20+ Years of CL_s

1-6 August 2018

XIII'th Quark Confinement and the Hadron Spectrum

A. Read (U. Oslo)

UiO : University of Oslo

The Research Council of Norway

2 main limit approaches

- Bayesian: probabilty(theory|data) $p(\theta|x)$
  - well-defined accounting for beliefs
  - prior-probability for the theory must be given
  - prior-dependence should be studied
- Frequentist/classical - probability(data|theory) $p(x|\theta)$
  - says nothing about probability of theory
  - typically used in HEP to report experimental results "objectively" (as possible)
  - can lead to subset of individual results which are obviously wrong but consistent with methodology

Bayes vs. freq.

- In many data-dominated situations hardly any difference in reported results, eg. M_Z=91.1876±0.0021 GeV
  - But interp. not the same!
    1) P(|MZ-91.1876|<0.0021)=68%
    2) 68% of such intervals contain the true M_Z
- Small data samples, physical boundries typically lead to differences
- Doing both analyses and studying the differences can give insights

Counting experiment

$$L(n|\mu s + b) = \frac{e^{-(\mu s + b)}(\mu s + b)^n}{n!}$$

Likelihood ratio of marked Poissons and combined channels

$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i} b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_i S_i(x_{ij}) + b_i B_i(x_{ij})}{s_i + b_i}}{\prod_{j=1}^{n_i} B_i(x_{ij})}$$

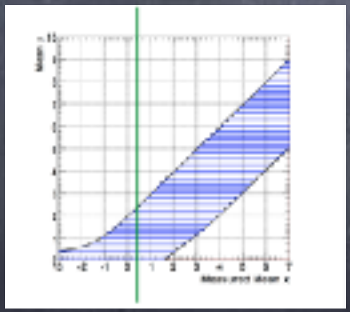
Brief (!) history of HEP-limits

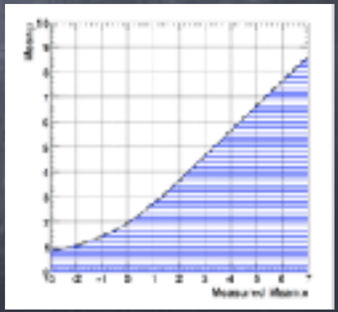
- O. Helene (1983) - Bayesian limit with flat prior on signal for counting expt.
- G. Zech (1988) - frequentist interpretation of Helene
- A. Read (1997) - rederived Zech from likelihood ratio and "background conditioning"; CL_S ≈ "confidence in the signal-only hypothesis"
- Feldman and Cousins (1998) - auto 2-sided frequentist confidence intervals - "coverage is king" (but tests signal+background hypothesis)
- Birnbaum (1961!) - concept of *statistical evidence* resembles CL_S - discovered in literature by O. Vitells

$$CL = \frac{\int_s^\infty \mathcal{L}(s', b)ds'}{\int_0^\infty \mathcal{L}(s', b)ds'}.$$

$$CL = 1 - \frac{\sum_{n=0}^{n_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{e^{-b}b^n}{n!}}.$$

$$CL_s \equiv CL_{s+b}/CL_b.$$

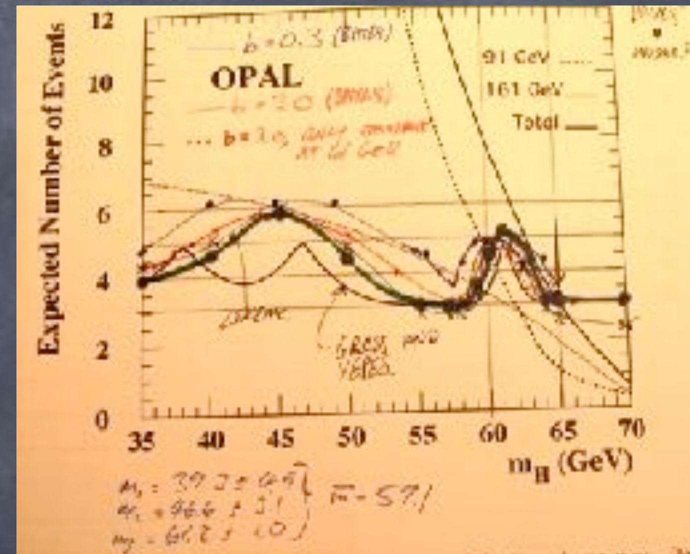


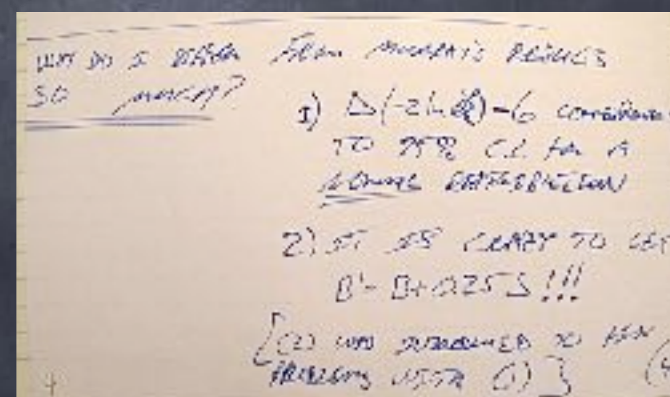


"A concept of statistical evidence is not plausible unless it finds 'strong evidence for H2 as against H1' with small probability (alpha) When H1 is true, and with much larger probability (1 -beta) when H2 is true."

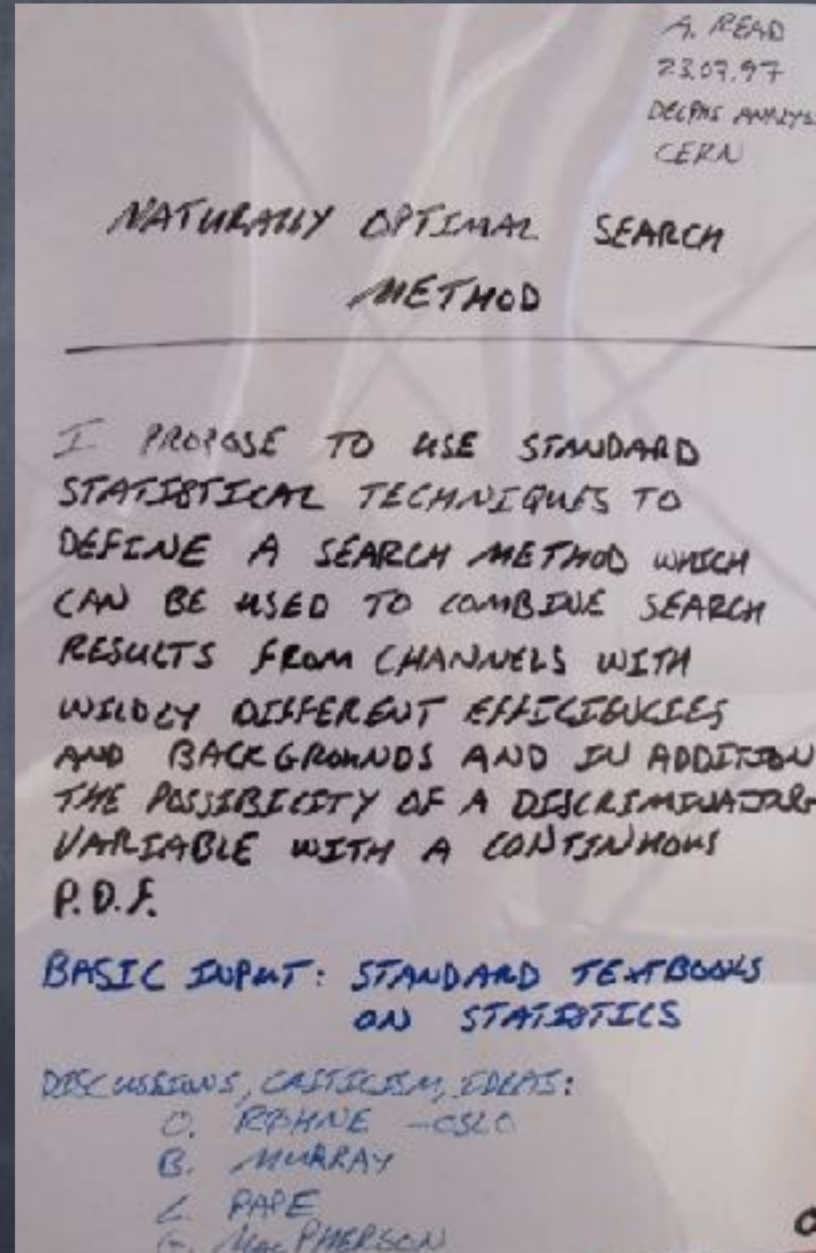
Origins of CL_s

- Almost background-less Higgs searches at LEP1, many different statistical treatments, combination not obvious, LEP2 data was coming
- I proposed simple LR, frequentist approach, combination simply adding channels to LR
- Exclusion with CL_s, invented to
  - Deal robustly with deficits
  - Adding low-sensitivity channels gives marginal improvement to overall sensitivity
  - Increasing uncertainty doesn't improve sensitivity
- Prepared discovery with CL_b, never got to ML for measurement
- Cousins&Highland (hybrid Bayes-frequentist treatment) for (generally small) systematics





First presentation of Q_{LEP} and CL_S (to DELPHI)



A. READ
23.07.97
DELPHI ANALYSIS
CERN

NATURALLY OPTIMAL SEARCH METHOD

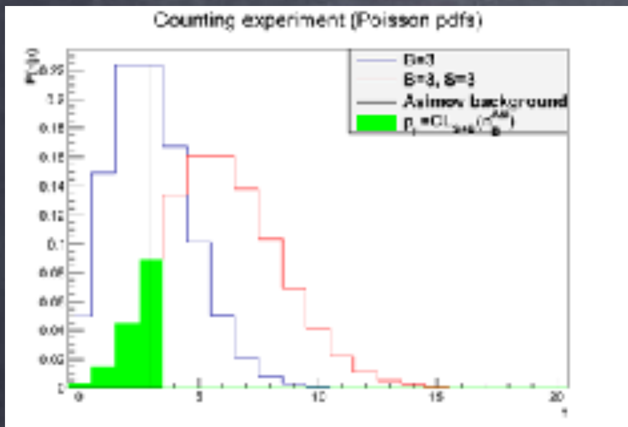
I PROPOSE TO USE STANDARD STATISTICAL TECHNIQUES TO DEFINE A SEARCH METHOD WHICH CAN BE USED TO COMBINE SEARCH RESULTS FROM CHANNELS WITH WILDLY DIFFERENT EFFICIENCIES AND BACKGROUNDS AND IN ADDITION THE POSSIBILITY OF A DISCRIMINATING VARIABLE WITH A CONTINUOUS P.D.F.

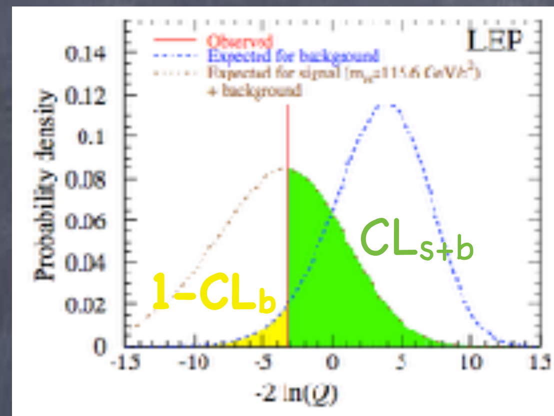
BASIC INPUT: STANDARD TEXTBOOKS ON STATISTICS

DISCUSSIONS, CRITICISM, IDEAS:
O. RØHNE - OSLO
B. MURRAY
L. PAPE
G. MacPHERSON

0

CLs





$$CL_{s+b} = P_{s+b}(X \le X_{obs}),$$

$$P_{s+b}(X \le X_{obs}) = \int_0^{X_{obs}} \frac{dP_{s+b}}{dX}\,dX$$

$$CL_b = P_b(X \le X_{obs}),$$

$$P_b(X \le X_{obs}) = \int_0^{X_{obs}} \frac{dP_b}{dX}\,dX$$

$$CL_s = CL_{s+b}/CL_b.$$

$$1 - CL_s \le CL.$$

IOPscience Journals Login

Journal of Physics G: Nuclear and Particle Physics

Journal of Physics G: Nuclear and Particle Physics > Volume 28 > Number 10

A L Read 2002 J. Phys. G: Nucl. Part. Phys. 28 2693 doi:10.1088/0954-3899/28/10/313

Presentation of search results: the CL_s technique

A L Read

$$Q_i = \frac{\dfrac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i^{cand}}}{n_i^{cand}!}}{e^{-b_i}\dfrac{b_i^{n_i^{cand}}}{n_i^{cand}!}}$$

$$-2\ln Q_i = 2\,s_i - 2\,n_i \ln\left(1 + \frac{s_i}{b_i}\right)$$

CERN-OPEN-2000-205

Modified frequentist analysis of search results (the CL_s method)

Read, A L (U. Oslo)

CERN, Geneva

2000

1st Workshop on Confidence Limits, CERN, Geneva, Switzerland, 17 - 18 Jan 2000, pp.81-101

CERN Document Server

Access articles, reports and multimedia content in HEP

LEP combinations

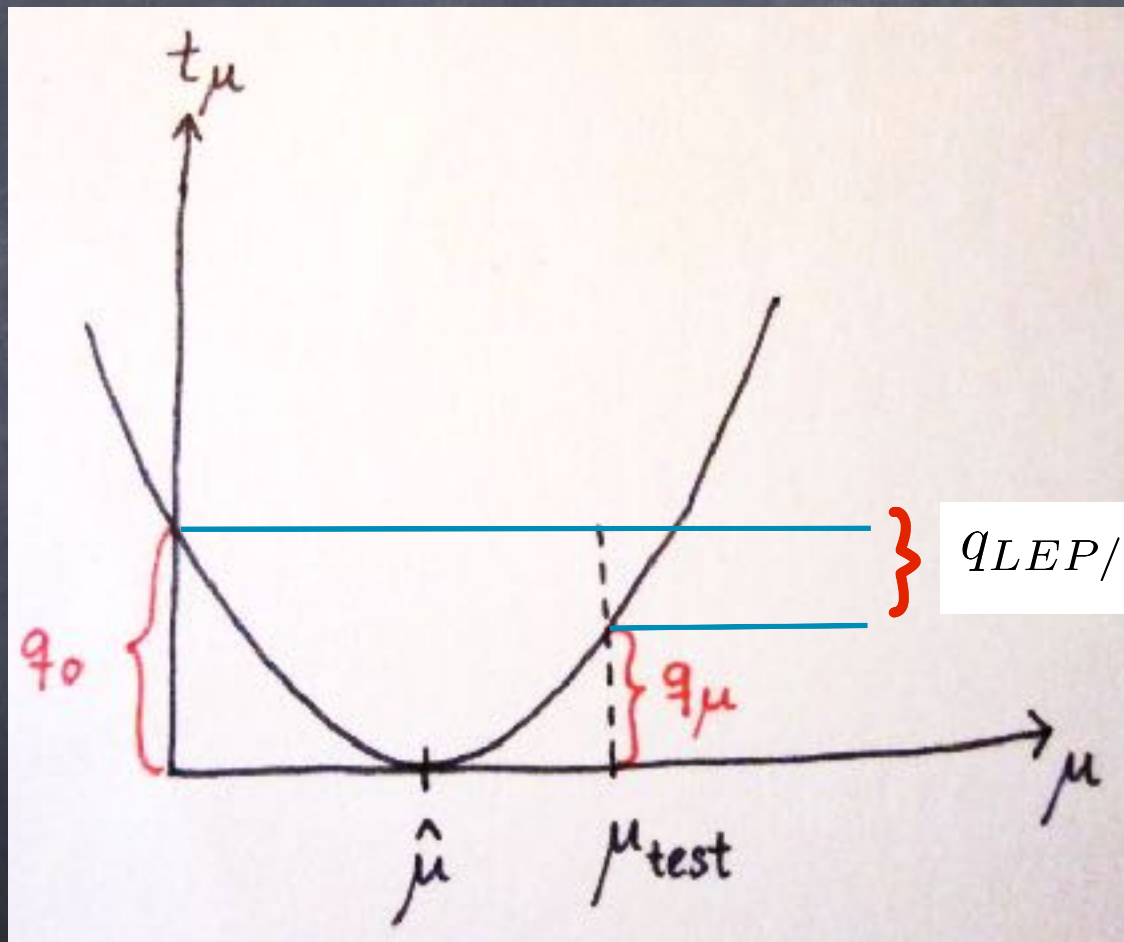
- Natural combination of channels, extension to discriminant (or counting) per channel
- Learned later Obraztsov (DELPHI 1992), L3 people proposed similar likelihood but Bayesian integration of likelihood (implicit uniform prior).
- At LEP eventually 4 experiments, O(10) center of mass energies, O(8) search topologies/channels combined

$$Q = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!} \prod_{j=1}^{n_i} \frac{s_i S_i(x_{ij})+b_i B_i(x_{ij})}{s_i+b_i}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i} b_i^{n_i}}{n_i!} \prod_{j=1}^{n_i} B_i(x_{ij})}$$

LR from LEP to Tevatron to LHC

	Test statistic	Nuisance parameters in LR	Randomized in toys	Sampling of test statistic
Q_{LEP}	$-2\ln\frac{L(\mu,\tilde{\theta})}{L(0,\tilde{\theta})}$	Fixed by MC	Nuisance parameters	Hybrid Bayes-frequentist
Q_{Tev}	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\hat{\theta}})}$	Profiled	Nuisance parameters	Hybrid Bayes-frequentist
"LHC" q_μ (q_0)	$-2\ln\frac{L(\mu(0),\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$	Profiled	External constraints	Frequentist

Q_{LEP} (Q_{TeV} w/o nuisances)



$$q_{LEP/TeV} = q_\mu - q_0$$

Profile likelihood (MINUIT)

lanl.arXiv.org > physics > arXiv:physics/0403059

Se

Physics > Data Analysis, Statistics and Probability

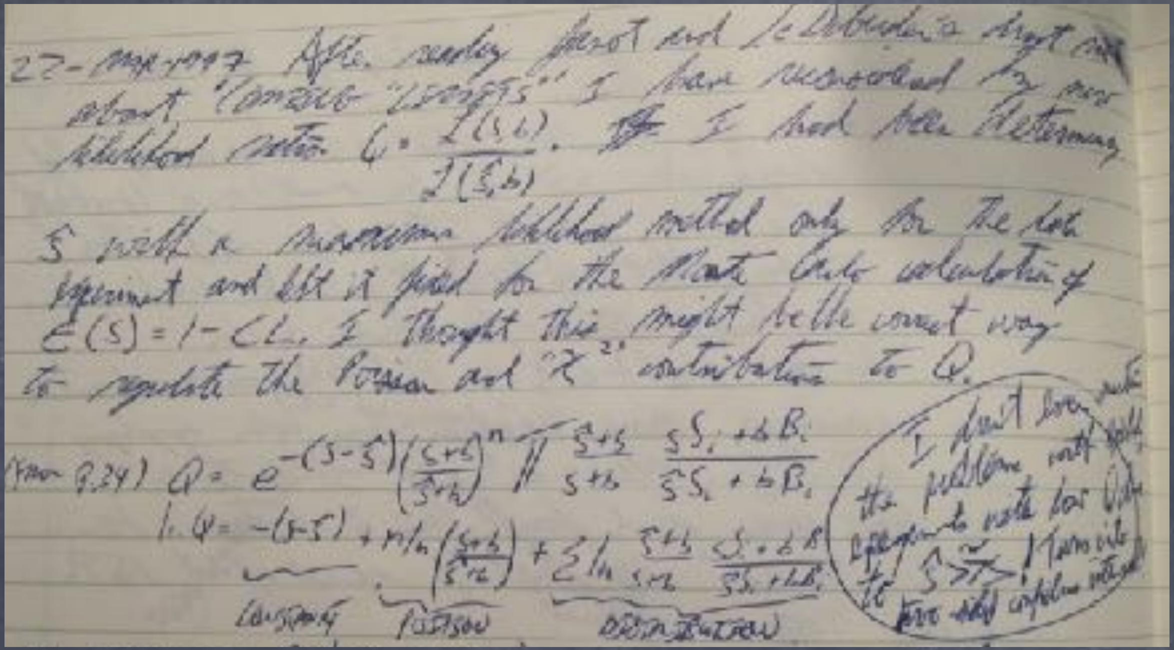
Limits and Confidence Intervals in the Presence of Nuisance Parameters

Wolfgang A. Rolke, Angel M. Lopez, Jan Conrad

(Submitted on 9 Mar 2004 (v1), last revised 19 Jan 2009 (this version, v5))

We study the frequentist properties of confidence intervals computed by the method known to statisticians as the Profile Likelihood. It is seen that the coverage of these intervals is surprisingly good over a wide range of possible parameter values for important classes of problems, in particular whenever there are additional nuisance parameters with statistical or systematic errors. Programs are available for calculating these intervals.

Curiousity: PL considered at LEP times



- I abandoned it to avoid 2-sided intervals (Feldman&Cousins!) - don't want to exclude if there is a nice fat excess!
- ~10 years later <u>CCGV</u> elegant solution:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

AA - Asymptotics and Asimov dataset

arXiv.org > physics > arXiv:1007.1727

Search

Physics > Data Analysis, Statistics and Probability

Asymptotic formulae for likelihood-based tests of new physics

Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells

(Submitted on 10 Jul 2010 (v1), last revised 3 Oct 2010 (this version, v2))

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.

Subjects:	**Data Analysis, Statistics and Probability (physics.data-an)**; High Energy Physics - Experiment (hep-ex)
Journal reference:	Eur.Phys.J.C71:1554,2011
DOI:	10.1140/epjc/s10052-011-1554-0

AA - Asymptotics and Asimov dataset

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$

$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right]$$

$$V_{jk}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_j \partial\theta_k}\right] = -\frac{\partial^2 \ln L_{\mathrm{A}}}{\partial\theta_j \partial\theta_k}$$

$$\sigma^2 = V_{00}$$

$$\begin{aligned} n_{i,\mathrm{A}} &= E[n_i] = \nu_i = \mu' s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}) \,, \\ m_{i,\mathrm{A}} &= E[m_i] = u_i(\boldsymbol{\theta}) \,. \end{aligned}$$

$$q_0 = \begin{cases} \hat{\mu}^2/\sigma^2 & \hat{\mu} \geq 0 \,, \\ 0 & \hat{\mu} < 0 \,, \end{cases}$$

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}$$

Compact formulae for both observed results and expectations (including fluctuation bands)

Curiosity: Precursor to Asimov dataset in LEP (DELPHI) Higgs combination code

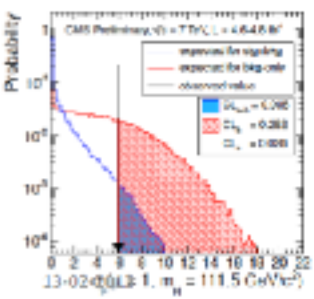
```
      SUBROUTINE explnQnom(s)
*.-----------------------------------------------------------------------
*  Compute the expected Likelihood Ratio for the combined counting and
*  invariant mass (or other discriminating variable) measurement experiment in
*  multiple channels. This only works for combinations where for each
*  channels the number of background and signal bins is indentical. This
*  is fast and simple to compute and can serve as a precise check
*  of Monte Carlo and semi-analytic computations.
*
*  The expected -2lnQ (Q is likelihood ratio) is computed both for
*  background-only and signal+background hypotheses.
*
*  10.12.99 Add the RMS of the distributions of -2lnQ for signal+background
*           and background-only experiments.
*.-----------------------------------------------------------------------
```

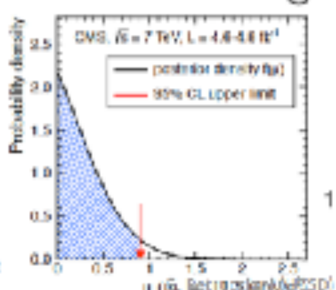
```
lrwt = log(1. + si*bkgprdx(i)/bi/sigprdx(i))
lnqisb = -si + (si+bi)*lrwt
lnqib   = -si +      bi*lrwt
avg2lnqsb8 = avg2lnqsb8 + lnqisb
avg2lnqb8  = avg2lnqb8  + lnqib

r2lnqisb = 4.*(si+bi)*lrwt**2
r2lnqib  = 4.*(   bi)*lrwt**2
avgr2lnqsb8 = avgr2lnqsb8 + r2lnqisb
avgr2lnqb8  = avgr2lnqb8  + r2lnqib
```

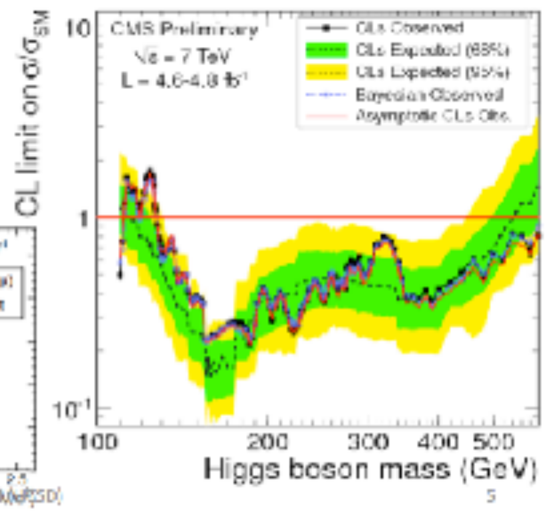
- But unlike CCGV not possible to treat nuisance parameters

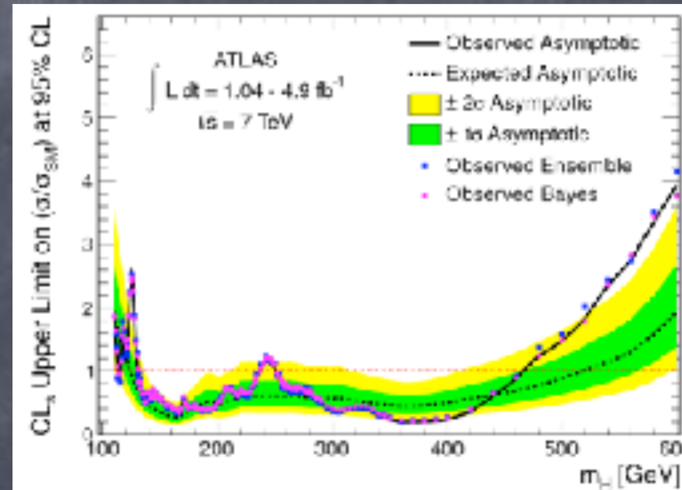
What about Bayesian methodology in LHC Higgs boson searches?

- Up to Moriond 2012, CMS produced limits with three prescriptions, to check robustness.
  - CLs using Toy MC
  - CLs using asymptotics
  - Bayesian w/ flat prior









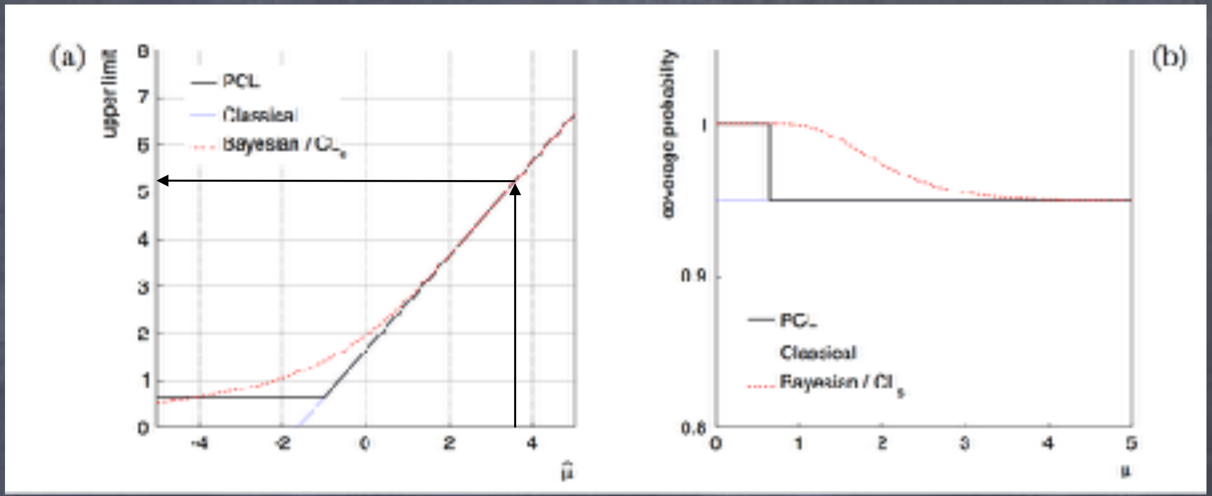
$$L(\mu) = \frac{1}{C}\int_{\theta} p(\text{data}|\mu s + b)\ \rho_\theta(\theta)\ \pi_\mu(\mu)\ d\theta$$

$$\int_0^{\mu_{95\% CL}} L(\mu)\, d\mu = 0.95$$

- Limits, with flat prior, very consistent with CLs limits derived in frequentist framework
- No attempt (yet!) to quantify excess at 125/6 GeV with Bayes factors

Challenge: Replace CLs?

- Proposal of "Power-constrained limits" in 2011 gave CLs a second wind



The choice of the minimum power threshold is a matter of convention. We prefer to use $M_{\mathrm{min}} = 0.16$, or more precisely, $M_{\mathrm{min}} = \Phi(-1) = 0.1587$, where Φ is the standard normal cumulative distribution (i.e., the cumulative distribution for Gaussian with a mean of zero and unit standard deviation). As shown below, this corresponds to applying the power constraint if the unconstrained limit fluctuates one standard deviation below its median value under the background-only hypothesis.

Challenge: Discreteness

- Discrete test-statistic, small samples and frequentist treatment can give unintuitive "better than zero" results - anything, like a nuisance parameter or additional insensitive channel that breaks discreteness ~halves the nominal probability of observing a particular outcome.

Study q_0 (simpler than q_μ)

$$L = \frac{e^{-(s+b)}(s+b)^{n_o}}{n_o!} \frac{e^{\frac{-(s-s_0)^2}{2\sigma_s^2}}}{\sqrt{2\pi\sigma_s^2}} \frac{e^{\frac{-(b-b_0)^2}{2\sigma_b^2}}}{\sqrt{2\pi\sigma_b^2}}$$

$$-2\ln L = 2(s+b) - 2n_o \ln(s+b) + \left(\frac{s-s_o}{\sigma_s}\right)^2 + \left(\frac{b-b_o}{\sigma_b}\right)^2$$

$\sigma_s \rightarrow \infty$ for unconstrained fit for s

$$q_0 = -2\ln\left[\frac{L(s=0,\hat{\hat{b}})}{L(\hat{s},\hat{b})}\right]$$

$$\hat{b} = b_o,\ \hat{s} = n_o - b_o,\ \boxed{\hat{\hat{b}} = \frac{b_o - \sigma_b^2 + \sqrt{(b_o-\sigma_b^2)^2 + 4n_o\sigma_b^2}}{2}}$$

Closed Form

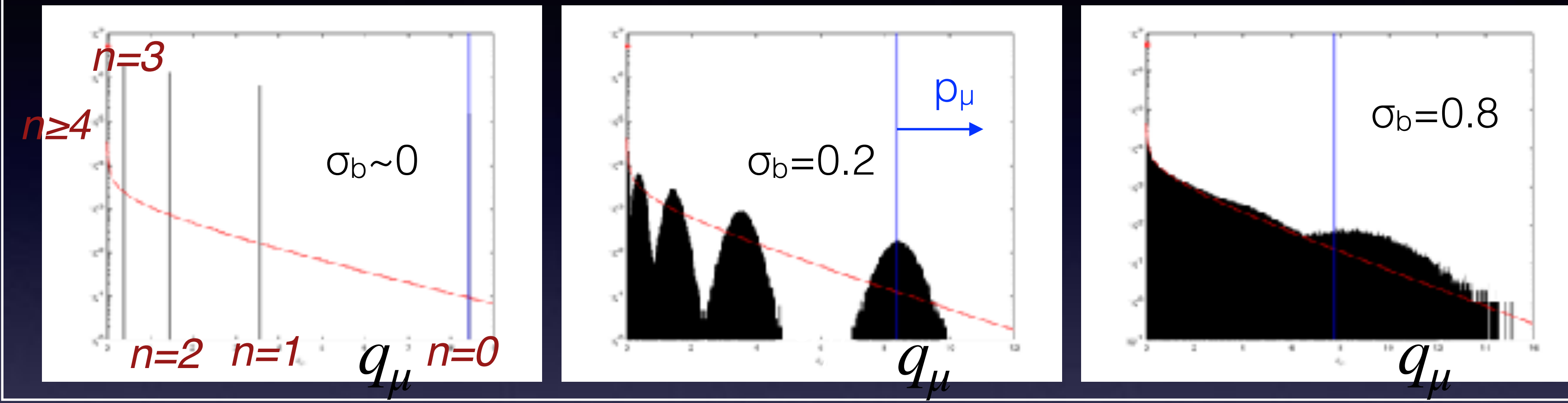
$$q_0 = 2(n_o \ln\frac{n_o}{\hat{\hat{b}}} + \hat{\hat{b}} - n_o) + \left(\frac{\hat{\hat{b}} - b_o}{\sigma_b}\right)^2 \qquad \text{for } n_o > b_o$$

$$= 0 \qquad \text{for } n_o \le b_o$$

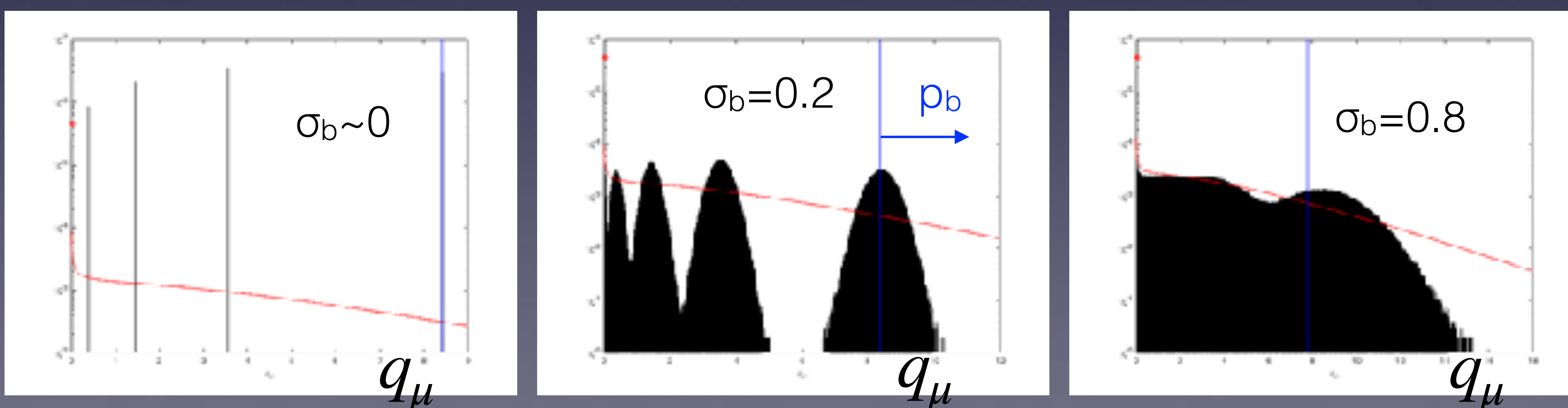
(Checked against L. Demortier, PHYSTAT 2003)

Asymptotics and exclusion?

Signal+background pdf's of q_μ



Background-only pdf's of q_μ



What about σ_s?

- Preliminary indication (not my work) is that profiling for n=0 (i.e. n<s+b) can lead to upper limits below "gold standard" of 3 events.

Summary

- CLs for limits is despised by both professional Bayesians and Frequentsists
- It has a lot of nice properties, not the least important of which is robustness
- It survived, to my surprise, a direct challenge just before the Higgs boson discovery
- Interesting features and questions still pop up in this tiny, almost dataless, corner of statistics