Systematic Errors: Facts and Fictions

Terascale Statistics School

Roger Barlow

Int. Inst. for Accelerator Applications
The University of Huddersfield

7th March 2017

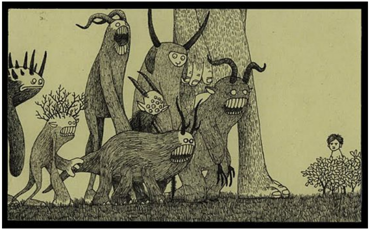
University of HUDDERSFIELD
International Institute for Accelerator Applications

Why?

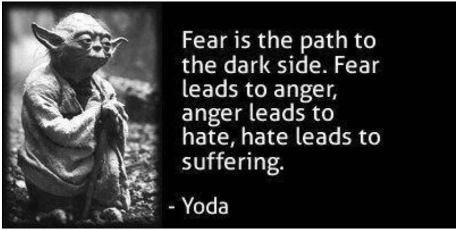
Caution! This talk contains material some people may find offensive.

There is a lot of bad practice out there. Muddled thinking and following traditional procedures without understanding.

When statistical errors dominated, this didn't matter much. In the days of particle factories and big data samples, it does.



People are ignorant - ignorance leads to fear. They follow familiar rituals they hope will keep them safe.



Outline?

- What is a Systematic Error?
- How to deal with them
- How to evaluate them
- Checking your analysis
- Conclusions and recommendations

What is a Systematic Error?

Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.

Bevington

Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.

Orear

What is a Systematic Error?

Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.

Bevington

Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.

Orear

These are contradictory

What is a Systematic Error?

Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.

Bevington

Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.

Orear

These are contradictory

Orear is RIGHT

What is a Systematic Error?

Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.

Bevington

Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.

Orear

These are contradictory

Orear is RIGHT

Bevington is WRONG

What is a Systematic Error?

Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.

Bevington

Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.

Orear

These are contradictory

Orear is RIGHT

Bevington is WRONG

So are a lot of other books and websites

An error is not a mistake

We teach undergraduates the difference between *measurement errors*, which are part of doing science, and *mistakes*.

If you measure a potential of 12.3 V as 12.4 V, with a voltmeter accurate to 0.1V, that is fine. Even if you measure 12.5 V

If you measure it as 124 V, that is a mistake.

Bevington is describing *Systematic mistakes*

Orear is describing *Systematic uncertainties* - which are 'errors' in the way we use the term.

Old usages die hard

Much work in BaBar, CDF etc to make this point. But still ....

- In 2011 the OPERA collaboration produced a measurement of neutrino travel times from CERN to Gran Sasso which appeared smaller by 6σ than the travel time of light in vacuum[15]. The effect spurred lively debates, media coverage, checks by the nearby ICARUS experiment and dedicated beam runs. It was finally understood to be due to a single large source of systematic uncertainty – a loose cable[16]

Dorigo's talk, Dec 16th

Not an uncertainty - a mistake!

Avoid using 'systematic error' and always use 'uncertainty' or 'mistake'? Probably impossible. But should always know which you mean

Examples

Track momenta from $p_i = 0.3B\rho_i$ have statistical errors from ρ and systematic errors from B

Calorimeter energies from $E_i = \alpha D_i + \beta$ have statistical errors from light signal D_i and systematic errors from calibration α, β

Branching ratios from $Br = \frac{N_D - B}{\eta N_T}$ have statistical error from N_D and systematics from efficiency η, background B, total N_T

Bayesian or Frequentist?

Can be either

Frequentist: Errors determined by an *ancillary experiment* (real or simulated)

E.g. magnetic field measurements, calorimeter calibration in a testbeam, efficiency from Monte Carlo simulation

Sometimes the ancillary experiment is also the main experiment - e.g. background from sidebands.

Bayesian: theorist thinks the calculation is good to 5% (or whatever). Experimentalist affirms calibration will not have shifted during the run by more than 2% (or whatever)

Some analysis techniques use hybrid of frequentist and Bayesian.

Beware of over-conservative experts

An error is not a tolerance

Another great fiction, c.f. Taylor *Systematic errors must be added linearly, not in quadrature.* Wrong!!!!

Experts have a tendency to be 'conservative'. Worry that they may quote an error which turns out to be smaller than the actual effect.

Remember 1/3 of quoted values should lie outside their error bars

A rose by any other name...

To define a systematic error as a ‘bias’ is basically a tautology.

Known or unknown? If known, compensate. If unknown, estimate uncertainty

‘Nuisance parameter’ is a very useful way to think of a systematic effect. One can integrate it out (Bayesian) - also known as ‘marginalisation’

Or use a profile likelihood: $\hat{L}(x; a) = L(x; a, \hat{\nu})$

How to handle them: Correlation

Actually quite straightforward. Systematic uncertainties obey the same rules as statistical uncertainties

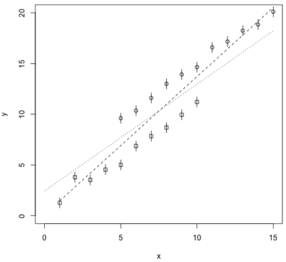
We write $x = 12.2 \pm 0.3 \pm 0.4$ but we could write $x = 12.2 \pm 0.5$.
For single measurement extra information is small.

For multiple measurements e.g. $x_a = 12.2 \pm 0.3, x_b = 17.1 \pm 0.4, all \pm 0.5$ extra information important, as results correlated.

Example: cross sections with common luminosity error, branching ratios with common efficiency ...

Application: fitting several experiments

Problem: combine datasets where data within each experiment share a systematic error



Method 1: minimise χ^2

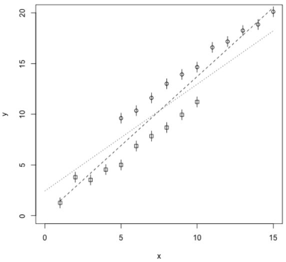
$\sum_{i,j}(y_i - f(x_i;a))V_{ij}^{-1}(y_j - f(x_j;a))$
with $V_{ij} = \delta_{ij}\sigma_i^2 + \delta_{E_i,E_j}S_{E_i}^2$

Method 2 : minimise χ^2

$\sum_i \frac{(y_i - f(x_i;a) - z_{E_i})^2}{\sigma_i^2} + \sum_E \frac{z_E^2}{S_E^2}$

More variables to minimise over - but no matrix inversion. Finds offsets z_i explicitly

Application: fitting several experiments



Problem: combine datasets where data within each experiment share a systematic error

Method 1: minimise χ^2

$\sum_{i,j}(y_i - f(x_i;a))V_{ij}^{-1}(y_j - f(x_j;a))$
with $V_{ij} = \delta_{ij}\sigma_i^2 + \delta_{E_i,E_j}S_{E_i}^2$

Method 2 : minimise χ^2

$\sum_i \frac{(y_i - f(x_i;a) - z_{E_i})^2}{\sigma_i^2} + \sum_E \frac{z_E^2}{S_E^2}$

More variables to minimise over - but no matrix inversion. Finds offsets z_i explicitly

These are the same: RJB, arXiv:1701.03701

The scary bit

Consequence: taking more measurements and averaging does not reduce the error.

Consequence: no way to estimate σ_{sys} from the data - hence no check from χ^2 test etc

If you have an unknown statistical error the data will show this by its spread. If you have an unknown systematic then it won't, as they're all affected the same way. So your data can look beautiful but still be wrong.

Actually this is normal behaviour. It's just that statistical errors are user-friendly and we get used to it.

My definition

Any uncertainty in the process whereby your raw data is converted into a published result is a systematic error.

Any numerical values (efficiency, background...)

Fitting procedures?

MC models

Handling Systematic Errors in your analysis

3 types

1) Uncertainty in an explicit continuous parameter:

E.g. uncertainty in efficiency, background and luminosity in branching ratio or cross section

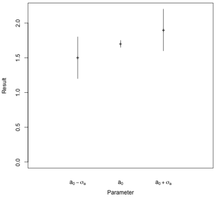
Standard combination of errors formula and algebra, just like undergraduate labs

Handling Systematic Errors (2)

Uncertainty in an implicit continuous parameter

Example: MC tuning parameters (σ_{p_T}, polarisation......)

Not amenable to algebra



Method: vary parameter by $\pm\sigma$ and look at what happens to your analysis result (directly, or through efficiency, background etc.)

Note 1: Hopefully effect is equal but opposite - if not then can introduce asymmetric error, but avoid if you can. Rewrite $^{+0.5}_{-0.3}$ as ± 0.4

Note 2. Your analysis results will have errors due to e.g. MC statistics. Some people add these (in quadrature). This is wrong. Technically correct thing to do is subtract them in quadrature, but this is not advised.

Handling Systematic Errors (3)

Discrete uncertainties, typically in model choice

Situation depends on status of model. Sometimes one preferred, sometimes all equal (more or less)

With 1 preferred model and one other, quote $R_1 \pm |R_1 - R_2|$

With 2 models of equal status, quote $\frac{R_1+R_2}{2} \pm |\frac{R_1-R_2}{\sqrt{2}}|$

N models: take $\overline{R} \pm \sqrt{\frac{N}{N-1}(\overline{R}^2 - \overline{R}^2)}$ or similar mean value

2 extreme models: take $\frac{R_1+R_2}{2} \pm \frac{|R_1-R_2|}{\sqrt{12}}$

These are just ballpark estimates. Do not push them too hard. If the difference is not small, you have a problem - which can be an opportunity to study model differences.

Checking the analysis

"As we know, there are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."

Donald H Rumsfeld

Checking the analysis: Errors are not mistakes - but mistakes still happen.

Statistical tools can help find them - though not always give the solution.
Check by repeating analysis with changes which *should* make no difference:

- Data subsets
- Magnet up/down
- Different selection cuts
- Changing histogram bin size and fit ranges
- Changing parametrisation (including order of polynomial)
- Changing fit technique
- Looking for impossibilities
- ...

Example: the BaBar CP violation measurement ".. consistency checks, including separation of the decay by decay mode, tagging category and B_{tag} flavour... We also fit the samples of non-CP decay modes for $\sin 2\beta$ with no statistically significant difference found."

What is a significant difference?

Results will not be spot-on identical. How much is allowed?

'Within the statistical error' (call it σ) is too generous if analyses share data

Sometimes the test can be formulated as two distinct subsamples. Then statistical errors are OK

If the test is a subsample, giving error σ' then testing difference against $\sqrt{\sigma'^2 - \sigma^2}$ is correct

For other tests this is also (usually) true

Less than 2 σ passes. More than 4 σ fails.
Between 2 and 4 you have to make a judgement .

If it passes the test

Tick the box and move on

Do not add the discrepancy to the systematic error

- It's illogical
- It penalises diligence
- Errors get inflated

The more tests the better. You cannot prove the analysis is correct. But the more tests it survives the more likely your colleagues[1] will be to believe the result.

[1] and eventually even you

If it fails the test

Worry!

- Check the test. Very often this turns out to be faulty.
- Check the analysis. Find mistake, enjoy improvement.
- Worry. Consider whether the effect might be real. (E.g. June's results are different from July's. Temperature effect? If so can (i) compensate and (ii) introduce implicit systematic uncertainty)
- Worry harder. Ask colleagues, look at other experiments

Only as a last resort, add the term to the systematic error. Remember that this could be a hint of something much bigger and nastier

Clearing up a possible confusion

What's the difference between?

Evaluating implicit systematic errors: vary lots of parameters, see what happens to the result, and include in systematic error

Checks: vary lots of parameters, see what happens to the result, and don't include in systematic error

(1) Are you expecting to see an effect? If so, it's an evaluation, if not, it's a check

(2) Do you clearly know how much to vary them by? If so, it's an evaluation. If not, it's a check.

Cover cases such as trigger energy cut where the energy calibration is uncertain - may be simpler to simulate the effect by varying the cut.

So finally:

1. Thou shalt never say 'systematic error' when thou meanest 'systematic effect' or 'systematic mistake'.
2. Thou shalt know at all times whether what thou performest is a check for a mistake or an evaluation of an uncertainty.
3. Thou shalt not add uncertainties on uncertainties in quadrature. If they are larger than chickenfeed thou shalt generate more Monte Carlo until they shrink to become so.
4. Thou shalt not incorporate successful check results into thy total systematic error and make thereby a shield to hide thy dodgy result.
5. Thou shalt not incorporate failed check results unless thou art truly at thy wits' end.
6. Thou shalt say what thou doest, and thou shalt be able to justify it out of thine own mouth; not the mouth of thy supervisor, nor thy colleague who did the analysis last time, nor thy local statistics guru, nor thy mate down the pub.

Do these, and thou shalt flourish, and thine analysis likewise.