Uncertainties and Errors

Roger Barlow
The University of Huddersfield

Terascale Statistics School, DESY, Hamburg

4^{th} April 2024

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

What do we mean by 'the error'?

Not what an engineer means
1200 ± 1 mm refers to the **tolerance**.
Guaranteed to be within this range
(Essential if you're building something)
This use lingers if you ask experts to estimate errors
and leads to a sneaking idea that errors should be added linearly

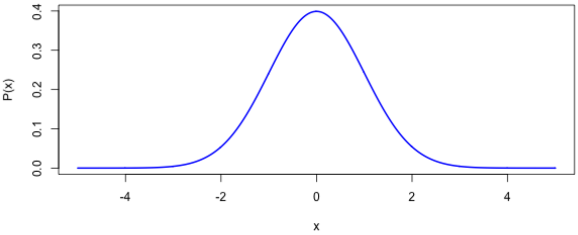
Not what a statistician means
They write $y_i = mx_i + c + \epsilon_i$
The error is the difference between truth and prediction
Different language - hence differences and similarities between statistician's "regression" and physicist's "straight line fit".

Physicist's "error" is some sort of "most probable statistician's error"

The Gaussian (or Normal) distribution

By 'the error is σ' we mean: *Measured values x are distributed about the true value a with a Gaussian pdf*

$$G(x,a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

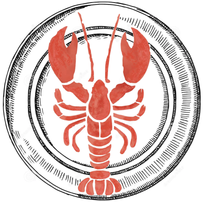


This rests on the Central Limit Theorem: if N uncertainties are combined the result tends to a Gaussian for large N, and our 'errors' are due to multiple causes.
Obviously problematic if N not large - but even before...

Example: simple weighing

Simple subtraction

A balance has an error of 5 g. It gives the weight of an object plus container as 234 ± 5 g. The empty container weighs 30 ± 5 g. So the weight of the object is 204 ± 7 grams

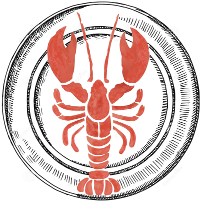


Interpretation: it may well be within 197 and 211 g, is fairly certain to be between 190 and 218 g, and is virtually sure to be between 183 and 225 g.

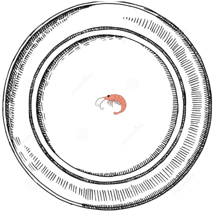
Example: simple weighing

Simple subtraction

A balance has an error of 5 g. It gives the weight of an object plus container as 234 ± 5 g. The empty container weighs 30 ± 5 g. So the weight of the object is 204 ± 7 grams



Interpretation: it may well be within 197 and 211 g, is fairly certain to be between 190 and 218 g, and is virtually sure to be between 183 and 225 g.



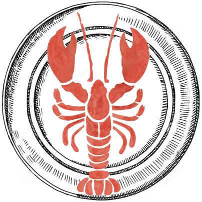
Simple subtraction

A smaller object, with the same container, weighs 34 ± 5 g. The same logic and arithmetic says it weighs 4 ± 7 g. How do interpret this?

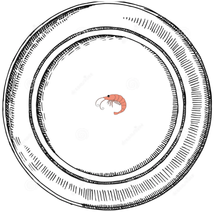
Example: simple weighing

Simple subtraction

A balance has an error of 5 g. It gives the weight of an object plus container as 234 ± 5 g. The empty container weighs 30 ± 5 g. So the weight of the object is 204 ± 7 grams



Interpretation: it may well be within 197 and 211 g, is fairly certain to be between 190 and 218 g, and is virtually sure to be between 183 and 225 g.



Simple subtraction

A smaller object, with the same container, weighs 34 ± 5 g. The same logic and arithmetic says it weighs 4 ± 7 g. How do interpret this?

This is quite plausible. Indeed, measurement might have been 29 ± 5 g
Similar (Poisson) example: number of signal+background events exceeds background estimate

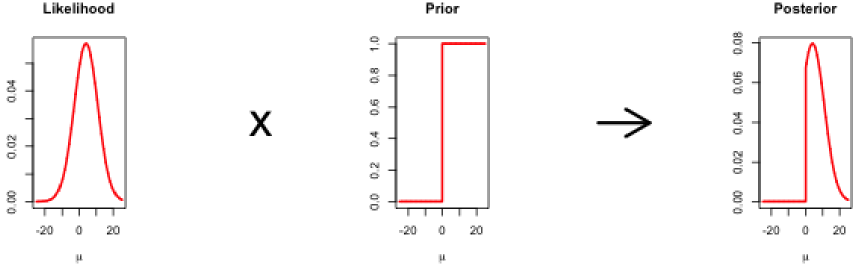
Bayesian Interpretation

No problems!

"A balance has an error of 5 g" means for a measurement m of a true value μ, $P(m|\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2}$ with $\sigma = 5g$

Bayes theorem: $\mathcal{P}'(\mu|m) \propto P(m|\mu)\mathcal{P}(\mu)$

Take $\mathcal{P}(\mu)$ as constant for all $\mu > 0$[1] and don't worry about the normalisation. Gives $\mathcal{P}'(\mu) \propto G(m, \mu, \sigma)$ for $\mu > 0$, else 0



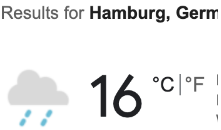
From this you can obtain the probability for m being in any given region

[1]Other priors are available

Frequentist Version

Go back a few steps

"It will rain here tomorrow with 90% probability" is a nonsense statement. There is only one tomorrow. It will either rain or not. P is either 0 or 1.



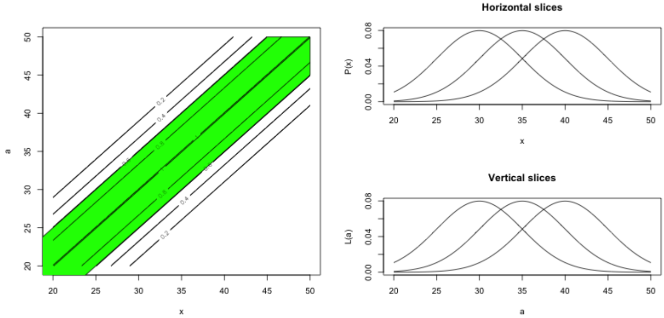
"The forecast predicts rain here tomorrow, and these forecasts are 90% accurate" is admissible. Then the statement "It will rain here tomorrow" has a 90% probability of being true.

We say "It will rain here tomorrow with 90% confidence". The core statement is either true or false, but it belongs to an ensemble of statements of which at least 90% are true.

"A balance has an error of 5 g" means for a measurement m of a true value μ, $P(m|\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2}$ with $\sigma = 5g$. We say with 68% confidence for any measured m that $m - 5 < \mu < m + 5$.

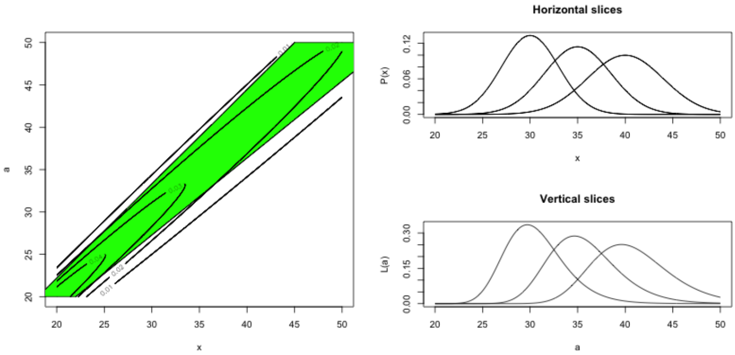
Confidence belts

Neyman Construction



Green area shows the 68% central confidence region, for a Gaussian with resolution 5.0. (x is measured value, a is true value.)
The statement "the result will lie in the green region" has a 68% probability of being true.
We say our result lies in the green region with 68% confidence, and for a measured x that gives a_{lo}, a_{hi}

The Proportional Gaussian



Measurement is accurate to 10% - quite plausible. $\sigma = 0.1a$

A measurement of, x=100 could have come from 90 ± 9 of 110 ± 11.

Equal distances but different probabilities (just over 1 sigma, just under one sigma)

Likelihoods - vertical slices - are asymmetric

Confidence belts- limits

Neyman construction

Drawing confidence belts you have a choice of both value (68%, 90%, 95%,...) and strategy (central, shortest, upper bund, lower bound, etc)

Uncertainties and limits are all part of the same picture

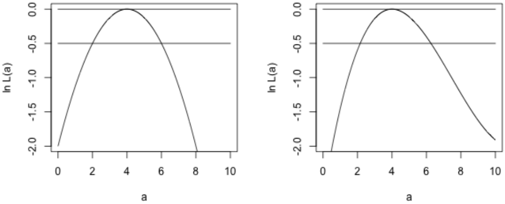
Can choose your strategy to change from 1-sided upper bound to 2-sided measurement if you like (Feldman-Cousins)

This strategy is applied to the horizontal pdfs and results read out vertically (likelihood)

$\Delta \ln L = -\frac{1}{2}$ errors

Or $\Delta\chi^2 = 1$ if you prefer

Everybody uses them



Argument: for large N the ML estimator saturates the MVB, so has variance $-\left\langle \frac{\partial^2 \ln L}{\partial a^2} \right\rangle^{-1}$. Use actual value instead of expectation value, log likelihood is, neglecting high order terms

$$\ln L = L_0 + \frac{1}{2}\frac{\partial^2 \ln L}{\partial a^2}(a - \hat{a})^2 = L_0 - \frac{1}{2}\left(\frac{a-\hat{a}}{\sigma}\right)^2$$

falls by $\frac{1}{2}$ at $\hat{a} \pm \sigma$. If $\ln L$ is not parabolic, one could always transform it to a parameter for which it was, and then back.

So at least 3 loopholes. But everybody uses them

$\Delta \ln L = -\frac{1}{2}$ for the proportional Gaussian

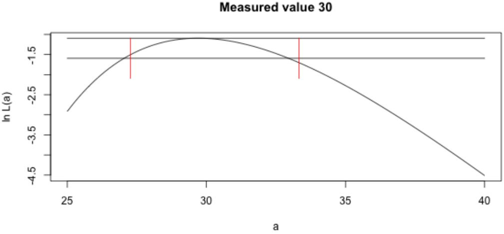
x measured from Gaussian with mean a, standard deviation $\sigma = \alpha a = 0.1a$

Suppose $x = 30$ is observed
Correct Neyman 68% central region from $\frac{x}{1+\alpha}$ to $\frac{x}{1-\alpha}$

$[27.27, 33.33]$

$\delta \ln L = -\frac{1}{2}$ gives

$[27.04, 32.95]$



Red lines show Neyman limits

Quoted errors differ in the second significant figure!

Same story for Poisson likelihood and exponential: See RB: arXiv:physics/0403046v1 for details

But everyone uses $\Delta \ln = -\frac{1}{2}$ - and you will never be criticised for doing so. So carry on.

Statistical v. Systematic Errors

Statistical	**Systematic**
Decreases with more data	More data does not help
Independent between results	Correlated between results
The main experimental result	Often from 'ancillary expt'.
Miscalculation → bad χ^2	χ^2 no help

ATLAS data results in a Higgs boson mass measurement of $125.22 \pm 0.11(\text{stat.}) \pm 0.09(\text{syst.})$ GeV. W

CMS: $m_{\text{H}} = 125.35 \pm 0.12\ (\text{stat}) \pm 0.09\ (\text{syst})\ \text{GeV}$.

Why do we have to quote them separately?

We don't have to. 125.22 ± 0.14 or 125.32 ± 0.15 is quite legitimate.
Separation (i) shows how much could be gained by taking more data, and (ii) enables combination with another result that shares systematics.

Either type can be frequentist or Bayesian.
Sometimes the ancillary experiment is also the main experiment, e.g. uncertainty in some branching ratio which is measured in another channel.
What you call such errors is ambiguous but irrelevant.

This slide intentionally left blank

What if an error is asymmetric?

and so the function is not Gaussian?

$$G(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

3 questions you have to tackle (answers depend on the particular case)

1. Are you dealing with a Pdf or likelihood? The Gaussian looks the same as a function of x or as a function of a, but other functions don't
2. What do you mean by σ? The square root of the variance or boundary of the 68% central region. These are the same for the Gaussian but not otherwise.
3. What are you doing? Combining results (ATLAS says... CMS says ..., together ..). or combining errors (there is an error on the muon ID and on the kaon ID...)?

Pdfs or likelihoods?

A likelihood
Any 'error ' σ must define a confidence region as likelihoods tell you nothing about expectation values.
You are probably combining results (but combining errors is possible using profile likelihoods)
This is typical for statistical errors

A Pdf
This is typical for systematic errors
You are probably combining errors
σ is $\sqrt{V} = \sqrt{< x^2 > - < x^2 >}$.
Variances add – even for non Gaussian pdfs. (And the CLT is your friend)
(Combining results is possible, as a special case using a weighted sum)

Likelihoods

"In conics I can floor peculiarities parabolous" – W.S. Gilbert, The Major General's song

Model an approximate parabola, using σ_+ and σ_- from $\Delta \ln L = -\frac{1}{2}$

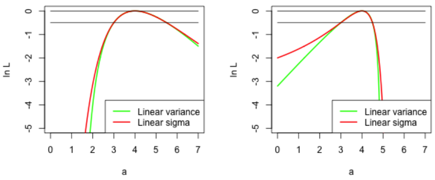
Many available. Recommend in particular

Linear-sigma : $\ln L(a) = -\frac{1}{2}\left(\frac{a-\hat{a}}{\sigma+\sigma'(a-\hat{a})}\right)^2 \qquad \sigma = \frac{2\sigma_+\sigma_-}{\sigma_++\sigma_-}, \sigma' = \frac{\sigma_+-\sigma_-}{\sigma_++\sigma_-}$

Linear-variance: $\ln L(a) = -\frac{1}{2}\frac{(a-\hat{a})^2}{V+V'(a-\hat{a})} \qquad V = \sigma_+\sigma_-, V' = \sigma_+ - \sigma_-$

Reasonable motivation. Simple relations between σ_+, σ_- and parameters.

(Can give crazy results in cases when denominator goes through zero.)



Shows fitted curves for $4.0^{+1.5}_{-1.0}$ and $4.0^{+0.5}_{-1.0}$

Go through the 3 points, by design

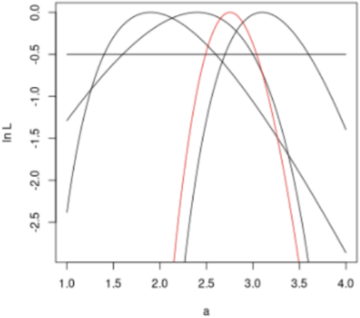
Agree very well in central region between the points, reasonably well outside

Comparisons against known likelihoods, e.g. Poisson, log of Gaussian variable, strength in signal+background, is reasonably acceptable.

Working with Likelihoods

But it is always better to give the full likelihood curves

To combine results:
If you knew the log likelihoods you would sum them and determine peak+errors (and goodness-of-fit)
If you only have the errors, you model the likelihoods, and then as before.



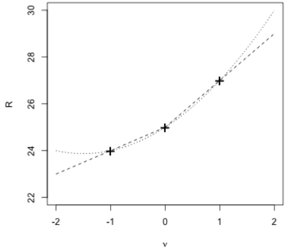
Combine $1.9^{+0.7}_{-0.5}, 2.4^{+0.6}_{-0.8}$ and $3.1^{+0.5}_{-0.4}$:
Linear variance gives $2.754^{+0.286}_{-0.263}$.
Linear sigma gives $2.758^{+0.293}_{-0.272}$
Can also check GOF

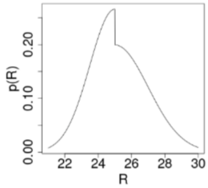
Combination-of-errors through profiling, using models of lnL function
Checks on cases with known answers do very well (see paper for details)

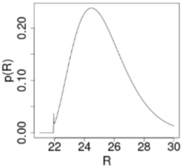
Pdfs

Model an approximate Gaussian - many available

Typical case is OPAT ("one parameter at a time") analysis of systematic errors.
With (Gaussian) nuisance parameter ν, evaluate result R at $\nu, \nu + \sigma_\nu, \nu - \sigma_\nu$
If upward and downward shifts are different, get asymmetric error on R







Two possible models, "dimidiated" and "distorted" Gaussians, by drawing 2 straight lines or one parabola through the 3 points.

The Bad Method

"To combine Asymmetric Errors, add the positive and negative errors in quadrature, separately."

Why it must be wrong

Suppose you combine N errors, all with the same σ_+ and σ_-. This method gives ${\sigma_+}^{tot} = \sqrt{N}\sigma_+, {\sigma_-}^{tot} = \sqrt{N}\sigma_-$
This has the same shape as the originals.
But The Central Limit Theorem says that at large N the shape becomes Gaussian

What's going on

Given 2 error sources, they may both fluctuate upwards, and this is described by adding σ_+ values in quadrature.
Likewise for downward fluctuations and σ_-
But they may (50% chance) have one upward and one downward fluctuation. And this is what drives the distribution towards being symmetric.

Working with Pdfs

Combination of Errors

Problem: Convolution of 2 pdfs in some class of model does not give a pdf in that class. (That's a unique property of the Gaussian)
Solution:

1. Calculate mean μ and variance V and (unnormalised) skew, $\gamma = \left\langle (x-\mu)^3 \right\rangle$ for each model.
2. Under convolution these add, so you have total μ, V, γ
3. Find model parameters corresponding to these moments

PDF modelling is more complicated than lnL modelling as you have to convert between the moments μ, V, γ, the 68% pdf quantile parameters R, σ_+, σ_-, and the actual model parameters, in both directions.

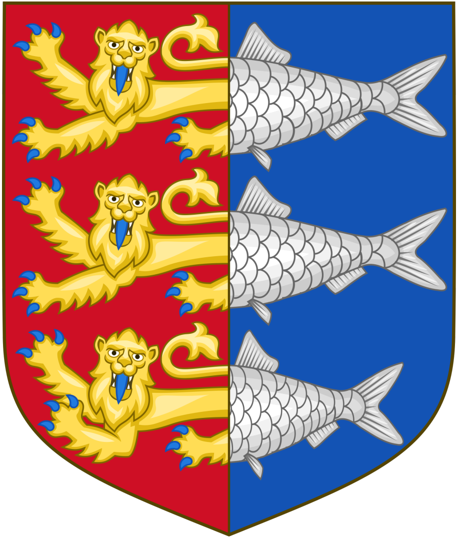
Method gives sensible results in practice - examples in forthcoming paper. Thought needed about shifts to the central value.

Conclusions

- Get acquainted with the Neyman construction way of thinking
- Don't be frightened of systematic uncertainties
- Treat $\Delta \ln L = -\frac{1}{2}$ errors with some scepticism
- Avoid asymmetric errors if at all possible
- If unavoidable, decide whether they are pdf-related rms spreads or likelihood-related quantile limits
- For likelihood-related errors, use two models and check compatibility. Plot the likelihood curves
- For pdf-related errors, use two models and check compatibility. **Do not add separately in quadrature**

Backup slides

The arms of Great Yarmouth



Per pale Gules and Azure three Lions passant guardant in pale Or dimidiated with as many Herrings naiant in pale Argent.