Two Frequentist(-Inspired) Methods: Feldman-Cousins and CLs

Physics 252C - Lecture 13

Prof. John Conway

The "Karmen Effect"

- Karmen: reactor neutrino experiment looking for neutrino oscillations, mid '90s
- expected ~2 events but observed none
- this allowed them to set quite stringent limits...but was this "fair" ?
- more generally, what happens when background systematic uncertainty increases?
- in some methods, the worse the background uncertainty, the better the limit, if there is a negative fluctuation!

  ⇒ my view is that this is <u>not desirable</u>

Feldman - Cousins

- in 1998 Gary Feldman (Harvard) and Bob Cousins (UCLA) wrote what has become a key paper on confidence intervals
- start with standard Neyman belt for a Gaussian parameter
- they ask "what happens if our physicist decides whether to claim a discovery or set a limit based on what she observes?
- leads to flip-flopping

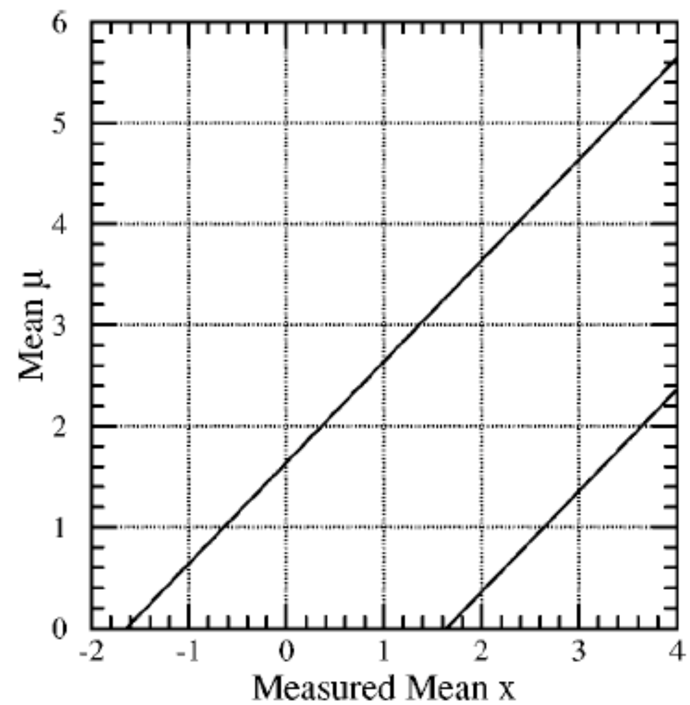


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

note that the axes are reversed from before...

Flip-flopping

- suppose we have a gaussian-distributed quantity, bounded at zero
- “If I have more than a 3σ excess I will say that, otherwise I will publish a 90% CL limit”
- what is the coverage?
- in the region where $1.36 < \mu < 4.28$ the coverage is only 85%
- similar problem in Poisson case

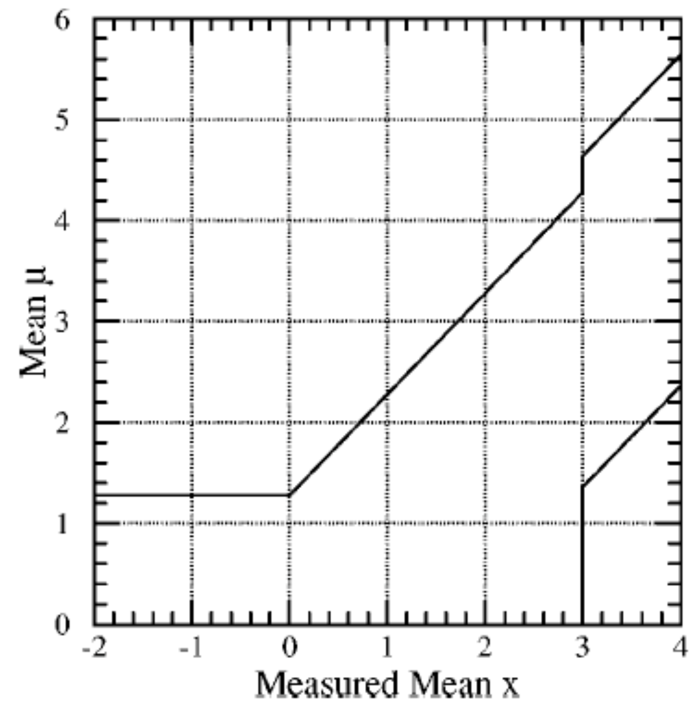


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence level. For $1.36<\mu<4.28$, the coverage (probability contained in the horizontal acceptance interval) is 85%.

Flip-flopping

- in Poisson case one has jumps in the confidence belt due to the discreteness of the distribution
- this leads to overcoverage as we've seen
- this is unfortunate, not simply "conservative"
- here as well if one decides whether to set a limit or quote an interval based on the outcome, the approach undercovers
- bad!

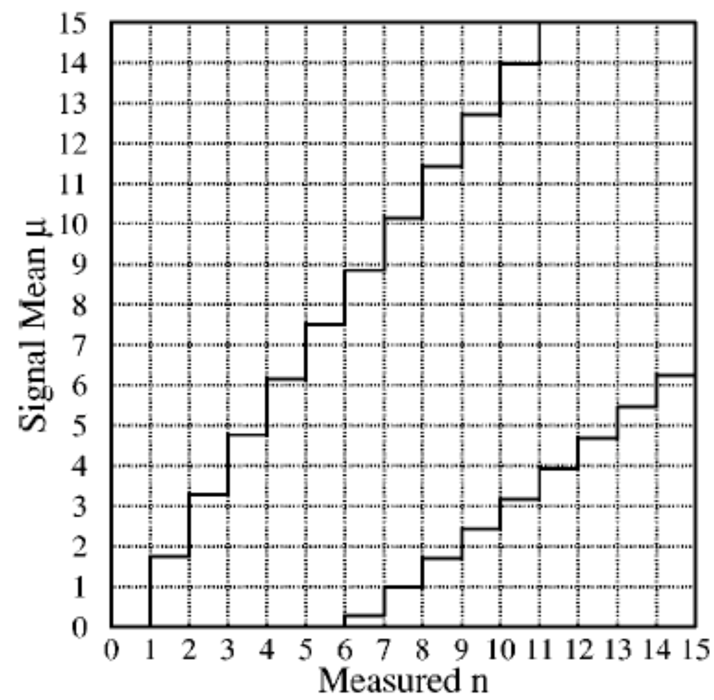


FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b = 3.0$.

likelihood ratio ordering

- Feldman and Cousins chose the interval in μ based on a likelihood ratio ordering
  - assume some value of μ
  - for each outcome n, construct the likelihood ratio of the Poisson probability to the best alternative hypothesis for μ

$$R = \frac{\mathcal{P}(n;\mu)}{\mathcal{P}(n;\mu_{best})}$$

  - we add those outcomes with the highest values of R into the band until the sum of their probabilities $\mathcal{P}(n;\mu)$ is at least the desired confidence level
  - F-C claim “minimal overcoverage” by this method

likelihood ratio ordering

- illustration of the LR ordering for b=3, μ = 0.5

TABLE I. Illustrative calculations in the confidence belt construction for signal mean μ in the presence of known mean background $b=3.0$. Here we find the acceptance interval for $\mu=0.5$.

n	$P(n\|\mu)$	μ_{best}	$P(n\|\mu_{\text{best}})$	R	rank	U.L.	central
0	0.030	0.0	0.050	0.607	6		
1	0.106	0.0	0.149	0.708	5	√	√
2	0.185	0.0	0.224	0.826	3	√	√
3	0.216	0.0	0.224	0.963	2	√	√
4	0.189	1.0	0.195	0.966	1	√	√
5	0.132	2.0	0.175	0.753	4	√	√
6	0.077	3.0	0.161	0.480	7	√	√
7	0.039	4.0	0.149	0.259		√	√
8	0.017	5.0	0.140	0.121		√	
9	0.007	6.0	0.132	0.050		√	
10	0.002	7.0	0.125	0.018		√	
11	0.001	8.0	0.119	0.006		√	

F-C intervals

- here is a comparison of the LR ordering intervals with the “standard” ones

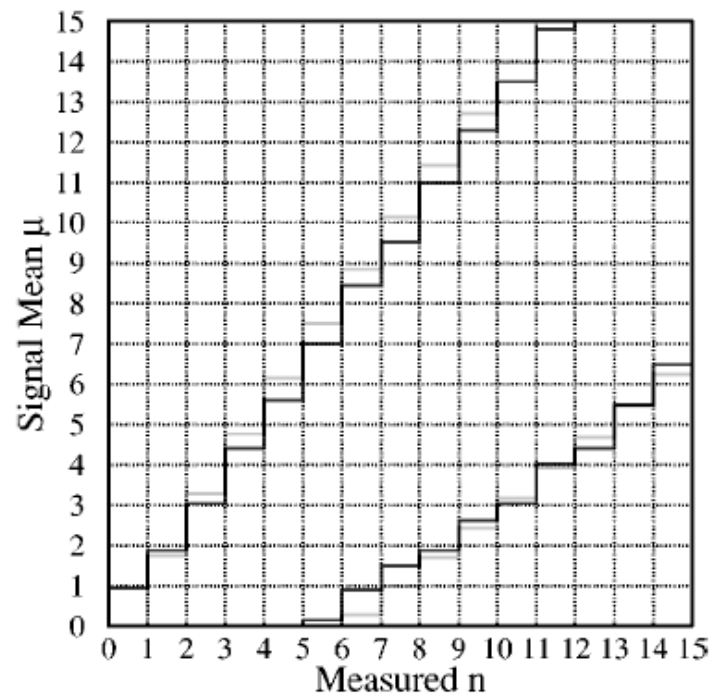


FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean μ in the presence of a Poisson background with known mean $b = 3.0$.

F-C inervals - Gaussian case

- F-C intervals pass smoothly from one-sided to two sided naturally
- observed data drives which is chosen
- all you can decide (a priori) is the CL
- for this case coverage is exact

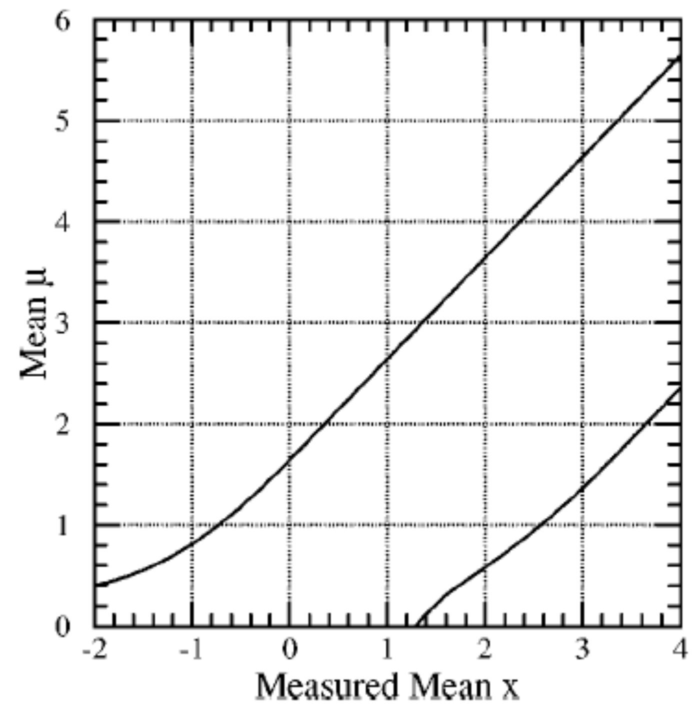
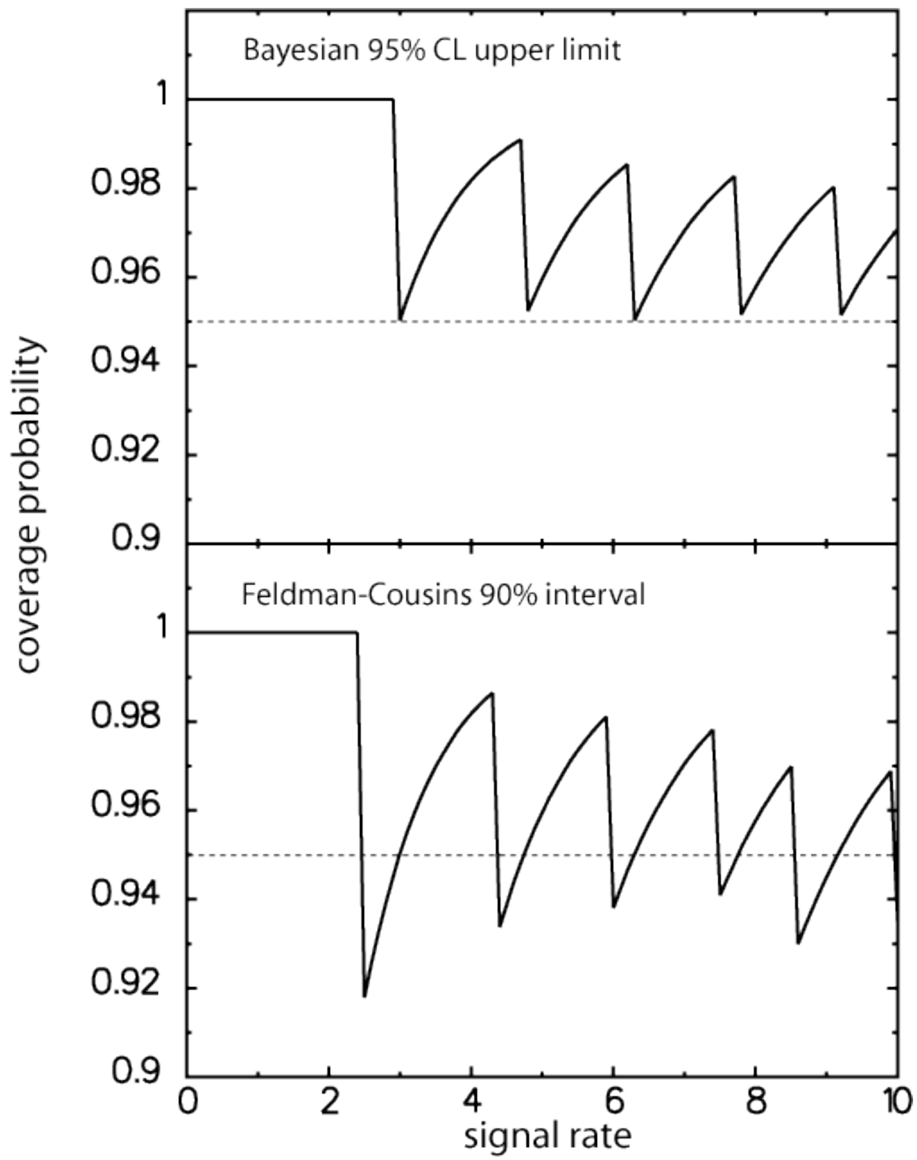


FIG. 10. Plot of our 90% confidence intervals for the mean of a Gaussian, constrained to be non-negative, described in the text.

coverage of F-C intervals

- how do we compare the coverage for, say, a 95% CL Bayesian upper limit with a F-C one?
- must decide on CL for F-C, since at low observed values we get a one sided limit and at high observed values we get a central interval



F-C intervals and systematics

- the obvious way to incorporate systematics into the F-C method is to put them into the likelihoods as nuisance parameters
- then, however, one must remove them either by marginalization or profiling
- what does this accomplish and how do we know if it does what we want?
- frequentists have a gold standard: coverage
- coverage here means coverage at a certain point in the n-dimensional space of all parameters including nuisance ones...computationally tractable?

people are trying to do just this...not common, however

LEP Higgs and CL_s

- LEP 2 Higgs search: combine ALEPH, DELPHI, OPAL, and L3 results:
  - six different search channels
  - use all information available
  - use event-by-event information (likelihoods)
  - include systematic errors and correlations
- the LEP Higgs WG invented a new method: CL_s
- "frequentist inspired"

definition of CL_s

- method assumes that there are two hypotheses, loosely dubbed “signal+background” and “background”
- the idea is to form an optimized discriminant and use p-values to characterize the decision between the hypotheses
- ultimately relies almost totally on distributions estimated using pseudoexperiments (though analytic calculations are in principle possible)
- the method was used extensively at LEP 2 and is now getting use at the Tevatron

definition of CL_s

- see A. Read note posted on line
- define two likelihoods, given observed quantities

$$\mathcal{L}(\bar{x}; s+b) \qquad \mathcal{L}(\bar{x}; b)$$

- form the statistic Q defined as

$$Q \equiv \frac{\mathcal{L}(\bar{x}; s+b)}{\mathcal{L}(\bar{x}; b)}$$

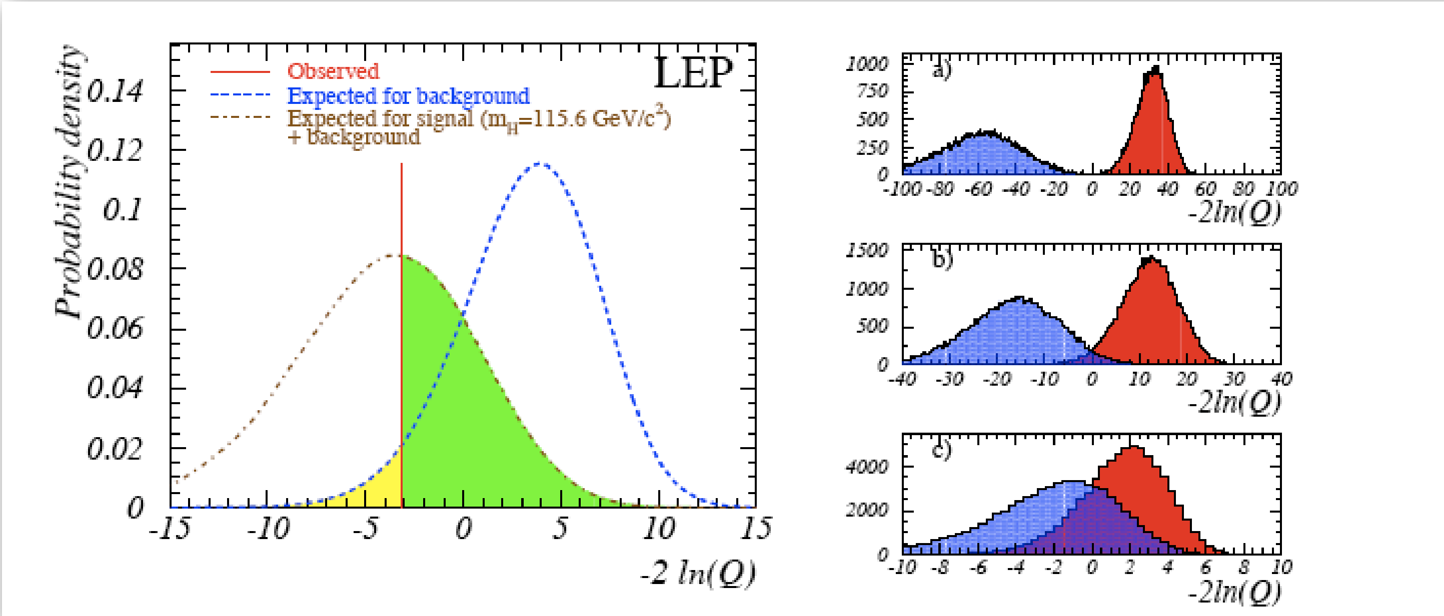
analogy to Helene formula!

- make distributions (find pdf's for) -2lnQ based on two situations:

1) s+b is true

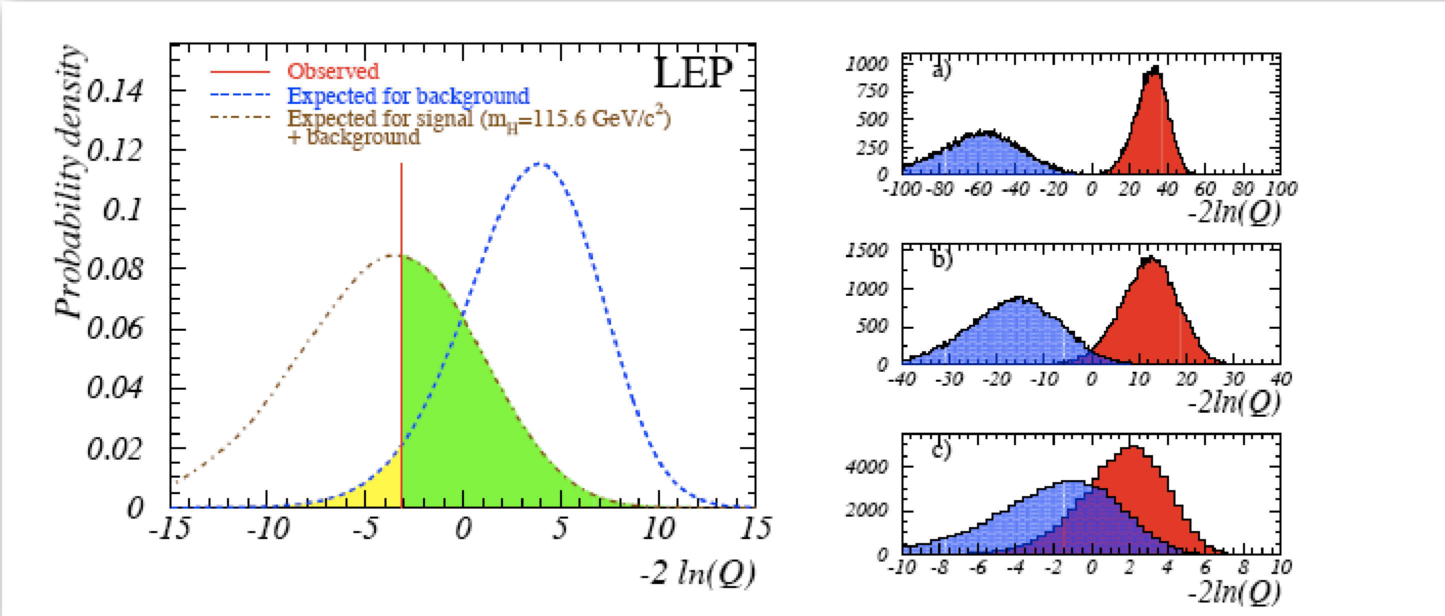
2) b is true

definition of CL_s



- in the experiment we observe one value for Q
- we compare with the distributions for s+b and b
- a lot of overlap: not much sensitivity (power)
- not much overlap: high sensitivity (power)

definition of CL_s



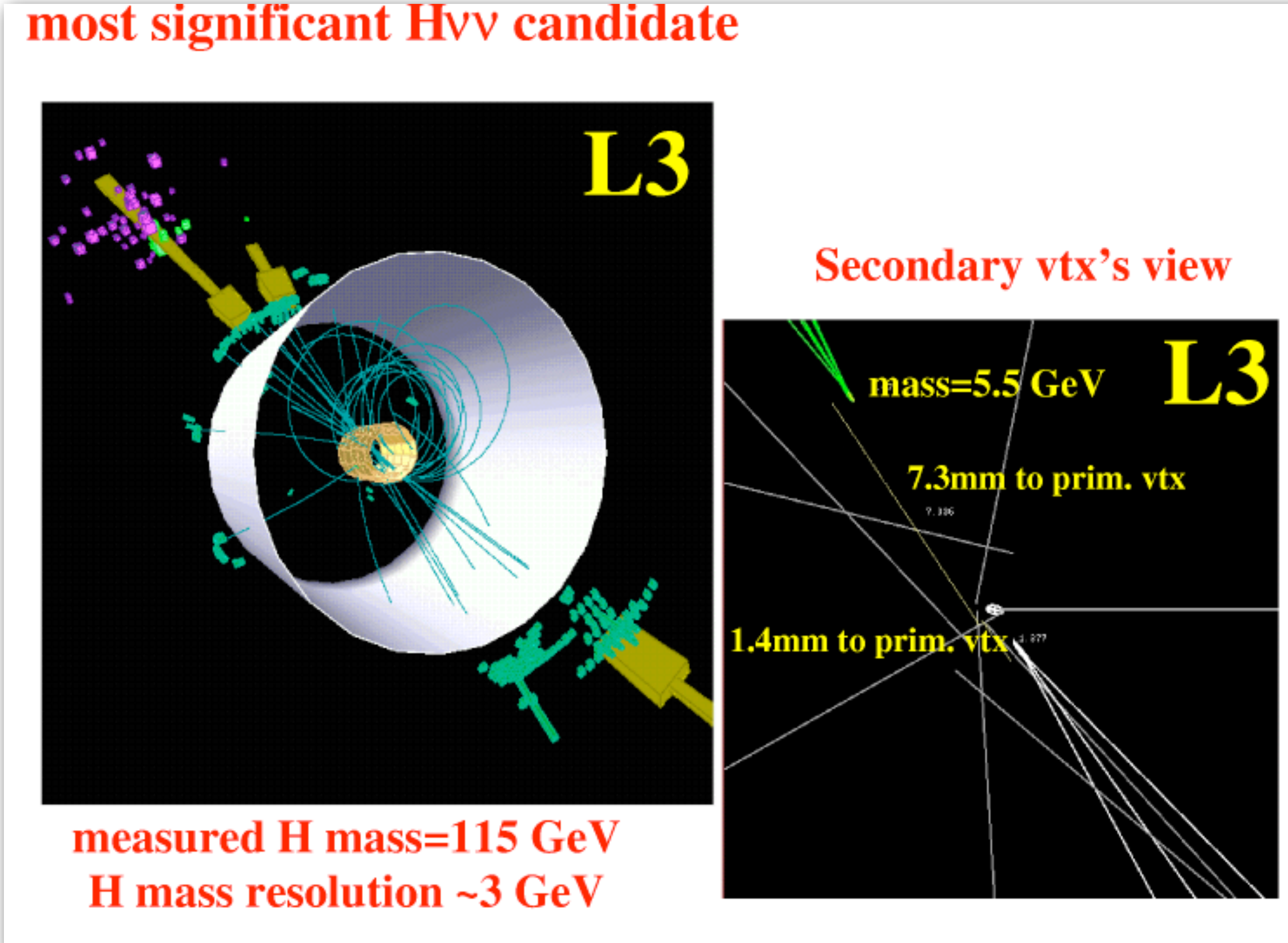
$$1 - CL_b \equiv \int_{-\infty}^{Q_{obs}} \mathcal{P}(Q'; b)\,dQ'$$

$$CL_{s+b} \equiv \int_{Q_{obs}}^{+\infty} \mathcal{P}(Q'; s+b)\,dQ'$$

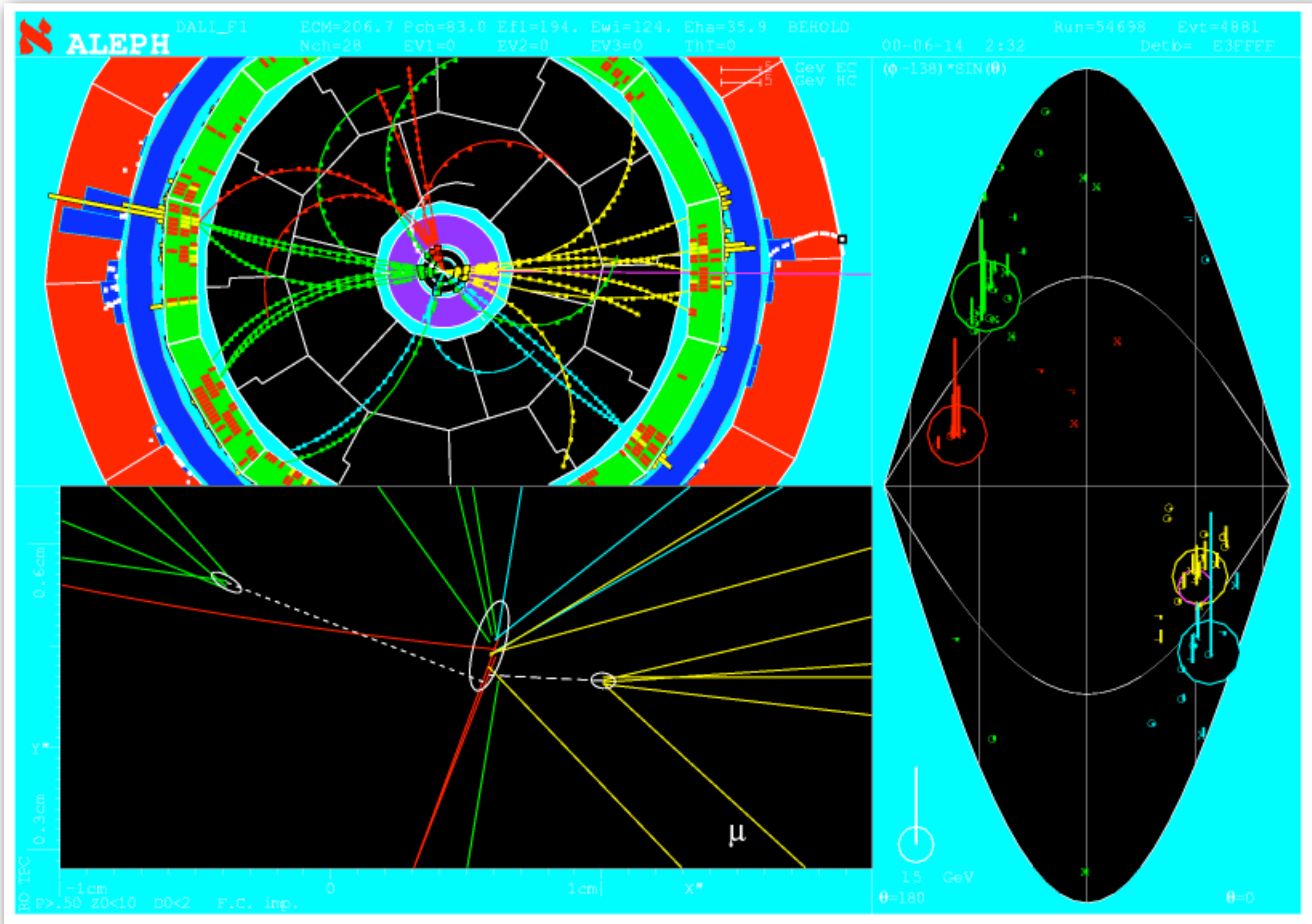
$$CL_s \equiv \frac{CL_{s+b}}{CL_b}$$

denominator penalizes insensitivity

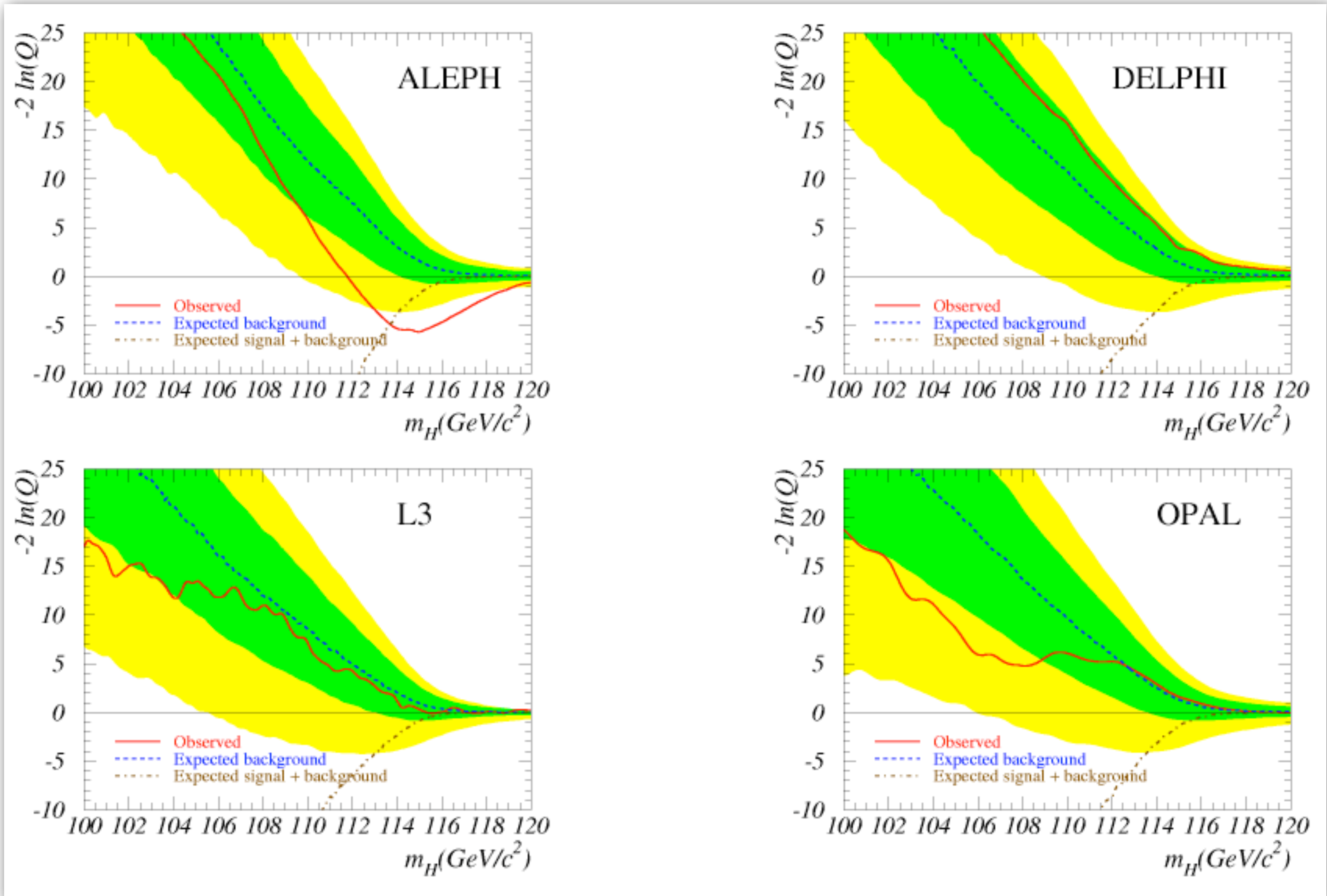
LEP 2 Higgs search



LEP 2 Higgs search

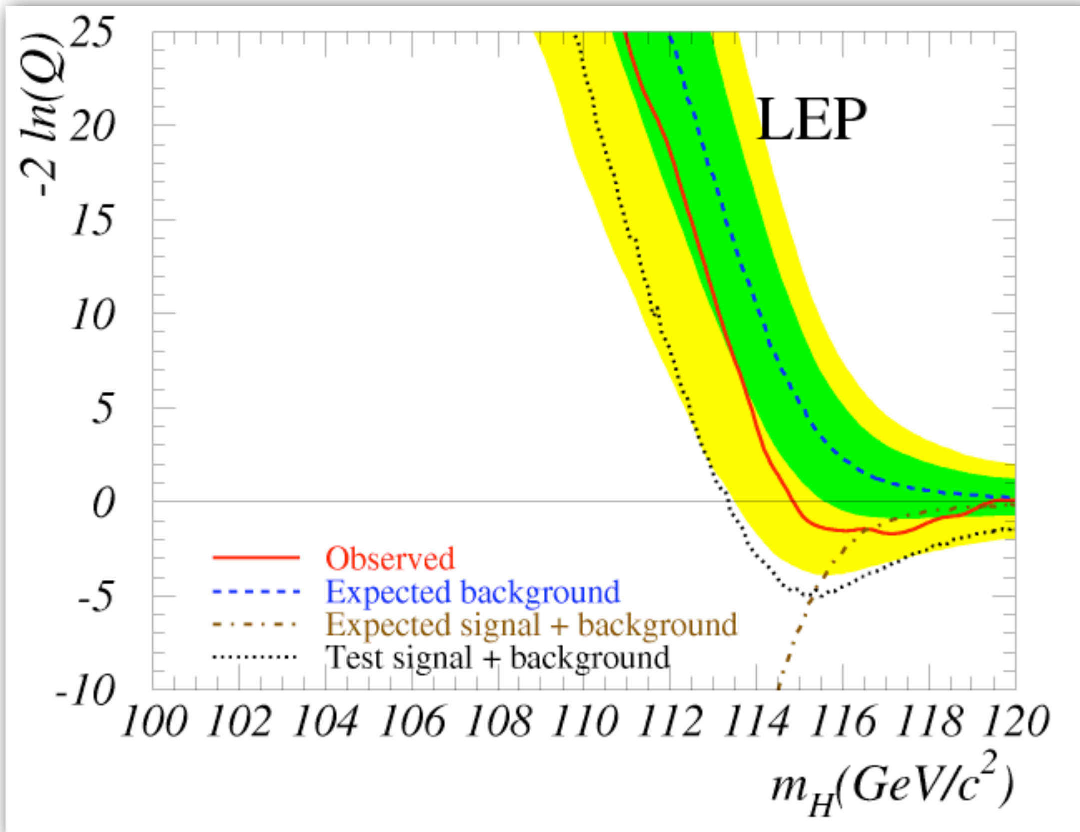


LEP 2 Higgs search



green band: expectation for no Higgs signal

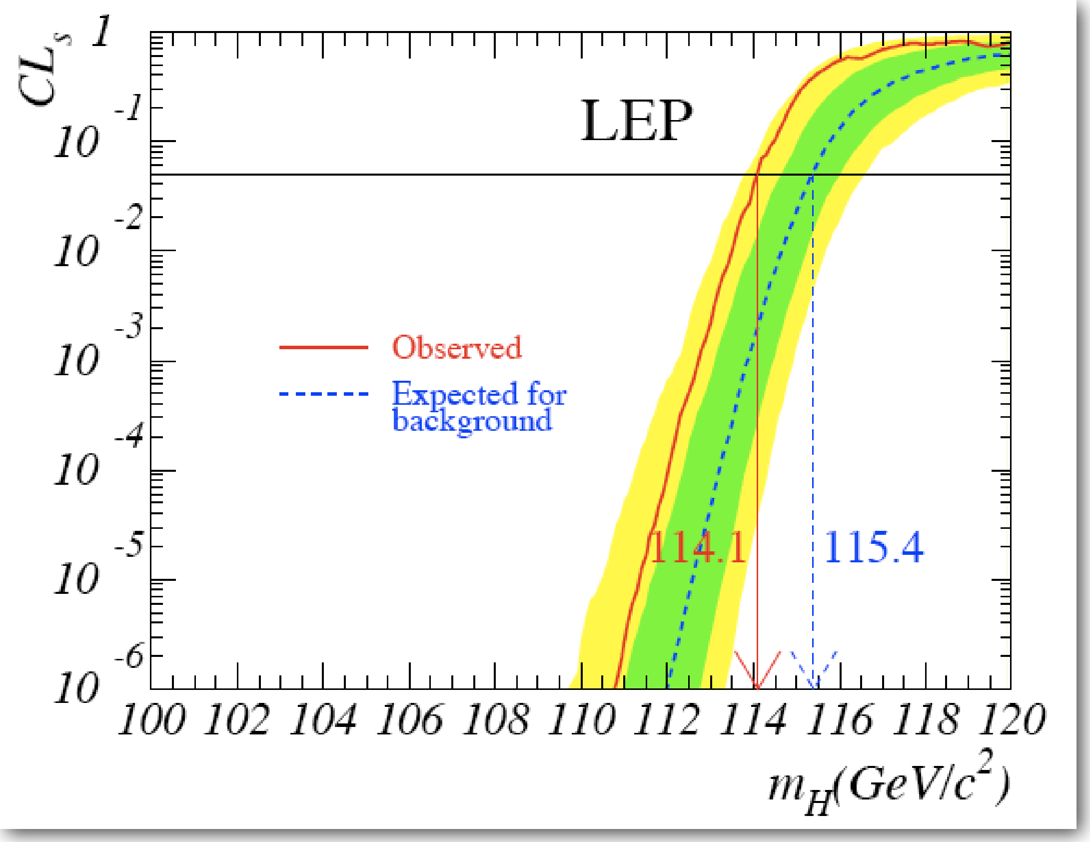
LEP 2 Higgs search - combined



Higgs mass > 114.3 GeV at 95% CL

but you cannot really tell that from this plot...

LEP 2 Higgs limit



CL_s and systematic uncertainties

- ok, fasten your seat belts - this is the weird part
- your first inclination is that you include systematic uncertainties as nuisance parameters in the likelihood
- but all that does is broaden the -2lnQ distribution for both the b and s+b cases!
- no change in sensitivity!
- so what do we do?
- in CL_s, systematic uncertainties are included at the level of the *pseudoexperiment generation* by varying the nuisance parameters