Setting Limits and Making Discoveries (I) Counting Experiments and Poisson Statistics

Roger Barlow
Huddersfield University

Aachen Online Statistics School

17^{th} March 2023

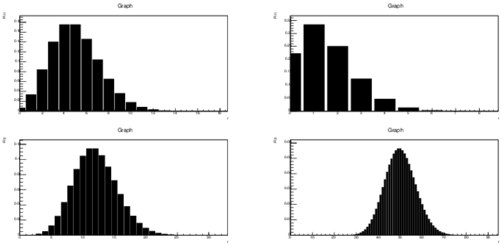
Reminder(1): The Poisson Distribution

Number of events occurring at random rate μ

$$P(r;\mu) = e^{-\mu}\frac{\mu^r}{r!}$$

As r increases $P(r;\mu)$ rises until r gets past μ, then it falls.

Poisson distributions for
(1) $\mu = 5$
(2) $\mu = 1.5$
(3) $\mu = 12$
(4) $\mu = 50$



Mean μ, Variance $V = \mu$, Standard Deviation $\sigma = \sqrt{\mu}$
Positive skew: upward fluctuations larger than downward fluctuations
Tends to Gaussian as μ becomes large

Reminder (2) Frequentist probability

Can't talk about the probability for a particular value of a quantity.
Can make statements about the probability of statements for particular values. (Confidence level statements).

CAN'T SAY

50% probability of rain tomorrow

$M_H = 125.18 \pm 0.16$ so there's a 68% probability $125.02 \leq M_H \leq 125.34$

CAN SAY:

The statement 'It will rain tomorrow' has a 50% probability of being true

$M_H = 125.18 \pm 0.16$ so if I assume $125.02 \leq M_H \leq 125.34$ I have a 68% probability of being correct.

Reminder (3) Confidence

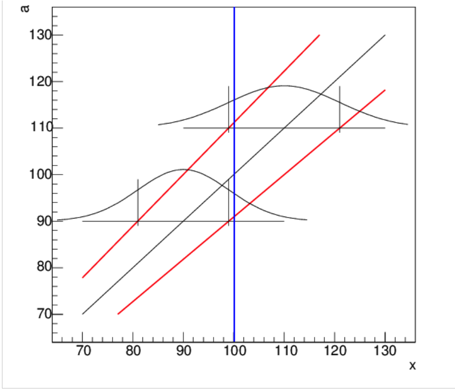
" X is true, with 95% confidence" means "X belongs to a collection of statements, of which at least 95% are true"
The 'at least' is there so that:

- Higher confidence levels embrace lower confidence levels. If it's true with 95% confidence then it's true with 90% confidence.
- Discrete data can be treated. If $\mu = 1.5$ then at 95% confidence there will not be more than 4 events. Actually this is true with 98% confidence - but 3 events gives only 93%.
- Composite hypotheses can be discussed. If $\mu \leq 1.5$ then at 95% confidence there will not be more than 4 events...

The actual probability is the coverage which depends on the hypothesis but must be at least as great as the confidence level.

Confidence Belts

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

The Overall Picture

n events pass your cuts, but you expect an average background b

What can you say?

If $n >> b$ - Discovery?

Example: $b = 4.3, n = 16$

$\sum_{16}^{\infty} P(r; 4.3) = 0.0012\%$ ($p-$value)
Under H_0 (there is no signal) the probability of getting a signal this large is only 0.0012%.
You say with 99.9988% confidence that pure background would not give this big a signal.
Or, equivalently, that it has 4.2 sigma significance (using 1 sided Gaussian)

If $n \sim b$ - Set a Limit

Example: $b = 4.3, n = 5$

Signal s must be pretty small
Choose CL and find limit s^+ for which $\sum_0^n P(r; b + s^+) = 1 - CL$
$\sum_0^5 P(r; 10.5) = 0.05$. If s is 6.2 or more the probability if getting 5 events or less is only 5%
Under H_0 ($s = s^+$) the probability of getting a signal this small is only 5%
With 95% confidence: s is not more than $s^+ = 6.2$
Or at 90%, $s \leq 5.0$ Or...

Discoveries: why 5 sigma?

Conventionally require 5 sigma to announce discovery
$p-$value 3×10^{-7}
Seems unduly cautious.
Reasons

1. The 'look elsewhere' effect. With many bins in many histograms plotted by many hard-working physicists, lots of low-probability results will be found. Blind analysis helps keep us honest.
2. Minor under-estimation of an error can lead to inflation of the significance
3. We have learnt the lessons of history! The Digamma is only the most recent in a long line of peaks that went away when more data was taken

Limits: Some simple examples

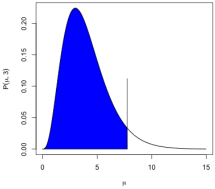
$$\sum_{0}^{n} e^{-s^+} \frac{s^{+r}}{r!} = 1 - CL \tag{1}$$

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

See 0 means 3 at 95%. CL

The Bayesian version

No conceptual problems
Prior pdf $\mathcal{P}(s)$
Observe n events. Ignore b for now...
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

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^+?

Frequentist: 7.75 - 3.40 = 4.35

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

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

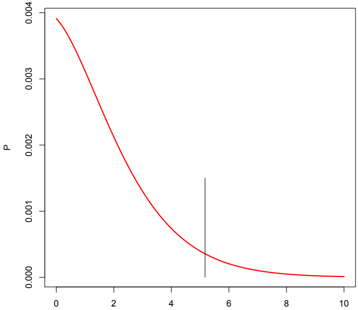
$$0.95 = \frac{\int_0^{s^+} e^{-(s'+3.4)} \frac{(s'+3.4)^3}{3!}\, ds'}{\int_0^{\infty} e^{-(s'+3.4)} \frac{(s'+3.4)^3}{3!}\, ds'}$$

Integrate by parts as before

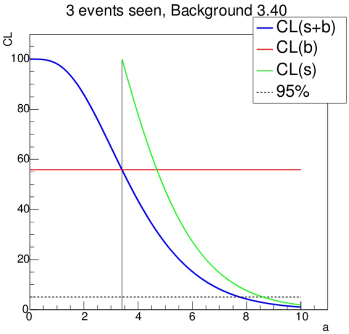
$$0.95 = 1 - \frac{\sum_0^3 P(r; s^+ + 3.4)}{\sum_0^3 P(r; 3.4)}$$

(The Helène Formula)

Limit is 5.21



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

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

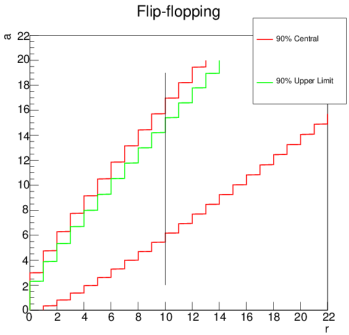
Apply as if confidence level $1 - CL_s$

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

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

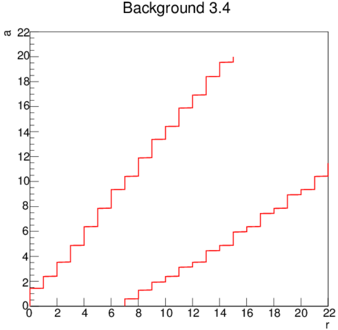
Plot $r \equiv n$ horizontally as before, but s vertically. So different $b \rightarrow$ different plot. Probability values $P(r;s) = e^{-(s+b)}\frac{(s+b)^r}{r!}$

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

Method 3: Feldman-Cousins 3: Example



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

From numbers to physics

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

$$Br = \frac{s^+}{\eta N} \tag{2}$$

Cross section

$$\sigma = \frac{s^+}{\eta L} \tag{3}$$

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.

Including Systematic Uncertainties

So far this has all been about statistical errors on n
Also (systematic) errors on b, η, L, N etc. (nuisance parameters)
Cousins and Highland: integrate over Gaussian σ
analytically/approximately or numerically.
(This is a hybrid frequentist-Bayes approach, but no-one worries)

Sometimes appropriate to use profile likelihood $\mathcal{L}(s; d, \hat{\hat{b}})$

Conclusions

- Claiming Discoveries and setting limits are linked
- but different
- Claiming a discovery means establishing a small $p-$value, usually translated into N-sigma significance. Please use blind analysis, beware the Look Elsewhere Effect and the Prosecutor's fallacy.
- Many analyses are based counting numbers and Poisson statistics (this lecture)
- Many analyses are more sophisticated, not just counting numbers but looking at the signal/background nature of events (next lecture)