Chapter 5

Uncertainty and errors

5.1 The nature of errors

We now have enough background information to revisit the topic of quantifying the uncertainty on a single, or set, of measurement(s). The concepts developed in this chapter are central to how scientific method is able to test theory against experimental observations. In practice the words error and uncertainty are often used interchangeably when describing how well we are able to constrain or determine the value of an observable.

5.1.1 Central limit theorem

If one takes N random samples of a distribution of data that describes some variable x, where each sample is independent and has a mean value μ_i and variance σ_i^2, then the sum of the samples will have a mean value M and variance V where

$$M = \sum_{i=1}^{N} \mu_i, \tag{5.1.1}$$

$$V = \sum_{i=1}^{N} \sigma_i^2. \tag{5.1.2}$$

As N tends to infinity, the distribution of the sum of the samples tends to a Gaussian distribution. This is called the ***central limit theorem*** (CLT) and it implies that independent measurements of some observable will have uncertainties that are Gaussian in nature. Indeed empirical tests of this theorem show that this is indeed the case. As shown for example in Barlow (1989), Eqns (5.1.1) and (5.1.2) can be verified by expanding the right hand sides of those equations and substituting in Eqns (3.2.1) and (3.3.1), respectively.

5.1.2 The Gaussian nature of statistical uncertainty

Statistical uncertainties are Gaussian in nature for sufficiently large samples of data. This result can be obtained both empirically and via the CLT. The ramification of this result will be discussed in more detail in the rest of this section, however before looking at this in detail, we briefly return to our expression of a measurement $\widehat{x} \pm \widehat{\sigma}$, where $\widehat{x}$ is our best estimate of the true mean value, and $\widehat{\sigma}$ is the estimate of the uncertainty on $\widehat{x}$ which is given by the standard deviation of the result. We interpret this result as saying that if we were to repeat the measurement, there is a certain probability that the second measurement would

be within $\pm 1\sigma$ of the original result. This probability is given by

$$P = \int_{-1\widehat{\sigma}}^{+1\widehat{\sigma}} G(x; \widehat{x}, \widehat{\sigma}) dx, \tag{5.1.3}$$

where $G(x; \widehat{x}, \widehat{\sigma})$ is a Gaussian PDF with mean $\widehat{x}$ and width $\widehat{\sigma}$. This probability is approximately 68.3%, and is sometimes referred to as the ***coverage*** of the interval $(x - \widehat{x}) \in [-1\sigma, +1\sigma]$. Similarly we can consider how often we would expect the second result to be within $\pm n\widehat{\sigma}$ of the original experiment. These results are summarised in Table 5.1 for $n = 1, 2, \ldots 5$.

Table 5.1: The Gaussian probability for some repeat measurement of an observable x to lie between $+n\sigma$ and $-n\sigma$ of the original measurement for values of n up to $n = 5$, shown to a precision of 4 decimal places.

$\pm n\widehat{\sigma}$	coverage (%)
1	68.2689
2	95.4500
3	99.7300
4	99.9937
5	99.9999

If we perform a measurement on some observable O, the result of which is given by $O_1 = \widehat{x}_1 \pm \widehat{\sigma}_1$, then at some later time we repeat the measurement with the same experimental conditions we would expect that by definition, we can find the new measurement $O_2 = \widehat{x}_2 \pm \widehat{\sigma}_2$ which can have a central value $\widehat{x}_2$ that is more than σ_1 away from $\widehat{x}_1$. If we perform an infinite number of subsequent measurements of the observable O we would expect that 68.2689% of the time, the results will fall between $\widehat{x}_1 - \widehat{\sigma}_1$ and $\widehat{x}_1 + \widehat{\sigma}_1$, and thus 31.7311% of the time the results will fall outside of this range. Implicitly when doing this we assume that $\widehat{x}_1$ is a sufficiently good estimate of the true mean value. By construction we are expecting that when we repeat a measurement quite a lot of the time our new result will be slightly different than our original one. Normally we consider two results to be compatible if a subsequent measurement is within $\pm 3\sigma$ of the original on the basis that we normally make many measurements. Therefore it is not unreasonable to expect that less than one in a hundred results has a large ($\geq \pm 3\sigma$) deviation from our expectations. Much more than this level of discrepancy and we would want to investigate the compatibility of the two results more rigorously in order to verify that any differences can be attributed to statistical variations in the measurements.

5.1.3 Repeating measurements with a more precise experiment

Often in science, we attempt to repeat the measurement of an observable in order to obtain a more stringent constraint on that observable than obtained by any previous experiment. An example of this would be to measure a physical observable, such as the mass of an electron, with greater precision. When doing so, in addition to preparing or constructing a more robust experiment, it may be necessary to increase the statistical power of the experiment. By 'increase the statistical power' we mean, increase the number of data entries recorded by the repeat experiment.

So the question arises, what would constitute a minimum increase in the number of data entries or events recorded at an experiment in order to justify the time and expense of building a new one. To answer this question, you have to understand how much better your new experiment would be relative to the old one(s) and offset the cost and time required to construct a new experiment and repeat the measurement against the benefit obtained in terms of increased precision. The increase in precision expected can be determined from the definition of standard deviation given in Eq. (3.3.10). The statistical precision of a new experiment with ten times the data of an old one would have an error $\sqrt{10} \sim 3.2$ times smaller than the old experiment. So if the statistical uncertainty dominated the precision of the measurement you wanted to make, then it would be

worth considering making the new experiment. Much less than this factor of increase in precision, and one could spend a lot of time working on a new measurement that provides only a very marginal improvement over the previous result. If one were to build an experiment that could collect 100 times the data of any existing experiment, then that new endeavour would be able to make a measurement 10 times better than any existing one. In general we follow the rule that statistical precision will scale with the square-root of the number of data, or $\sqrt{N}$.

If a great deal of care is taken in the design and construction of the new experiment, then it may be possible to obtain a better than $\sqrt{N}$ reduction in the error. But that would require the new experiment to be significantly better than the old one (for example better energy or momentum resolution, better calibration, etc.). If there is no such improvement and someone claims to be able to do better than the $\sqrt{N}$ scaling in uncertainty by collecting more data with an equivalent method to that used in an existing experiment, then one should seriously question how this new measurement can 'beat statistics'. It is not possible to 'beat statistics', hence this scaling rule can be used to make, and check the validity of, extrapolations from an old experiment to a new one.

5.2 Combination of errors

In order to understand how to combine uncertainties we can first consider the case of a function dependent on two measured inputs: x and y each with with mean values $\overline{x}$ and $\overline{y}$, and errors σ_x and σ_y. We can Taylor expand this function $f(x,y)$ about $x=\overline{x}$, and $x=\overline{y}$ as

$$f(x,y) = f(\overline{x},\overline{y}) + \frac{\partial f}{\partial x}(x-\overline{x}) + \frac{\partial f}{\partial y}(y-\overline{y}) + \ldots, \tag{5.2.1}$$

where in the following we ignore higher order terms. If we now consider how to compute the variance of some quantity f as described in section 3.3, this is given by

$$V(f) = (f-\overline{f})^2, \tag{5.2.2}$$

where $\overline{f} = f(\overline{x},\overline{y})$. Thus, it follows that

$$V(f) = \left(\frac{\partial f}{\partial x}(x-\overline{x}) + \frac{\partial f}{\partial y}(y-\overline{y})\right)^2, \tag{5.2.3}$$

$$= \left(\frac{\partial f}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2\sigma_y^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\sigma_{xy} \tag{5.2.4}$$

where we have replaced $(x-\overline{x})^2$, $(y-\overline{y})^2$, and $(x-\overline{x})(y-\overline{y})$ by σ_x^2, σ_y^2, the covariance σ_{xy}, and neglected higher order terms in the last step. Equation (5.2.4) can be expressed in matrix form as

$$V(f) = \left(\tfrac{\partial f}{\partial x}, \tfrac{\partial f}{\partial y}\right)\begin{pmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{pmatrix}\begin{pmatrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\end{pmatrix}, \tag{5.2.5}$$

where the covariance matrix V (Eq. 3.6.4) is evident as the middle term in the equation.

If x and y are independent variables, the third term in Eq. (5.2.4) vanishes, and we obtain a general expression for combining uncertainties for functions of two independent variables

$$\sigma_f^2 = \left(\tfrac{\partial f}{\partial x}\right)^2\sigma_x^2 + \left(\tfrac{\partial f}{\partial y}\right)^2\sigma_y^2. \tag{5.2.6}$$

The previous discussion can be generalised in order to compute the uncertainty on some function with an arbitrary number of input variables $f = f(x, y, z, \ldots)$ where the observables $u_i = x, y, z, \ldots$ are independent, resulting in

$$\begin{aligned} \sigma_f^2 &= \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + \ldots, & (5.2.7) \\ &= \sum_i \left(\frac{\partial f}{\partial u_i}\right)^2 \sigma_{u_i}^2, & (5.2.8) \end{aligned}$$

and the sum is over the observables.

So given the variances and values of the variables u_i, as well as the form of the differentiable functional f it is possible to determine the variance of f. This general form can be used in order to determine some familiar results, as shown in the following. We shall return to the scenario of correlated observables in section 5.2.4.

5.2.1 Functions of one variable

If the function f has a form that depends only on one observable x, for example

$$f = ax + b \tag{5.2.9}$$

we can use Eq. (5.2.8), where here the sum is over a single term ($i = 1$), and $u_1 = x$, to determine how the error on x is related to the error on f. This is given by

$$\begin{aligned} \sigma_f &= \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2} & (5.2.10) \\ &= a\sigma_x. & (5.2.11) \end{aligned}$$

So for this function, the error on f, given by σ_f, is a times the error on x. This result is independent of any offset of the measured observable. This result can be cross-checked by starting from the definition of variance given in Eq. (3.3.1) and replacing x with $ax + b$ as follows

$$\begin{aligned} V(x) &= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - \overline{ax + b})^2, & (5.2.12) \\ &= \langle (ax+b)^2 \rangle - \langle ax_i + b \rangle^2, & (5.2.13) \\ &= \langle a^2x^2 + 2abx + b^2 \rangle - a^2\langle x \rangle^2 - 2ab\langle x \rangle - b^2, & (5.2.14) \\ &= a^2\langle x^2 \rangle + 2ab\langle x \rangle + b^2 - a^2\langle x \rangle^2 - 2ab\langle x \rangle - b^2, & (5.2.15) \\ &= a^2\langle x^2 \rangle - a^2\langle x \rangle^2, & (5.2.16) \\ \sqrt{V(x)} &= a\sigma_x = \sigma_f. & (5.2.17) \end{aligned}$$

In both cases we obtain the same result, as required.

5.2.2 Functions of two variables

Now consider the function $f = x + y$, where we have measured the mean and standard deviation of both x and y, and want to compute the standard deviation on their sum. We can use the general formula of

Eq. (5.2.8) to determine how to do this, hence

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2, \tag{5.2.18}$$

$$= \sigma_x^2 + \sigma_y^2, \tag{5.2.19}$$

$$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}. \tag{5.2.20}$$

Here we find the result that the variance on f is the sum of variances on x and y. We usually say that errors combined in this way have been ***added in quadrature***. This result is independent of the signs of x and y in f, so it is valid for $f = \pm x \pm y$. By following a very similar derivation we are able to show that the relative error on products or ratio of x and y is also given by the sum in quadrature of the errors on x and y, however this is a task left for the reader.

5.2.3 Functions involving powers of x

The final example considered here is that of the function $f = x^n$. If we have determined the value and uncertainty on x, we can use the general formula of Eq. (5.2.8) once again to determine the corresponding error on f as

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2, \tag{5.2.21}$$

$$\sigma_f = n x^{n-1} \sigma_x. \tag{5.2.22}$$

Here we find a relationship between the uncertainty on f and both the value and uncertainty on x.

5.2.4 Correlations revisited

The concept of correlation between variables was introduced in section 3.6.2. Just as the determination of one variable may depend on the value of another correlated variable, the error on the measurement of one observable or parameter may depend on the value of another correlated observable or parameter. It is necessary to understand in detail the covariance between correlated variables for such a problem in order to be able to correctly compute errors, or combining repeated measurements of an observable that may be correlated. Returning to Eq. (5.2.4), we now see that if the covariance between two variables is non-zero that the last term plays an important role in understanding the uncertainty on the combination.

The result in Eq. (5.2.4) can be generalised for the case where one has an N dimensional problem. Correlated variables will have terms depending on the covariance of the pairs, and uncorrelated terms will take the form of Eq. (5.2.8). This generalised result is

$$\sigma_f^2 = \sum_{i,j} \frac{\partial f}{\partial u_i} \frac{\partial f}{\partial u_j} \sigma_{u_i u_j}, \tag{5.2.23}$$

$$= \sum_i \left(\frac{\partial f}{\partial u_i}\right)^2 \sigma_{u_i}^2 + \sum_{i,j,\, i \neq j} \frac{\partial f}{\partial u_i} \frac{\partial f}{\partial u_j} \sigma_{u_i u_j}, \tag{5.2.24}$$

where the first sum corresponds to the uncorrelated error computation, and the second sum corresponds to the correlated error component for some pair of variables u_i and u_j. The indices i and j are over the N dimensions of the problem. The uncorrelated parts form the diagonal of a matrix, and the correlated parts

are off diagonal terms. Equation (5.2.24) can be expressed in matrix form as

$$\sigma_f^2 = \left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \ldots, \frac{\partial f}{\partial u_n}\right) \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} & & \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 & & \\ & & \vdots & \\ & & & \sigma_{u_n}^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial u_1} \\ \frac{\partial f}{\partial u_2} \\ \vdots \\ \frac{\partial f}{\partial u_n} \end{pmatrix}, \tag{5.2.25}$$

where we see the (symmetric) covariance matrix as the middle term of the right hand side of Eq. (5.2.25). This can be re-written in a compact form as

$$\sigma_f^2 = \Delta^T V \Delta, \tag{5.2.26}$$

where Δ is a column matrix of partial derivatives of f with respect to the u_i and V is the $n \times n$ covariance matrix. Thus if one knows the uncertainties on u_i, covariances between pairs of variables, and is also able to determine the partial derivatives it is possible to compute the error on f via the appropriate matrix multiplication. The error propagation formalism described here is applied to the problem of tracking a moving object, and determining the error on its trajectory in section **??**.

5.2.5 Correlated and uncorrelated uncertainties

The preceding discussion raised the issue that some or all of the uncertainties we are interested in may be correlated with each other. If we have some observable x that has a number of sources of uncertainty, all of which are uncorrelated, then we are able to combine the uncertainties in quadrature in order to obtain the total uncertainty on x using a generalisation of the result given in Eq. (5.2.20).

If however, the uncertainties on x were correlated, it would be wrong to assume that we can add them in quadrature. Instead correlated uncertainties should be added coherently (or linearly) in order to obtain the total uncertainty. This follows from Eq. (5.2.24) as can be seen from the following: Given $f = x + y$, where we have determined $\sigma_x = \sigma_y$, and $\rho_{x,y} = 1$, then it follows that the total error on f is given by

$$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho_{x,y}\sigma_x\sigma_y}, \tag{5.2.27}$$

$$= 2\sigma_x = 2\sigma_y. \tag{5.2.28}$$

Often a measurement will have sources of uncertainty that are both correlated and uncorrelated. In such cases we add the correlated errors linearly, and the uncorrelated parts in quadrature. Finally we can consider calculating the total uncertainty by combining the correlated and uncorrelated sums in quadrature, however this implicitly assumes that the uncertainties are all Gaussian in nature. This may not always be the case for systematic uncertainties discussed in section 5.5.

5.3 Binomial error

It is useful at this point to revisit the binomial probability distribution (section 4.2) in order to discuss the binomial error and its uses. Let us consider the case where we are trying to measure the efficiency of a detector. Our experiment has two states: the first where we detect what we want to measure, and a second when we don't detect a signal. Knowing in advance that we want to measure the detector efficiency we design our experiment with sufficient redundancy so as to be able to compute this. Then we embark on a detailed study to accumulate data over a period of time in order to estimate the detector efficiency. This efficiency is simply the fraction of detected events normalised by the total number of events measured by a

reference system. If our detector is particularly good and only fails to detect one in a thousand events, then the efficiency is $\epsilon = 0.999$. To convince ourselves that we understand the detector efficiency fully we must also determine what the error on ϵ is. We know that the efficiency is

$$\epsilon = \frac{n_d}{n}, \tag{5.3.1}$$

where n_d is the number of detected events, and n is the total number of trials. The detection of events is a binomial quantity, so the variance is given by Eq. (4.2.4) and is $n\epsilon(1-\epsilon)$, where here $p = \epsilon$. From Eq. (5.2.8) it follows that the uncertainty on ϵ is given by

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{n}}. \tag{5.3.2}$$

This uncertainty is sometimes referred to as the ***binomial error*** on the efficiency. A practical way to measure the efficiency of a detector for particles that would pass through it is to sandwich that between two other 'trigger' devices, and correlate the detection of some event in both triggers with the presence or absence of an event being detected in the device under study. The underlying reasoning is that if the two triggers are appropriately set up, then one can be certain that when they both record the passage of a particle, that the test device should also have done so. This method can be adapted to the scenario of estimating the efficiency of a set of planar detectors each of unknown efficiency in a stack.

Example: If an efficiency is measured as $\epsilon = 0.999$, from a sample of 1000 events, then the uncertainty on the efficiency is $\sigma_\epsilon = 0.001$, and so the efficiency should be quoted as $\epsilon = 0.999 \pm 0.001$. It can be seen from Eq. (5.3.2) that a precision determination of an efficiency that is extremely high, or low will require the accumulation of a large number of events.

If we consider the extreme cases where $\epsilon = 0$ or 100%, then the above formula would suggest that $\sigma_\epsilon = 0$. However if one thinks about this, then it is unreasonable to expect that a result where an efficiency is trivially good or bad (maximal or minimal) is a perfect measurement. This is a reflection of our ignorance, or lack of information. In order to compute a reasonable estimate of the uncertainty that may be associated with such an efficiency estimate, one must place an upper or lower limit on the efficiency, a subject discussed in section 6.5.

5.4 Averaging results

5.4.1 Weighted average of a set of measurements for a single observable

If we perform an ensemble of independent measurements to obtain an estimate of the mean and uncertainty of an observable, $x \pm \sigma$, we are free to repeat our experiment in order to accumulate a larger ensemble of data. In such circumstances we can recompute $x \pm \sigma$ for the combined original and new ensembles of data. Usually we don't have the luxury of being able to access all of the data available, for example when previous or competing experiments have made a measurement that we are performing a more precise cross check of. If this is the case, we need to find a way of combining the two results $x_1 \pm \sigma_1$ and $x_2 \pm \sigma_2$. Naively one might be tempted to average x_1 and x_2 by computing $\overline{x} = (x_1 + x_2)/2$, however that is not a particularly useful measure of the combination if one of the measurements is more precisely determined than the other. Only in the limit where $\sigma_1 \simeq \sigma_2$ is a good approximation might we consider the arithmetic mean of x_1 and x_2 a good representation of the average value of x.

In order to overcome this problem we can compute a ***weighted average*** derived using a least squares statistic (see chapter 8) of two uncorrelated measurements, that takes into account the knowledge of uncertainties in

each individual measurement using

$$\overline{x} \pm \sigma_x = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} \pm \left(1/\sigma_1^2 + 1/\sigma_2^2\right)^{-1/2}, \tag{5.4.1}$$

$$= \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \pm \left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)^{1/2}. \tag{5.4.2}$$

When considering n such measurement this generalises to

$$\overline{x} \pm \sigma_x = \frac{\sum\limits_{i=1}^{n} x_i/\sigma_i^2}{\sum\limits_{i=1}^{n} 1/\sigma_i^2} \pm \left(\sum_{i=1}^{n} 1/\sigma_i^2\right)^{-1/2}. \tag{5.4.3}$$

Here each measurement is weighted by the reciprocal of its variance in order to compute the average. The corresponding uncertainty is the square root of the inverse of the sum in quadrature of the reciprocals of the variances of the measurements.

5.4.2 Weighted average of a set of measurements for a set of observables

If one wants to compute a weighted average of a set of correlated observables x from a number of measurements M, then the procedure follows a generalisation of the method used for the uncorrelated case. For the general case (again obtained using the least squares method) one obtains a matrix equation relating the covariance matrix V and the difference between the set of observables x_j from the j^{th} measurement and the mean value of that measurement set $\overline{x}$. As shown in James (2007) the mean values $\overline{x}$ and covariance matrix obtained for the ensemble of measurements are given by

$$\begin{aligned} \overline{x} &= \left[\sum_{j=1}^{M} V_j^{-1}\right]^{-1} \cdot \left[\sum_{j=1}^{M} V_j^{-1} x_j\right], \\ V &= \left[\sum_{j=1}^{M} V_j^{-1}\right]^{-1}. \end{aligned} \tag{5.4.4}$$

If we consider the case of M measurements of a set of two observables a and b, then Eq. (5.4.4) becomes

$$\begin{aligned} \overline{x} &= \left[\sum_{j=1}^{M} \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}_j^{-1}\right]^{-1} \cdot \left[\sum_{j=1}^{M} \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}_j^{-1} \begin{pmatrix} a \\ b \end{pmatrix}_j\right], \\ V &= \left[\sum_{j=1}^{M} \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}_j^{-1}\right]^{-1}, \end{aligned} \tag{5.4.5}$$

where

$$\begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}_j^{-1} = \left[\frac{1}{\sigma_a^2\sigma_b^2 - \sigma_{ab}^2} \begin{pmatrix} \sigma_b^2 & -\sigma_{ab} \\ -\sigma_{ab} & \sigma_a^2 \end{pmatrix}\right]_j. \tag{5.4.6}$$

From this one can see that for uncorrelated observables in the measurement set we can reduce Eq. (5.4.4) to a set of calculations for each of the individual observables. On doing this we recover a set of weighted averages of the form given in Eq. (5.4.3), one for each observable. A worked example of how to compute a weighted average of correlated observables is given in section **??**.

5.5 Systematic errors and systematic bias

Until now we have only considered errors that are statistical in nature. These are uncertainties on a measurement that result solely from our ability to extract information about an observable assuming that the analysis method is perfect. In reality there are other types of errors that have to be considered when making measurements. These are all called ***systematic errors*** and are related to uncertainties in our measuring device (i.e. the scale on a meter rule or its equivalent) as well as anything else that may impact upon our result.

Systematic errors should not be confused with a systematic bias on a measurement. What's the difference between bias and error? A ***bias*** is a systematic offset on a measured observable that results from the procedure used. An error is an uncertainty attributable to an effect. The knowledge of a systematic bias on a measurement also has a systematic uncertainty associated with it. Having determined the value and uncertainty of a bias, one can make a correction to a measured observable to remove the effects of the bias. Such a correction becomes an integral part of the analysis method, and may also provide insight on how to improve the method for future measurements. This can be illustrated through the following example.

> *Having prepared an experiment to measure the lifetime of cosmic ray muons one makes an assumption that the equipment used by one of your colleagues was sufficiently well calibrated. After taking your data you have a measurement with a relative precision of 1%, but you find that this is 10% smaller than anticipated based on the literature. This difference is significant, so you investigate possible sources of bias in your procedure. Finally the problem is traced back to a poor calibration. In an ideal world at this point you would consider re-running the experiment, discarding your existing data, however you have a deadline in a few days and only have time to use the data already collected. In this case, you determine a way that you can quantify the bias from the poor calibration, which corresponds to* $-(10 \pm 0.5)\%$ *on the measured lifetime. So now you can correct your original value by a shift of* $+10\%$*, and combine the error on the bias, with the statistical uncertainty for your report. Your final result is compatible with the literature, and the precision of your measurement is 1.5% adding the statistical and systematic errors assuming they may both be correlated, and highlight that one may improve the technique by re-calibrating more often for future measurements.*

It is important to note that in the above illustration, the experimenter finds a problem with the method, and quantifies how this affects the result, prior to correcting the result and including an additional uncertainty from the correction. One thing that must <u>never</u> be done is to 'massage' or fabricate the data! From time to time experiments make more precise measurements of well known quantities and do find that they are significantly different from those reported in the literature. Such differences are often an indication that the previous method had a systematic effect that was unknown or overlooked and sometimes these will simply be the result of statistical fluctuations. If the people doing those repeat experiments had massaged their data in order to agree with the previous result, they would obtain the wrong results and would have wasted considerable resources in doing so. It is important that you remain an objective scientist when making measurements, and treat surprising results with the same (or more) diligence that you would an expected result.

Sometimes the systematic error associated with the correction will turn out to be small, sometimes it will dominate the total uncertainty reported. Similarly sometimes the bias from an effect will be negligible, and sometimes, as in the above illustration it will not.

Statistical errors are Gaussian in nature, which is why we add them in quadrature. Systematic errors are by definition almost certainly not. When dealing with systematic uncertainties there is no firm rule on how to combine them. Often one will find that they are combined in quadrature, sometimes they will be added linearly, and the other option is to add some in quadrature and some linearly depending on their nature. Kirkup and Frenkel (2006) discuss systematic uncertainties on measurement for a number of concrete examples. The treatment of systematic errors has to be considered carefully on a case by case basis, and experience will help significantly in determining what to do with them. One definite case is outlined in the following example.

> *On measuring the probability for a particle to decay into a final state with two neutral pions (π^0), we introduce a systematic uncertainty based on our ability to reconstruct and identify both particles. This uncertainty is typically 3% per π^0 for the BABAR experiment, which was built at the SLAC National Laboratory in California during the late 1990s, and took data between 1999 and 2008. With a $2\pi^0$ final state, we have to add the uncertainty on the reconstruction efficiency for each π^0 linearly because the effect of reconstructing one π^0 is completely correlated with the effect of reconstructing the second. So the systematic uncertainty on measuring such a final state on the BABAR experiment has a total contribution from π^0 reconstruction of 6%. This uncertainty can now be combined in quadrature with any other uncertainties that are considered to be uncorrelated with it in order to obtain the total uncertainty on the measurement.*

5.6 Blind analysis technique

The concept of blind analysis is not new, for example Dunnington (1933) measured the value of e/m for the electron this way. This technique has gained popularity in recent years. The previous section discussed the need to remain objective when making a measurement, and to be as diligent as possible so as not to introduce experimenter bias. That is by favouring one outcome over another the experimenter could unwittingly perform a measurement that enables them to draw the favoured conclusion, irrespective of what the data is actually telling them. The most natural way to unwittingly introduce experimenter bias in a measurement would be to stop investigating possible systematic effects and sources of bias when one has obtained a result in agreement with a previous experiment. This is not a sufficient requirement to ensure that the result you have obtained is correct.

The method behind performing a ***blind analysis*** is to obscure the observable x being measured in an unknown, but repeatable, way such that a measurement can be performed giving the result x_{blind}. In many circumstances, in addition to extracting x_{blind}, it is possible to extract the systematic errors on x_{blind} (and subsequently x) and to perform all necessary cross checks and validation studies without the need to determine the actual value of the observable being measured. Only once the analysis of data has been completed is the observable 'un-blinded' to give the result $x \pm \sigma_x$, where σ_x is the combined statistical and systematic uncertainty on both x and x_{blind}.

This may seem a rather strange way to perform a measurement, however it does remove any potential for experimenter bias to influence a result. Of course if a mistake is found in the procedure once the observable has been un-blinded, that mistake has to be rectified, otherwise an incorrect un-blind result would be reported. The benefit of performing a blind analysis is that experimenter bias will be absent. The detriment of performing such an analysis is that sometimes it is can be impossible or extremely impractical to perform all necessary checks on a particular analysis in order to extract the observable. Where it is possible to perform

a blind analysis, the experimenter needs to realistically weigh the advantages against the disadvantages of doing so, including understanding how one can perform any and all necessary cross-checks on the analysis. An example of such a case is one where you search for a particular signal in a sample of limited statistics, and the signal itself depends on a number of parameters, all of which are unknown. How can you perform a blind analysis of something you don't even know exists? If there are a limited number of unknowns it may be feasible to perform a blind analysis when searching for a new effect. When it is not possible to adopt a blind analysis approach, the experimenter needs to rely on remaining objective and ensure that there are sufficient processes in place in an analysis chain to avoid unwittingly biasing the end result. Ultimately as long as the experimentalist remains objective then the procedures imposed by a blind analysis technique are unnecessary.