CL_s : Physicist Presentation

Alex Read
University of Oslo

Tom Junk
Fermilab

Setting the Stage: What Problem CL_s is Intended to Solve

Nearly all observations in experimental particle physics are well described by the "Standard Model" (Some might say "all" here)

We run experiments to discover new particles and processes, as well as to measure the known ones as precisely as we can, so we can learn about how Nature works.

Most speculative ideas are in fact false, and searching for evidence of them usually comes up empty-handed (null results).

Even searching for processes predicted by the Standard Model can be difficult if the rates are small, or if there are unmeasured parameters.

We need a way to express what we sought but did not find.

We set limits – and compute expected limits

Discovery and Exclusion

The discovery criterion is usually pre-assigned: $p < \alpha$, where $\alpha = 2.78\text{E-}7$ is customary in HEP. Outside of HEP, $\alpha = 0.05$ is common. Here, p is the null hypothesis (H_0) p-value. Call it p_0 for the rest of the talk.

Example for counting experiment: p_0 is the probability, under H_0, of getting n_{obs} counts in the detector or more (*). More observed counts → smaller p_0. (* some "signals" subtract from the background, but the logic is similar).

We discover something that adds to our event count rate by excluding the null hypothesis if we see more events than the background.

But we are interested in also excluding test hypotheses! In HEP, we can, and usually do, publish (many!!) papers doing exactly this.

A naive extension: p_1 is the probability of getting n_{obs} counts in the detector or less, assuming the test hypothesis H_1. If $p_1 < 0.05$, declare H_1 excluded.

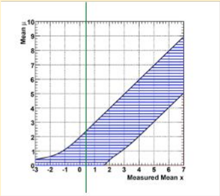
One brief history of HEP limits

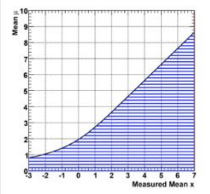
- O. Helene (1983) - Bayesian limit with flat prior on signal for counting experiment
- G. Zech (1988) - Frequentist re-interpretation of Helene with "background conditioning"
- A. Read (1997) - generalized Zech for any likelihood; $CL_s \approx$ "confidence in the signal-only hypothesis"
- Feldman and 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 ca. 2012)

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

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

$$CL_s \equiv CL_{s+b}/CL_b.$$ (LHC-speak: $CL_S = p_\mu / p_b$)





"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."

Typical likelihoods (w/o nuisances)

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

Poisson, counting (no background)

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

Counting, known bkg

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

Counting "on/off"

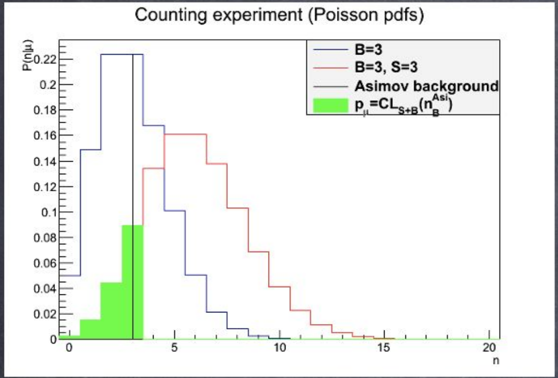
$$L(x|x_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

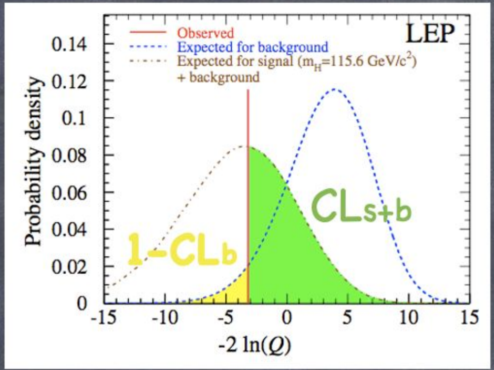
Gaussian

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

Likelihood ratio of marked Poissons in combined channels

CL_s





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

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

$$\mathrm{CL}_b = P_b(X \le X_{obs}),$$

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

$$\mathrm{CL}_s \equiv \mathrm{CL}_{s+b}/\mathrm{CL}_b.$$

$$1 - \mathrm{CL}_s \le \mathrm{CL}.$$

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

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

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

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

Origins of CL_S (1997)

- Almost backgroundless searches for Higgs at LEP, many different statistical treatments, combination of counting and bump searches not immediately obvious, data arriving
- I initially tried product of LR's, standard frequentist approach
- CL_S "procedure" or "modified frequentist" to deal with
  - Deficits - don't strongly exclude searches with no sensitivity
    - Avoid "better than zero": e^{-s} for no counts, zero background
  - Adding low-sensitivity channels → marginally better sensitivity
  - Uncertainty → worse sensitivity
  - (Cousins 1993) https://inspirehep.net/literature/355322 addresses points 2,3
    - Mid p-values too (https://www.jstor.org/stable/2348891)

Limits on Signal Strength

H_1 usually has a signal-strength parameter μ, whose value is unknown for many exotic models.

If the model can predict the signal strength, as is the case for the SM Higgs or top-quark production, we still invent a parameter μ to scale the signal by if we cannot yet see it.

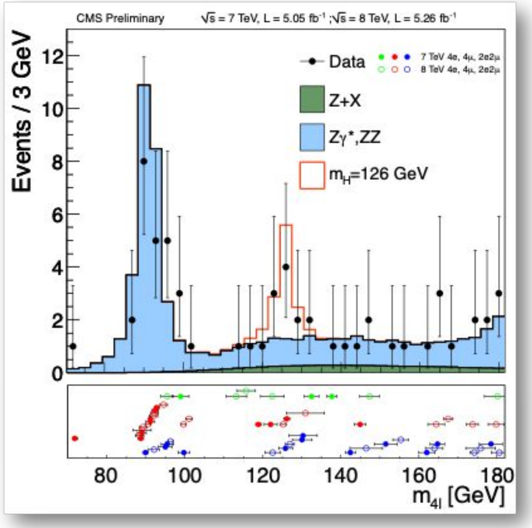
How big does μ have to be before we can exclude $H_1(\mu)$? Call that an *upper limit* on μ. For additive mixture models, p_1 is a monotonic function of μ. For other models, one has to work a bit harder.

Usually this is a function of some other parameter, such as the mass of the new particle being sought.

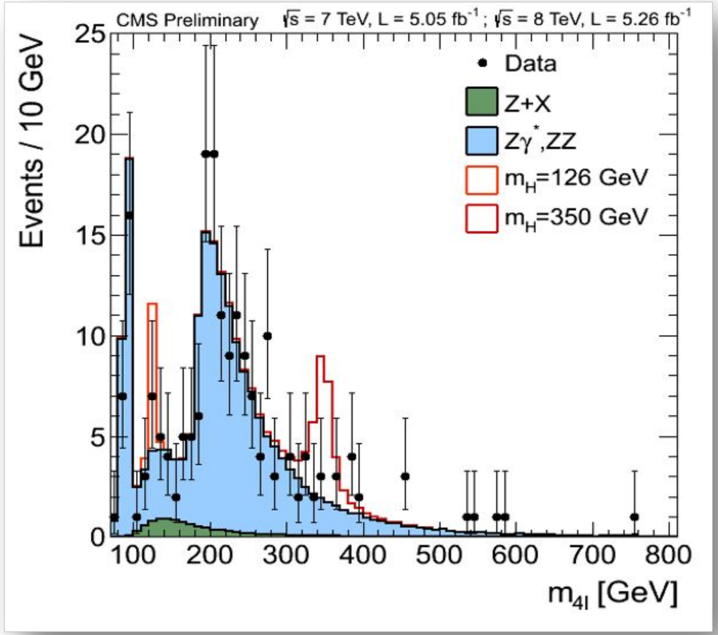
Sometimes μ has a physical meaning and can affect decay widths too, making just a scaling of the signal impossible.

Example Mixture-Model Search: Higgs Boson Plus Backgound

From Joe Incandela's slides at the CERN Higgs boson discovery seminar, July 4, 2012



Looking at more of this distribution.



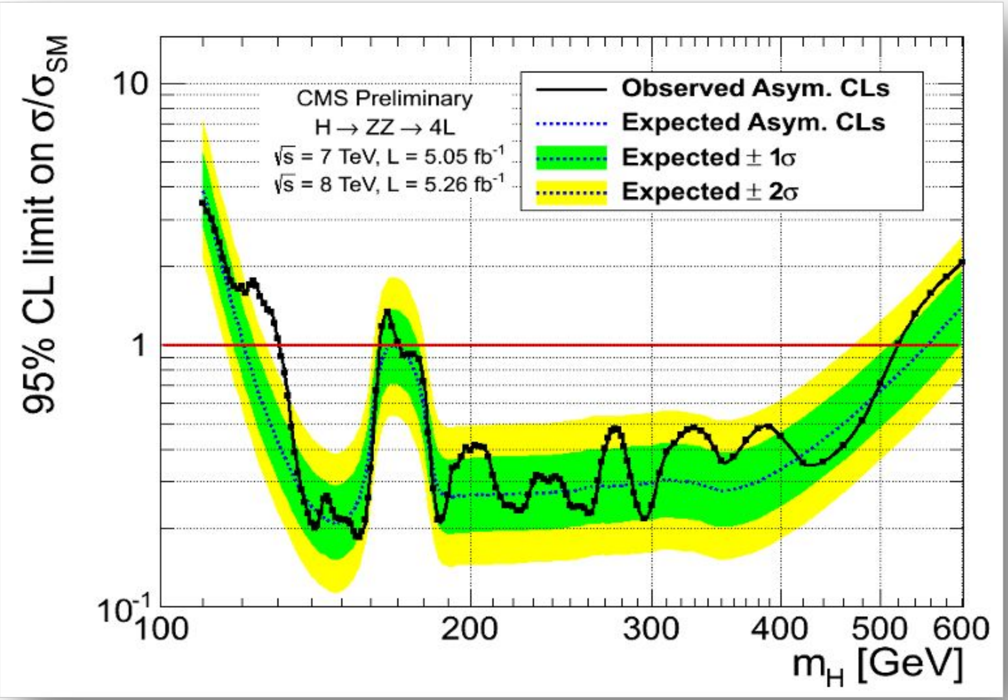
Example Limit on Signal Strength vs m_H

From Joe Incandela's slides at the CERN Higgs boson discovery seminar, July 4, 2012

This one has a true signal in it! (at $m_H \approx 125$ GeV).

Sensitivity of search weakens around $m_H = 165$ GeV due to lower Br($H \to ZZ$) ($H \to WW$ dominates the decay there)

Expected limits shown with dashed black, with green and yellow bands to indicate 68% and 95% quantiles.



The Zero Signal Strength Problem

Using $p_1 < 0.05$ to exclude $H_1(\mu)$ has proper coverage, but the error rate is 5%.

If $\mu = 0$, in 5% of outcomes, one would exclude $H_1(0)$. You can get "lucky" (or unlucky, depending on your wishes). You might as well toss real coins.

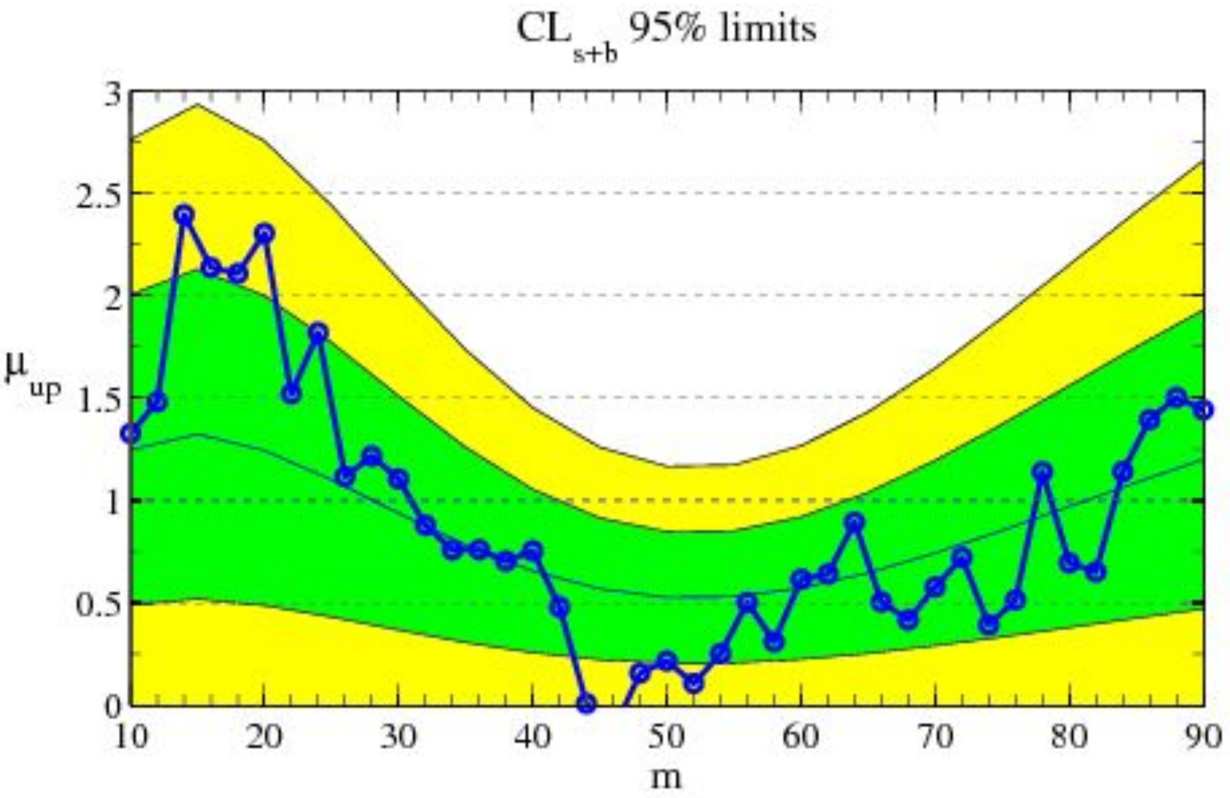
Analyses often scan over many values of some parameter like m_H. The 5% error cases will be sampled and shown in headline results plots.

The limit curve, normally shown on a log scale, will dip to zero at 5% of values of m_H.

The bottom yellow band, at a quantile of 2.5%, will also extend to the bottom of any limit plot.

These are "recognizable subsets" in which we know we've produced a result that is incorrect – we claimed to test an untestable model (μ=0) and excluded it at 95% CL

What a Limit Plot Looks Like Just using $p_1 < 0.05$



We usually show plots like these on a logarithmic scale

Cowan, Cranmer, Gross and Vitells, *Eur.Phys.J.C* 71 (2011) 1554,*Eur.Phys.J.C* 73 (2013) 2501 (erratum) e-Print: 1007.1727

Empty Intervals

This is also known as the “empty interval problem”.

5% of the time, one can rule out all values of a parameter, including zero.

Particle decay branching fractions are between 0 and 1.

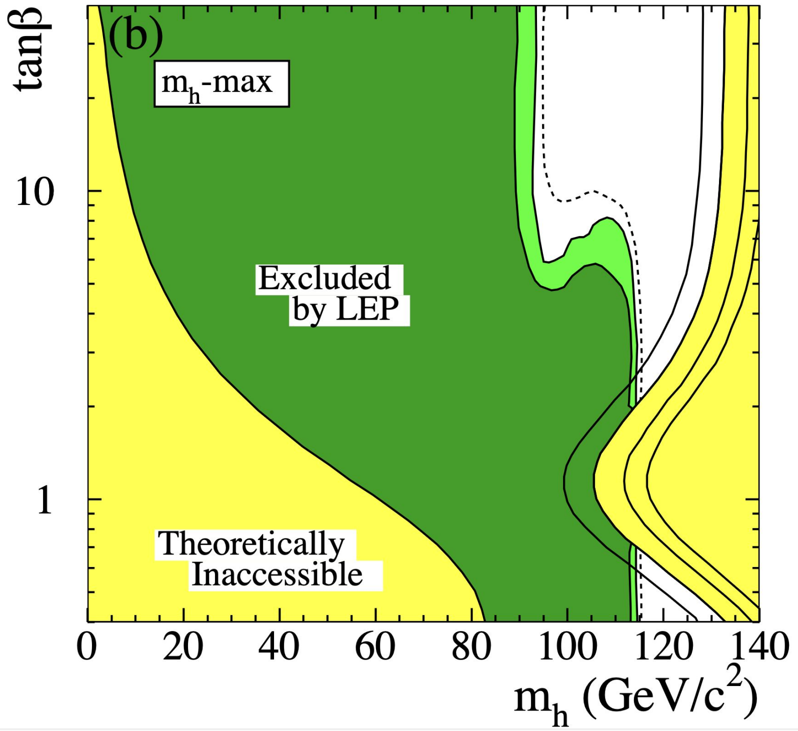
Ruling out the entire interval including the endpoints means we don’t believe there is a branching ratio, even though an exotic process that does not exist has B=0.

Physicists don’t like to publish results they know are wrong, even if just unlucky.

Blind analyses and aversion to flip-flopping mean physicists don’t like to publish results that could possibly be wrong, even if they aren’t in any particular case.

Example: Exclusions in a Multi-Dimensional Parameter Space

Unexcluded area – white would have a 5% dusting of
falsely excluded models if we didn’t protect it somehow.



ALEPH, DELPHI, L3 and OPAL Collbs and the LHWG, *Eur.Phys.J.C* 47 (2006) 547-587

Prescriptions Used in High-Energy Physics

- Modified Frequentist: $CL_s = p_1/(1 - p_0)$.

A. Read, *J.Phys.G* 28 (2002) 2693-2704.
http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings//read.pdf
A. Read, DELPHI 97-158 phys 737 (29 October, 1997). http://cds.cern.ch/record/2627667
A. Read, CERN 2000-005, p. 81 (2000)
T. Junk, *Nucl.Instrum.Meth.A* 434 (1999) 435-443 e-Print: hep-ex/9902006

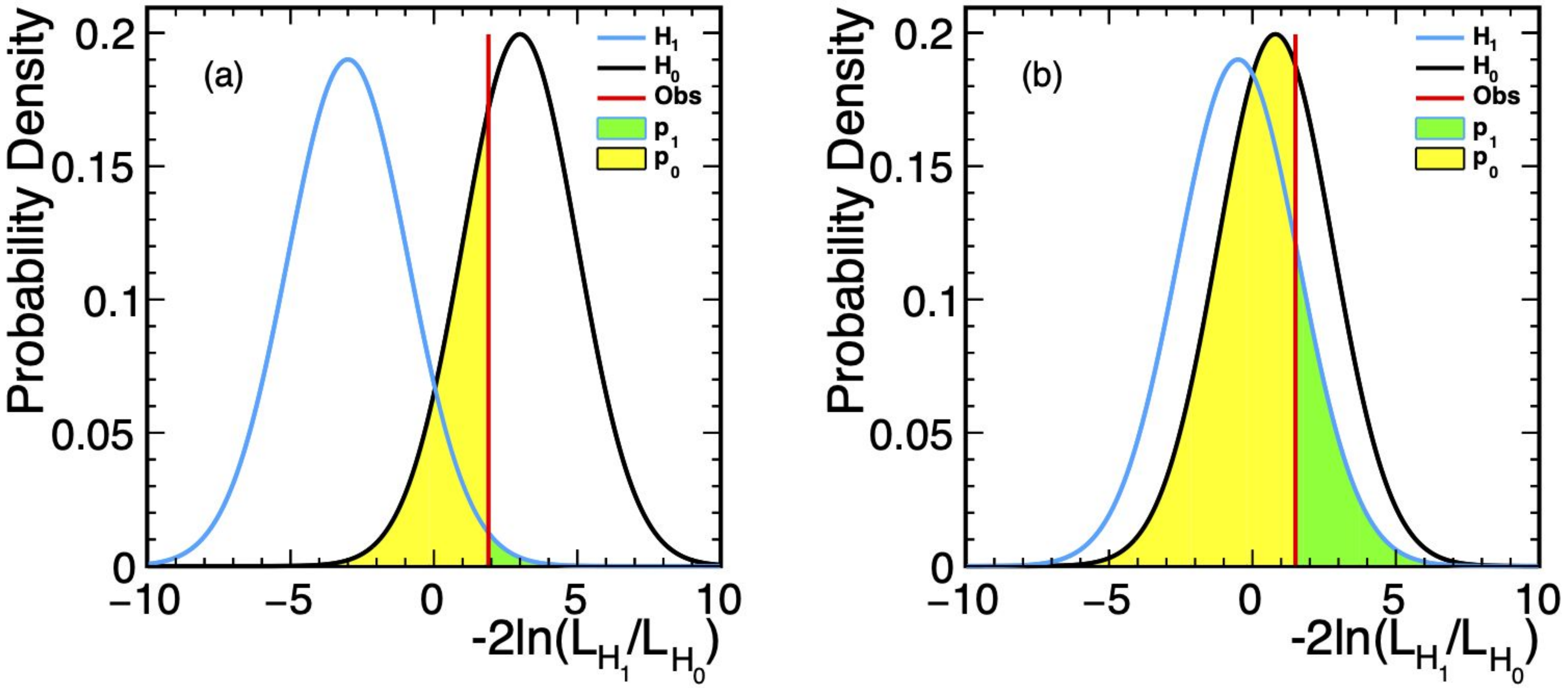
- Do a Fully Bayesian Calculation
  Integrate the posterior probability as a function of μ until you get 0.95.
  The limit will always be credible.
  Systematic uncertainties often only have subjective belief functions for the nuisance parameters.

G. Cowan, in R.L. Workman *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2022**, 083C01 (2022) and 2023 update

- Use the method of Feldman and Cousins – LR-ordered Neyman construction.

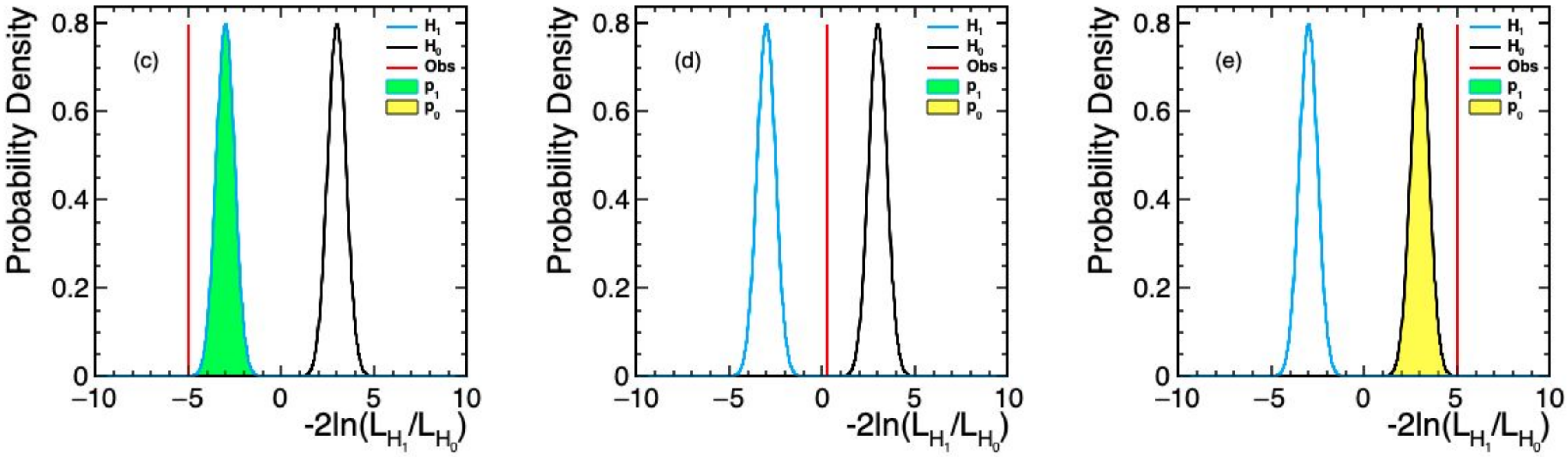
G. Feldman and R. Cousins, *Phys.Rev.D* 57 (1998) 3873-3889 e-Print: physics/9711021

More Sensitivity Or Less

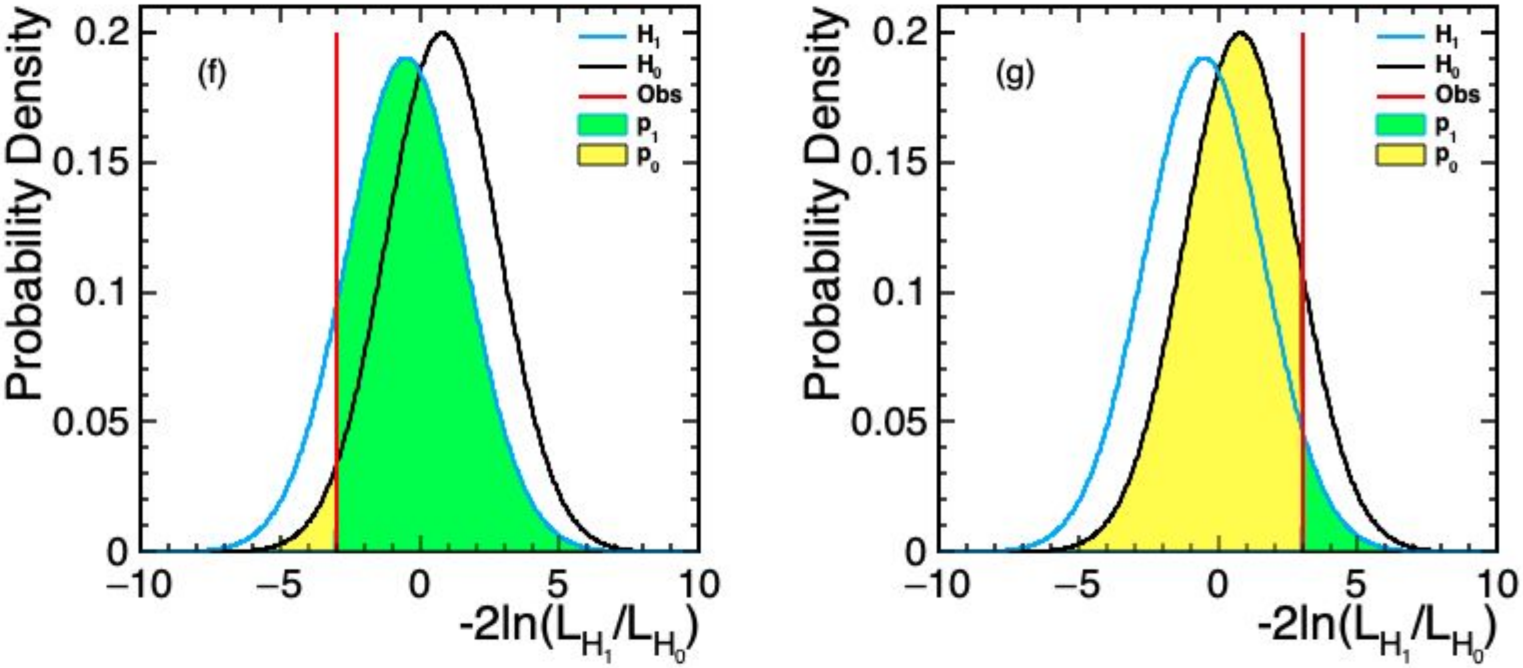


Junk and Lyons, Reproducibility and Replication of Experimental Particle Physics Results, Harvard Data Science Review, Fall 2020
https://hdsr.mitpress.mit.edu/pub/1lhu0zvn/release/4?readingCollection=c6cf45bb

“Uncooperative” Data: High-Sensitivity Case



“Uncooperative” Data: Low-Sensitivity Case



Why Conditioning On The Number of Background Events Doesn't Always work

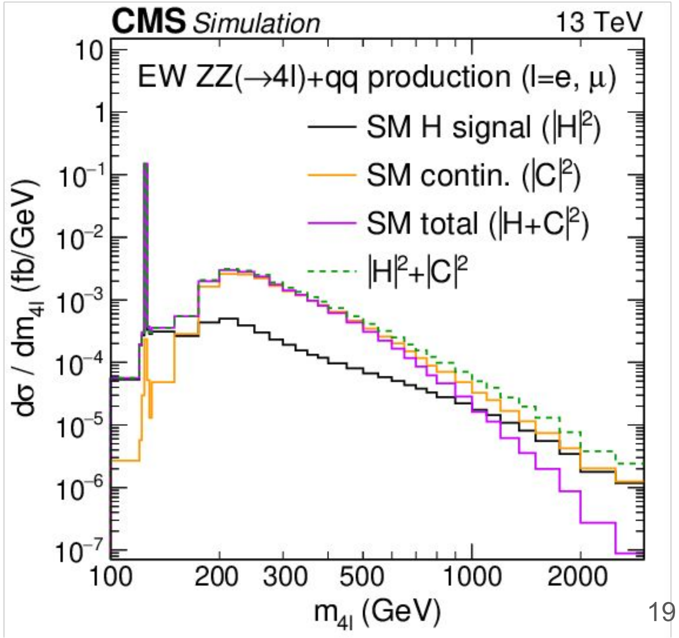
Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production

https://arxiv.org/abs/2202.06923

Expected signal + background < background only!

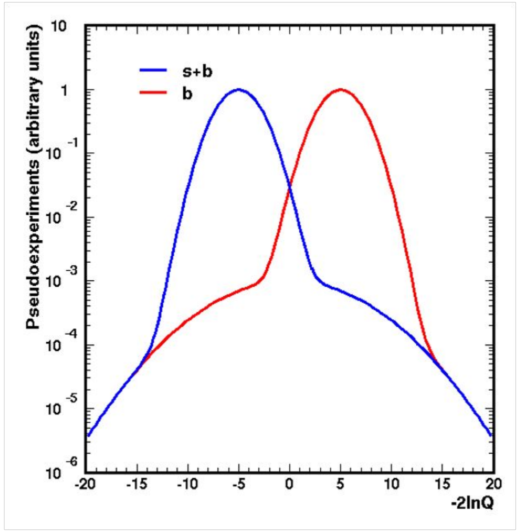
Blame quantum mechanics!

CL_s was not used in this paper – Higgs boson was already discovered. Hypothesis of no signal is more exotic than presence of signal.



CL_s May not be a Monotonic Function of -2ln(Q)

Tails in the -2lnQ distribution shared in the s+b and b-only hypothesis (fit failures)



CL_s=1 for
-2lnQ < -15 or
-2lnQ > +15

Not really a pathology of the method, but rather a reflection that the test statistic isn't always doing its job of separating s+b-like outcomes from b-like outcomes in some fraction of the cases.

Power-Constrained Limits

Cowan, Cranmer, Gross and Vitells, https://arxiv.org/abs/1105.3166

Comment by R. Cousins, https://arxiv.org/abs/1109.2023

Main objection is that the power constraint is arbitrary – observed limits depend not on the data but on a choice made a priori in the case of data that are in the recognizable subset.

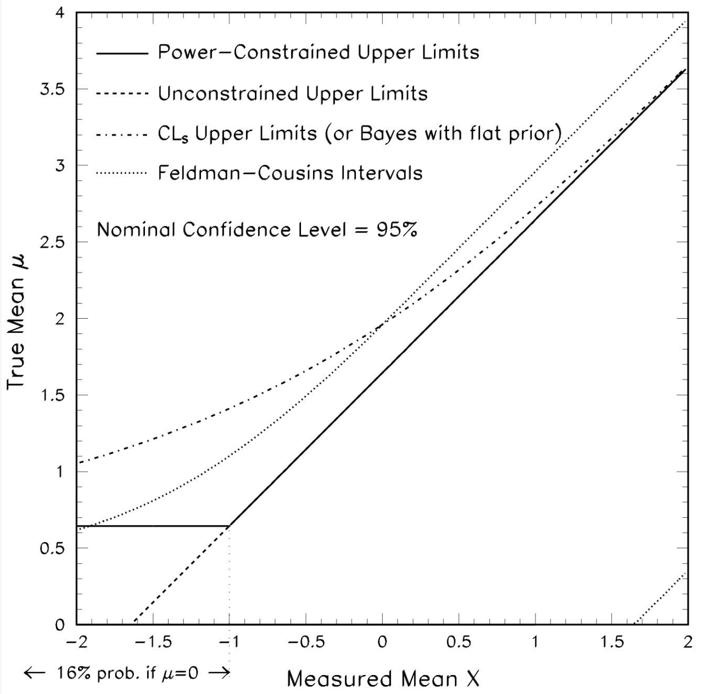


Figure from a CMS note by R. Cousins, summarizing figures in https://arxiv.org/abs/1109.2023

Why Not 1-CL_b'?

Bill Murray floated an idea at the BANFF PHYSTAT meeting. Should we protect the null-hypothesis p-value 1-CL_b in the same way we "protect" CL_{s+b}?

$$1\text{-}CL_b' = p_0/(1-p_1)$$

Some discussion among physicists concluded it is not necessary:

- We ought to be able to discover signals for which our model underestimates the intensity
- We have already discovered the background
- Ruling out the null hypothesis does not mean the test hypothesis is right.
- See "uncooperative data" in plot f earlier

It is not Always Clear Which Hypothesis is the Null

Example: Neutrino mass ordering. Each neutrino mass eigenstate is a mixture of the three flavor eigenstates. Two are close in mass, a third is very different.

We have two hypotheses – “normal” (more or less aligned with the masses of the leptons with the largest contributions, but two of the eigenstates are rather evenly split, and “inverted”.

Implicit prejudice in the words is cheerfully ignored by experimentalists.

Also: top quark charge: 2/3 or 4/3? One is indeed more “exotic” than the other.

No Look-Elsewhere Effect for Limits

Each alternate hypothesis is tested against a null.

We are not cherry-picking the “most-excluded” alternate hypothesis, like we normally are with a discovery, finding the smallest p-value among many tests.

That said, the error rate is on a per-test basis. A plot with many limits (e.g. limits as a function of a parameter like m_H) can have a 5% error rate.

In practice, a lot of these models only assume one value of the parameter is possible. The number of false exclusions is thus limited.

Another way to “test” a model for which there is no sensitivity: Make a discovery!

$m_H \approx 125$ GeV $\rightarrow$ $m_H = 2$ TeV is excluded. In the SM, yes. Multiple Higgs bosons?

Curiosity: “Confidence distributions” by N. Hjort et al.

Journal of Statistical Planning and Inference
Volume 195, May 2018, Pages 1-13

Editorial overview

Confidence distributions and related themes

Nils Lid Hjort [a], Tore Schweder [b]

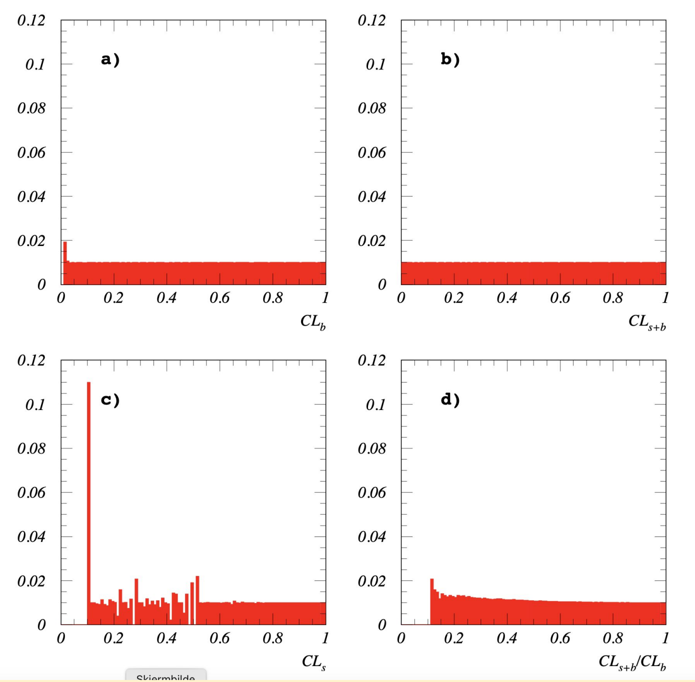
Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.jspi.2017.09.017   Get rights and content

Abstract

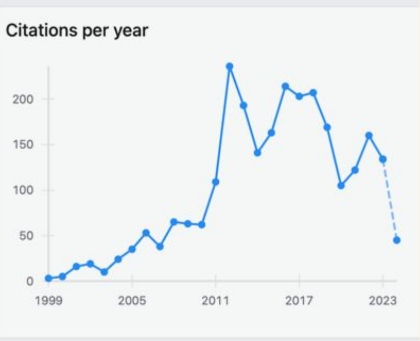
This is the guest editors’ general introduction to a Special Issue of the Journal of Statistical Planning and Inference, dedicated to confidence distributions and related themes. Confidence distributions (CDs) are distributions for parameters of interest, constructed via a statistical model after analysing the data. As such they serve the same purpose for the frequentist statisticians as the posterior distributions for the Bayesians.

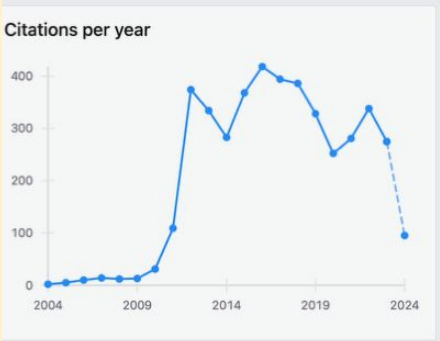


Workshop on Confidence Limits, 2000, Read

Summary

- CL_S for exclusion limits is neither frequentist nor Bayesian, but
  - Has some elements of both
  - Has several robust properties
  - Is popular among HEP colleagues (despite some skepticism)





Extras

The Likelihood Ratio with Unconstrained μ

Cowan, Cranmer, Gross and Vitells, *Eur.Phys.J.C* 71 (2011) 1554, *Eur.Phys.J.C* 73 (2013) 2501 (erratum) e-Print: [1007.1727](1007.1727)