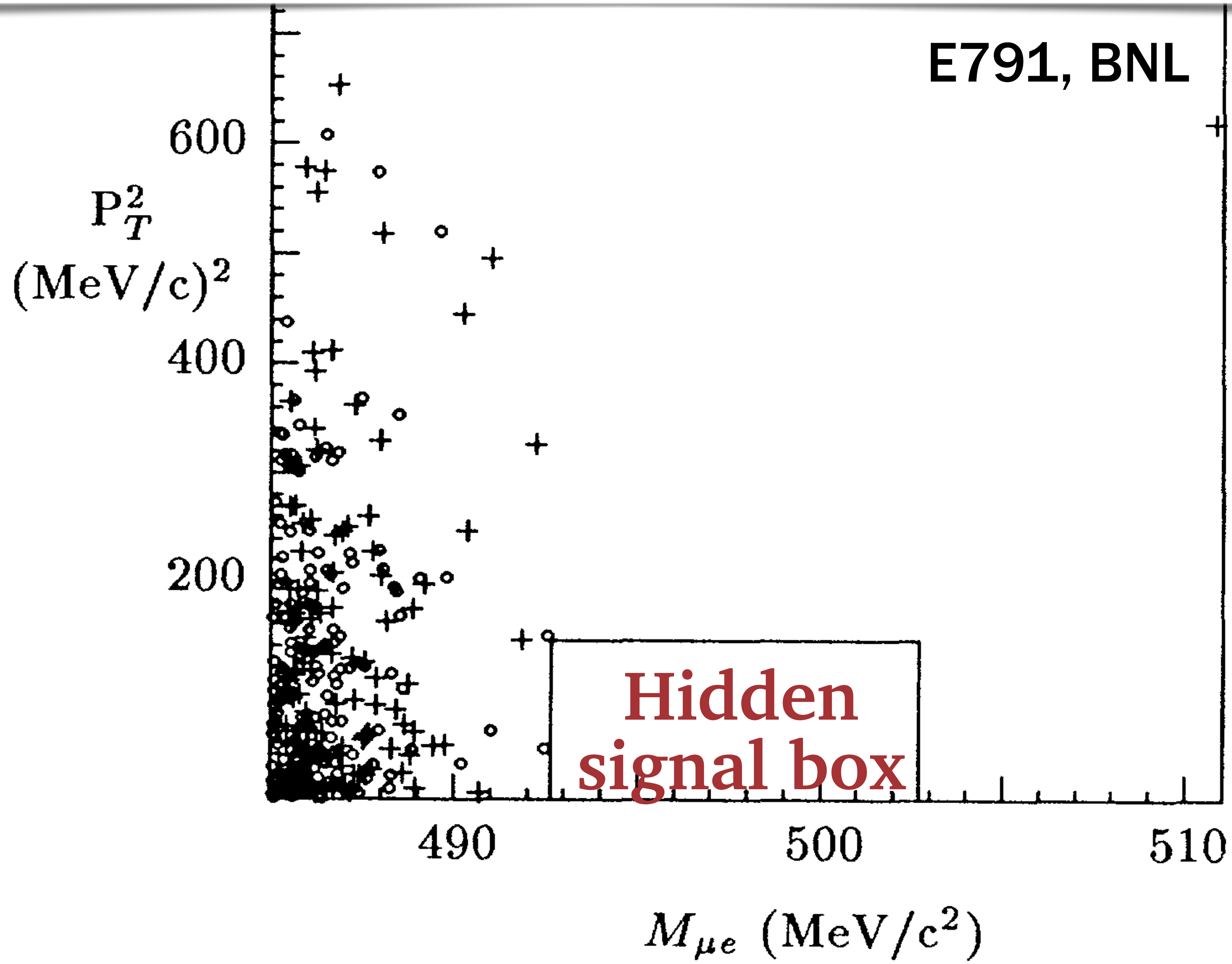
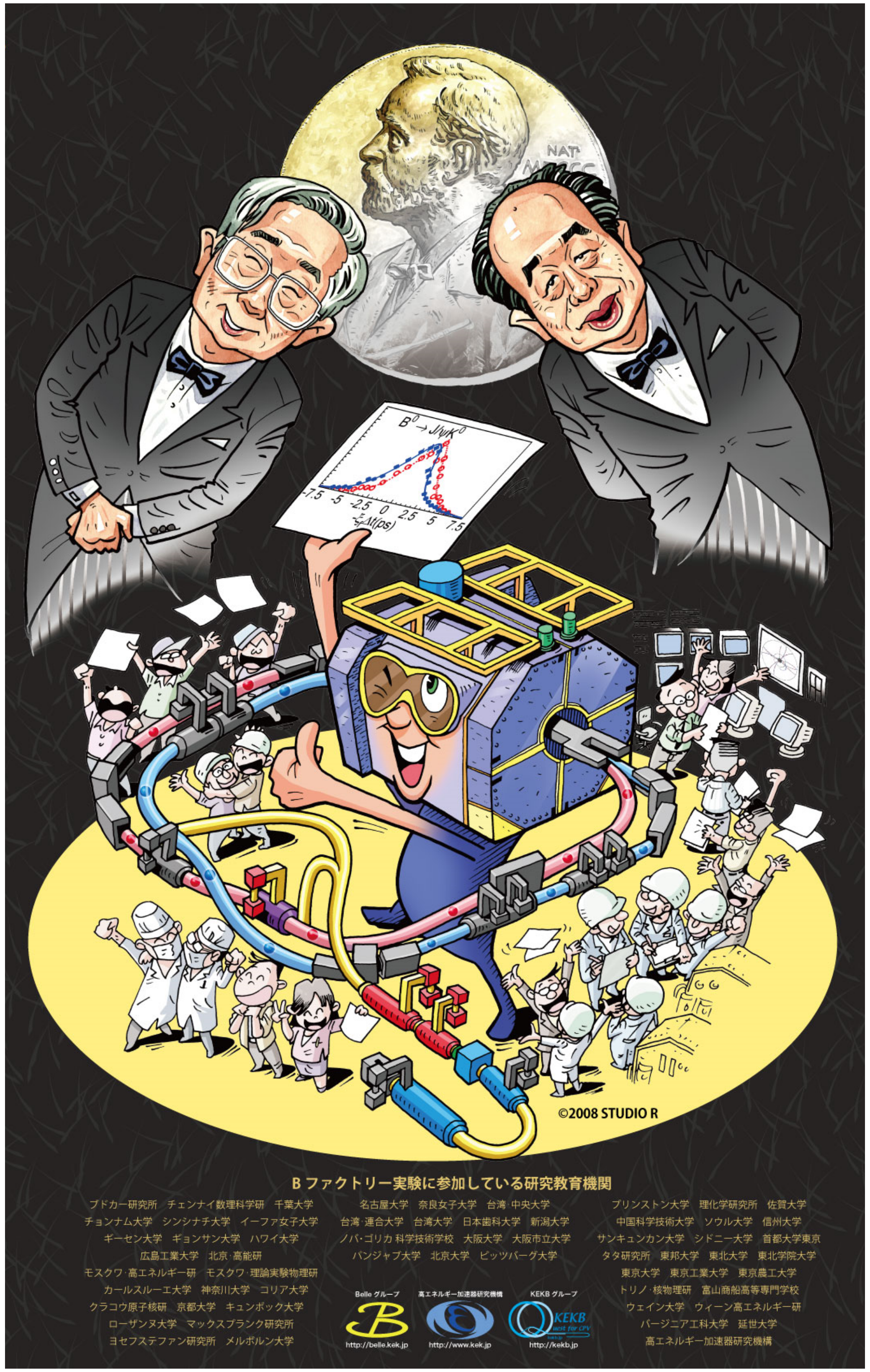
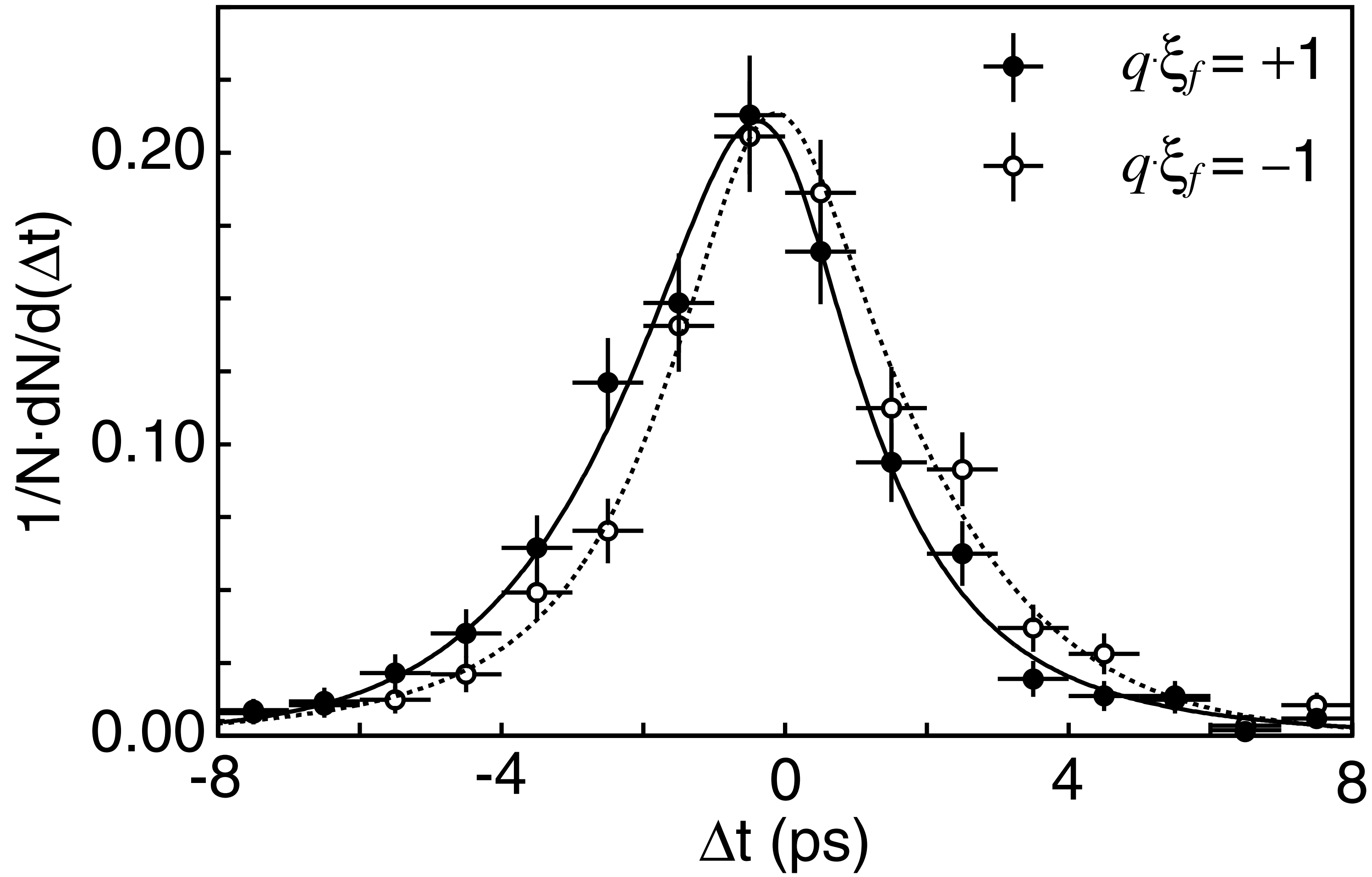


FIG. 2. Plot of P_T^2 vs $M_{\mu e}$. Plus signs are 1989 data and circles are 1990 data.

Blind analysis — examples



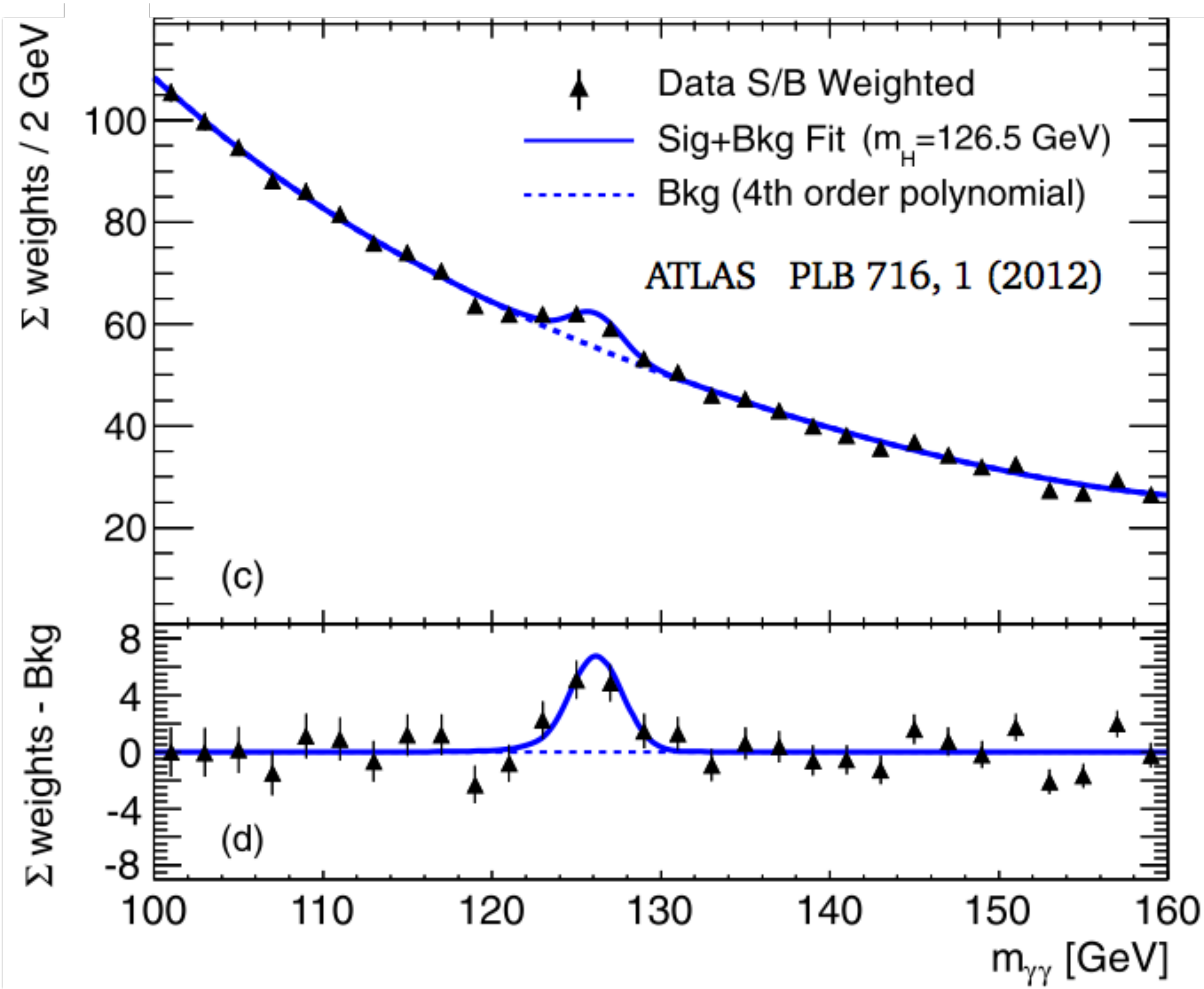


The value of $q\xi_f$ was hidden until all the procedure was firmly established and the collaboration agreed on un-blinding.

Belle (and BaBar, too)

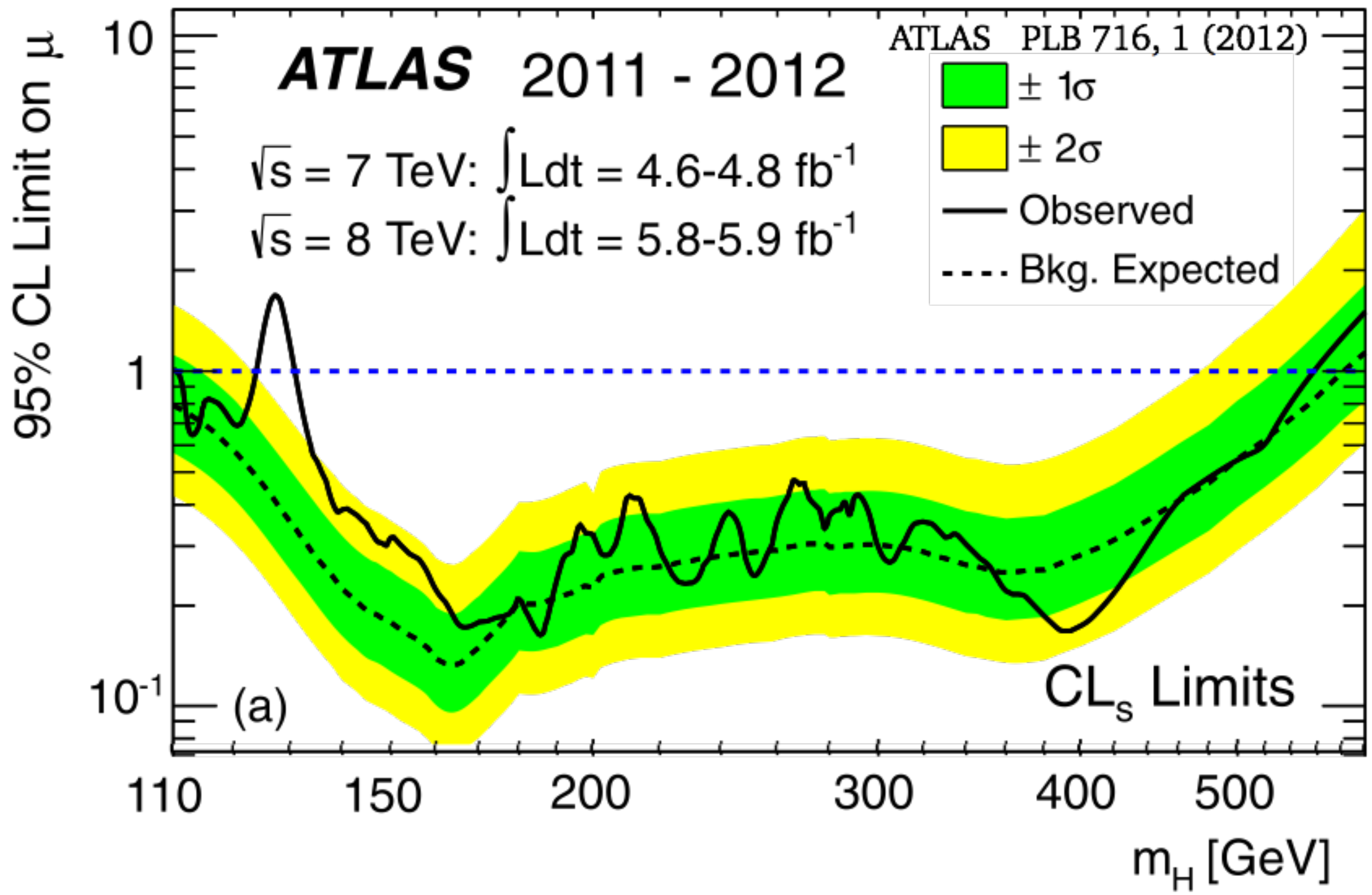
for first observation of CPV in B^0 to confirm the KM mechanism

Wrapping-up *& test what you've learned*



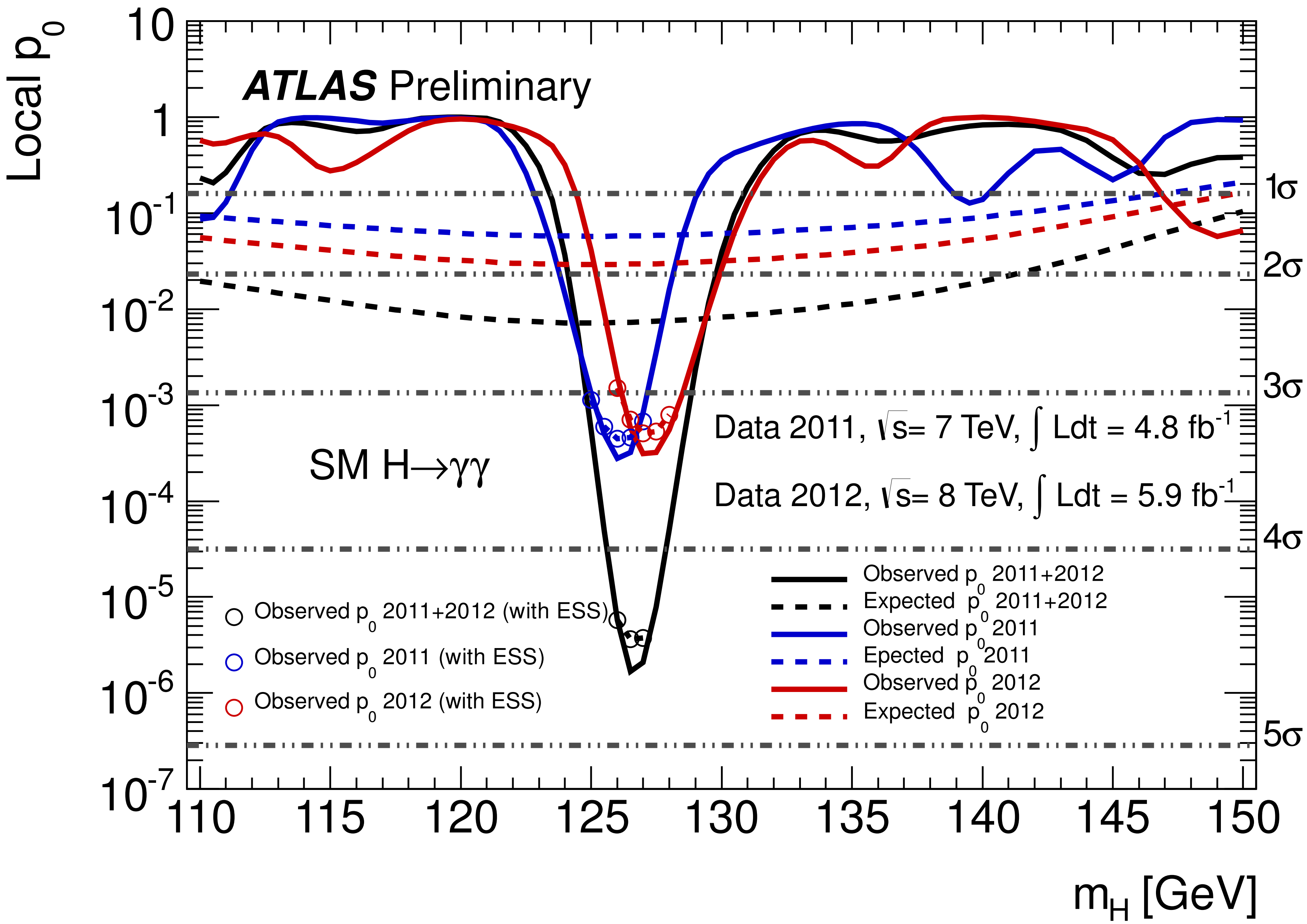
- How to determined the (local and global) significance of the signal?
- How to estimate the parameter, e.g. mass of the new resonance?

the green & yellow plots



- For every (assumed) value of m_H, we want to find the CL$_s$ upper limit on $\mu \equiv \sigma(H)/\sigma_{\rm SM}(H)$ (*solid curve*)
- Also shown is the 'expected upper limit', determined for each assumed m_H value, under the assumption that we see no excess above background.

the p_0 plots



- The **local** p_0 values for a SM Higgs boson as a function of assumed m_H.
- The minimal p_0 (observed) is 2×10^{-6} at $m_H = 126.5$ GeV.
  $\Rightarrow$ local significance of 4.7σ $\rightarrow$ reduced to 3.6σ after LEE


Now that you have the language
to talk about stat. interpretation of
HEP results (e.g. LHC),
it’s your job to explore & enjoy!
Thank you!

a suggested reading for advanced/interested students

Eur. Phys. J. C (2011) 71: 1554
DOI 10.1140/epjc/s10052-011-1554-0

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

Asymptotic formulae for likelihood-based tests of new physics

Glen Cowan[1], **Kyle Cranmer**[2], **Eilam Gross**[3], **Ofer Vitells**[3,a]

[1]Physics Department, Royal Holloway, University of London, Egham TW20 0EX, UK
[2]Physics Department, New York University, New York, NY 10003, USA
[3]Weizmann Institute of Science, Rehovot 76100, Israel

Abstract

We describe likelihood-based statistical tests for use in high energy physics for the discovery of new phenomena and for construction of confidence intervals on model parameters. We focus on the properties of the test procedures that allow one to account for systematic uncertainties. Explicit formulae for the asymptotic distributions of test statistics are derived using results of Wilks and Wald. We motivate and justify the use of a representative data set, called the "Asimov data set", which provides a simple method to obtain the median experimental sensitivity of a search or measurement as well as fluctuations about this expectation.