Making Discoveries and Setting Limits

Roger Barlow
Huddersfield University

LHC physics school, NCP Islamabad

22^{nd} August 2025

1. Claiming a discovery using simple counting
2. Claiming a discovery from a fitted signal strength
3. Expected performance for a discovery
4. Setting an upper limit using simple counting
5. Setting an upper limit from a fitted signal strength
6. Expected performance for an upper limit

Claiming a discovery using simple counting

You expect 2.3 background events and you see 11

What can you say?

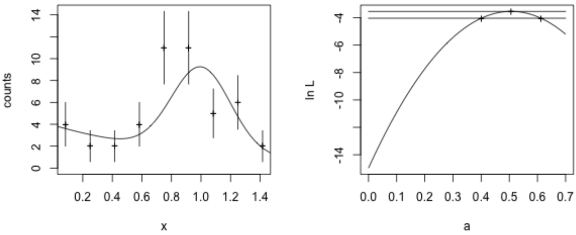
Poisson says $P(r;\mu) = e^{-\mu}\frac{\mu^r}{r!}$ so $P(11;2.3) = 0.0000239$

Also need to consider $P(12;2.3) = 0.0000046$ etc.
$\sum_{11}^{\infty} P(r, 2.3) = 1 - \sum_{0}^{10} P(r;2.3) = 0.0000295$

That's the p-value. Under the null hypothesis – that there is no signal and this is just a statistical fluke – the probability of getting a result that looks are much (or more) like a signal as this one is only 0.003 %.

The corresponding $Z = 4.017$. So you can quote 'evidence for'

Claiming a discovery by fitting a signal strength



$P(x) =$
$(1 - a) \times exp(-x) + a \times$
$Gauss(x; 1.0, 0.2)$

Use $t = \chi^2 or - 2\Delta \ln L$. Plot $\ln L$ (actually $-\frac{1}{2}\chi^2$) as function of signal strength a
Can read off $\hat{a} = 0.507 \pm 0.106$
This is 4.7832 sigma from zero - discovery.

Alternatively: $\Delta \ln L = \ln L(\hat{a}) - \ln L(0) = 11.44$, $\Delta\chi^2 = 22.88$,
$\sqrt{\Delta\chi^2} = 4.7834$. Compatible answer!
Using $\Delta \ln L$ is quicker and (probably) better.

Expected performance

Your result may be more significant or less significant, depending on luck.

Expected Performance

Suppose the signal is present at the level expected. How significant a result do we expect to get?

Useful in 2 cases

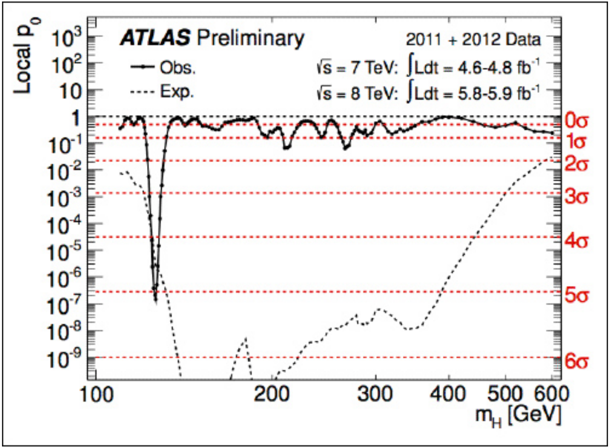
1. When applying for funding, machine time, etc before the experiment
2. After the experiment, to show whether you just got lucky

Suppose a Poisson background is 2.3 and the model says your signal strength is 5.4. Your result will be random Poisson of strength 7.7.

Each possible result $r = 0, 1, 2.....$ has a probability $P(r; 7.7)$ and a p-value $1 - \sum_{r'=0}^{r} P(r', 2.2)$

You find the average p-value, or average Z – and for 'average' you take the median rather than the (more usual) mean, as the median p-value matches the median Z, but this doesn't hold for means.

Expected performance - example from ATLAS



Calculation for each M_H separately

Dashed line shows median expected p-value for different M_H.

Shows 5 sigma expected if M_H happened to be between $\sim$ 140 and 400 GeV

At 125 GeV would have expected 4.5 sigma - but ATLAS got (a bit) lucky.

Setting an upper limit using simple counting

You do a search and there is no discovery or anything exciting. Don't whinge. This is (a) quite common and (b) useful science.

You expect 2.30 background events and you see 3

What can you say?

There is no evidence for any signal - indeed, if there is a signal it's small.
To say something useful we use the same language as for discovery, but the null hypothesis is now that there is some signal s^+.
If (say) $s^+ = 5.60$ then the total is 7.90, and the probability of seeing a signal this small or smaller is $\sum_{r=0}^{3} P(r; 7.9) = 0.045$.
The statement "If $s \geq 5.60$ then you would see more than 3 events" is true 95.5% of the time.
From our 3 events we can say $s <= 5.6$ with 95.5% confidence.
It's more helpful to pick a confidence level and then find the matching limit

Counting Limits: Some simple examples

$$\sum_{0}^{n} e^{-s^+} \frac{{s^+}^r}{r!} = 1 - CL \quad (1)$$

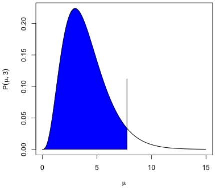
n	0	1	2	3	4	5
90% limit	2.30	3.89	5.32	6.68	7.99	9.27
95% limit	3.00	4.74	6.30	7.75	9.15	10.51

Handy fact: If you see 0 that means total ≤ 3 at 95%. CL

In the previous example we could use $s^+ = 5.45$ for a 95% limit, or 4.38 for a 90% limit, or...

The Bayesian version

Prior pdf $\mathcal{P}(s)$
Observe n events. Suppose (for now) $b = 0$.
Posterior pdf $\mathcal{P}(s|n) \propto P(n,s)\mathcal{P}(s)$.
Fix constant by normalising to 1.
From posterior select credible intervals
(analogous to confidence regions)



Suppose $\mathcal{P}(s)$ is constant and you want a 95% upper limit
Posterior $\mathcal{P}(s|n) = e^{-s}\frac{s^n}{n!}$
Require $0.95 = \int_0^{s^+} e^{-s}\frac{s^n}{n!}\,ds$
Integration by parts gives
$\left[-e^{-s}\frac{s^n}{n!}\right]_0^{s^+} + \int_0^{s^+} e^{-s}\frac{s^{n-1}}{(n-1)!}\,ds = -e^{-s^+}\frac{s^{+n}}{n!} + \int_0^{s^+} e^{-s}\frac{s^{n-1}}{(n-1)!}\,ds$
$= 1 - \sum_0^n e^{-s^+}\frac{s^{+r}}{r!}$ Same as Equation 1

So frequentists and Bayesians agree on the answer even though they don't agree on the question

From numbers to physics

For limits, s^+ itself is not what matters
Branching Ratio:

$$Br = \frac{s^+}{\eta N} \quad (2)$$

Cross section

$$\sigma = \frac{s^+}{\eta L} \quad (3)$$

η is the efficiency, N is the total number. L is the integrated luminosity
Other quantities (Masses, couplings...) obtained through formula for $\sigma(M)$ etc and Equations 2 or 3
Two not-very-complicated complications:

1. η may vary with M etc. So may the cuts, and thus the value of s^+
2. If two parameters involved, you get contour plot.

The Low data problem

Suppose $b = 4.30$ and $n = 1$. What do you do?
You check the calculation of b but it really is correct
Table gives 90% upper limit on $(s + b)$ as 3.89. So quote $s^+ = -0.41$
This is clearly crazy

Table gives 95% upper limit on $(s + b)$ as 4.74. So quote $s^+ = 0.44$
This is clearly very shaky. It's a very good result from rather poor data

This happens! If there really is no signal, Poisson predicts $n < b$ about half the time.

In a sense this is not a problem

10% of your 90% CL statements are allowed to be wrong.

In a sense it is

It's absurd

A question and 3 answers

Example: Given $n = 3$ observed events, and an expected background of $b = 3.4$ events, what is the 95% upper limit s^+?

Methid 1– Frequentist: 7.75 - 3.40 = 4.35

Method 2 – Bayesian: Assign a uniform prior to s, for $s > 0$, zero for $s < 0$.

The posterior is then just the likelihood, $P(s|n,b) = e^{-(s+b)}\frac{(s+b)^n}{n!}$

Required Limit from integrating $\int_0^{s^+} P(s|n,b)\,ds = 0.95$

$P(s) \propto e^{-(s+3.4)}\frac{(s+3.4)^3}{3!}$

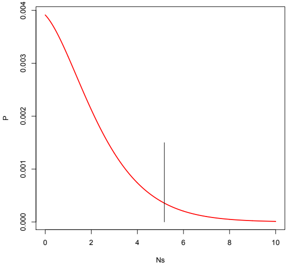
$$0.95 = \frac{\int_0^{s^+} e^{-(s'+3.4)}\frac{(s'+3.4)^3}{3!}\,ds'}{\int_0^{\infty} e^{-(s'+3.4)}\frac{(s'+3.4)^3}{3!}\,ds'}$$

Integrate by parts as before

$$0.95 = 1 - \frac{\sum_0^3 P(r;s^+ + 3.4)}{\sum_0^3 P(r;3.4)}$$

(The Helène Formula)

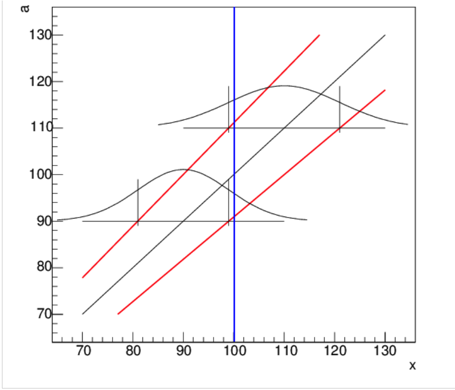
Limit is 5.21



Necessary diversion: Neyman Confidence Belts

How to use a measurement x give a confidence region for a parameter a when the pdf $P(x|a)$ is not a simple constant Gaussian.

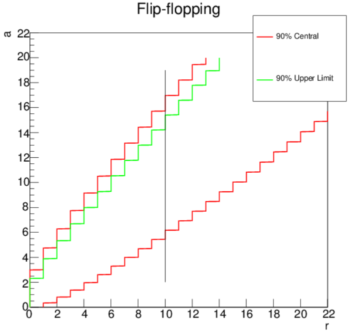
Constructed horizontally and read vertically



For all values of parameter a, construct a confidence region for result x at the desired CL, with the desired strategy (central, upper limit, ...).
The statement 'the result lies within the belt' is true with probability CL
Then when you get a result, read off the limit(s) on a
The curve for the lower x limit gives the upper a limit, and vice versa

Method 3: Feldman-Cousins 1: Motivation

The Unified Approach



In principle, can use 90% central or 90% upper limit, and the probability of the result lying in the band is at least 90%.

In practice, you would quote an upper limit if you get a low result, but if you get a high result you would quote a central limit. Flip-flopping.

Break shown here for $r = 10$

Confidence belt is the green one for $r < 10$ and the red one for $r \geq 10$.

Probability of lying in the band no longer 90%. Undercoverage. Method breaks down if used in this way

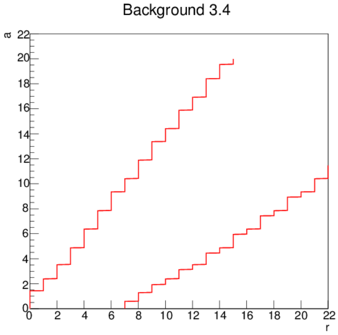
Method 3: Feldman-Cousins 2: Method

Plot $r \equiv n$ horizontally as before, but s vertically. So different $b \rightarrow$ different plot. Probability values $P(r; s) = e^{-(s+b)} \frac{(s+b)^r}{r!}$

For any s have to define region R such that $\sum_{r \epsilon R} P(r; s) \geq 90\%$.

First suggestion: rank r by probability and take them in order (would give shortest interval)
Drawback: outcomes with $r << b$ will have small probabilities and all s will get excluded. But such events happen - want to say something constructive, not just 'This was unlikely'

Better suggestion: For each r, compare $P(r; s)$ with the largest possible value obtained by varying s. This is either at $s = r - b$ (if $r \geq b$) or 0 (if $r \leq b$). Rank on the ratio



Flip-flopping incorporated! Coverage is correct.
For $r = 3$ get limit 4.86

Have to re-compute confidence belt specifically for each background number. Not a problem.

Method 3: Feldman-Cousins 4: Discussion

There are two arguments raised against the method
It deprives the physicist of the choice of whether to publish an upper limit or a range. Could be embarrassing if you look for something weird and are 'forced' to publish a non-zero result. *But isn't this the point?*

If two experiments with different b get the same small n, the one with the higher b will quote a smaller limit on b. The worse experiment gets the better result!
But for an event with large background to get a small number of events is much less likely.

Summary so far

Given 3 observed events, and an expected background of 3.4 events, what is the 95% upper limit on the ‘true’ number of signal events?

Answers:

Strict Frequentist	4.35
Bayesian (uniform prior)	5.21
Feldman-Cousins	4.86

Take your pick!

All are correct. (Well, not wrong.)

Golden Rule

Say what you are doing, and if possible give the raw numbers

Setting an upper limit from a fitted signal strength

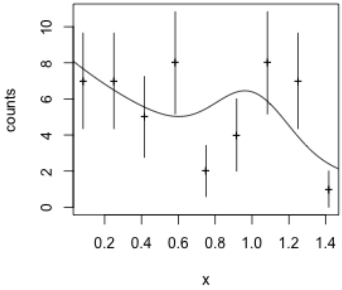
Same algebra as for a discovery

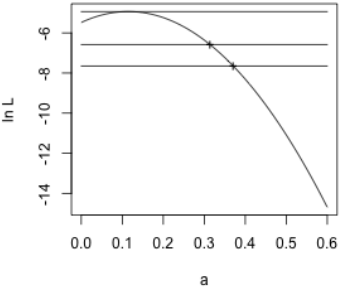
Use $t(s^+) = -2\Delta \ln L(\hat{s})/L(s^+)$

Given by χ^2 for 1 degree of freedom(WIlks' theorem)

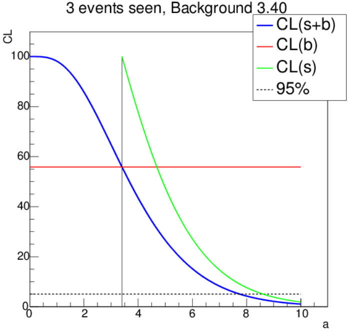
For(say) 90% confidence we want Z=1.28, and $\chi^2 \equiv Z^2 = 1.64$, and this data gives $s^+ = 0.314$

For 95% confidence, need $\Delta \ln L = 2.71$, and get $s^+ = 0.371$





From the Helène Formula to CL_s



CL_{s+b}: Probability of getting a result this small (or less) from $s+b$ events. Same as strict frequentist.

CL_b: CL_{s+b} for $s=0$ - no signal, just background

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

Apply as if confidence level $1 - CL_s$

Result larger than strict frequentist ('conservative') ('over-covers')
In our example 8.61 for $s+b$, 5.21 for s

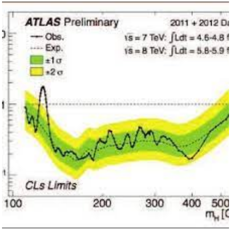
The "CL_s method" applies this to confidence levels from likelihoods

Expected performance

Brazil flag plots

Same rationale as for discovery

But maths is different
For full details see Cowan, Cranmer, Gross & Vitells EPJC **71**, 1554, (2011)



1. Choose a confidence level - 90% or 95% or whatever - and set Z
2. Assume that there is no signal.
3. Scan over the mass (or whatever)
    1. Ask what is the average (median) upper limit I expect to set
    2. Find the $\pm 1\sigma$ and $\pm 2\sigma$ values by changing Z by ± 1 and ± 2
4. Colour it green and yellow
5. Add the limits you get from the data (optional)

Conclusions

- Claiming discoveries and setting limits are linked
- but different
- Claiming a discovery means establishing a small $p-$value, usually translated into Z-sigma significance. You see a large signal, which is unlikely to be an upward fluctuation from the null hypothesis $s = 0$
- In setting a limit, you see a small signal, which is unlikely to be a downward fluctuation from the null hypothesis $s = s^+$, and you adjust s^+ to get a desired p-value (typically 0.10 or 0.05).
- Many analyses are based on counting numbers and Poisson statistics
- Many analyses are more sophisticated, not just counting numbers but looking at the signal/background nature of events and fitting
- A successful search may be worth a Nobel rprize. A successful limit will be worth a PhD.