Statistical Methods of Data Analysis - WS 00/01
Chapter 4

Ian C. Brock

14th September 2006 – 09 : 33

Contents

4 Errors

Now know some of the expected distributions, as well as ways to characterise the data. Next want to consider how to treat sources of error, how to combine different sources and how to combine different measurements, properly taking into account their errors.
First step is to find out why errors are almost always assumed to be Gaussian distributed.

4.1 Central Limit Theorem

Theorem states that if we take the sum, X, of N *independent* variables, x_i, where $i = 1, 2, 3, ..., n$, and each of the x_i is taken from a distribution with mean μ_i and variance V_i (or σ_i^2) then the distribution of X has the following characteristics:

1. Expectation value:

$$\langle X \rangle = \sum_{i=1}^{N} \mu_i$$

2. Variance:

$$V(X) = \sum_{i=1}^{N} V_i = \sum_{i=1}^{N} \sigma_i^2$$

3. tends to a Gaussian distribution as $N \to \infty$, independent of the form of then individual distributions.

It is the 3rd characteristic that makes the Gaussian so important. Proof of the first 2 characteristics is straightforward:

1. Expectation Value:

$$\begin{aligned} \langle X \rangle &= \langle \sum_i x_i \rangle \\ &= \sum_i \langle x_i \rangle \\ &= \sum_i \mu_i \end{aligned}$$

2. Variance:

$$\begin{aligned} V(X) &= \langle (X - \langle X \rangle)^2 \rangle \\ &= \langle (\sum_i x_i - \sum_i \mu_i)^2 \rangle \\ &= \left\langle \left[\sum_i (x_i - \mu_i) \right]^2 \right\rangle \\ &= \langle \sum_i (x_i - \mu_i)^2 \rangle + \langle \sum_i \sum_{j \neq i} (x_i - \mu_i)(x_j - \mu_j) \rangle \\ &= \sum_i \langle (x_i - \mu_i)^2 \rangle + \sum_i \sum_{j \neq i} \langle (x_i - \mu_i)(x_j - \mu_j) \rangle \end{aligned}$$

   So the variance is given by the sum of the variances plus the sum of the covariances of every measurement. But, the covariance of independent measurements is 0. Stated requirement for CLT to work is that the measurements must be independent. Thus variance of the sum is $\sum V_i$.

Proof of 3rd part is much harder. See Appendix in Barlow or Kendall and Stuart for a more thorough treatment.
Maybe better than a proof is a demonstration. Already done one in the exercises.

⇒ *Transparency* LAPTOP clt.mnf: Uniform distribution

⇒ *Transparency* LAPTOP clt.mnf: Uneven distribution

⇒ *Transparency* LAPTOP clt.mnf: Exponential distribution

See that with enough variables (enough depends on the distribution), Gaussian form appears, even for asymmetric or very uneven distributions.
In order for CLT to work it is important that $V \gg V_i$
Consider a measurement which has a number of error sources. The measured value is then given by:

$$x_i = \langle x_i \rangle + \delta x_i^{(1)} + \delta x_i^{(2)} + \delta x_i^{(3)} + ... + \delta x_i^{(N)}$$

where δx_i are the different error sources. For each of the error sources:

$$\langle \delta x_i^{(k)} \rangle = 0$$

and $\sigma_i^{(k)}$ is the measurement error due to source k.
The CLT says that x_i is Gaussian distributed around $\langle x_i \rangle$ with a variance given by the sum of the individual variances.
Where can the CLT be applied?

4.1.1 Repeated Measurements

Assume that we repeat a measurement many times. Can use the CLT to predict what the results of such measurements would be. In this case all the expectation values, μ_i, are the same μ and all the σ_i are also the same σ.
Thus:

$$\langle X \rangle = \sum \mu = N\mu$$

The mean value of x is then given by $\overline{x} = X/N$:

$$\langle \overline{x} \rangle = \mu$$

Although we are measuring the same thing, if the errors are random, then each measurement can and should be treated as an independent measurement:

$$\begin{aligned} V(\overline{x}) &= \frac{1}{N^2} \sum V_i \\ &= \frac{\sigma^2}{N} \end{aligned}$$

What is $\langle \overline{x} \rangle$ and its variance? It means that if we make a series of measurements and then average them, $\langle \overline{x} \rangle$ is the value we expect to get. Of course the actual measured value will differ from the expectation value. $V(\overline{x})$ says how much we expect it to differ. Thus if we

look at the above we see that when we average a series of measurements we expect to get closer and closer to true value as a function $1/\sqrt{N}$. Repeated measurements win with the square root of the number of measurements.
One also sees that improving the accuracy of a measurement by a factor of 2 is as good as repeating the measurement 4 times!
Take a series of N measurements with an error σ:

$\Rightarrow$ *Transparency* LAPTOP emean.mnf: c.f. Error on measurement with error on mean

4.1.2 Weighted Mean

In above assumed a series of measurements, each with the same error. How to we combine measurements with different errors and how can we use the CLT to help here?
Measure the speed of a car with 2 different radar devices:

$$\begin{aligned} v_1 &= 67 \pm 4\,\mathrm{ms}^{-1} \\ v_2 &= 63 \pm 2\,\mathrm{ms}^{-1} \end{aligned}$$

Simple average of measurements would give:

$$\begin{aligned} \overline{v} &= (63+67)/2 = 65\,\mathrm{ms}^{-1} \\ V(v) &= (4^2+2^2)/4 = 5 \\ \sigma(v) &= \sqrt{5} = 2.24\,\mathrm{ms}^{-1} \end{aligned}$$

which is somehow very unsatisfactory, as error on mean is larger than resolution on best measurement.
If we had taken 4 measurements with the device that has a resolution of $4\,\mathrm{ms}^{-1}$, then error on average would be $2\,\mathrm{ms}^{-1}$. Thus measurement with resolution of ± 2 should be given 4 times the weight of measurement with error ± 4.

$$\begin{aligned} \overline{v} &= \frac{4\times 63 + 1\times 67}{(1+4)} = 63.8\,\mathrm{ms}^{-1} \\ V(v) &= \frac{4^2\times 5}{5^2} = 3.2 \\ \sigma(v) &= 1.8\,\mathrm{ms}^{-1} \end{aligned}$$

or more generally we should give each measurement a weight that is inversely proportional to the square of the resolution:

$$\begin{aligned} \overline{x} &= \frac{\sum_i x_i^2/\sigma_i^2}{\sum_i 1/\sigma_i^2} \\ V(\overline{x}) &= \frac{1}{\sum_i 1/\sigma_i^2} \end{aligned}$$

What would you do if the measurements were:

$$\begin{aligned} v_1 &= 67 \pm 4\,\mathrm{ms}^{-1} \\ v_2 &= 53 \pm 2\,\mathrm{ms}^{-1} \end{aligned}$$

Could just follow standard procedure and obtain:

$$\overline{v} = 55.8 \pm 1.8\,\mathrm{ms}^{-1}$$

Hope you would be very unhappy with such a result. Mean with error where neither of the 2 measurements is within 1σ of the mean! You could decide that one of the measurements is just wrong! A very risky procedure. You can even pretend to be scientific and let the computer throw out bad measurements for you. BUT:

> *So unexpected was the hole that for several years computers analysing ozone data had systematically thrown out readings that should have pointed to its growth*
>
> – New Scientist, 31 March 1988

⇒ *Transparency* New Scientist on ozone hole

Thus be very careful doing such things. You may even miss your chance for a Nobel prize! Also always remember that while you expect 2/3 of the measurements to be within 1σ of the mean, the other 1/3 should be outside. If you have much more than 5% outside 2σ you should start to get suspicious!

4.1.3 Particle Data Group Averaging Procedure

What can one do then? Or to turn the question around what is a reasonable estimate of the error on the mean, when the measurements vary by more than their errors indicate that they should? The "Particle Data Group" have developed a procedure, which they use to combine the various measurements of particle properties etc. into a so-called "world average". They first calculate the weighted mean as described above. They then calculate:

$$\chi^2 = \sum_i \frac{(x_i - \overline{x})^2}{\sigma_i^2}$$

which is called the χ^2 for the mean. Will come back later to use χ^2 to fit data to theoretical distribution. For now, if you expect each measurement to differ from the mean by about 1σ see that we expect $\chi^2 \approx N$.
There are 3 cases to consider:

1. $\chi^2/(N-1) < 1$: Everything is OK – use simple weighted average, assuming that there are no known problems with the data.

2. $\chi^2/(N-1) \gg 1$: Depending on reason, can choose not to make an average at all, or quote calculated average and then make an educated guess of the error, taking into account known problems with the data.

3. $\chi^2/(N-1) > 1$: This indicates that the errors on some or all of the measurements may have been underestimated. As we have no further information, a reasonable procedure is to scale all the errors by a factor

$$S = \sqrt{\chi^2/(N-1)}$$

so that $\chi'^2/(N-1)$ calculated using S is 1.

Can also go a step further and consider what to do when combining large and small errors. Evaluate S using only those experiments with small errors. Make a cut and only use those errors for which:

$$\sigma_i < \sigma_0 = 3\sqrt{N}\,\sigma_{\overline{x}}$$

when calculating S, i.e. only use those measurements that have an error that is smaller than 3 times the error one would have on each measurement, if all the measurements had the same error.

Reason is that while large errors do not contribute to value of $\overline{x}$ and $\sigma_{\overline{x}}$, they can make significant contributions to S.

Property of this procedure is that if we have 2 experiments with values separated by much more than their errors, then the scaled-up error is approx. half the interval between the 2 measurements.

As a aid one can make a so-called ideogram.

⇒ *Transparency* PDG ideogram

Each measurement is represented by a Gaussian with mean x_i and error σ_{x_i} and area proportional to $1/\sigma x_i$. Choice of area is somewhat arbitrary, but may be better when some experiments have underestimated their systematic errors.
Note that this procedure works if the errors on all measurements are independent. If this is not the case, then one must evaluate the correlations between the measurements and take them into account when averaging. Come to this topic in a later lecture.

4.2 Combination of Errors

Have discussed how to combine several measurements with their errors. But how should one combine the different errors within an experiment. Probably already know all the standard simple formulae, but I will go through them anyway, as a first step towards the more general formulation.
In normal case we measure a particular quantity, but in order to get to the physical parameter that we want we have to calculate or function of that quantity, or a function of a series of measurements of different quantities.

4.2.1 Functions of a Single Variable

Simplest case, f is a linear function of x:

$$f = ax + b$$

where a and b are constants and x has a distribution with variance V, or error σ. We need to calculate the variance on f:

$$\begin{aligned} V(f) &= \langle f^2 \rangle - \langle f \rangle^2 \\ &= \langle (ax+b)^2 \rangle - \langle (ax+b) \rangle^2 \\ &= a^2\langle x^2 \rangle + 2ab\langle x \rangle + b^2 - a^2\langle x \rangle^2 - 2ab\langle x \rangle - b^2 \\ &= a^2(\langle x^2 \rangle - \langle x \rangle^2) \\ &= a^2 V(x) \end{aligned}$$

or $\sigma_f = |a|\sigma_x$, as one would expect, as a is a multiplicative constant and b cannot affect the width.
In more general case when f is an arbitrary function of x we can make a Taylor expansion about a point x_0:

$$f(x) \approx f(x_0) + (x - x_0)\left(\frac{df}{dx}\right)_{x=x_0}$$

Thus:

$$\begin{aligned} V(f) &\approx \left(\frac{df}{dx}\right)^2 V(x) \\ \sigma(f) &\approx \left|\frac{df}{dx}\right| \sigma(x) \end{aligned}$$

However, note that in order for this to be valid, the Taylor expansion has to be valid. That implies that df/dx should be a constant within 2-3 σ of the value x_0. In theory, the expansion should also be made about the "true" value of x. If this is not known, one has to use the measured value.
For example, if one can measure an angle θ with an accuracy of 0.01, then $\sin\theta$ is known with an accuracy of $0.01|\cos\theta|$.

4.2.2 Functions of Two or More Variables

Start with a linear function again:

$$\begin{aligned} f &= ax + by + c \\ V(f) &= a^2(\langle x^2 \rangle - \langle x \rangle^2) + b^2(\langle y^2 \rangle - \langle y \rangle^2) + 2ab(\langle xy \rangle - \langle x \rangle\langle y \rangle) \\ &= a^2 V(x) + b^2 V(y) + 2ab\, cov(x, y) \end{aligned}$$

More generally we can again do a Taylor expansion of each of the variables about a point (x_0, y_0) and find:

$$\begin{aligned} V(f) &= \left(\frac{df}{dx}\right)^2 V(x) + \left(\frac{df}{dy}\right)^2 V(y) + \left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right) cov(x,y) \quad \text{or} \\ \sigma_f^2 &= \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + \left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right) \rho\, \sigma_x \sigma_y \end{aligned}$$

Note that again the approximation is only valid for *small* errors. Also note that the signs of the derivatives have to be properly taken into account here.
If x and y are independent of each other then we arrive at the famous Gaussian error propagation formula:

$$\begin{aligned} V(f) &= \left(\frac{df}{dx}\right)^2 V(x) + \left(\frac{df}{dy}\right)^2 V(y) \quad \text{or} \\ \sigma_f^2 &= \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 \end{aligned}$$

Can easily extend formula to include further variables:

End of Lecture 3

$$\sigma_f^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + \left(\frac{df}{dz}\right)^2 \sigma_z^2$$

Use a trigonometric example again to see how it works:

$$y = A\sin\theta + B\cos\theta$$

where in this case A, B and θ are measurements with errors. Then:

$$\sigma_y^2 = \sin^2\theta\,\sigma_A^2 + \cos^2\theta\,\sigma_B^2 + (A\cos\theta - B\sin\theta)\sigma_\theta^2$$

assuming that the A, B and θ are independent of each other.

4.2.3 Standard Formulae for Error Propagation

Go through some standard formulae to combine errors. Some are easier to remember if expressed as fractional or percentage errors.

- Addition and subtraction:

$$\begin{aligned} f &= x \pm y \\ V(f) &= V(x) + V(y) \quad \text{or} \\ \sigma_f^2 &= \sigma_x^2 + \sigma_y^2 \end{aligned}$$

- Product:

$$\begin{aligned} f &= xy \\ V(f) &= y^2V(x) + x^2V(y) \quad \text{or} \\ \left(\frac{\sigma_f}{f}\right)^2 &= \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 \end{aligned}$$

  i.e. you some the squares of the fractional errors in the case of the product of 2 measurements.

- Quotient:

$$\begin{aligned} f &= x/y \\ V(f) &= \left(\frac{1}{y}\right)^2 V(x) + \left(-\frac{x}{y^2}\right)^2 V(y) \quad \text{or} \\ \left(\frac{\sigma_f}{f}\right)^2 &= \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 \end{aligned}$$

- Powers of x:

$$\begin{aligned} f &= x^2 \\ V(f) &= (2x)^2V(x) \\ \sigma_f &= 2x\sigma_x \quad \text{or} \\ \frac{\sigma_f}{f} &= 2\frac{\sigma_x}{x} \end{aligned}$$

$$\begin{aligned} f &= 1/x^2 \\ V(f) &= \left(-\frac{2}{x^3}\right)^2 V(x) \\ \frac{\sigma_f}{f} &= 2\frac{\sigma_x}{x} \end{aligned}$$

- Logarithm:

$$\begin{aligned} f &= \ln x \\ V(f) &= \left(\frac{1}{x}\right)^2 V(x) \\ \sigma_f &= \frac{\sigma_x}{x} \end{aligned}$$

- Efficiency: What about when you want to measure an efficiency or acceptance? Could say this has already been discussed. If a particle goes through a detector, have either

signal or no signal, so the number of cases in which you have a signal must follow a binomial distribution:

$$\begin{aligned} \epsilon &= n/N \\ V(\epsilon) &= N\epsilon(1-\epsilon)/N^2 \\ \sigma_\epsilon &= \sqrt{\frac{\epsilon(1-\epsilon)}{N}} \end{aligned}$$

Should be able to get the same result using error propagation. However, this is not a simple quotient. The n signals seen are part of N. Let $N_1 = n$ and $N_1 + N_2 = N$. N_1 and N_2 are then independent measurements. They are the number of times I see a signal and the number of times I do not.

$$\begin{aligned} \epsilon &= \frac{N_1}{N_1+N_2} \\ V(\epsilon) &= \left(\frac{\partial\epsilon}{\partial N_1}\right)^2 V(N_1) + \left(\frac{\partial\epsilon}{\partial N_2}\right)^2 V(N_2) \\ &= \left(\frac{1}{N_1+N_2} - \frac{N_1}{(N_1+N_2)^2}\right)^2 V(N_1) + \left(\frac{-N_1}{(N_1+N_2)^2}\right)^2 V(N_2) \\ &= \frac{1}{(N_1+N_2)^4}\left(N_2^2 V(N_1) + N_1^2 V(N_2)\right) \end{aligned}$$

What are $V(N_1)$ and $V(N_2)$? In other words how much should the measured values of N_1 and N_2 vary about their true values? The fluctuations about the true values are random. Given an efficiency ϵ we expect $N_1 = N\epsilon$ signals (or events). For N_1 sufficiently large previously showed that a Binomial distribution tends to a Poisson. In order for error propagation to be valid, the error on N_1 should be "small", which implies that N_1 is "large". In the case the distribution of N_1 about this value is then a Poisson and as such has mean $N\epsilon$ and variance $N\epsilon = N_1$. Similarly for N_2. It should have a mean of $N(1-\epsilon)$ and a variance of $N(1-\epsilon) = N_2$. Thus:

$$\begin{aligned} V(\epsilon) &= \frac{1}{(N_1+N_2)^4}\left(N_1 N_2^2 + N_2 N_1^2\right) \\ &= \frac{N_1 N_2}{(N_1+N_2)^3} \\ &= \frac{\epsilon(1-\epsilon)}{N} \end{aligned}$$

as required.

4.2.4 General Case

How do we treat things if we have several functions of several variables and want to know the errors on the functions? Take n variables:

$$x_{(1)}, x_{(2)}, ..., x_{(n)}$$

Parentheses denote n elements of a single measurement, rather than the N measurements that comprise the sample. In earlier lecture defined covariance for 2 variables in a sample:

$$\begin{aligned} cov(x_{(i)}, x_{(j)}) &= \overline{(x_{(i)} - \overline{x_{(i)}})(x_{(j)} - \overline{x_{(j)}})} \\ &= \overline{x_{(i)}x_{(j)}} - \overline{x_{(i)}}\,\overline{x_{(j)}} \end{aligned}$$

In this case we have 2 variables in a joint probability distribution function $P(x_{(1)}, x_{(2)}, ..., x_{(n)})$, which describes the probability to measure values $x_{(1)}, x_{(2)}, ..., x_{(n)}$. The expectation value for each of the variables is given by $\mu_1, \mu_2, ..., \mu_n$ and $\mu_i \equiv \langle x_{(i)} \rangle$. The covariance can then be defined as:

$$\begin{aligned} cov(x_{(i)}, x_{(j)}) &= \langle (x_{(i)} - \mu_i)(x_{(j)} - \mu_j) \rangle \\ &= \langle x_{(i)}x_{(j)} \rangle - \mu_i\mu_j \end{aligned}$$

These are elements of the covariance matrix (or error matrix):

$$V_{ij} = cov(x_{(i)}, x_{(j)})$$

Diagonal elements are just the variances. Define correlation matrix as the dimensionless equivalent:

$$\rho_{ij} = \frac{cov(x_{(i)}, x_{(j)})}{\sigma_i\sigma_j}$$

Elements lie between -1 and $+1$ and tell you how much the two variables are correlated. If not obvious that these are the limits consider a simple example:

$$\begin{aligned} y &= x \\ z &= \pm 2x \end{aligned}$$

i.e. y and z are 100% either correlated or anticorrelated (if I know one then I know the other). Thus:

$$\begin{aligned} \sigma_y &= \sigma_x \\ \sigma_z &= 2\sigma_x \\ \Rightarrow cov(y,z) &= \langle \pm 2x^2 \rangle - \langle x \rangle \langle \pm 2x \rangle \\ &= \pm \int 2x^2 P(x)\,dx \mp 2\left(\int xP(x)\,dx\right)^2 \\ \Rightarrow \rho_{yz} &= \frac{\pm 2\sigma_x^2}{2\sigma_x^2} \\ &= \pm 1 \end{aligned}$$

Clearly the same applies whatever constant I put in front of x and as mentioned before an offset has no effect on the width of a distribution.

Get more general and suppose that we have m functions $f_1, f_2, ..., f_m$
of n variables $x_{(1)}, x_{(2)}, ..., x_{(n)}$. Each of the $x_{(i)}$ has an associated error or variance and therefore so do the f_k. The f_k are also correlated, independent of whether the $x_{(i)}$ are correlated, because the different f_k share the same values $x_{(i)}$. Calculate the variances on the f_k in the usual way:

$$V(f_k) \;=\; \langle f_k^2 \rangle - \langle f_k \rangle^2$$

Then expand f_k as a Taylor series around the expectation values for the $x_{(i)}$:

$$f_k \;\approx\; f_k(\mu_1, \mu_2, ..., \mu_n) + \left(\frac{\partial f_k}{\partial x_{(1)}}\right)(x_{(1)} - \mu_1) + \left(\frac{\partial f_k}{\partial x_{(2)}}\right)(x_{(2)} - \mu_2) + \cdots$$

Thus the f_k are now a linear combination of the $x_{(i)}$ and we can put this expansion into the variance formula, as was done before:

$$\begin{aligned} V(f_k) \;&=\; \left(\frac{\partial f_k}{\partial x_{(1)}}\right)^2 \langle (x_{(1)} - \mu_1)^2 \rangle + \cdots + 2\left(\frac{\partial f_k}{\partial x_{(1)}}\right)\left(\frac{\partial f_k}{\partial x_{(2)}}\right)\langle (x_{(1)} - \mu_1)(x_{(2)} - \mu_2)\rangle + \cdots \\ &=\; \sum_i \left(\frac{\partial f_k}{\partial x_{(i)}}\right)^2 V(x_i) + \sum_i \sum_{j \neq i} \left(\frac{\partial f_k}{\partial x_{(i)}}\right)\left(\frac{\partial f_k}{\partial x_{(j)}}\right) cov(x_{(i)}, x_{(j)}) \end{aligned}$$

which looks just like the standard combination of errors formula given in previous lecture. But, f_k and f_l are also correlated, so we have to determine their covariance:

$$\begin{aligned} \langle f_k f_l \rangle - \langle f_k \rangle \langle f_l \rangle \;\approx\; &\langle (x_{(1)} - \mu_1)(x_{(1)} - \mu_1)\rangle \left(\frac{\partial f_k}{\partial x_{(1)}}\right)\left(\frac{\partial f_l}{\partial x_{(1)}}\right) + \cdots \\ &+ \langle (x_{(1)} - \mu_1)(x_{(2)} - \mu_2)\rangle \left(\frac{\partial f_k}{\partial x_{(1)}}\right)\left(\frac{\partial f_l}{\partial x_{(2)}}\right) + \cdots \end{aligned}$$

Getting horribly complicated to follow all the indices. Try expressing things as sums: **Keep on board**

$$cov(f_k, f_l) \;=\; \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_{(i)}}\right)\left(\frac{\partial f_l}{\partial x_{(j)}}\right) cov(x_{(i)}, x_{(j)})$$

which also includes the variance formula if $k = l$.
Even this is not so easy to remember. However if we introduce a matrix:

$$G_{ki} \;=\; \left(\frac{\partial f_k}{\partial x_{(i)}}\right)$$

and $\mathbf{V_x}$, $\mathbf{V_f}$ are the error matrices for $\mathbf{x}$ and $\mathbf{f}$ respectively, then: **Keep on board**

$$\mathbf{V_f} \;=\; \mathbf{G V_x \tilde{G}}$$

which is a nice easy formula to remember. It includes everything we have mentioned to do with error propagation. You just have to remember that $\mathbf{V_f}$ and $\mathbf{V_x}$ are symmetric with dimensions $m \times m$ and $n \times n$ respectively, while $\mathbf{G}$ is rectangular with dimension $m \times n$. Apply it to a couple of examples:

- A function f that is a function of 2 variables x and y, who have errors σ_x and σ_y and a correlation coefficient ρ. Then

$$\mathbf{G} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

$$\begin{aligned}\mathbf{V_f} &= \left(\tfrac{\partial f}{\partial x}, \tfrac{\partial f}{\partial y}\right) \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \\ V_f &= \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right)\rho\sigma_x\sigma_y\end{aligned}$$

  which is the standard error propagation formula with two correlated variables.

- Changing variables in 3-dimensional coordinates. Typical cylindrical tracking chamber measures the track of a particle in 3 dimensions. Natural coordinates for for such a chamber are cylindrical (r, ϕ, z). Construction of chamber is such that r is known to high accuracy, so that we independent errors on ϕ and z to consider. Want to know what are the errors on Cartesian coordinates (x, y, z)?

  Using $x = r\cos\phi$ and $y = r\sin\phi$ we have:

$$G = \begin{pmatrix} \cos\phi & -r\sin\phi & 0 \\ \sin\phi & r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned}V_{xyz} &= \begin{pmatrix} \cos\phi & -r\sin\phi & 0 \\ \sin\phi & r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -r\sin\phi & r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} r^2\sigma_\phi^2\sin^2\phi & -r^2\sigma_\phi^2\sin\phi\cos\phi & 0 \\ -r^2\sigma_\phi^2\sin\phi\cos\phi & r^2\sigma_\phi^2\cos^2\phi & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_\phi^2 y^2 & -\sigma_\phi^2 xy & 0 \\ -\sigma_\phi^2 xy & \sigma_\phi^2 x^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}\end{aligned}$$

  Formula also looks reasonable. In region close to $\phi = 0$, i.e. $y = 0$ error is only on y and not on x, as x error is radial there. Also see that the correlation between the error on x and the error on y is -1.

  Could clearly also include an error on r, but formulae would start getting pretty complicated.

4.3 Systematic Errors

Above treatment was for random errors (fluctuations). How should we handle systematic errors, both within a single measurement and when we combine measurements? First what do we mean by a systematic error? Typically it is an error due to some flaw or inaccuracy in the measuring apparatus that is common for all measurements that we make. For example if I am measuring a voltage and the meter has an offset of 0.1 V, then all my measurements will have a systematic error of 0.1 V. This I should be able to see by plotting voltage against current and seeing that the line does not go through zero. If the meter also has a scale error then this is a bigger problem, as the line goes nicely through zero, but its slope is wrong!
First feature of systematic errors is that repeating the measurements does not help, i.e. the usual effect seen in random errors that the errors decrease with $(1/\sqrt{N})$ does not apply. You can turn the argument around and say that the measurements are not independent of each other, so the CLT does not apply when we try to combine such measurements.

4.3.1 Evaluating Systematic Errors

Before we consider how to combine systematic errors with random errors, or how to combine measurements that have systematic errors, think a bit on how to evaluate them, or even better, how to avoid them.
Most important is to consider what possible sources of error can occur at the planning stage of your experiment. You then have to make intelligent guesses on their possible size. It is particularly important to consider how one can calibrate the devices that you want to use to do the measurement and to consider how you can check that the calibration is still valid as your measurement proceeds. Do not forget environmental factors such as temperature, pressure and humidity than can all have an effect on your apparatus. You should always record the values of such quantities, so that you can afterwards look to see if they influence your measurement, by looking for correlations for example.
Suppose you have a voltmeter and are using it to measure current through a resistor. Clearly the temperature of the resistor can change as the current changes, but also the room temperature could change. In order to see the effect of this you should repeat your measurements under the same nominal conditions at random times and then plot the results vs. room temperature. You can look to see if they are randomly scattered and/or evaluate the correlation coefficient.
What do you do with tolerances? A meter might be quoted to measure voltage to ± 0.1V, or your technician has constructed your chamber with an accuracy of 0.2mm. Should you then just plug these values in as your systematic error? No! The accuracy of the construction means that if a particular dimension is 10 mm, then your technician guarantees that it is between 9.8 and 10.2 mm. Given no further information, the probability distribution for the size within this range is flat. You should therefore use an error of $0.4/\sqrt{12} = 0.12$ mm.
Also do not forget to include the errors on "constants" when evaluating the total error. Suppose you are measuring temperature with a thermocouple. To calibrate the device you

measure the voltages at a couple of standard temperatures (ice water and steam) and then use the formula:

$$T = \frac{(T_2 - T_1)}{(V_2 - V_1)}(V - V_1) + T_1$$

In order to evaluate error on T, you need to know resolution of your voltmeter, but you should also remember to include this in the errors on the values of V_1 and V_2. Best would be to repeat the measurements of V_1 and V_2 several times to reduce the size of the error and also to check that the resolution is really that which you expect. While this will not help you if there if a scale problem, it will reduce the error due to the resolution.
Also important to consider how to proceed with your measuremements. Try to randomise the order of your data as much as possible. Do not measure the current at voltages of 0,1,2,...,10 V. Much better to measure at 7,3,4,6,9,1,... V, i.e. in a random order to randomise the effect of drifts in the apparatus. You don't get rid of effects by doing this, but you do randomise the effect. If you have hysterisis effects and are interested in an absolute value, measure sometimes coming from below and sometimes from above. If you are interested in a slope, then always come from the same side. Then the differences are always the same, even if the absolute value is not correct.

4.3.2 Combining Systematic Errors

Once you have evaluated and/or estimated the size of your systematic errors, you have to combine the various sources. As the errors are independent of each other they should be combined in quadrature. In many cases one or two errors dominate, so those are the ones you have to work on most to reduce and/or estimate. Just remember that most of the time the values that you give for the systematic errors are not much better tha educated guesses.
You now have the statistical and systematic errors. Given that they are again independent of each other you can also use the CLT and combine them in quadrature. However, it is very often helpful to the reader to quote them separately:

$$A = -10.2 \pm 0.4(\text{stat}) \pm 0.3(\text{sys})$$

This way of quoting the result means that one sees at a glance the relative size of the errors. It is also common (although statistically not strictly correct) for the systematic error to be estimated *conservatively*; or in other words to be overestimated. If you have forgotten a possible source this may not be so bad, but who says that your *conservative* estimate compensates correctly for a totally different source that you have not considered. If you want to combine several experiments, that often have correlated systematic errors, then it is essential that the statistical and systematic errors be quoted separately.

4.3.3 Propagating Systematic Errors

Can use the techniques learnt earlier to combine measurements with common systematic errors. Suppose we have two measurements, x_1 and x_2, each with a common systematic

error, S, and random errors σ_1 and σ_2. How to make a covariance matrix? Treat x_1 as if it were made of 2 parts: x_1^R with error σ_1 and x_1^S with common systematic error S. Similarly for x_2. x_1^R and x_2^R are thus uncorrelated, while x_1^S and x_2^S are completely correlated. Variance of x_1 is given by:

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

as x_1^R and x_2^S are independent. Note that this also justifies adding the 2 terms in quadrature to form the total error. What is the covariance?

$$\begin{aligned} cov(x_1, x_2) &= \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle \\ &= \langle (x_1^R + x_1^S)(x_2^R + x_2^S) \rangle + \langle (x_1^R + x_1^S) \rangle \langle (x_2^R + x_2^S) \rangle \\ &= \langle x_1^S x_2^S \rangle - \langle x_1^S \rangle \langle x_2^S \rangle \\ &= S^2 \end{aligned}$$

as 3 of the 4 cross products involve x^R, which are by definition uncorrelated and therefore 0. On the other hand x_1^S and x_2^S are absolutely correlated. We can now write down the variance (error) matrix:

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

Once we have this matrix we are set! We can propagate errors, calculate correlation coefficients, etc. etc.

Sometimes the systematic error is a fraction of the value rather than being an absolute value. In this case $S_1 = \epsilon x_1$ and $S_2 = \epsilon x_2$. Following the same procedure as before we can again calculate the covariance between x_1 and x_2 They are still completely correlated so:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + \epsilon^2 x_1^2 & \epsilon^2 x_1 x_2 \\ \epsilon^2 x_1 x_2 & \sigma_2^2 + \epsilon^2 x_2^2 \end{pmatrix}$$

Even more generally, if we have several measurements and some of the errors are correlated between all the measurements and some are only correlated between a subset of the measurements, we can still follow the above procedure, and just split the measurements into more parts. Suppose we have 3 measurements. S is the correlated systematic error between all 3 measurements and T is correlated between measurement 1 and measurement 2. Then:

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

Note that this is also the standard procedure. You always try to split the measurements into parts that are either independent or completely correlated. This is much easier to work with, than trying to determine correlation coefficients!

End of Lecture 4

4.3.4 Example

Suppose we use a pendulum to measure how the length of a bar changes with temperature. Measure period, t, and then use known value (with error) of g to determine the length of the pendulum, l. Then:

$$\begin{aligned} t &= 2\pi\sqrt{\frac{l}{g}} \\ \Rightarrow l &= \frac{gt^2}{4\pi^2} \end{aligned}$$

g is given with an error S_g and the measurement of the period has a random error σ_t and a systematic error S_t. Error on g is common to both measurements, so is denoted with S_g. Using the standard error propagation:

$$\begin{aligned} \sigma_l^2 &= \left(\frac{1}{4\pi^2}\right)^2 \left[(t^2)^2 S_g^2 + (2gt)^2(\sigma_t^2 + S_t^2)\right] \\ \left(\frac{\sigma_l}{l}\right)^2 &= \left(\frac{S_g}{g}\right)^2 + 4\left(\frac{\sigma_t^2 + S_t^2}{t^2}\right) \end{aligned}$$

If we make 2 measurements of l then there is a non-zero covariance between the 2. Best way to determine it is to write down the error matrices and the derivatives and then use the form:

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

where:

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} \frac{\partial l_1}{\partial t_1}, \frac{\partial l_1}{\partial t_2}, \frac{\partial l_1}{\partial g} \\ \frac{\partial l_2}{\partial t_1}, \frac{\partial l_2}{\partial t_2}, \frac{\partial l_2}{\partial g} \end{pmatrix} \\ &= \begin{pmatrix} 2\frac{l_1}{t_1} & 0 & \frac{l_1}{g} \\ 0 & 2\frac{l_2}{t_2} & \frac{l_2}{g} \end{pmatrix} \end{aligned}$$

and the error matrix on the measurements is given by:

$$\mathbf{V}_{t_1 t_2 g} = \begin{pmatrix} \sigma_t^2 + S_t^2 & S_t^2 & 0 \\ S_t^2 & \sigma_t^2 + S_t^2 & 0 \\ 0 & 0 & S_g^2 \end{pmatrix}$$

and we can then multiply these together to get:

$$V(l_1, l_2) = \begin{pmatrix} 4\frac{l_1^2}{t_1^2}(\sigma_t^2 + S_t^2) + \frac{l_1^2}{g^2}S_g^2 & 4\frac{l_1 l_2}{t_1 t_2}S_t^2 + \frac{l_1 l_2}{g^2}S_g^2 \\ 4\frac{l_1 l_2}{t_1 t_2}S_t^2 + \frac{l_1 l_2}{g^2}S_g^2 & 4\frac{l_2^2}{t_2^2}(\sigma_t^2 + S_t^2) + \frac{l_2^2}{g^2}S_g^2 \end{pmatrix}$$

See that only the common systematic errors contribute to the covariance between the measurements, while all 3 error sources contribute to the variance on the length.

Assume that what we really want to know is the difference in length between the bar at 2 different temperatures. The variance on $l_1 - l_2$ can then be determined in the same way, again using the matrix form with:

$$\begin{aligned}\mathbf{G} &= \left(\frac{\partial(l_1-l_2)}{\partial t_1}, \frac{\partial(l_1-l_2)}{\partial t_2}, \frac{\partial(l_1-l_2)}{\partial g}\right) \\ &= \left(2\frac{l_1}{t_1}, -2\frac{l_2}{t_2}, \frac{l_1-l_2}{g}\right)\end{aligned}$$

$$\mathbf{V}_{t_1 t_2 g} = \begin{pmatrix} \sigma_t^2 + S_t^2 & S_t^2 & 0 \\ S_t^2 & \sigma_t^2 + S_t^2 & 0 \\ 0 & 0 & S_g^2 \end{pmatrix}$$

$$\begin{aligned}V(l_1-l_2) &= 4\sigma_t^2\left[\left(\frac{l_1}{t_1}\right)^2 + \left(\frac{l_2}{t_2}\right)^2\right] + 4S_t^2\left[\left(\frac{l_1}{t_1}\right) - \left(\frac{l_2}{t_2}\right)\right]^2 + S_g^2\left(\frac{(l_1-l_2)}{g}\right)^2 \\ &= \frac{g^2}{16\pi^4}\left[4\sigma_t^2(t_1^2+t_2^2) + 4S_t^2(t_1-t_2)^2\right] + \left(\frac{S_g}{g}\right)^2 (l_1-l_2)^2 \\ &= \frac{g^2}{16\pi^4}\left[4\sigma_t^2(t_1^2+t_2^2) + 4S_t^2(t_1-t_2)^2 + \left(\frac{S_g}{g}\right)^2 (t_1^2-t_2^2)^2\right]\end{aligned}$$

Note that if length were proportional to period, rather than period squared, then dependence on systematic error on t would drop out. This is independent of whether systematic error is fractional or absolute.

See that calculations get quite long and easily prone to error, but procedure is relatively straightforward. The matrix form is by far the easiest to remember and use in most cases. You see which partial derivatives you have to evaluate and it is usually easy to write down the error matrix of the actual measurements with any correlations.

Come back to this topic later when talking about different methods to combine several measurements and how to decide what the best way to do this is, when errors are both uncorrelated and when they are correlated.