

Part I. Experimental Error

1 Types of Experimental Error. There are always blunders, mistakes, and screwups; such as: using the wrong material or concentration, transposing digits in recording scale readings, or arithmetic errors. There are no fancy techniques that we know about which will save you from these sorts of errors; you just have to be careful and keep your wits about you. In general, careful recording and repeating of measurements helps. All recorded numbers should go directly into your notebook which helps to find and fix some kinds of errors such as arithmetic ones. The following types of errors are ones that we can address quantitatively.

1.1 Systematic error. Systematic errors are consistent effects which change the system or the measurements made on the system under study. They have *signs*. They might come from: uncalibrated instruments (balances, etc.), impure reagents, leaks, unaccounted temperature effects, biases in using equipment, mislabelled or confusing scales, seeing hoped-for small effects, or pressure differences between barometer and experiment caused by air conditioning. Systematic error affects the *accuracy* (closeness to the true value) of an experiment but not the *precision* (the repeatability of results). Repeated trials and statistical analysis are of no use in eliminating the effects of systematic errors.

Whenever we recognize systematic errors, we correct for them. Sometimes systematic errors can be corrected for in a simple way; for example, the thermal expansion of a metal scale can easily be accounted for if the temperature is known. Other errors, such as those caused by impure reagents, are harder to address. The most dangerous systematic errors are those that are unrecognized, because they affect results in completely unknown ways. The quality of results depends critically on an experimenter's ability to recognize and eliminate systematic errors.

1.2 Random error. There are always random errors in measurements. They do not have signs, i.e. fluctuations above the result are just as likely as fluctuations below the result. If random errors are not seen, then it is likely that the system is not being measured as precisely as possible. Random error arises from mechanical vibrations in the apparatus, electrical noise, uncertainty in reading scale pointers, the uncertainty principle, and a variety of other fluctuations. It can be characterized, and sometimes reduced, by repeated trials of an experiment. Its treatment is the subject of much of this course. Random error affects the precision of an experiment, and to a lesser extent its accuracy. Systematic error affects the accuracy only. Precision is easy to assess; accuracy is difficult. All of the following methods concern themselves with the treatment of random error.

2 Measurements of a Single Quantity. If you measure some quantity N times, and all the measurements are equally good as far as you know, the best estimate of the true value of that quantity is the mean of the N measured values:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1).$$

A measure of the uncertainty in the true value obtained in this way is the estimated standard deviation (e.s.d.) of the mean,

$$S_m = \frac{S}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} \quad (2),$$

where S is the estimated standard deviation of the distribution. We use the word "estimated" because this relation only becomes the standard deviation as N goes to ∞ . When N is small, the estimate is worse. Note the distinction between S_m and S . As N goes to ∞ , S_m goes to zero, but S goes to a fixed value, i.e. the uncertainty of the mean goes to zero, but there will still be a fixed width in the distribution of values.

2.1 Significant Figures and Reporting Results. The estimated error of a result (such as S_m above) determines the number of significant figures to be reported in the result. All results should be reported with units, uncertainties, the appropriate number of significant figures, and the kind of uncertainty used, i.e. number \pm uncertainty units (type of uncertainty). Since this is very important, let's reiterate: the uncertainty in a quantity determines the number of significant figures in the quantity. As a general rule, **uncertainties are given with one significant figure if the most significant figure in the uncertainty is greater than 2, and two significant figures if the most significant figure in the uncertainty is less than or equal to 2.** The quantity is then reported with the same number of decimal places as the least significant figure in the uncertainty. One should also specify the type of uncertainty, such as an estimated first standard deviation (e.s.d.) or a 95% confidence level. Usually the result and uncertainty are reported with a common exponent. For example, suppose that the mean of a set of measurements was given on a calculator as 2.167598×10^{-4} in units of seconds and the estimated standard deviation of the mean was 2.27349×10^{-5} s. This result would be reported as $(2.17 \pm 0.23) \times 10^{-4}$ s (e.s.d.).

It is even better to report *confidence limits* that carry information about the number of measurements. A 95% confidence limit is a range within which 95% of randomly varying measurements will fall. The standard deviation of the distribution S (as defined in eq. 2) is technically a 68.3% confidence limit in the case of large N . There are methods and tabulations that can give us better confidence limits based on estimated standard deviations and specific values of N when N is small. For example, the mean of eight measurements might be reported as

$$\bar{x} \pm tS_m \text{ units (95\% confidence, } \nu=7) \quad (3)$$

where $t=2.36$ and $\nu=N-1$ for $N=8$ at 95% confidence (see Table 2 on p. 46 of Shoemaker, Garland, and Nibbler).

3 Propagation of Error

3.1 Formula Approach. The quantity of interest often cannot be measured directly, but is calculated from other quantities which are measured. In this case we need a way to obtain the uncertainty in the final quantity from the uncertainties in the quantities directly measured. A quantity F is calculated from several measured quantities, for instance x , y , z , so $F = F(x, y, z)$. The total differential of F is by the chain rule

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial z} dz \quad (4).$$

The total differential gives the infinitesimal change in F caused by infinitesimal changes in x , y , or z . If we approximate the change in F brought about by small but finite changes in x , y , and z (such as small, well-behaved errors) by a similar formula, we obtain

$$\Delta F = \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y + \frac{\partial F}{\partial z} \Delta z \quad (5).$$

This is equivalent to saying that the surface $F(x, y, z)$ is a plane over the region in space $[x \pm \Delta x, y \pm \Delta y, z \pm \Delta z]$ and curvature over that small region is not important. If the errors in x , y , and z are small and known in both sign and magnitude, i.e. systematic, the above equation can be used to propagate the errors and find the resulting error in F . In practice when the systematic errors are known, it is both easier and more accurate to simply recalculate F using corrected values of x , y , and z . In other words we don't ordinarily propagate systematic errors, just random errors.

To handle random errors, we must perform averages. The average random error in F is zero, so instead we calculate $\langle (\Delta F)^2 \rangle$. Squaring both sides of eq. 5

$$(\Delta F)^2 = \left(\frac{\partial F}{\partial x} \right)^2 (\Delta x)^2 + \left(\frac{\partial F}{\partial y} \right)^2 (\Delta y)^2 + \left(\frac{\partial F}{\partial z} \right)^2 (\Delta z)^2 + 2 \left(\frac{\partial F}{\partial x} \right) \left(\frac{\partial F}{\partial y} \right) \Delta x \Delta y + 2 \left(\frac{\partial F}{\partial x} \right) \left(\frac{\partial F}{\partial z} \right) \Delta x \Delta z + 2 \left(\frac{\partial F}{\partial y} \right) \left(\frac{\partial F}{\partial z} \right) \Delta y \Delta z \quad (6).$$

Now we need to average $(\Delta F)^2$ over many determinations of F , each with different values of the errors in x , y , and z and the cross terms will tend to cancel. Given that the errors are small, symmetrically distributed about 0, and uncorrelated, you may use the "full random error propagation equation" given below

$$(\Delta F)^2 = \left(\frac{\partial F}{\partial x} \right)^2 (\Delta x)^2 + \left(\frac{\partial F}{\partial y} \right)^2 (\Delta y)^2 + \left(\frac{\partial F}{\partial z} \right)^2 (\Delta z)^2 \quad (7),$$

where F is the interesting quantity calculated from the directly measured quantities x , y , and z . Δx , Δy , and Δz are the uncertainties in x , y , and z as calculated before. The partial derivatives should be evaluated at the mean values of the quantities. Note that you may use any form of error measure you like; standard deviation, 95% confidence interval, etc.,

so long as you use the same form for each of the independent variables. The result ΔF will then be of that form.

3.1.1 Example for Small, Uncorrelated, Random Errors. As an example, let's calculate the number of moles of a gas from measurements of its pressure, volume, and temperature:

$$n = \frac{pV}{RT} \quad (8)$$

Assume that the gas is ideal (a possible source of systematic error). Perhaps our measurements of p , V , and T are: $p = 0.268 \pm 0.012$ atm (e.s.d), $V = 1.26 \pm 0.05$ L, and $T = 294.2 \pm 0.3$ K. Recall that $R = 0.082056 \pm 0.000004$ L atm mol⁻¹ K⁻¹. If we punch these numbers into a calculator, we find that $n = 0.013987894$ moles. But how many of these digits are significant? To find out, we propagate the errors in p , V , R , and T using eq. 7

$$\Delta n = \sqrt{\left(\frac{\partial n}{\partial p}\right)^2 (\Delta p)^2 + \left(\frac{\partial n}{\partial V}\right)^2 (\Delta V)^2 + \left(\frac{\partial n}{\partial R}\right)^2 (\Delta R)^2 + \left(\frac{\partial n}{\partial T}\right)^2 (\Delta T)^2} \quad (9).$$

To use this, we must determine the partial derivatives:

$$\frac{\partial n}{\partial p} = \frac{V}{RT}, \quad \frac{\partial n}{\partial V} = \frac{p}{RT}, \quad \frac{\partial n}{\partial T} = -\frac{pV}{RT^2}, \quad \frac{\partial n}{\partial R} = \frac{-pV}{R^2 T} \quad (10)$$

At this point it is very important to realize that a separate calculation of each term in eq. 9 can be used to determine which variable contributes the most to the error in the number of moles. So we calculate each term separately

$$\begin{aligned} \left(\frac{\partial n}{\partial p}\right)^2 (\Delta p)^2 &= \left(\frac{V}{RT}\right)^2 (\Delta p)^2 = \left(\frac{1.26}{0.082056 \cdot 294.2}\right)^2 (0.012)^2 = 3.92 \times 10^{-7} \text{ mol}^2 \\ \left(\frac{\partial n}{\partial V}\right)^2 (\Delta V)^2 &= \left(\frac{p}{RT}\right)^2 (\Delta V)^2 = \left(\frac{0.268}{0.082056 \cdot 294.2}\right)^2 (0.05)^2 = 3.08 \times 10^{-7} \text{ mol}^2 \\ \left(\frac{\partial n}{\partial R}\right)^2 (\Delta R)^2 &= \left(\frac{-pV}{R^2 T}\right)^2 (\Delta R)^2 = \left(\frac{0.268 \cdot 1.26}{0.082056^2 \cdot 294.2}\right)^2 (0.000004)^2 = 4.65 \times 10^{-13} \text{ mol}^2 \\ \left(\frac{\partial n}{\partial T}\right)^2 (\Delta T)^2 &= \left(\frac{-pV}{RT^2}\right)^2 (\Delta T)^2 = \left(\frac{0.268 \cdot 1.26}{0.082056 \cdot 294.2^2}\right)^2 (0.3)^2 = 2.03 \times 10^{-10} \text{ mol}^2 \end{aligned} \quad (11).$$

So the pressure error is the most important error in this determination, i.e. it is the limiting error and the volume error is also important. Considering that it always costs

time and energy to determine uncertainties, if you were to halve the error in temperature, it would hardly affect the result, i.e. you wasted your time, energy, and probably someone's money. Note that the error in the gas constant R is the smallest contribution and a negligible contribution. It usually would not even be included among the variables propagating error, but this is how you know such things. Finally, the error in n is

$$\Delta n = \sqrt{3.92 \times 10^{-7} + 3.08 \times 10^{-7} + 4.65 \times 10^{-13} + 2.03 \times 10^{-10}} = 0.0008367 \text{ mol} \quad (12).$$

Using our significant figure rules, the most significant figure in the uncertainty is "8", so we will have only one sig. figure in the uncertainty. Noting the proper rounding, the result would be reported as $n = (1.40 \pm 0.08) \times 10^{-2} \text{ mol (e.s.d.)}$. A Mathcad template of this example follows.

4 Strategies when Error Analysis gets Tough.

4.1 Numerical Differentiation. If the function $F(x, y, z)$ is complicated and it is very difficult to determine the partial derivatives analytically, you can evaluate them numerically. In general, evaluation of numerical derivatives is less desirable, but some problems become analytically intractable. Program your calculator or Mathcad to evaluate F given input variables x, y, z . Then evaluate the derivatives as

$$\frac{\partial F}{\partial x} \approx \frac{F(x + \Delta x, y, z) - F(x - \Delta x, y, z)}{2\Delta x} \quad (13),$$

with similar formulas for y and z . For each derivative you evaluate F twice, so if you have three independent variables you will evaluate it six times. It's therefore very useful to have a programmable way to evaluate $F(x, y, z)$. Then insert the numerical values of the partial derivatives you found into the propagation of random errors formula (eq. 7) to evaluate the propagated error.

4.2 Full Covariant Propagation. If the independent variables in use did not result directly from different independent measurements, then the assumption of uncorrelated errors may be violated. For instance, two parameters of a fit, such as the slope and intercept of a line fitted to some pairs of data, might both be used to determine the final result. If these parameters are correlated, then so are their errors. In this case, the cross terms of eq. 6 will not cancel and we have (for the two variable case)

$$(\Delta F)^2 = \left(\frac{\partial F}{\partial x} \right)^2 (\Delta x)^2 + \left(\frac{\partial F}{\partial y} \right)^2 (\Delta y)^2 + 2 \left(\frac{\partial F}{\partial x} \right) \left(\frac{\partial F}{\partial y} \right) \sigma_{xy} \quad (14).$$

In this expression the errors are standard deviations and their squares are known as variances. The quantity σ_{xy}^2 is known as the *covariance* of x and y . A good least squares

Mathcad Example Propagating Random Error

input data (all errors are e.s.d.)

Pressure and its error in atm	$p := 0.268$	$\delta p := 0.012$
Volume and its error in L	$V := 1.26$	$\delta V := 0.05$
Temperature and its error in K	$T := 294.2$	$\delta T := 0.3$
Gas constant and its error in L atm mol ⁻¹ K ⁻¹	$R := 0.082056$	$\delta R := 0.000004$

From the ideal gas law we have $n := \frac{p \cdot V}{R \cdot T}$ $n = 0.013988$ mol

Finding the Error in n

To propagate these errors we need the derivatives which can be done with symbolics

the derivative of $\frac{p \cdot V}{R \cdot T}$ with respect to p is $\frac{V}{(R \cdot T)}$ with respect to V is $\frac{p}{(R \cdot T)}$
 with respect to R is $-p \cdot \frac{V}{(R^2 \cdot T)}$ with respect to T is $-p \cdot \frac{V}{(R \cdot T^2)}$

we need the individual terms to evaluate the relative importance of various errors, so

$$\sqrt{\left(\frac{V}{R \cdot T}\right)^2 \cdot \delta p^2} = 6.263 \times 10^{-4} \quad \text{most important}$$

$$\sqrt{\left(\frac{p}{R \cdot T}\right)^2 \cdot \delta V^2} = 5.551 \times 10^{-4} \quad \text{also important}$$

$$\sqrt{\left(\frac{-p \cdot V}{R^2 \cdot T}\right)^2 \cdot \delta R^2} = 6.819 \times 10^{-7} \quad \text{not important, usually not included}$$

$$\sqrt{\left(\frac{-p \cdot V}{R \cdot T^2}\right)^2 \cdot \delta T^2} = 1.426 \times 10^{-5} \quad \text{Not important unless other errors reduced}$$

$$\delta n := \sqrt{\left(\frac{V}{R \cdot T}\right)^2 \cdot \delta p^2 + \left(\frac{p}{R \cdot T}\right)^2 \cdot \delta V^2 + \left(\frac{-p \cdot V}{R^2 \cdot T}\right)^2 \cdot \delta R^2 + \left(\frac{-p \cdot V}{R \cdot T^2}\right)^2 \cdot \delta T^2} \quad \delta n = 8.37 \times 10^{-4}$$

final result is $(1.40 \pm 0.08) \times 10^{-2}$ mol (e.s.d)

fitting routine will provide the covariance between x and y as one of its outputs. Usually the covariances are small and can be ignored; however when things are getting tricky and the normal error analysis is not making sense or perhaps the theory of the problem is not well established, it can be informative to consider correlations.

4.3 Monte Carlo Approach. When the errors are large, correlated, or otherwise problematic, the best approach to error analysis is *Monte Carlo simulation*. This method is a superior way to evaluate errors, but involves more computation. The idea is simple. Use a computer to generate many (perhaps 100-1000) synthetic data sets, just as we averaged over many hypothetical data sets in the section above. The data sets should have values of the independent variables drawn as closely as possible from the same parent distribution that applied in your experiment. Then, for each synthetic data set, calculate a value of F the same way you did for your real data set. Make a histogram of the resulting values of F in order to examine the resulting distribution, i.e. divide the range of F values into bins and plot the number of results that fall within each bin. To determine the 95% confidence limit, you might find two values in that list which enclose 95% of all the values, i.e. sort the list of values and find the values at each end that are in 2.5% of the values from the ends.

The Monte Carlo method has several advantages over classical propagation of error: 1) Often the errors in F are not normally distributed, particularly when $F(x,y,z)$ is nonlinear, even though the raw data x,y,z are normally distributed. The Monte Carlo method gives the correct distribution for F . That means it works correctly even when the assumption that the errors are small and/or normal is violated. 2) Even if the errors in the independent variables are not symmetrically distributed, this method works correctly as long as the simulations are done with the correct input (Δx , Δy , Δz) distributions. 3) Monte Carlo analysis can be done even when the evaluation of F from the data involves very complicated calculations such as nonlinear fits and Fourier transforms, so long as the calculations are automated.

The disadvantage is that one must both (1) have a computer with a good random number generator available, and (2) be pretty handy with it.

4.3.1 Generating Data Sets with Synthetic Errors. The idea is to use the computer to simulate many experiments. Perhaps you measured the pressure, temperature, and volume of a gas sample in order to calculate the number of moles. You perform each of the measurements ten times, so that you have some idea of the extent of random error in the measurements, and can estimate the input error distribution. In simple measurements like these, the normal or Gaussian distribution is often a good one to use, and you can calculate the average values of p , V , T , and the estimated standard deviations of their distributions. The procedure is to generate a huge set of (p,V,T) triples to subject to analysis. Each generated value should be drawn randomly from a normal distribution with the appropriate mean (the mean value from your experiment) and standard deviation (the esd of the distribution from your experiment).

How can we obtain such randomly drawn numbers? This question belongs to an interesting branch of computer science called "seminumerical algorithms." The rather deep mathematics behind the construction of good random number generators will not be discussed, but we will begin by assuming that you have available a routine which

produces on each call a "random" number between 0 and 1. In Mathcad the function is $\text{rnd}(1)$. These "random numbers" are essentially drawn from the continuous *uniform* distribution, $P(x)$, between 0 and 1:

$$P(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (15).$$

Notice that $P(x)$ is normalized, since its integral is 1. To get random numbers drawn from other distributions, we can scale or transform the output from the available routine. For instance, to obtain uniformly distributed random numbers between -5 and 5, we calculate $r_i = 10(x_i - 1/2)$, where x_i is a random number drawn from the uniform distribution described above. To obtain numbers drawn from distributions other than uniform, transformations are required. For example, to obtain random numbers drawn from an exponential distribution with decay constant τ ,

$$P(r) = \begin{cases} \frac{1}{\tau} e^{-\frac{r}{\tau}} & \text{if } r \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (16),$$

calculate $r_i = -\tau \ln(x_i)$. The most common distribution function of errors is the normal or Gaussian distribution. To obtain random numbers drawn from the normal distribution, you must draw two random numbers x_1 and x_2 from the uniform distribution between 0 and 1, and then calculate

$$r = \sqrt{-2\ln x_1} \cos(2\pi x_2) \quad (17)$$

To get normal random values for the parameters of interest, for example parameter s with a standard deviation of the distribution of σ and a mean value of μ , one uses

$$s_i = \sigma r_i + \mu. \quad (18).$$

Now we have enough information to generate synthetic data sets with normally distributed errors. For each independent variable, which in your experiment had the mean, \bar{p} , and standard deviation σ_p , generate lots of random numbers r_i from the normal distribution and then calculate lots of simulated data points

$$p_i = r_i \sigma_p + \bar{p} \quad (19).$$

Do this for each of your independent variables, and you now have many computer-generated data sets. Analyzing synthetic data is easy: you just do exactly the same things to each synthetic data set that you did to your real data set. Since you have hundreds or thousands of synthetic data sets, clearly the analysis procedure should be automated. In the end you should have a long list of values of F .

4.3.2 Analyzing Synthetic Errors in Final Result. There are two ways of using the list of F values you obtain to get confidence limits. If both are available to you, we recommend you do both. One is good at giving you a feeling for the importance of errors in your experiment and the probability distribution of the result, and the other is more convenient for getting out numerical error limits.

4.3.2.1 Histogram Method. You now have a collection of calculated Fs. You can make a histogram of them, by making a plot of the number of values which fall within bins or intervals along the possible range of F values. Many programs can do this for you; it's usually called a "histogram" or a "frequency plot." Smaller bin widths give better characterization of the distribution, but more noise; so a compromise is involved in the choice of bin width. The histogram gives you a visual description of the probability distribution of F. If the histogram looks like a normal distribution, i.e. it is symmetric and bell shaped, you can meaningfully extract a mean value and standard deviation of F using eqs. 1 and 2. The primary reason for making the histogram is to see whether F really is distributed normally; often it isn't.

4.3.2.2 Sorting Method. This method is quick and easy, and gives good confidence limits, but doesn't give as good a feel for the F distribution as the histogram method. Get the computer to sort the collection of Fs from lowest to highest (use the sort function in Mathcad). Then just read confidence limits from that list. If you want 95% confidence limits, for example, and you have 1000 simulated points, the 25th value from each end of the list (values 25 and 975) give the lower and upper confidence limits.

4.3.3 Mathcad Monte Carlo Example. A Monte Carlo error propagation using the same example as above follows.

