Advanced Statistics 2

Setting upper limits

Roger Barlow

Huddersfield University

September 17, 2013

The general scenario

An experiment is searching for

- Charged lepton flavour violation
- A particular supersymmetry decay channel
- Neutrinoless double beta decay
- *insert your topic here*

It sees nothing. Or perhaps a signal so small it could come from background processes.
What can you say?

A basic case

to set the scene

Suppose zero background. Plan to measure number of events N. Convert to physics R (cross-section, decay-rate, wimp density...) by $R = AN$

Run experiment - get $N = 0$

Poisson formula is $P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$

What can you say about λ? Large λ implausible as $P(0,\lambda)$ very small.

Frequentist says:

I note that $P(0; 2.3) = 0.10$.

I quote $\lambda_{UL} = 2.3$ as the upper limit at 90% confidence

Bayesian says:

I take prior flat in λ: $P(\lambda) = const$.

The posterior is $P'(\lambda) = P(N;\lambda) \times const/const'$, so for 0 events, $P'(\lambda) = e^{-\lambda}$

Most likely value is 0, but I note that $\int_0^{2.3} e^{-\lambda}\, d\lambda = 0.9$.

I quote $\lambda_{UL} = 2.3$ as the upper limit at 90% confidence

More details

on the Frequentist

They said:

"I note that $P(0; 2.3) = 0.10$. I quote $\lambda_{UL} = 2.3$ as the upper limit at 90% confidence."

The full description runs

For a result N the $(1 - \alpha)$ upper limit λ_{UL} is chosen such that for $\lambda = \lambda_{UL}$ *or more* the probability of obtaining N *or less* is α *or less*.

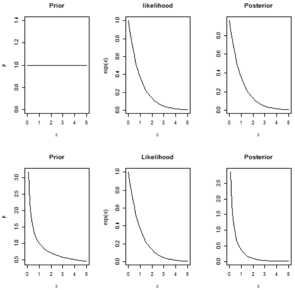
Notes on the 3 bits of small print:

1. The statement about λ_{UL} is not about a value but about the beginning of a range. 2.3 and above are ruled out.
2. If N is nonzero but small you may still want to quote a limit. Then you have to include the probability of getting an even smaller result.
3. So if a statement is true with 95% confidence, it is true with 90% confidence, etc. Also needed if the variable has discrete values.

To say "$\lambda < 2.3$@90% confidence" means: "The statement '$\lambda < 2.3$' belongs to an ensemble of statements of which (at least) 90% are true."

More details

on the Bayesian



This is easy to calculate and seems to make intuitive sense - but the 'flat prior in λ' is a fudge.
Mathematically, it can't be normalised.
In reality, you don't believe it.

There is no reason to choose λ rather than $\sqrt{\lambda}$, or maybe $cos\lambda$, as the variable to be flat in. These will give different results. A good analysis will try several priors (or, equivalently, several variables in which the prior is flat) and check that the result is not sensitive. ("Robustness under changes of prior")

More details

on both

There are other useful values for zero counts. Principally λ_{UL} @ 95% is 3.0
For any N, Frequentist and Bayesian (with prior flat in λ) quote the same values, but they mean different things. This is basically a coincidence. (It does not work for lower limits.)

Experiments with background

Things begin to get tricky

If there is background, call it B, then $N > 0$ does not mean an unambiguous discovery. You need $N >> B$ for a discovery, otherwise you quote a limit.

1. The Strict frequentist
   Observing N gives a limit T_{UL} on the total number $T = \lambda + B$.
   Subtract B to get the limit on λ.
   Works fine unless $N < B$, or even $N \approx B$
   E.G. $B = 2.9, N = 0$
   At 95%, $\lambda_{UL} = 0.1$ At 90%, $\lambda_{UL} = -0.6$
   The first is correct but dishonest. The second is ridiculous but correct.
2. The Bayesian
   No problems - integrate $\int_0^{\lambda_{UL}} e^{-\lambda+B}(\lambda + B)^N/N!\, d\lambda$
   (But remember the ambiguity of the flat prior)
3. CL_s
4. Feldman Cousins

Some technical stuff (1)

p-values

You have some data value(s) x and a hypothesis H
(H is often the null hypothesis H_0. H_0 is often the Standard Model)
The p-value is the probability, according to the hypothesis H, for getting a result as extreme as x (or worse).

Notes

1. This 'or worse' may need care. Standard example is 1-tailed and 2-tailed Gaussian significances.
2. The formula is the same as for the power of a test - but p values are computed *after* seeing the data, powers are calculated *before* (In principle, anyway)
3. This is not the probability that H is true. But try telling that to journalists...

Some technical stuff (2)

coverage

How often is a 90% CL statement true? Depends on actual value.

Example

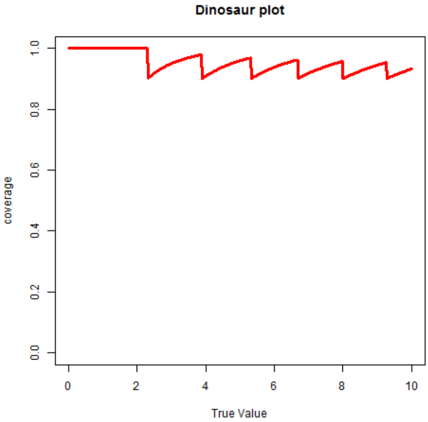
Poisson with true mean 4.5678
Probabilities of 0,1,2,3... counts are 0.010, 0.047, 0.108, 0.165 ...
Upper limits 2.30, 3.89, 5.32, 6.68... are false, false, true, true...
Statement has a 5.7% probability of being false, 94.3% of being true
Coverage 94.3% in this particular case.

The R code

```
dpois(0:10,4.5678)
f<-function(x){sum(dpois(0:N,x))-0.10}
N=3
uniroot(f,c(0,10))
```

More on coverage

Why not exactly 90%?
Because of discrete data.
For Frequentists, overcoverage is allowed but inefficient.
Undercoverage is not allowed.
For Bayesians: coverage is strictly irrelevant but gives very useful insights.



Coverage for Poisson as a function of λ

Some technical stuff (3)

Realistic experiments

So far: just count events. use Poisson statistics.

In many experiments: events may be 'signal-like' and 'background-like' in varyng degrees. (Especially with Neural Network or BDT outputs involved) Simple cut-and-count loses precision.
Use Monte-Carlo to generate likelihood functions for signal and background.

In some searches (e.g. Higgs) the parameter being studied effects both the rate and the likelihood functions.

Same limit-setting principles apply

CL_S

or 'modified frequentist', developed by Alex Reid and the LEP Higgs Hunters

Define $CL_s = CL_{s+B}/CL_B$
with $CL_{s+B} = \sum_{r=0}^{N} P(r; s+B)$ $\qquad CL_B = \sum_{r=0}^{N} P(r; B)$
Normally $CL_B = 0.9...$. Small CL_B betrays downward fluctuation.
So CL_s is bigger than CL_{s+B}, by an amount which depends sensibly on the plausibility background dipped. To reduce to 5% (or 10%...), increase λ.
Makes you more honest, but destroys the frequentist coverage.

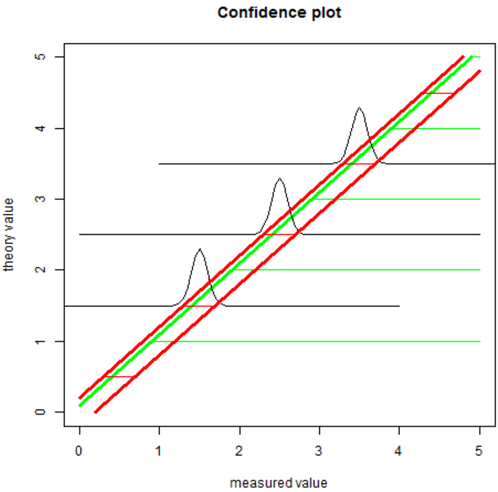
Example

Observe 3 events. Calculated background 1.2. Work with 90% limits
p value $p(\lambda) = \sum_0^3 e^{-\lambda}\lambda^r/r!$ gives 0.1 for $\lambda = 6.68$
Straight frequentist: $\lambda \equiv s + B$ so 90% upper limit on s is 5.48
Modified frequentist: $p(1.2) = 0.966$, so want $CL_s = p(\lambda)/0.966 = 0.1$
Actually $p(6.74) = 0.0966$, so limit on s is 5.54.
Oops! Background recalculated. Now 5.2. Frequentist adjusts to 1.48
$p(5.2) = 0.238$, so you want $p(\lambda) = 0.0238$, $\implies \lambda = 8.84$, and limit 3.64

Feldman-Cousins

Or: 'Unified Method'. A technique cunningly solves one problem by attacking another.

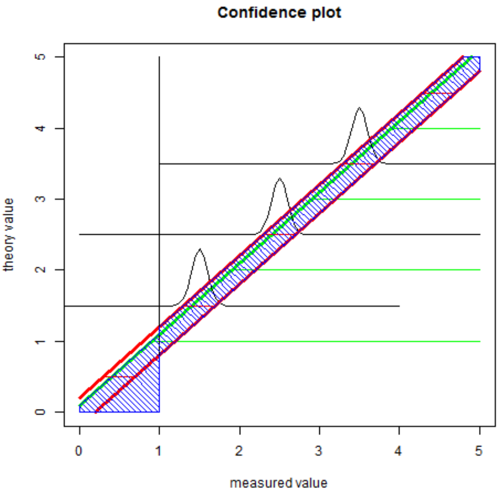
Confidence plot - 90% limits.
Horiz. axis. Measured x.
Vert. axis.True λ. Construct $p(x;\lambda)$ for each λ
Find x value for which $\int_x^\infty p(x')\,dx' = 0.9$. Green.
Find x values for which $\int_{-\infty}^x p(x')\,dx' = \int_x^\infty p(x')\,dx' = 0.05$, for central limits. Red .
Given x, get upper limit from thick green line, or central liits from thick red lines, and all is well.



Feldman-Cousins 2

the story continues

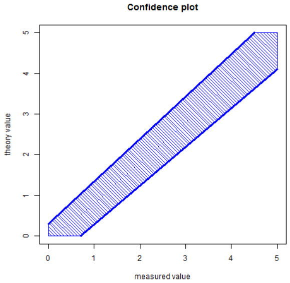
Real life practice ('flip-floppng'): If x is small (say, 1.0), quote an upper limit. For larger x quote a measurement.
(Means using the shaded area)
This undercovers and is therefore evil and wrong.



Feldman-Cousins 3

The method

Choose any limits with
$\int_{x_1}^{x_2} p(x;a)\,dx = 0.90$
First guess: choose highest probabilities till 90% reached. (Gives shortest interval).
Minor glitch: some values - e.g. $P(0;3.2)$ unlikely and never get chosen - even though you would want to do so if, say, 0 events and $B = 3.1$.
Second guess: For a given a, rank values of x according to $P(x;a)/P(x,a_{best})$ and choose the highest ranked till 90% reached.
For Poisson $p(x;s+B)$, : s_{best} is either $x - B$, or 0

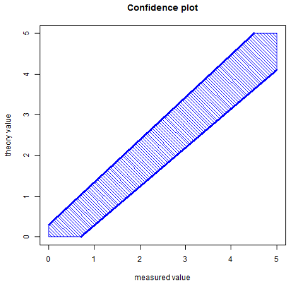


Feldman-Cousins 4

How it works

For small x, get an upper limit.
For larger x, get range.
Both are good frequentist resulrs.
'Flip-flop' is automatic.
Need to calculate limits anew for each B.
Not a problem.
Objections raised

1. May quote range when you don't believe there's a real signal.
   You can live with it!
2. For zero events, experiments with larger backgrounds quote better limits.
   so what?

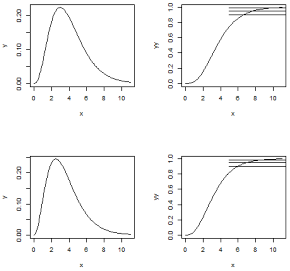


Exercise

Your experiment detects 3 events.
Calculate the 90%, 95% and 99% upper limits using (1) a frequentist approach (2) A Bayesian approach wth a prior flat in λ and (3) A Bayesian approach wth a prior flat in $\sqrt{\lambda}$

Answer...

Frequentist: define `f <- function(x){ppois(3,x)-.1}` then `uniroot(f,c(1,11))` Get 6.68, 7.75, 10.05 for 90,95,99%
Bayes: (normalised) posterior is $e^{-\lambda}\lambda^3/6$
Can integrate algebraically - same as frequentist. Or read off graph



Bayes with prior flat in $\sqrt{\lambda}$ - proportional to $1/\sqrt{\lambda}$
Plot and read off - approx 6.0,.7.0,.9.1