SLUO Lectures on Statistics and Numerical Methods in HEP
Lecture 5: Systematic Errors

Roger Barlow
Manchester University
10^{th} August 2000

"There is another class of errors which are not so easy to detect and for which statistical analysis is not generally useful."

R. Bevington: Data Reduction and Analysis for the Physical Sciences

Abstract

The above statement is only half true.

1. Introduction

Consider the formula:

$$P_T = 0.3B\rho \tag{1}$$

The transverse momentum is obtained from three quantities of very different numerical status.

- 0.3 is a hard number - the speed of light times 10^{-9}. It is known absolutely exactly.
- The radius of curvature ρ is obtained from each track, and is subject to errors. Measurements are sometimes too high, and sometimes too low. Over lots of tracks, on average it balances.
- The field B has been measured once, with some finite accuracy. For every track, the value of B is always too high, or always too low. Averaging over lots of tracks doesn't help. The same error is applied systematically to all the tracks.

  There are lots of examples like this.

Example 1: A calorimeter reconstructs energy E from digitisation value D.

$$E = aD + b$$

The constants a and b, obtained from calibration and applied to all energy measurements, have associated errors.

Example 2: A silicon detector (or RPC) gives the position X from strip-value S

$$X = a + bS$$

The pitch b is probably well established. a is subject to survey errors etc.

If you only have one track, then this doesn't matter. If you know B to 3% and ρ to 4% then you know P_T to 5%. But if you have two tracks then it does. Although you know both the P_T values to 5%, if you do anything with the two of them together (like forming a mass or an average) then the standard Law of Combination of Errors does not apply because the measurements are correlated. The familar form

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

is only true if x and y are independent

2. Covariance and Correlation

The good news, however, is that once you know what your systematic errors are, they can be handled with standard statistical methods: covariance and correlation.

2.1 Basic definition

Consider a joint distribution $P(x, y)$ of random variables x and y. Its properties are described by the means $< x >$ and $< y >$, and variances

$$V(x) \;=\; \Big\langle (x - < x >)^2 \Big\rangle \;= < x^2 > - < x >^2$$

$$V(y) \;=\; \Big\langle (y - < y >)^2 \Big\rangle \;= < y^2 > - < y >^2$$

the standard deviations σ_x and σ_y are just the square roots of these variances.

But there is more information there. You can also look at the two variables together – are they independent or do they depend on one another, and if so in what way? This is done by the *Covariance* of x and y:

$$Cov(x, y) \;=\; \Big\langle (x - < x >)(y - < y >) \Big\rangle \;= < xy > - < x > < y > \tag{3}$$

If values of x that are above average have a tendency to go along with above average y values (which means that small x go with small y) then the signs of the two terms will tend to be the same, and the net total will be positive. Likewise if large x tend to go with small y, the covariance is negative. Notice that:

i) Covariance involves only differences from the mean x and y, so it does not change if the origin is shifted.

ii) Covariance is a generalisation of the Variance, in that $Cov(x, x) = V(x)$

The covariance is useful, but has dimensions. A more meaningful measure of the relation between two variables is the correlation coefficient, ρ. This has the dimensions taken out.

$$\rho = \frac{Cov(x, y)}{\sigma_x \sigma_y} \tag{4}$$

ρ is a number between -1 and $+1$. If ρ is zero then x and y are uncorrelated. A ρ of 1 signifies complete correlation, -1 signifies complete anticorrelation. ρ is dimensionless, and is unaffected by shifts in the origin or by changes in the scale for x or y.

Example 3: The two P_T values will be positively correlated: if B is too high/too low then both P_T values are too high/too low. We can write

$$Cov(P_{T1}, P_{T2}) = < (0.3B\rho_1)(0.3B\rho_2) > - < (0.3B\rho_1) > < (0.3B\rho_2) >$$

Expanding ($B = < B > + \delta_B$ etc) and dropping third and higher order terms (the linear terms vanish due to symmetry) gives

$$Cov(P_{T1}, P_{T2}) \approx 0.3^2 \rho_1 \rho_2 \sigma_B^2$$

2.2 Combining Systematic and Random errors

If a final result has a random error and a systematic error, then the total error is given by adding the two in quadrature. Even so, it is often helpful and useful to others to quote the two errors separately, and you'll often see statemens like

$$A = -10.2 \pm 1.2 \pm 2.3 \qquad \text{or} \qquad \sigma = 45 \pm 4 \pm 1mb$$

where the first error is statistical and the second systematic. This can enable the combining and comparing of experiments which share systematic effects. It also shows the relative importance: here the first result is dominated by systematics, so that there is nothing to be gained by taking more data.

If two measurements x_1 and x_2 have a common systematic error S, and also individual random errors σ_1 and σ_2, this can be treated by considering x_1 as having two parts, x_1^R with random error σ_1 and x_1^S with systematic error S, and similarly x_2 is divided into x_2^R and x_2^S. x_1^R and x_2^R are independent of each other and of x_1^S and x_2^S, whereas x_1^S and x_2^S are absolutely correlated. The Variance of x_1 is given by

$$\begin{aligned} V(x_1) &= < x_1^2 > - < x_1 >^2 \\ &= < (x_1^R + x_1^S)^2 > - < x_1^R + x_1^S >^2 \\ &= \sigma_1^2 + S^2 \end{aligned}$$

which shows that systematic and random errors are added in quadrature to give the total.

Similar treatment gives $V(x_2) = \sigma_2^2 + S^2$ and the covariance

$$\begin{aligned} Cov(x_1, x_2) &= < x_1 x_2 > - < x_1 > < x_2 > \\ &= < (x_1^R + x_1^S)(x_2^R + x_2^S) > - < x_1^R + x_1^S > < x_2^R + x_2^S > . \end{aligned}$$

3 of the 4 cross products involve the x^R and therefore cancel, as the x^R are independent of everything else. The fourth involves x_1^S and x_2^S, which are absolutely correlated, leaving

$$Cov(x_1, x_2) = Cov(x_1^S, x_2^S) = S^2.$$

Thus the variance matrix for x_1 and x_2 has the random and systematic errors added in quadrature along the diagonal; off the diagonal the covariances are given by the squared systematic error.

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

This can all be generalised in the obvious way. If there are several independent sources of systematic error then they are added in quadrature. If there are more variables the matrix is larger. For example, consider 3 variables, x_1, x_2, x_3, and suppose that in addition to their random errors $\sigma_1, \sigma_2, \sigma_3$ they have a common systematic error S and also another independent systematic error T which is shared by x_1 and x_2 but not x_3. The covariance matrix is then (as you can show for yourself by multiplying out the covariances)

$$\begin{pmatrix} \sigma_1^2 + S^2 + T^2 & S^2 + T^2 & S^2 \\ S^2 + T^2 & \sigma_2^2 + S^2 + T^2 & S^2 \\ S^2 & S^2 & \sigma_3^2 + S^2 \end{pmatrix}.$$

It's worth pointing out that all this makes very few assumptions. It does not assume that the errors are small, and it does not assume that the distributions are Gaussian.

2.3 Combination of Errors (revised)

The equivalent of Eq. 2 (which does assume small errors) when x and y are not independent is

$$V(f) = \left(\frac{df}{dx}\right)^2 V(x) + \left(\frac{df}{dy}\right)^2 V(y) + 2\left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right) Cov(x,y) \tag{5}$$

Incidentally, if you do have correlated variables it may be possible to find a change of variables to make the covariance vanish. It's always worth a few minutes thought.

Example 4: Suppose $X = x_1 + x_2$, and x_1 and x_2 are completely correlated. Then

$$\begin{aligned} V(X) &= V(x_1) + 2Cov(x_1, x_2) + V(x_2) \\ &= \sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2 \\ &= (\sigma_1 + \sigma_2)^2 \end{aligned}$$

i.e. for completely correlated errors, add arithmetically rather than in quadrature.

2.4 Extension to several variables

If there are n variables $x_1, x_2, ...x_n$ † the Covariance between two of them is

$$Cov(x_i, x_j) = < x_i x_j > - < x_i >< x_j > .$$

These are the elements of the *Covariance Matrix*, $\mathbf{V}$, where

$$V_{ij} = Cov(x_i, x_j).$$

The diagonal elements of this matrix are just the variances

$$V_{ii} = Cov(x_i, x_i) = V(x_i) = \sigma_i^2$$

Incidentally, the χ^2 for correlated variables uses the inverse of this matrix

$$\chi^2 = \sum_i \sum_j d_i V_{ij}^{-1} d_j$$

where the d_i are the discrepancies between data and fit (or whatever).

The *Correlation matrix* is the dimensionless equivalent

$$\rho_{ij} = \frac{Cov(x_i, x_j)}{\sigma_i \sigma_j}$$

Now suppose there are m different functions $f_1, f_2...f_m$, of n different variables $x_1, x_2...x_n$. The x_i have errors, therefore so do the f_k. The f_k will be correlated with one another (even if the x_i aren't) because the different f_k share the same x_i.

† The subscripts are used to distinguish between the n elements of a single measurement, not to the N measurements that comprise the sample. Just remember that x_2 is another name for y.

The general combination-of-errors formula is

$$Cov(f_k, f_l) = \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_i}\right)\left(\frac{\partial f_l}{\partial x_j}\right) Cov(x_i, x_j) \tag{6}$$

which includes the formula for the variance as $Cov(f_i, f_i) \equiv V(f_i)$.

It can be written compactly using Matrices - if

$$G_{ki} = \left(\frac{\partial f_k}{\partial x_i}\right)$$

and $\mathbf{V_x}$ and $\mathbf{V_f}$ are the error matrices for $\mathbf{x}$ and $\mathbf{f}$ then Eq. 6 can be written

$$\mathbf{V_f} = \mathbf{G}\mathbf{V_x}\tilde{\mathbf{G}} \tag{7}$$

a neat little formula which contains all there is to know about error propagation.

$\mathbf{V_x}$ and $\mathbf{V_f}$ are symmetric and square, $n \times n$ and $m \times m$. $\mathbf{G}$ is rectangular, $m \times n$.

Example 5: With

$$\mathbf{f} = \begin{pmatrix} P_{T\,1} \\ P_{T\,2} \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ B \end{pmatrix}$$

Equation 1 gives

$$\mathbf{G} = \begin{pmatrix} 0.3B & 0 & 0.3\rho_1 \\ 0 & 0.3B & 0.3\rho_2 \end{pmatrix}$$

and applying Equation 7 to

$$\mathbf{V_x} = \begin{pmatrix} \sigma^2_{\rho_1} & 0 & 0 \\ 0 & \sigma^2_{\rho_2} & 0 \\ 0 & 0 & \sigma^2_B \end{pmatrix}$$

gives the error matrix for the two P_T measurements as

$$\begin{pmatrix} 0.3^2B^2\sigma^2_{\rho_1} + 0.3^2\rho_1^2\sigma^2_B & 0.3^2\rho_1\rho_2\sigma^2_B \\ 0.3^2\rho_1\rho_2\sigma^2_B & 0.0.3^2B^2\sigma^2_{\rho_2} + 0.3^2\rho_2^2\sigma^2_B \end{pmatrix}$$

Example 6: If you regard the two P_T variables as functions of 4 variables, ρ_1, ρ_2, B_1 and B_2, with B_1 and B_2 completely correlated, this gives the same answer.

Example 7: Tracking chambers measure the position of a particle in cylindrical polar co-ordinates, (r, ϕ, z). r is known precisely and the error can be taken as zero. ϕ and z are measured with uncorrelated errors σ_ϕ and σ_z. What are the errors on the cartesian co-ordinates x, y, z?

Well, $x = rCos\phi$, and $y = rSin\phi$ (and $z = z$) so $\mathbf{V_{xyz}}$ is

$$\begin{pmatrix} Cos\phi & -rSin\phi & 0 \\ Sin\phi & rCos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \begin{pmatrix} Cos\phi & Sin\phi & 0 \\ -rSin\phi & rCos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma_\phi^2 y^2 & -\sigma_\phi^2 xy & 0 \\ -\sigma_\phi^2 xy & \sigma_\phi^2 x^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

This is sensible. The error and correlations involving z are not affected. The errors on x and y depend on where they are: for ϕ near 0, y is small, and x is along the radial direction and thus has a small error. The correlation between x and y is ± 1.

Example 8: Consider a straight line $y = mx + c$ where all the y values have a random error σ and share a common systematic error S.

$$m = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \qquad c = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2}$$

The full error formula gives, for the slope ($\Delta = \overline{x^2} - \overline{x}^2$)

$$V(m) = \frac{1}{N^2\Delta^2} \sum_i \sum_j (x_i - \overline{x})(x_j - \overline{x}) Cov(y_i, y_j)$$

And we have $Cov(y_i, y_j) = \delta_{ij}\sigma^2 + S^2$ so

$$V(m) = \frac{1}{N^2\Delta^2} \sum_i (x_i - \overline{x})^2 \sigma^2 + \sum_i \sum_j (x_i - \overline{x})(x_j - \overline{x}) S^2$$

The first summation gives the usual answer $\sigma^2/N\Delta$. The second gives zero, as $\sum x_i = \overline{x}$.

Now for the constant.

$$V(c) = \frac{1}{N^2\Delta^2} \sum_i \sum_j (\overline{x^2} - \overline{x}x_i)(\overline{x^2} - \overline{x}x_j) Cov(y_i, y_j)$$

The σ^2 values along the diagonal give the standard term $\sigma^2\overline{x^2}/N\Delta$, but the other term does not cancel. $\sum(\overline{x^2} - \overline{x}x_i)$ is just $N\Delta$, giving an additional term $(N\Delta)^2 S^2/(N^2\Delta^2)$ which is just S^2. The presence of an additional systematic (constant) error S on the y measurements does not affect the slope, but adds (in quadrature) to the uncertainty on the constant by an amount S.

This result may be obvious (though it can usefully be extended to a straight line through two datasets with different systematic errors) but it shows nicely how the whole apparatus of calculation works. It also shows two other features typical of systematic errors:

1) Taking further measurements does not improve the contribution of the systematic error

2) An unsuspected systematic error does not reveal itself through a bad χ^2.

3. Diagnosis of Systematic Errors

The term 'systematic error' is used inconsistently and inaccurately. In elementary textbooks one meets sentences such as 'Systematic errors arise from incorrectly calibrated equipment.' Not so. Any equipment is calibrated with a finite error. Conversion factors are there - even if they're taken as unity or zero. The errors on these conversion factors are the 'systematic errors'. If a volt meter reads high by 10% that means a conversion factor of 1.1 is taken as 1.0. This is not a 'systematic error' but a plain mistake; a systematic error of 0.1 should be included - or (better) a systematic *correction* of 0.1 applied to the factor, making the systematic *error* (legitimately) small. Systematic errors are the uncertainties in systematic correction factors.

The treatment of systematic errors is one of the most demanding parts of experimental physics. It requires ingenuity, experience, cunning and a fair degree of paranoia. These are well provided in particle physics. It also requires honesty and bravery, which are scarcer.

3.1 Finding Systematic errors

Think carefully about the whole analysis. Worry hard about the accuracy of anything applied to all your data, unless it's completely above suspicion like π or $\sqrt{2}$, as it affects all your data points in the same way and is thus a systematic effect. It is helpful to distinguish between errors from *manifest* sources, as when the uncertain calibration of a detector is applied to all energy measurements, and those from *unsuspected* sources; for example unsuspected changes in detector performance with external conditions.

Often there are checks you can do to satisfy yourself and others of the absence of systematic effects. Plots that should be linear, or go through the origin. Particle masses should come out right. But you have to invent each check yourself: plots may look perfectly healthy, with nice straight lines and good χ^2 values, although systematics have rendered the result worthless.

Having done all you can on your own, the next thing to do is to ask Grumpy. Most of us have a colleague whose chief talent lies in destructive criticism – ask him politely (it's usually a him) what he thinks of your experiment. (Do not despair at the response! Before you decide to drop physics and take up web-site design instead, have a stiff drink and a good night's sleep, after which his critique of your experiment will seem less devastating.)

Naturally, such analysis is most effective at the planning stage of an experiment. (Though with today's long timescales this advice is probably not very helpful.) Suitable design can remove foreseen systematic effects, or make arrangements for suitable monitoring measurements, or convert them into random errors. Various nasty effects (electronics drifts, temperature drifts, and even psychological changes in the experimenter) are largely functions of time. If you take your data in an orderly sequence then this will become a systematic effect. Chopping and changing the order does not destroy these effects, but it renders them random rather than systematic.

3.2 A friendly fact

Systematic errors are, by virtue of their independence, added in quadrature. One or two are usually bigger than the rest and dominate the total. This is useful, as although you have to work hard evaluating the large contributions, you don't have to sweat blood over the smaller ones.

Example 9: You have 3 sources of systematic error which contribute 3%, 4% and 8%. When added in quadrature, this gives an error of 9.4%. This is dominated by the largest error of 8%. If the 3% were actually 2% or 4%, the total changes to only 9.2% or 9.8%

3.3 Estimation of Manifest Systematic Errors

At the more benign end are factors with known errors used explicitly in analysing your data: efficiencies, calibrations, etc. Any such factor is embodied in a number, it is clearly uncertain, and the uncertainty may be well established. Any calibration run should give you conversion factor(s) *and* associated error(s). Or if you are analysing a τ pair cross section the predicted contribution of the background from multihadrons depends on the parameters used in the MC model; you use the values obtained by tuning the MC, which come with associated errors.

An ideal procedure would then be to repeat the analysis, with the constants reduced and increased by some step, so the dependence of the result on the variable constant can be established and the resulting error on the answer obtained by multiplying this slope by the error on the constant (using Eq. 7). Using the errors as the step is an obvious convenience, but larger values may be better numerically. It may be valid to move the goalposts. Suppose you count ψ production with a mass window of 3.00 and 3.20 GeV. Maybe you suspect systematic effects in your mass reconstruction at the level of 0.01 GeV. You can simulate these by moving the window, if that's easier than adjusting all the masses.

Less friendly are the numbers applied where you can't quantify the error. Here one is reduced to intelligent guesswork, or sometimes just guesswork. In such cases the 'systematic errors' may be untrustworthy, and the 'friendly fact' of quadrature relied on heavily.

A third class are errors where there is no number that quantifies the assumption. For example, a hadronic background contribution affects the parameter fit you are doing to your τ data in some way which cannot be wrapped in a simple number (other than 'background correction to result'). The effect may depend on the general model used for hadronisation, and you have a choice between HERWIG and JETSET (and other models). Two techniques can be used for this. One is to take several typical assumptions and look at the variation in the correction factor computed, taking its standard deviation to be the systematic error (or, in a cheap alternative justfied by friendly quadrature, take two different assumptions and use the difference in the result as a measure.) The second is to take the values given under two extreme assumptions, and argue that the true factor must lie somewhere between them, and thus that the error is conservatively given by the difference between them divided by $\sqrt{12}$, the standard deviation of a uniform distribution.

Example 10: Consider the nature of the 'error' in the following cases:

i) B is uniform and constant, and was measured with a flux meter good to 1%.

ii) B is uniform and constant. It was measured with a highly accurate flux meter. However changes in geometry since then (e.g. inserting a drift chamber, opening and closing doors) will have changed that field by an amount we estimate at 1%.

iii) B is uniform and was well measured, however it varies with time (temperature effects on power supplies) at a 1% level.

iv) B is constant but not uniform. Different tracks see different fields (but still obey $P_T = 0.3B\rho$.)

v) B is constant but not uniform. The same track sees different fields along its length. So the equation $P_T = 0.3B\rho$ is inadequate.

vi) B is constant, uniform and measured accurately, but you typed the number in wrong.

vii) tracks lose momentum in the beam pipe so all your measurements are low.

viii) tracks lose momentum in the drift chamber so they are not pure circular helices.

ix) For high P_T tracks ρ is non-Gaussian: $1/\rho$ is much better.

3.4 Unsuspected systematic errors

The most vicious systematic errors are the ones you didn't anticipate, such as an electronics component failing, or a new source of noise, or ..... (fill in your own horror story here.)

There will be checks you can do to satisfy yourself and others of the absence of systematic effects (plots that should be linear, etc.) It is also standard and sensible practice to repeat the analysis in different forms, for example

- varying the range of the data used to extract the result
- varying cuts applied to ensure the purity of the sample and the quality of the data
- including and excluding subsets of data taken under different experimental conditions
- fitting a function using histograms of different bin sizes
- determining quantities by simple counting and by fits to a parametrised curve

This analysis resembles the investigation of known error sources, as it involves repeating the analysis with minor differences. But it is fundamentally different. You are not measuring an effect, you are asking whether or not an effect exists.

Such changes are made in the expectation (or hope) that they will have no effect. If the values agree despite having been extracted using differing techniques, that argues that the result is stable and free of unsuspected errors. But if a supposedly innocent change produce a *large* shift in the value, that indicates the presence of an unsuspected systematic effect. If an effect is found then, as its origin is unknown, it has to be treated very seriously. Ideally it should be investigated and explained. If and only if this fails should it, as a last resort, be folded into the systematic error.

Example 11: You have data measured as a function of x in the range $0 < x < 1$ and will use it to predict the value at $x = 10$. You fit to a straight line $y = mx + c$. In fact, the data is described by the curve $y = (x + 5)^2$. 'Systematic' checks on the result by varying the range of fit will produce values of the slope between 10 and 12. No matter how this is included as a systematic error, the error on $y(10)$ will be grossly underestimated.

The action to be taken if the changes are small also deserves more consideration than is often given. Many analyses adopt a methodology of varying all possible fitting techniques, cuts, etc., with the resulting deviations being added in quadrature to the systematic error. This exhaustive technique must inflate the estimate of the systematic errors – indeed, this gives the pusillanimous researcher a comfortably large error estimate to hide behind, describing it as 'conservative'. By contrast, I would maintain that, if there is no *a priori* reason to suspect a systematic effect from a particular source, and if there is no numerical evidence for a significant variation, then there is no cause to include it as a contribution to the systematic error.

We now need to define 'large' and 'small'. Values obtained by different techniques will, even in the absence of systematic effects, disagree. Different treatments will in general produce *small* differences in numerical values: one cannot hope for agreement to all decimal places. We need to know the size of such deviations that may legitimately occur due to statistical fluctuations . This is not just given by comparison with the statistical errors on the results; the measurements share, or partially share, the same data sample, and are therefore correlated. (Some uncorrelated comparisons can be done by splitting the data into two independent subsamples, but not all checks can be made in this form).

3.4.1 Evaluation of 'large' and 'small'

Suppose the parameter a is estimated from a data sample by two different methods, giving estimates $\hat{a}_1$ and $\hat{a}_2$, with statistical errors σ_1 and σ_2. The difference between the estimates is

$$\Delta = \hat{a}_1 - \hat{a}_2$$

and it is required to find the statistical error σ_Δ on this, so that by comparing the actual value of Δ with σ_Δ, you can establish whether the estimates agree to a satisfactory extent: within two standard deviations, or three standard deviations, or whatever. The error is given by

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

where ρ is the correlation between the two estimates. It is this correlation which means that the simple statistical errors on the estimates cannot be used directly for comparison. If the estimates share some of the same data it is presumably positive. The evaluation of ρ will in general be complicated or even intractable. However limits can be placed on it [1] as follows.

Two estimates can in principle be weighted and combined to give an estimate more efficient (i.e. with smaller error) than either one separately. Let the joint estimator be

$$\hat{a} = W\hat{a}_1 + (1-W)\hat{a}_2$$

where W is some chosen weight. Then the error on a is given by

$$\sigma^2 = W^2\sigma_1^2 + (1-W)^2\sigma_2^2 + 2W(1-W)\rho\sigma_1\sigma_2.$$

Setting the derivative with respect to W to zero to find the minimum of σ gives $W = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$ and the resulting minimum error on $\hat{a}$ is given by

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}.$$

However no estimator can be more efficient than the Minimum Variance Bound (MVB) [2]

$$\sigma^2 \geq \frac{1}{N\int\left(\frac{dlnP}{da}\right)^2 Pdx}$$

where $P(x;a)$ is the probability density function for the distribution being measured. This bound can be calculated moderately easily in any problem. Denoting it as σ_0^2 and imposing $\sigma^2 > \sigma_0^2$ gives an expression for ρ which is satisfied only between the two roots of the quadratic

$$\rho = \frac{\sigma_0^2 \pm \sqrt{(\sigma_1^2-\sigma_0^2)(\sigma_2^2-\sigma_0^2)}}{\sigma_1\sigma_2}$$

The maximum value for σ_Δ comes from the minimum value for ρ, and vice versa

$$\sigma_\Delta^{max} = \sqrt{(\sigma_1^2-\sigma_0^2)} + \sqrt{(\sigma_2^2-\sigma_0^2)} \qquad \sigma_\Delta^{min} = \left|\sqrt{(\sigma_1^2-\sigma_0^2)} - \sqrt{(\sigma_2^2-\sigma_0^2)}\right| \tag{8}$$

In practice this range is usually fairly narrow, because an analysis will normally use a highly efficient estimator, which saturates the MVB. We then have

$$\sigma_\Delta = \sqrt{|\sigma_1^2 - \sigma_2^2|} \tag{9}$$

(If neither σ saturates the MVB then one still has the choice of making statements of the form: 'the results agree within at least $\Delta/\sigma_\Delta^{min}$ standard deviations.' or: 'there is no evidence that these results differ by more than $\Delta/\sigma_\Delta^{max}$ standard deviations.')

With this definition of 'large' and 'small' changes, one can make a rational decision as to whether the variations in results produced by variations in the analysis technique are large enough to merit further investigation.

Example 12: An experiment measures the angular distribution of μ pairs produced in Z decays[3]. The probability distribution is assumed to have the form

$$P(\cos\theta) = \frac{3}{8}(1+\cos^2\theta) + A\cos\theta$$

and A is the asymmetry between events produced in the 'forward' and 'backward' regions of $\cos\theta$. With 33 464 events, A is measured both by counting the number of forward and backward events and by fitting the form of the curve, with the result:

$$A = 0.0123 \pm 0.0055\ (counting) \qquad A = 0.0084 \pm 0.0051\ (fitting)$$

a difference of 0.0039. Does this indicate a systematic effect?

The Minimum Variance Bound is

$$\frac{1}{\sigma_0^2} = N\int_{-1}^{+1} P(x)\left(\frac{dlnP}{dA}\right)^2 dx \ = \frac{8N}{3}\int_{-1}^{+1}\frac{x^2}{1+x^2+8Ax/3}dx$$

The value of this integral depends on the unknown A, but this is ≈ 0, giving

$$\frac{1}{\sigma_0^2} = \frac{8N}{3}\int_{-1}^{+1}\frac{x^2}{1+x^2}dx = \frac{8N}{3}\frac{4-\pi}{2}$$

For 33 464 events that gives a value of σ_0 of 0.0051. (Varying A between -0.1 and +0.1 gives values in the range 0.005072 to 0.005110.) The fitting method's quoted error thus saturates the MVB – this is not surprising. The expression for σ_Δ is just $\sqrt{\sigma_1^2 - \sigma_0^2}$ which has the value 0.0021. The discrepancy between the two values is thus (just) less than 2 standard deviations.

References

[1] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics* Section 17.27, Vol II. 4th 3-volume edition. Charles Griffin and Co. 1979

[2] Ibid., Section 17.15, and many statistics textbooks

[3] OPAL collaboration (R. Akers et al) *Zeit. Phys.* **C61** p19 (1994)