Systematics

Roger Barlow
The University of Huddersfield

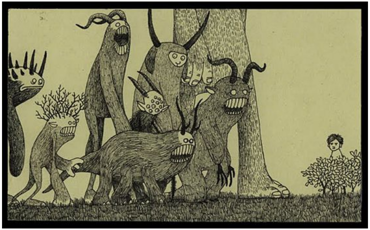
Terascale Statistics School, DESY, Hamburg

25^{th} February 2025

Systematic Errors

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.



- What is a Systematic Error?
- How to evaluate them
- How to handle 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.
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)

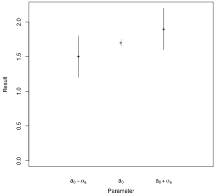
Evaluating Systematic Errors in your analysis

3 types

1) Uncertainty in an explicit continuous parameter. E.g. uncertainty in efficiency, background and luminosity in branching ratio or cross section
This is trivial. 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



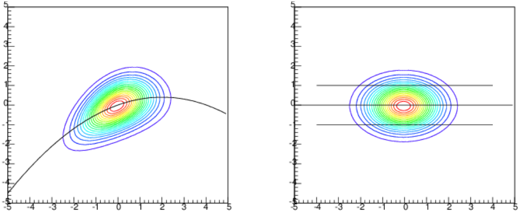
Method: vary one parameter at a time (OPAT) by $\pm\sigma$ and look at what happens to your analysis result

Note 1: Hopefully effect is equal but opposite - if not then can introduce asymmetric error, but avoid if you can.

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.

Note 3. Alternatively use more points regularly spaced

Note 4 Alternatively use more points chosen at random according to Gaussian distribution



You have a 2D likelihood plot with axes a_1 and a_2. You are interested in a_1 but not in a_2 ('Nuisance parameter')

Different values of a_2 give different results (central and errors) for a_1

Suppose it is possible to transform to $a_2'(a_1, a_2)$ so L factorises, like the one on the right. $L(a_1.a_2') = L_1(a_1)L_2(a_2')$

Whatever the value of a_2', get same result for a_1

So can present this result for a_1, independent of anything about a_2'.

Path of central a_2' value as fn of a_1, is peak - path is same in both plots

So no need to factorise explicitly: plot $L(a_1, \hat{\hat{a}}_2)$ as fn of a_1 and read off 1D values.

$\hat{\hat{a}}_2(a_1)$ is the value of a_2 which maximises $\ln L$ for this a_1

Nuisance Parameters 2

Marginalised likelihoods

Instead of profiling, just integrate over a_2.
Can be very helpful alternative, specially with many nuisance parameters
But be aware - this is strictly Bayesian

Frequentists are not allowed to integrate likelihoods wrt the parameter

$\int P(x; a)\, dx$ is fine, but $\int P(x; a)\, da$ is off limits

Reparametrising a_2 (or choosing a different prior) will give different values for a_1

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.

Why do we quote systematic errors separately?

Results are always given like

In conclusion, we have measured $m = 12.1 \pm 0.3 \pm 0.4$, where the first error is statistical and the second is systematic

Or even '$\pm$ statistical, $\pm$systematic, $\pm$luminosity uncertainty, $\pm$theory uncertainty, $\pm$branching ratio uncertainty'

Why quote them separately?

Why not just 12.1 ± 0.5?

Minor reason - shows whether result is statistics limited
Major reason - to enable combination of this result with others that share a systematic uncertainty
Systematic uncertainties obey the same rules as statistical uncertainties
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:X-sections with common lumi error, BRs with common efficiency

Combination of Errors

What is the error on $f(x, y)$

For undergraduates

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$$

For graduates

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\rho \left(\frac{\partial f}{\partial x}\right) \left(\frac{\partial f}{\partial y}\right) \sigma_x \sigma_y$$

If there are several functions and several variables this generalises to

$$\mathbf{V}_f = \tilde{\mathbf{G}} \mathbf{V_x} \mathbf{G} \qquad (1)$$

where V_f and V_x are the covariance matrices and $G_{ij} = \frac{\partial f_j}{\partial x_i}$

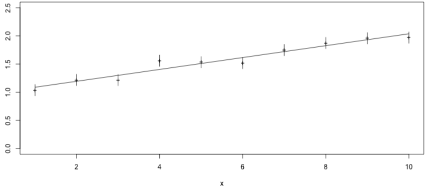
Example - the straight line fit

$y = mx + c$

$m = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\sum(x_i - \overline{x})y_i}{N(\overline{x^2} - \overline{x}^2)}$

$c = \overline{y} - m\overline{x} = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2} = \frac{\sum(\overline{x^2} - x_i\overline{x})y_i}{N(\overline{x^2} - \overline{x}^2)}$

$\mathbf{V_y} = \sigma^2 \mathbf{I}$



Equation 1 gives the usual errors, and also the correlation:

$V_m = \frac{\sigma^2}{N(\overline{x^2} - \overline{x}^2)}$ $\quad V_c = \frac{\sigma^2 \overline{x^2}}{N(\overline{x^2} - \overline{x}^2)}$ $\quad Cov = -\frac{\overline{x}\sigma^2}{N(\overline{x^2} - \overline{x}^2)}$ $\quad \rho = -\frac{\overline{x}}{\sqrt{\overline{x^2}}}$

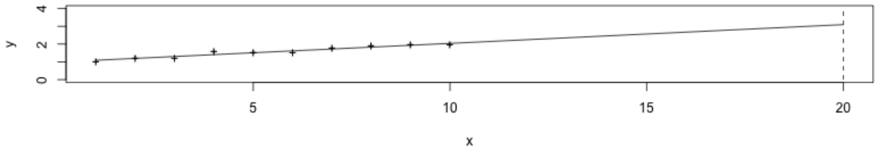
Note 1: Even though the y_i are independent, m and c are correlated

Note 2: Correlation vanishes if $\overline{x} = 0$. Or write $y = m(x - \overline{x}) + c'$

Note 3: in this example,

$m = 0.105 \pm 0.011, c = 0.983 \pm 0.068, \rho = -0.886$

Extrapolation of a straight line - what is y at $x = 20$?



$y = 0.983 + 20 \times 0.105$

Error from $\sqrt{0.068^2 + 20^2 \times 0.011^2} = 0.23$ Wrong

Correct Error from

$\sqrt{0.068^2 + 20^2 \times 0.011^2 - 2 \times 0.886 \times 20 \times 0.068 \times 0.011} = 0.16$

Building a covariance matrix

Matrix element $V_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$

Given correlated x_1 and x_2, model as $x_1 = y_1 + z, x_2 = y_2 + z$, where y_1, y_2, z independent with errors σ_1, σ_2, S. (Example: tracking detector where $y_i \pm \sigma_i$ are the measurements within the detector and $z \pm S$ is the position of the detector.)
$V_{11} = \langle (y_1 + z)(y_1 + z) \rangle - \langle (y_1 + z) \rangle^2 = \sigma_1^2 + S^2$.
V_{22} similar
$V_{12} = V_{21} = \langle (y_1 + z)(y_2 + z) \rangle - \langle (y_1 + z) \rangle \langle (y_2 + z) \rangle = S^2$

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 \end{pmatrix}$$

For more variables, build up larger matrix where off-diagonal elements come from shared features, on-diagonal gives total variance.

Building a correlation matrix
continued

Suppose experiment A measures x_1 and x_2 with shared systematic uncertainty S_A, and experiment B measures x_3 and x_4 with shared S_B

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + S_A^2 & S_A^2 & 0 & 0 \\ S_A^2 & \sigma_2^2 + S_A^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 + S_B^2 & S_B^2 \\ 0 & 0 & S_B^2 & \sigma_4^2 + S_B^2 \end{pmatrix}$$

Similar for (more common) shared multiplicative uncertainty - (e.g. efficiency, luminosity, normalisation...)
$x_1 \pm \sigma_1 \pm S_1$ and $x_2 \pm \sigma_2 \pm S_2$ with $S_1 = \xi x_1, S_2 = \xi x_2$

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + S_1^2 & S_1 S_2 \\ S_1 S_2 & \sigma_2^2 + S_2^2 \end{pmatrix}$$

PDG, HFLAV and similar groups do this on an industrial scale

Using the matrix

Independent measurements

Maximum Likelihood → Least Squares → minimise $\chi^2 = \sum_i \left(\frac{y_i - f(x_i)}{\sigma_i}\right)^2$

What if the y_i are not independent but correlated with non-diagonal covariance matrix V_y?

Rotate to $\mathbf{y}' = \mathbf{R}\mathbf{y}$ such that $Cov(y_i' y_j')$ is diagonal

V' diagonal by construction. $\mathbf{V}'^{-1} = \begin{pmatrix} 1/\sigma_1'^2 & 0 & 0 & \dots \\ 0 & 1/\sigma_2'^2 & 0 & \dots \\ 0 & 0 & 1/\sigma_3'^2 & \dots \\ \dots & & & \end{pmatrix}$

and $\mathbf{V}' = \mathbf{R}\mathbf{V}\tilde{\mathbf{R}}$

$\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}})\tilde{\mathbf{R}}[\mathbf{R}\mathbf{V}\tilde{\mathbf{R}}]^{-1}\mathbf{R}(\mathbf{y} - \mathbf{f}) = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{f})$

Forget about the primed system and use $\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{f})$

The famous Hessian matrix

$$\frac{\partial^2 \ln L}{\partial a_i \partial a_j}$$

$\hat{a}_1$ and $\hat{a}_2$ are functions of the data: maximise
$\ln L(a_1, a_2) = \sum_i \ln P(x_i; a_1, a_2)$

To first order about a^{true},
$\frac{\partial \ln L}{\partial a_1}|_{a=a^{true}} + \frac{\partial^2 \ln L}{\partial a_1^2}(\hat{a}_1 - a_1^{true}) + \frac{\partial^2 \ln L}{\partial a_1 \partial a_2}(\hat{a}_2 - a_2^{true}) = 0$
$\frac{\partial \ln L}{\partial a_2}|_{a=a^{true}} + \frac{\partial^2 \ln L}{\partial a_1 \partial a_2}(\hat{a}_1 - a_1^{true}) + \frac{\partial^2 \ln L}{\partial^2 a_2}(\hat{a}_2 - a_2^{true}) = 0$
Same as last lecture on ML errors, but matrix form
Various assumptions (no bias, large N, slow variation so use found values for expectation values...)
$V_{ij} = -\left\langle \frac{\partial^2 \ln L}{\partial a_j \partial a_k} \right\rangle^{-1}$
Covariance matrix is just minus the inverse of Hessian matrix, which is (typically) found by minimiser

Averaging

BLUE

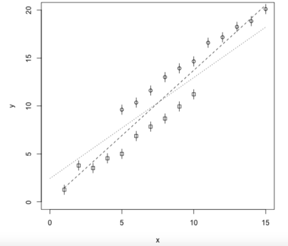
Given several (correlated) results y_i, how do you average them?

Best Linear Unbiased Estimator (L Lyons et al, NIM **A270** 110 (1988))

Minimise $\chi^2 = \sum_{i,j}(y_i - \hat{y}) V_{ij}^{-1} (y_j - \hat{y})$
$\hat{y} \sum_{i,j} V_{ij}^{-1} = \sum_{i,j} V_{ij}^{-1} y_j$
Write as $\hat{y} = \sum_i w_i y_i$ with $w_i = \frac{\sum_j V_{ij}^{-1}}{\sum_{i,j} V_{ij}^{-1}}$

Error on $\hat{y}$ given by $\sqrt{\tilde{\mathbf{w}}\mathbf{V}\mathbf{w}}$
Notice that $\sum_i w_i = 1$ which is intuitive
Notice that some w_i may be negative (if correlations are large) which is counterintuitve
This assumes the elements of $\mathbf{V}$ are known exactly. If not, care needed.

Equivalent alternative for additive systematics



Obvious method: Construct full covariance matrix **V** and minimise χ^2

Alternative: introduce explicit offsets $y'_{ij} = y_{ij} + \xi_j$ for value i of expt j.

ξ_j Gaussian with mean 0, sd S_j, included in χ^2

Fit the ξ_i and the parameter(s) a

Downside: n more parameters to fit

Upside (1) avoids matrix inversion

Upside (2): extracts the factors which can be useful to check behaviour

These two methods are actually (surprisingly!) equivalent

RB. *Combining experiments with systematic errors* **NIM A 987** 164864 (2021)

Conclusions(1)

Systematic errors can readily be handled - with the help of the correlation matrix and other techniques

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."

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 incorporate successful check results into thy total systematic error and make thereby a shield to hide thy dodgy result.
4. Thou shalt not incorporate failed check results unless thou art truly at thy wits’ end.
5. 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.
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

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