CLs: A heretic repents

Bill Murray
RAL, STFC
bill.murray@stfc.ac.uk

Banff
13th July 2010

- **Where is Cl_s from?**
- **What is Cl_s?**
- **What is it not?**

LEP Higgs search issues

- Typical analysis:
  - Very low background
    - Handful of events
  - Candidates come with mass estimate
    - Good resolution
- Lets look at what we said in 1997...

Report from the LEP working group for Higgs boson searches

Members

ALEPH: J. Carr, P. Janot
DELPHI: W. Murray, A. Read, V. Ruhlmann-Kleider
L3: M. Biasini, A. Kounine, F. diLodovico, M. Pieri
OPAL: P. Bock, S. De Jong, E. Gross, P. Igo-Kemenes
Theory: M. Carena, C. Wagner

Mandate / goals

Combine the results of the four experiments on Higgs boson searches (SM and MSSM)

→ **Increase in the overall statistical sensitivity**

Since January 1997 regular, (monthly) open meetings

- **Examination of existing / development of new statistical methods** → optimal procedure

different decay channels

at different c.m. energies — **SM:** $e^+e^- \rightarrow HZ$

from different experiments — $5 \times 2 \times 4 = 40$ channels

Complexity of input information :

Detection efficiencies, mass resolutions: $f(m_H)$

residual backgrounds, candidate events.

→ **Four methods proposed,** one per experiment

(adapted to LEP2 experimental environment)

Statistical methods proposed

(based on those used internally by each experiment)

ALEPH: *Combining confidence levels analytically*
P. Janot, F. Le Diberder, CERN PPE/97-053

DELPHI: *Modified frequentist approach* A. Read (soon ...)

L3: *Bayesian likelihood ratio* A. Favara, M. Pieri, DFF-278/4/1997

OPAL: *Fractional event counting* P. Bock, HD-PY-96/05 (1996)

Common features

- Construction of a single test-statistic (*"estimator"* X) allowing to *rank* the channels (or experiments) between *signal+background*-like and *background*-like.

  Usually X is calculated from (i) the event rate and (ii) from distributions in discriminating variables (e.g. m_{H}^{rec}, event shape, b-flavour content).

- Distribution of the same estimator for a large number of *"gedanken" experiments* incorporating all experimental features and the SM predictions for signal and background, for the *s+b* and *b only* hypotheses: X_{s+b} and X_b, for various m_{H} hypotheses;

- Definition of a *confidence level*, depending of the m_{H} hypothesis: fraction of gedanken experiments with estimators *less s+b*-like than observed:

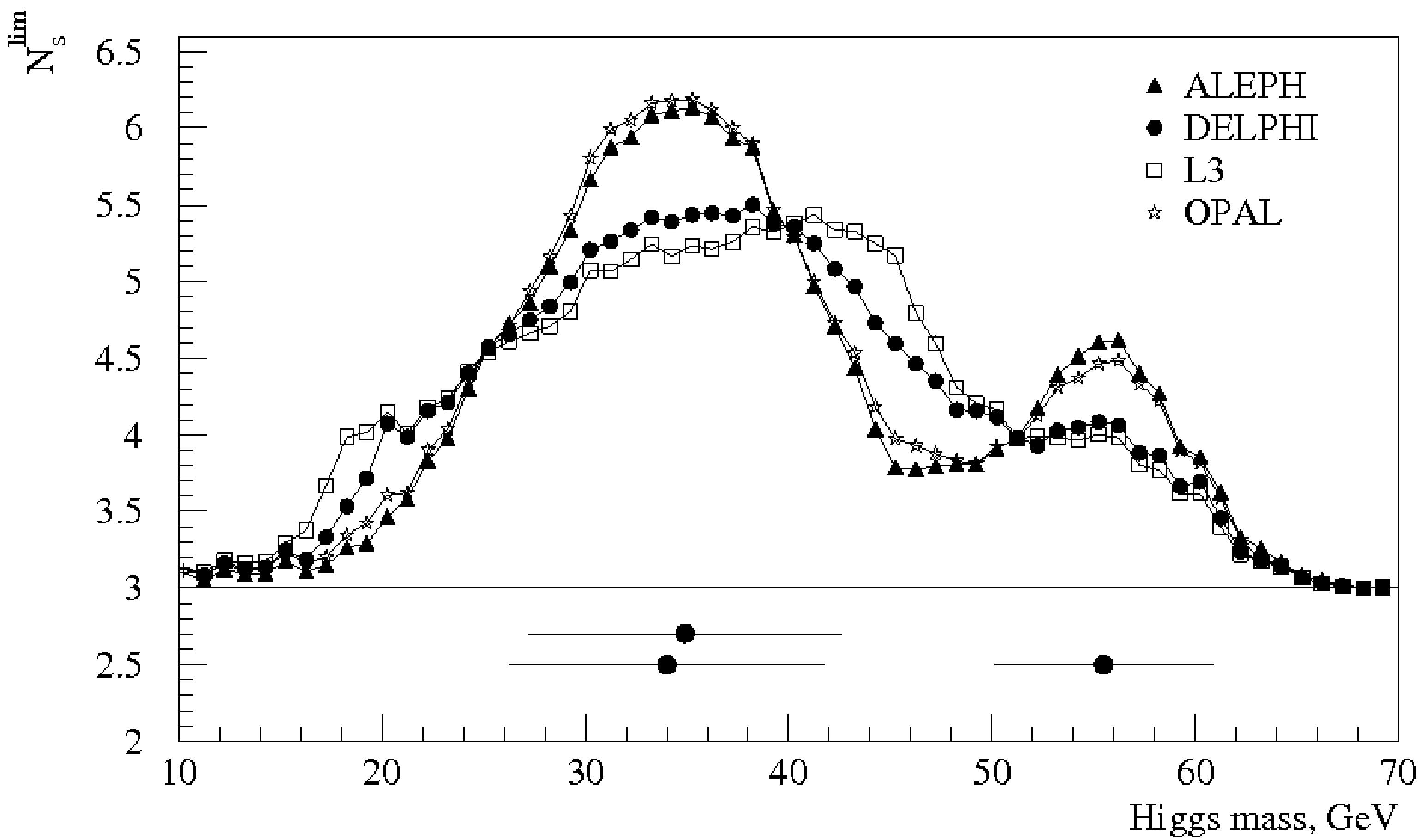
$$CL = \frac{P(X_{s+b} < X_{obs})}{P(X_b < X_{obs})} \qquad (1)$$

Comparative tests (Monte Carlo experiments)

1. Effect of candidate events: a fictive experiment

Background prediction: 4 events, flat in mass

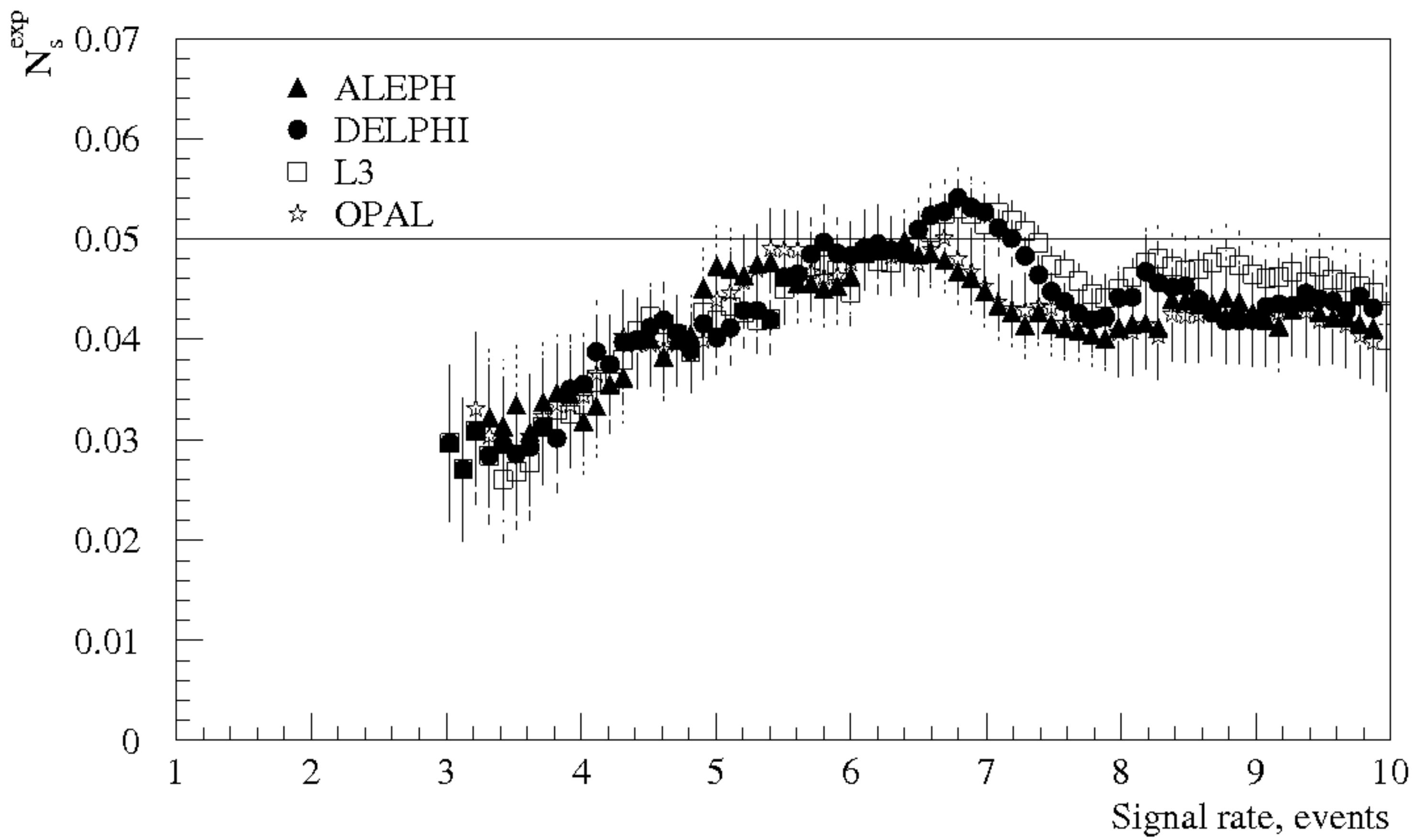
Observation: 3 candidates with measured masses;



2. False exclusion probability

Assume: Signal actually existing (m_{H}=77 GeV/c^2);

? Fraction of gedanken experiments which would exclude the signal, at 95% c.l. ?



Fraction ... as a function of the signal event rate expected.

→ *Convergence close to 5%; similar for all methods*

CL_s

- So...what LEP did not invent was CL_s
- That came from the Helene formula in the RRP
  - R.M. Barnett et al., Physical Review D54, 1 (1996).

$$\epsilon = 1 - \frac{e^{-\mu_b + N} \sum_{n=0}^{n_0} (\mu_b + N)^n / n!}{e^{-\mu_b} \sum_{n=0}^{n_0} (\mu_b)^n / n!}$$

- All we did was generalize it from Poisson

$$CL_s = CL_{sb} / CL_b$$

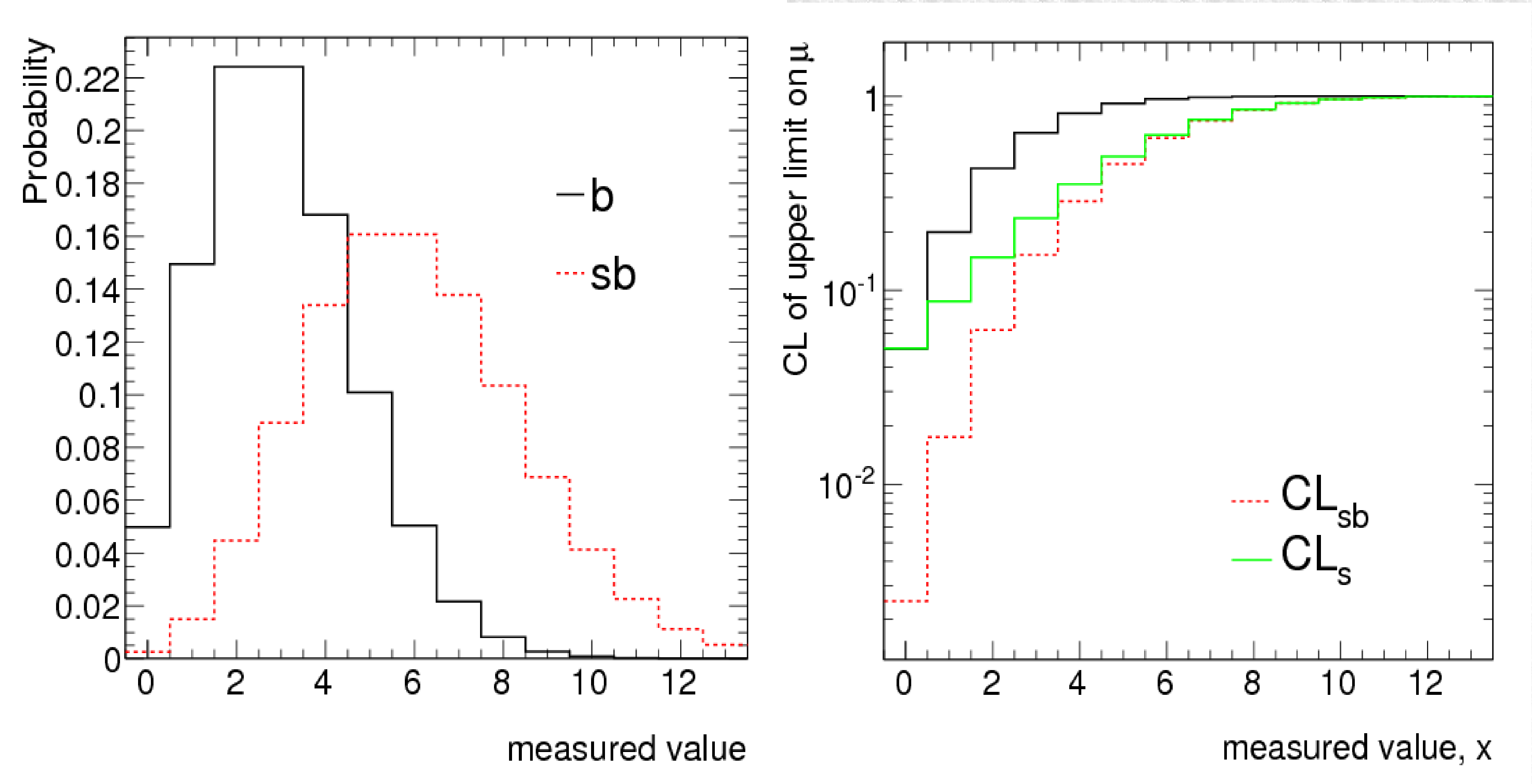
Properties of CL_s

- It was recommended by the RPP
- It produced (over)covering Frequentist limits
  - We knew 95% CL must mean SOMETHING
- It produced results Bayesians could use
  - "We cannot exclude a signal with less than 3 events"
- It gave more aggressive limits than Bayesian
    - (We didn't realise the flat prior was there!)
  - We didn't do this to be conservative

– E Gross

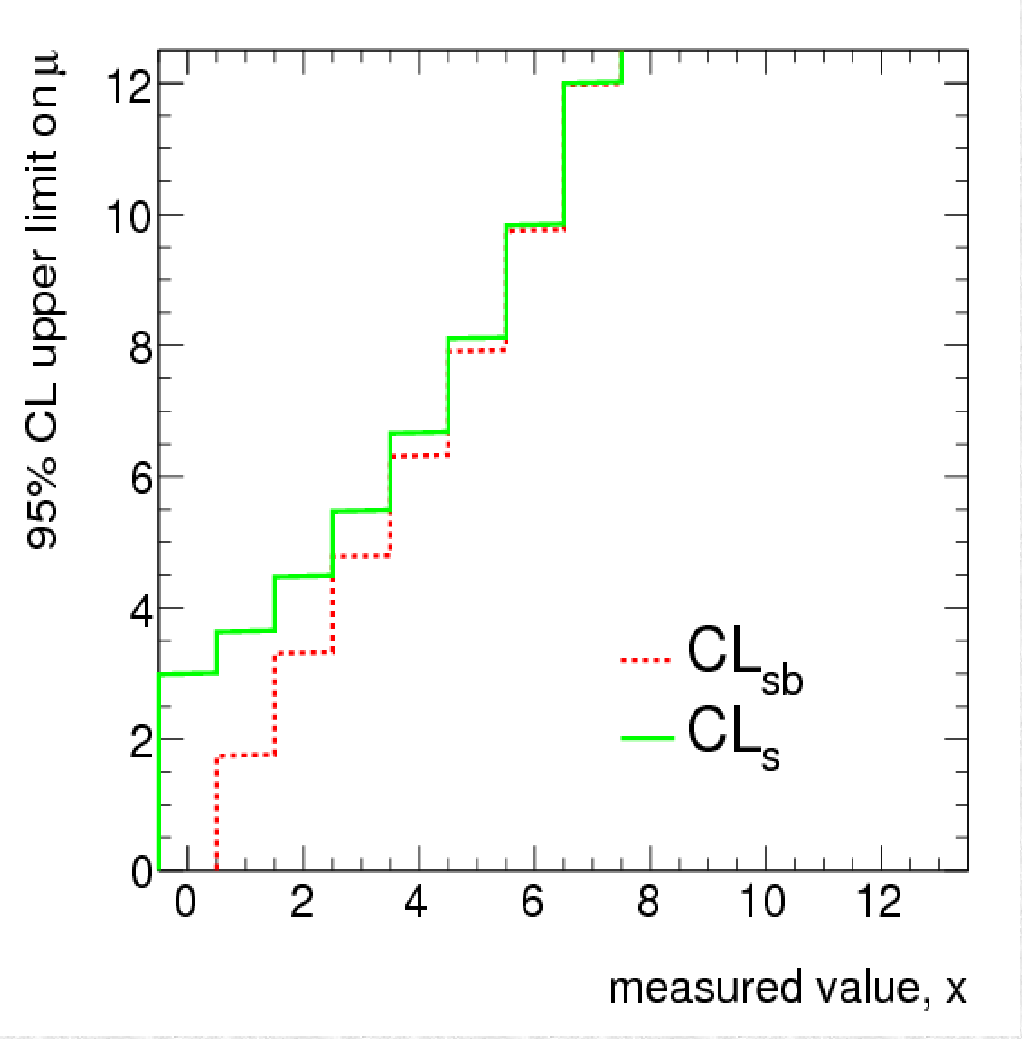
Demonstration for Poisson

- Background of 3, signal of 3



Poisson 95% CL upper limits

- The Frequentist Cl_{sb} for 0 observed has 95% UL on signal at 0
  - All signal excluded!
- Cl_s has minimum possible UL at 3
  - The same as a background-free analysis
  - **No way to profit from presence of background**



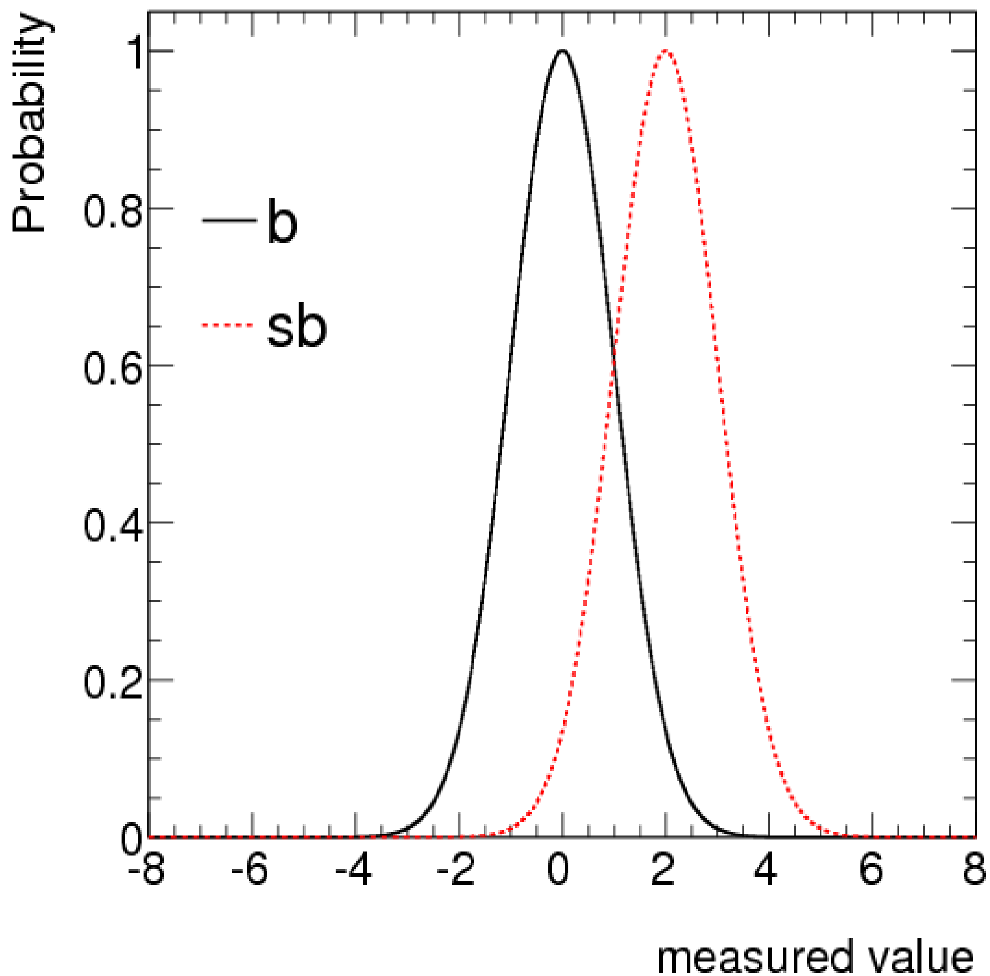
So far so good?

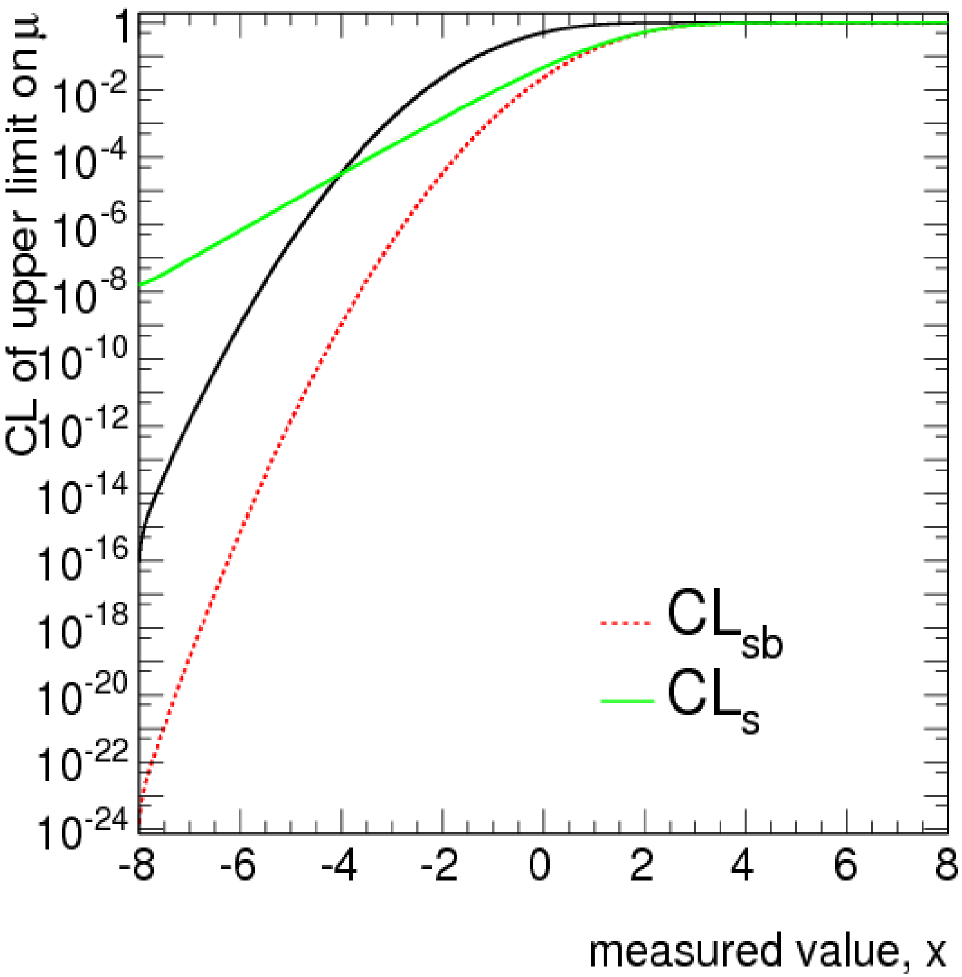
$$CL_s = CL_{sb}/CL_b$$

- This division somehow mimics LR = $\mathcal{L}_{sb}/\mathcal{L}_s$
  - While maintaining coverage
- But statisticians always disliked it...
- Cowan, Cranmer, Gross and Vitells propose power constrained limits
  - Sensitivity 'cut off' in over-sensitive regions.
- Demotier and Cousins exchange emails on their properties
- And suddenly it all goes wrong for me...

Properties of Gaussian

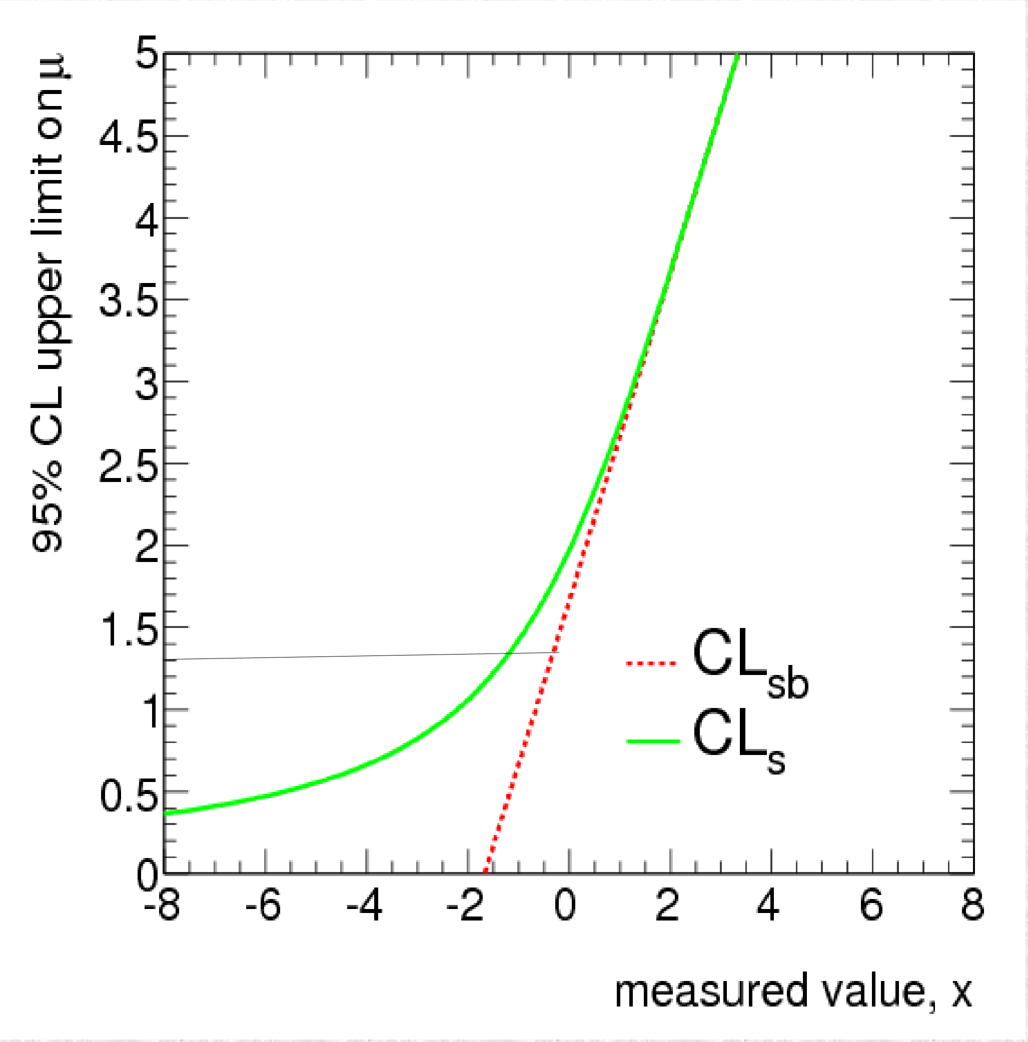
- Mean 0, width 1; signal of 2





Gaussian 95% CL upper limits

- Cl_{sb} for negative x can exclude all signal.
  - OK
- Cl_s always positive
  - But a signal of 1±1 can be excluded
  - **Can profit from background fluctuations**
  - I thought that was what CL_s prevented.



Why did Poisson work?

- Because it cheated
- The CL sum, for n=0, becomes just the Poisson probability, and the CL_{sb}/CL_b is just the LR
  - Thus it worked for one special case
  - If rather a common one

Conclusion

- CL_s appeared to provide limits acceptable to Bayesians and Frequentists
  - That was illusory
  - Although some protection was given
- There was no principled justification for extending the Helene formula beyond the Poisson case
  - And it doesn't work
- Limits (Discoveries?) which truncate results outside sensitivity should be pursued
  - Arbitrary nature is a feature, not a bug
  - Probably more aggressive: Happy Eilam?