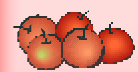
# Statistics for HEP

Roger Barlow
Manchester University

## Lecture 4: Confidence Intervals

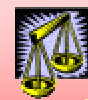
## The Straightforward Example

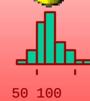


Apples of different weights

Need to describe the distribution

$\mu$ = 68g  $\sigma$ = 17 g





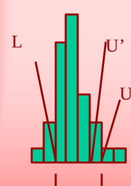

All weights between 24 and 167 g (Tolerance)

90% lie between 50 and 100 g
94% are less than 100 g
96% are more than 50 g

} Confidence level statements

Slide 2

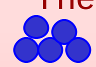
## Confidence Levels




- Can quote at any level (68%, 95%, 99%...)
- Upper or lower or twosided (x<U  x<L  L<x<U)
- Two-sided has further choice (central, shortest...)

Slide 3

## The Frequentist Twist

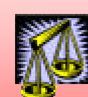


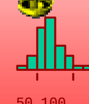
Particles of the same weight

Distribution spread by measurement errors

What can we say about M?

$\mu$ = 68  $\sigma$ = 17






"M<90" or "M>55" or "60<M<85" @90% CL

These are each always true or always false

Solution: Refer to ensemble of statements

Slide 4

## Frequentist CL in detail

You have a meter: no bias, Gaussian error 0.1.

For a value $X_T$ it gives a value $X_M$ according to a Gaussian Distribution

$X_M$ is within 0.1 of $X_T$ 68% of the time

$X_T$ is within 0.1 of $X_M$ 68% of the time

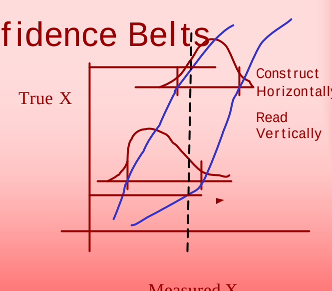
Can state $X_M-0.1<X_T<X_M+0.1$ @68% CL

Slide 5

## Confidence Belts

For more complicated distributions it isn't quite so easy

But the principle is the same




Slide 6

## Coverage

"L<x<U" @ 95% confidence
(or "x>L" or "x<U")

This statement belongs to an ensemble of similar statements of which at least* 95% are true

95% is the *coverage*

This is a statement about U and L, not about x.

*Maybe more. Overcoverage
EG composite hypotheses.
Meter with resolution < 0.1

Think about: the difference between a 90% upper limit and the upper limit of a 90% central interval.

Slide 7

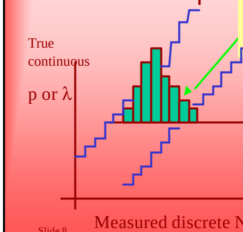
## Discrete Distributions

CL belt edges become steps

May be unable to select (say) 5% region

Play safe.

Gives overcoverage




Binomial: see tables

Poisson

| | | Upper | | | Lower | |
|---|---|---|---|---|---|---|
| | 90 | 95 | 99 | 90 | 95 | 99 |
| 0 | 2.3 | 3 | 4.61 | | | |
| 1 | 3.89 | 4.74 | 6.64 | 0.11 | 0.05 | 0.01 |
| 2 | 5.32 | 6.3 | 8.41 | 0.53 | 0.36 | 0.15 |
| 3 | 6.68 | 7.75 | 10.05 | 1.1 | 0.82 | 0.44 |

*Given 2 events, if the true mean is 6.3 (or more) then the chance of getting a fluctuation this low (or lower) is only 5% (or less)*

Slide 8

## Problems for Frequentists

Weigh object+container with some Gaussian precision

Get reading R

R-σ < M+C < R+σ @68%

R-C-σ < M < R-C+σ @68%

E.g. C=50, R=141, σ=10

81<M<101 @68%

E.g. C=50, R=55, σ=10

-5 < M < 5 @68%

E.g. C=50, R=31, σ=10

-29 < M < -9 @68%

Poisson: Signal + Background

Background mean 2.50

Detect 3 events:

Total < 6.68 @ 95%

Signal<4.18@95%

Detect 0 events

Total < 2.30 @ 95%

Signal < -0.20 @ 95%

These statements are OK.
We are allowed to get 32% / 5% wrong.
But they are stupid

Slide 9

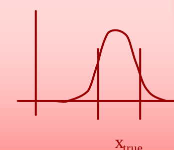
## Bayes to the rescue

$$P(Theory|Data) = \frac{P(Data|Theory)}{P(Data)} P(Theory)$$

Standard (Gaussian) measurement

- No prior knowledge of true value
- No prior knowledge of measurement result
- P(Data|Theory) is Gaussian
- P(Theory|Data) is Gaussian

Interpret this with Probability statements in any way you please




Gives same limits as Frequentist method for simple Gaussian

Slide 10

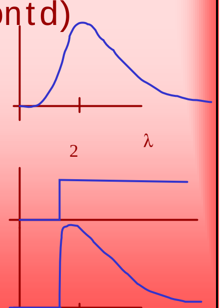
## Bayesian Confidence Intervals (contd)

Observe (say) 2 events
$P(\lambda;2) \propto P(2;\lambda) = e^{-\lambda} \lambda^2$

Normalise and interpret

If you know background mean is 1.7, then you know $\lambda > 1.7$

Multiply, normalise and interpret




Slide 11

## Bayes: words of caution

Taking prior in λ as flat is not justified

Can argue for prior flat in ln λ or $1/\sqrt{\lambda}$ or whatever

Good practice to try a couple of priors to see if it matters

Slide 12

## Feldman-Cousins Unified Method

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:*

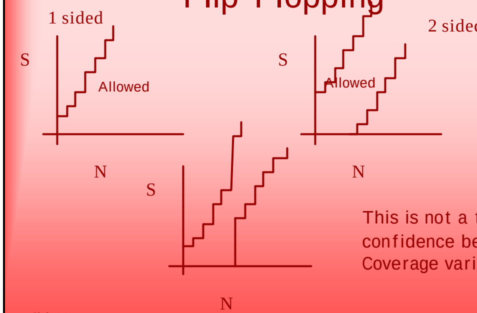
*You have a background of 3.2*

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

*Observe 25 events? Quote two-sided limits*

Slide 13

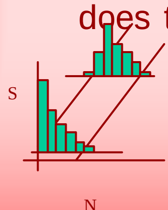
## "Flip-Flopping"




This is not a true confidence belt! Coverage varies.

Slide 14

## Solution: Construct belt that does the flip-flopping




For 90% CL

For every S select set of N-values in belt

Total probability must sum to 90% (or more): there are many strategies for doing this

Crow & Gardner strategy (almost right):

Select N-values with highest probability → shortest interval

Slide 15

## Better Strategy

N is Poisson from S+B

B known, may be large

E.g. B=9.2,S=0 and N=1

P=.1% - not in C-G band

But any S>0 will be worse

Fair comparison of P is with best P for this N

Either at S=N-B or S=0

To construct band for a given S:

For all N:

Find P(N;S+B) and

$P_{best}$=P(N;N) if (N>B) else P(N;B)

Rank on $P/P_{best}$

Accept N into band until $\Sigma$ P(N;S+B) ≥90%

Slide 16

## Feldman and Cousins Summary

- Makes us more honest (a bit)
- Avoids forbidden regions in a Frequentist way

- Not easy to calculate
- Has to be done separately for each value of B
- Can lead to 2-tailed limits where you don't want to claim a discovery
- Weird effects for N=0; larger B gives lower (=better) upper limit

Slide 17

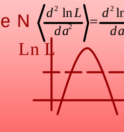
## Maximum Likelihood and Confidence Levels

ML estimator (large N) has variance given by MVB $s_{\hat{a}}^2 = V(\hat{a}) = \frac{-1}{\left\langle \frac{d^2 \ln L}{da^2} \right\rangle}$

At peak $\ln L \approx L_{max} + \frac{(a-\hat{a})^2}{2} \left. \frac{d^2 \ln L}{da^2} \right|_{a=\hat{a}}$ For large N $\left\langle \frac{d^2 \ln L}{da^2} \right\rangle = \left. \frac{d^2 \ln L}{da^2} \right|_{a=\hat{a}}$

Ln L is a parabola (L is a Gaussian)

$$\ln L = L_{max} - \frac{(a-\hat{a})^2}{2 s_{\hat{a}}^2}$$




Falls by ½ at $a = \hat{a} \pm s_{\hat{a}}$

Falls by 2 at $a = \hat{a} \pm 2 s_{\hat{a}}$

Slide 18

Read off 68% , 95% confidence regions

## MVB example

N Gaussian measurements: estimate μ

$$P(x_i;\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x_i-\mu)^2/2\sigma^2}$$

Ln L given by $-\sum_i \frac{(x_i-\mu)^2}{2\sigma^2} - N\ln\left(\sigma\sqrt{2\pi}\right)$

Differentiate twice wrt μ $\quad -\frac{N}{\sigma^2}$

Take expectation value – but it's a constant

Invert and negate: $V(\hat{\mu})=\frac{\sigma^2}{N}$

Slide 19

## Another MVB example

N Gaussian measurements: estimate σ

Ln L still given by $-\sum_i \frac{(x_i-\mu)^2}{2\sigma^2} - N\ln\left(\sigma\sqrt{2\pi}\right)$

Differentiate twice wrt σ

$$-\sum_i \frac{3(x_i-\mu)^2}{\sigma^4}+\frac{N}{\sigma^2}$$

Take expectation value $\langle(x_i-\mu)^2\rangle=\sigma^2\ \forall\ i$

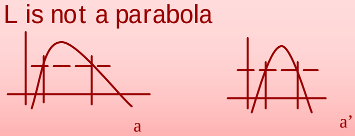
Gives $-\frac{2N}{\sigma^2}$

Invert and negate: $V(\hat{\sigma})=\frac{\sigma^2}{2N}$

Slide 20

## ML for small N

ln L is not a parabola




Argue: we could (invariance) transform to some a' for which it is a parabola

We could/should then get limits on a' using standard $L_{max}$-½ technique

These would translate to limits on a

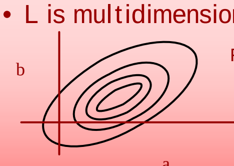
These limits would be at the values of a for which L= $L_{max}$-½

So just do it directly

Slide 21

## Multidimensional ML

- L is multidimensional Gaussian




For 2-d 39.3% lies within 1σ i.e. within region bounded by $L=L_{max}$-½

For 68% need $L=L_{max}$-1.15

Construct region(s) to taste using numbers from integrated $\chi^2$ distribution

Slide 22

## Confidence Intervals

- Descriptive
- Frequentist
  - Feldman-Cousins technique
- Bayesian
- Maximum Likelihood
  - Standard
  - Asymmetric
  - Multidimensional

Slide 23