MANCHESTER
1824

The University of Manchester

Statistics for Particle Physics: Limits

Roger Barlow

Karlsruhe: 12 October 2009

- Limits
- Neyman constructions
- Poisson data
- Poisson Data with low statistics
- Poisson Data with low statistcs and backgrounds
- Feldman Cousins methid
- Bayesian limits

Intervals and Limits

Quoting an interval gives choice – not just in probability/confidence level

Usual 'interval' is symmetric. E.g. 68% inside, 16% above, 16% below.

'Limit' is 1-sided. E.g.Less than 172.8 @ 68%

Usually upper limits. Occasionally lower limits.

The University of Manchester

MANCHESTER 1824

Generic Problem

You are searching for some process ("Cut and count". More sophisticated methods technically tougher but conceptually similar)

- Choose cuts (from Monte Carlo, sidebands, etc)
- Open the box (Blind analysis!)
- Observe N events

Report N ± √N events (Poisson Statistics)

Scale up to get Branching Ratio, Production Cross section, or more complicated things like particle mass

What happens if N is small (So Poisson ≠ Gaussian) ?

What happens if N =0?

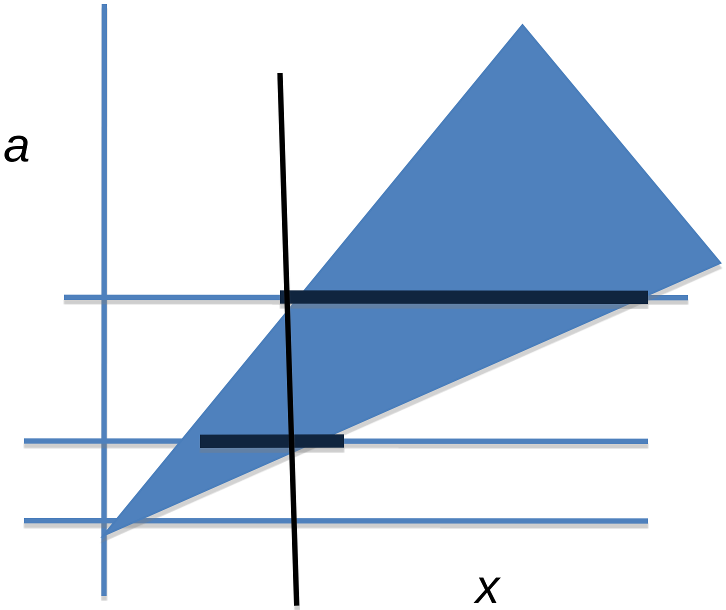
How do you handle background component?

Intervals (again)

Non-Gaussian distributions need care.

Confidence band (Neymann)

Construct horizontally. Read vertically.



The University of Manchester

Example

Open the box. See nothing! Cannot quote answer as μ= 0 ± 0

Poisson formula $P(n;\mu)=e^{-\mu}\mu^n/n!$

Plausibly μ≅0.1. Run 100x longer and you'll see something.
Possibly μ≈1 and you're just unlucky.

Find μ such that P(0; μ)=100-68% (or whatever)
In particular P(0;3.0) = 5%
p-values again: If μ=3, the probability of getting a result this bad (=small) is only 5%
So: when you see nothing, say "With 95% confidence, the upper limit on the signal is 3" – and calculate BR, cross section etc from that

Small numbers

See an event. Or a few events.

Discovery? Perhaps. But maybe you expect a background…

Repeat p-value calculation. See n, find μ such that $\Sigma_0^n e^{-\mu}\mu^r/r!$ is 0.05 (or whatever).

n	90%	95%
0	2.30	3.00
1	3.89	4.74
2	5.32	6.30
3	6.68	7.75
4	7.99	9.15

Where it gets sticky…

You expect 3.5 background events.

Suppose you open the box and find 15 events.
Measurement of signal 11.5±√15.

Suppose you find 4 events. Upper limit on total 9.15. Upper limit on signal 5.65 @ 95%

Suppose you find 0 events. 3.00 → -0.5

Or 1 event: 4.74→ 1.24

MANCHESTER 1824

The University of Manchester

Problem

Technically these are correct. 5% of '95%CL' statements are allowed to be wrong.

But we don't want to make crazy statements. Even though purists argue we should to avoid bias.

Problem is that the background clearly has a downward fluctuation*. But we have no way of including this information in the formalism

* Assuming you calculated it right.

Similar problems

- Expected number of events must be non-negative
- Mass of an object must be non-negative
- Mass-squared of an object must be non-negative
- Higgs mass from EW fits must be bigger than LEP2 limit of 114 GeV

3 Solutions

- Publish a ‘clearly crazy’ result
- Use Feldman-Cousins technique
- Switch to Bayesian analysis

Bayesian limits from small number counts

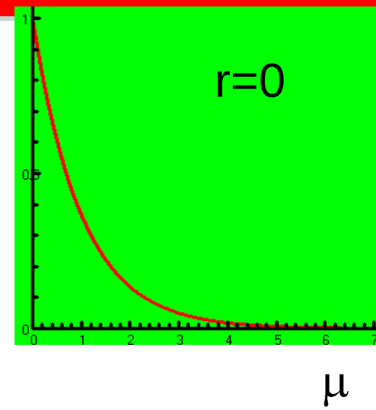
$P(r,\mu)=e^{-\mu}\mu^{r}/r!$

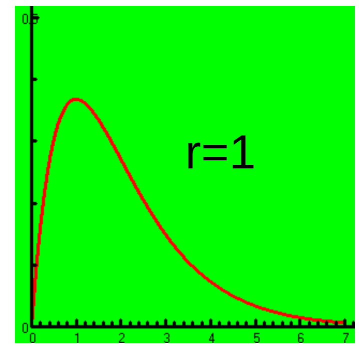
With uniform prior this gives posterior for μ

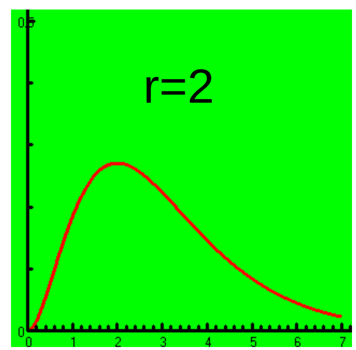
Shown for various small r results

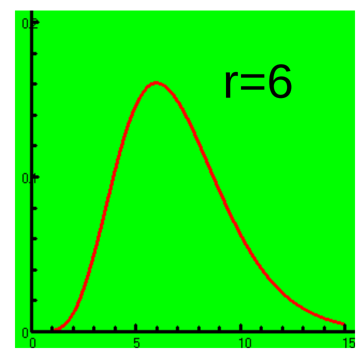
Read off intervals...

$P(\mu)$









Upper limits

Upper limit from n events

$$\int_0^{\mu_{HI}} \exp(-\mu)\, \mu^n/n!\ d\mu = CL$$

Repeated integration by parts:

$$\Sigma_0^n \exp(-\mu_{HI})\, \mu_{HI}^r/r! = 1\text{-}CL$$

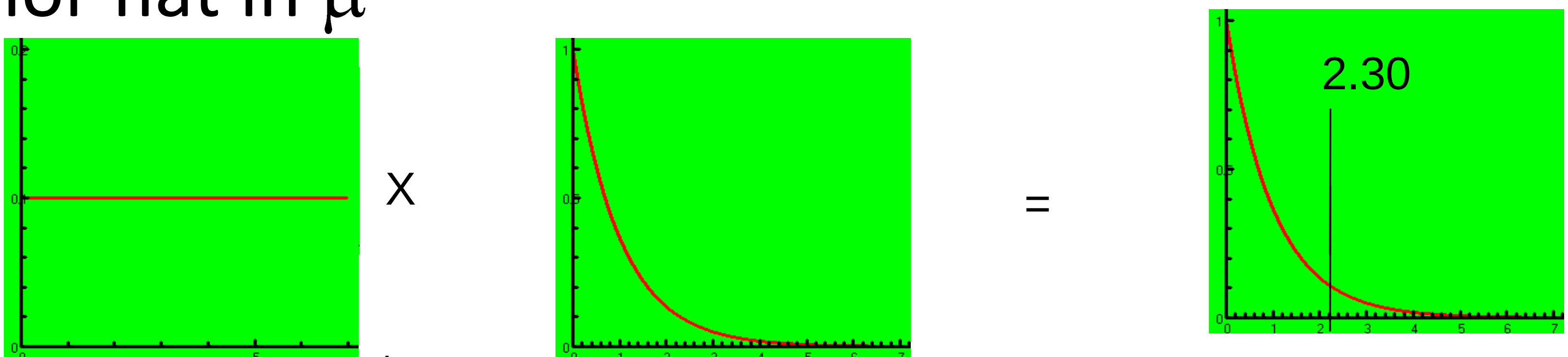
Same as frequentist limit

This is a coincidence! Lower Limit formula is not the same

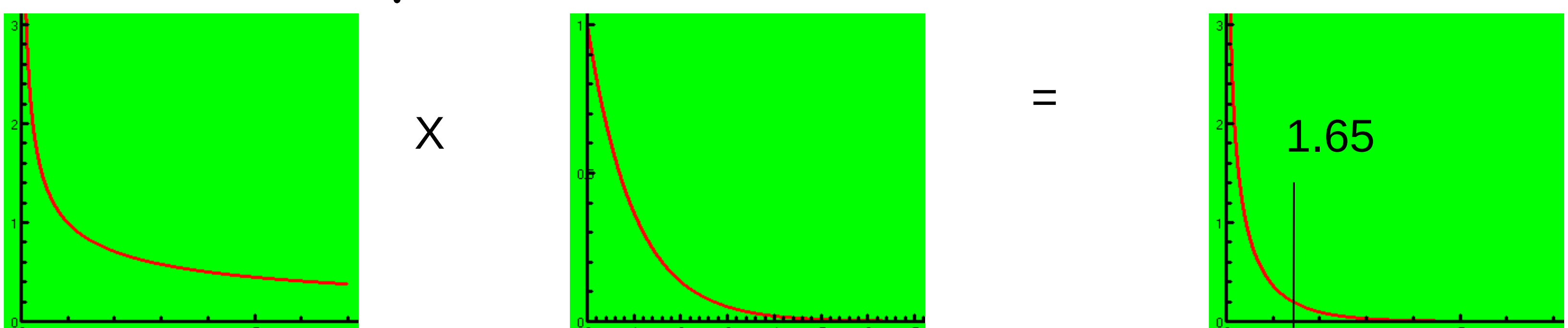
Result depends on Prior

Example: 90% CL Limit from 0 events

Prior flat in μ



Prior flat in √μ

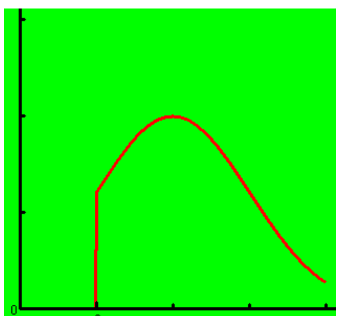


Aside: "Objective" Bayesian statistics

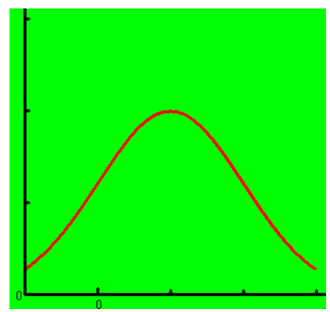
- Attempt to lay down rule for choice of prior
- 'Uniform' is not enough. Uniform in what?
- Suggestion (Jeffreys): uniform in a variable for which the expected Fisher information $<d^2 \ln L/dx^2>$ is minimum (statisticians call this a 'flat prior').
- Has not met with general agreement – different measurements of the same quantity have different objective priors

μ=S+b for Bayesians

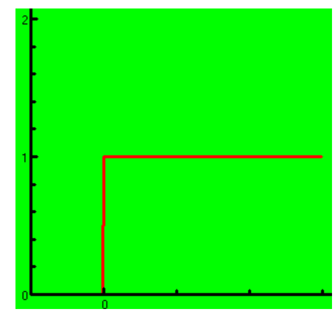
- No problem!
- Prior for μ is uniform for S≥b
- Multiply and normalise as before



=



X



Posterior Likelihood Prior

Read off Confidence Levels by integrating posterior

Another Aside: Coverage

Given $P(x;\mu)$ and an ensemble of possible measurements $\{x_i\}$ and some confidence level algorithm, coverage is how often '$\mu_{LO} \leq \mu \leq \mu_{HI}$' is true.

Isn't that just the confidence level? Not quite.

- Discrete observables may mean the confidence belt is not exact – move on side of caution
- Other 'nuisance' parameters may need to be taken account of – again erring on side of caution

Coverage depends on μ. For a frequentist it is never less than the CL ('undercoverage'). It may be more ('overcoverage') – this is to be minimised but not crucial

For a Bayesian coverage is technically irrelevant – but in practice useful

Feldman Cousins Method

Works by attacking what looks like a different problem...

Also called* 'the Unified Approach'

Physicists are human

Ideal Physicist

1. Choose Strategy
2. Examine data
3. Quote result

Real Physicist

1. Examine data
2. Choose Strategy
3. Quote Result

Example:

You have a background of 3.2

Observe 5 events? Quote one-sided upper limit (9.27-3.2 =6.07@90%)

Observe 25 events? Quote two-sided limits

* by Feldman and Cousins, mostly

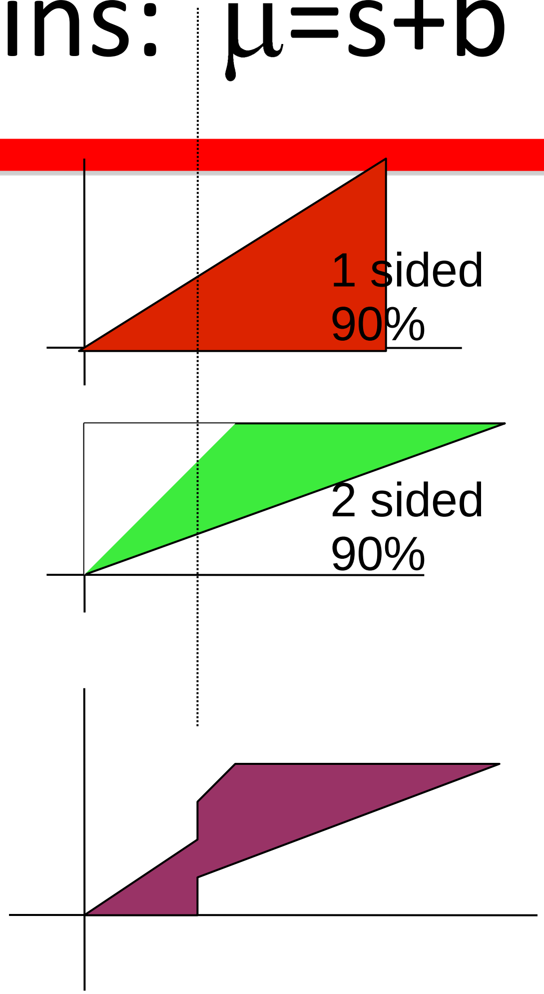
Feldman Cousins: μ=s+b

This is called 'flip-flopping' and BAD because is wrecks the whole design of the Confidence Belt

Suggested solution:

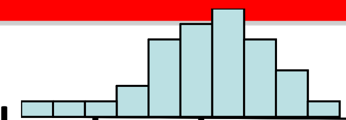
1) Construct belts at chosen CL as before (for s,N and given b)

2) Find new ranking strategy to determine what's inside and what's outside



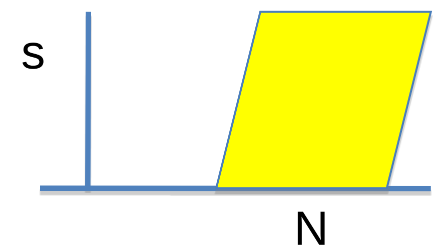
Feldman Cousins: Ranking

First idea (almost right)



Sum/integrate over range of values with highest probabilities.

(advantage: this is the shortest interval)



Glitch: Suppose N small. (low fluctuation)

P(N;s+b) will be small for any s and never get counted

Instead: compare to 'best' probability for this N, at s=N-b or s=0 and rank on that number

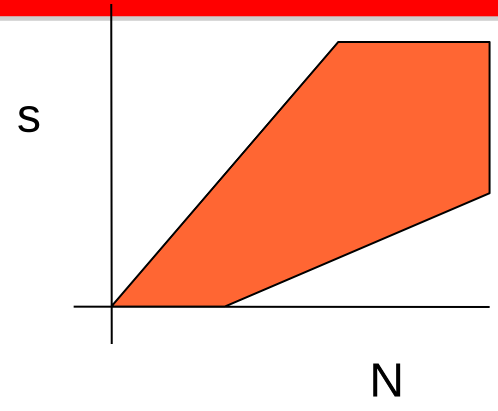
Such a plot does an automatic ‘flip-flop’

N~b single sided limit (upper bound) for s

N>>b 2 sided limits for s

How it works

- Has to be computed for the appropriate value of background b. (Sounds complicated, but there is lots of software around)
- As N increases, flips from 1-sided to 2-sided limits – but in such a way that the probability of being in the belt is preserved



Means that sensible 1-sided limits are quoted instead of nonsensical 2-sided limits!

Arguments against using Feldman Cousins

- Argument 1

It takes control out of hands of physicist. You might want to quote a 2 sided limit for an expected process, an upper limit for something weird

- Counter argument:

This is the virtue of the method. This control invalidates the conventional technique. The physicist can use their discretion over the CL. In rare cases it is permissible to say ”We set a 2 sided limit, but we're not claiming a signal”

Feldman Cousins: Argument 2

- Argument 2

If zero events are observed by two experiments, the one with the higher background b will quote the lower limit. This is unfair to hardworking physicists

- Counterargument

An experiment with higher background has to be 'lucky' to get zero events. Luckier experiments will always quote better limits. Averaging over luck, lower values of b get lower limits to report.

Example: you reward a good student with a lottery ticket which has a 10% chance of winning €10. A moderate student gets a ticket with a 1% chance of winning €20. They both win. Were you unfair?

Including Systematic Errors

$$\mu = aS + b$$

μ is predicted number of events

S is (unknown) signal source strength. Probably a cross section or branching ratio or decay rate

a is an acceptance/luminosity factor known with some (systematic) error

b is the background rate, known with some (systematic) error

1) Full Bayesian

Assume priors

- for S (uniform?)
- For a (Gaussian?)
- For b (Poisson or Gaussian?)

Write down the posterior P(S,a,b).

Integrate over all a,b to get marginalised P(s)

Read off desired limits by integration

2) Hybrid

Assume priors

- For a (Gaussian?)
- For b (Poisson or Gaussian?)

Integrate over all a,b to get marginalised P(r,S)

Read off desired limits by $\Sigma_0^n P(r,S) = 1\text{-}CL$ etc

Done approximately for small errors (Cousins and Highland). Shows that limits pretty insensitive to σ_a, σ_b

Numerically for general errors (RB: java applet on SLAC web page). Includes 3 priors (for a) that give slightly different results

And more…

- Extend Feldman Cousins
- Profile Likelihood: Use $P(S)=P(n,S,a_{max},b_{max})$ where a_{max},b_{max} give maximum for this S,n
- Empirical Bayes
- And more…

And another things…

- Using information as well as numbers can use χ^2 or likelihood

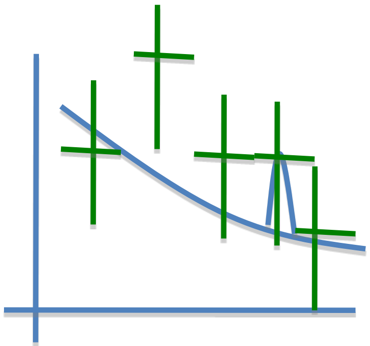
$P(x;s,a)=B(x)+s\,S(x)$

Can obtain intervals for s and limits on s

Warning: these limits are not valid if $S(x)=S(x,a)$. Even if you reduce the number of degrees of freedom of χ^2

Example

Suppose S(x) is a Breit-Wigner of known mass and small width. All OK. Minimisation neutralises the relevant bin. Big improvement means its doing something real



Suppose S(x) is a Breit-Wigner but mass is allowed to float. Minimisation will neutralise the worst bin. Massive improvement anyway. Large $\Delta\ \chi^2$ but so what?

Use a toy Monte Carlo!!

Summary

- Straight Frequentist approach is objective and clean but sometimes gives ‘crazy’ results
- Bayesian approach is valuable but has problems. Check for robustness under choice of prior
- Feldman-Cousins deserves more widespread adoption
- Lots of work still going on
- This will all be needed at the LHC