Ed Eyler 2003

Notes on experimental uncertainties and their propagation

These notes are not intended as a complete set of lecture notes, but instead as an enumeration of some of the key statistical ideas needed to obtain quantitative results from the analysis of straightforward measurements. We will gradually cover these issues in much more detail as the course progresses.

I. References

A number of good texts on data analysis for the physical and engineering sciences are available. One of the best is *Data Reduction and Error Analysis for the Physical Sciences*, Third Edition, by Philip R. Bevington and D. Keith Robinson (McGraw Hill, 2003). Much of this short book is accessible to a second-year student, but it nevertheless presents a wealth of techniques and information that are sufficient to see many scientists through their entire careers. Most of the numerical techniques are accompanied by examples in the form of computer programs in Fortran, or in the Web supplement to the text, in C++. At a more introductory level, there are numerous books, particularly *An Introduction to Error Analysis,* Second Edition, by John R. Taylor (University Science Books, 1997), This books is oriented towards the student laboratory, so it is easier to follow, but has less to say about the analysis of real-world experiments.

II. Population Statistics

It is useful to devise a few statistical measures of the properties of experimental data that are independent of any presumptions about the model to be used in analyzing the experiment. The most widely-used such measures are the **mean** or **average** and the **standard deviation**. The mean can be determined whenever a minimum of two individual measurements of the same quantity are available, according to

$$\textbf{mean:} \quad \bar{x} \equiv \frac{1}{N}\sum_i x_i$$

The standard deviation is a measure of the scatter of repeated measurements, and can be estimated only if you take at least three measurements under nominally identical conditions. Thus in a quantitative experiment subject to random fluctuations it is usually wise to take each data point *at least* this many times. The sample standard deviation is often denoted by s to distinguish it from the (population) standard deviation σ of the underlying probability distribution, though sometimes σ is used for both. It is best to keep them distinct, since s is actually an experimental *estimate* for σ, which becomes exact in the limit of a large number of measurements N. It is calculated according to,

$$\textbf{standard deviation:} \quad s \equiv \sqrt{\frac{1}{N-1}\sum_i (x_i - \bar{x})^2}$$

Most calculators have this or a similar formula pre-programmed. Sometimes a factor of $1/N$ instead of $1/(N\text{-}1)$ is used; the version given here is preferable when the mean has been determined from experimental data and not from previous knowledge of the underlying parent distribution.

The value of s gives an estimate of the uncertainty σ associated with each individual measurement. To determine from this the uncertainty of the mean or any other derived quantity, we must use the techniques for error propagation described in section V.

III. Statistical distributions--"parent" vs. "experimental" statistics

The population statistical measures in section II only describe the fluctuations of the measured values. To say anything more we must make some assumptions about the statistical distribution that describes these fluctuations. By way of illustration let's look at a particular example. One of the most common types of physical measurement is a counting measurement, where we measure the pulses from a detector, or some other signal that depends on the number of electrons or photons arriving at a particular point. Such a process, where the time of arrival of the next count is independent of when the previous count arrived, is called a *Poisson process*. The probability of observing a particular number x of counts in a unit time interval is governed by the *Poisson probability distribution* with mean μ,

$$P(x) \equiv \frac{\mu^x}{x!} e^{-\mu}$$

The mean μ is the average number of counts that are observed if the measurement is repeated many times. Thus if we take a sample of experimental data and find the sample mean $\bar{x}$ as in section II, we determine an *approximate* value for the underlying mean μ, which will become increasingly accurate as the sample size is increased. A similar result applies for the standard deviation. It can be shown that if data are accumulated under identical conditions until N counts are expected, the standard deviation in the number of counts is given by $\sigma = \sqrt{N}$. In particular, if we sample for a unit time interval we expect μ counts, so we say that the standard deviation of the Poisson distribution is $\sqrt{\mu}$. If the population standard deviation is determined as described in section II, it should yield a result close to the actual value of σ. Because it is estimated from fluctuating data, it will not give an exact result, although for large N the estimate will become quite good. In discussing this situation we say that the Poisson distribution is the *parent distribution*, whose parameters are exact constants and must be clearly distinguished from the population statistics obtained from experimental results sampled from this distribution.

It is worth pointing out that the fractional accuracy of a measurement, $N/\sigma = 1/\sqrt{N}$, improves only as the square root of the number of counts accumulated. If it takes one day to acquire data giving a measurement accurate to 10%, it will take 100 days to reach an accuracy of 1%! This discouraging property is an unavoidable accouterment of life as an experimentalist.

Other distributions also arise commonly in physics and engineering. One is the binomial distribution, which characterizes coin flips and many thermodynamic processes. Another is the Lorentzian distribution, which describes atomic line shapes. It has the unique peculiarity that there is no finite value for the standard deviation, so it is instead characterized by its full width at half-maximum! Most important of all is the Gaussian distribution, which is a limiting case of both Poisson and binomial distributions.

IV. The Gaussian or normal distribution

This is by far the most commonly used statistical distribution, since it is the limiting case of other important distributions. It has two parameters, the mean μ and the standard deviation. The Gaussian probability function is given by

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right]$$

This is one of the most ubiquitous functions in physics and engineering. An indication of its importance is given by observing that this function and its graph are printed on the German 10-Deutschmark bill! It

is a *continuous* probability distribution, so the probability of observing a value for x lying within a given range is given by its integral. More specifically, the probability dP of measuring a value within a range dx around x is given by,

$$dP = P(x)dx$$

Just as for the Poisson distribution, the population statistics for the mean and standard deviation can be used to estimate the underlying mean and standard deviation if they are not already known. The standard distribution is a measure of the width of the distribution. The probabilities are approximately 68% and 95% that a measurement will fall within one and two standard deviations of the mean, respectively. Though this distribution is exceedingly common, it can be overused: it is sometimes incorrectly assumed that virtually all random variables obey Gaussian statistics! Thus we often see absurd discussions of quantities like IQ scores or economic fluctuations in which detailed predictions are made by assuming strict adherence to the functional form given above.

V. Propagation of uncertainties to derived quantities

Assume that we have measured a number of quantities $u, v, \cdots$ and that we know the standard deviations $\sigma_u, \sigma_v, \cdots$ for each. Ordinarily we know these standard deviations by estimating them from the population standard deviations s as described in section II, although we may know their values in some other way. How do we find the standard deviation σ_x for a derived quantity $x = f(u, v, \cdots)$? This can be solved by using the *error propagation equation*, which can be derived using a combination of the chain rule for derivatives and a Taylor expansion. The resulting formula, the most important result given in this entire discussion, is

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \cdots + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \cdots$$

Terms like the last one are called *covariance* terms, and are only needed when the variables u and v are not independent of one another. This type of correlated fluctuation usually does not arise in a simple experiment. The covariance σ_{uv} of two variables can be estimated by taking N repeated samples much like the standard deviation. The estimate is given by

$$s_{uv}^2 = \frac{1}{N-1}\sum_i (u_i - \bar{u})(v_i - \bar{v}).$$

Examples:

1. **Sums:** $x = au \pm bv$, where a, b are constants. From the above, we find that

$$\sigma_x^2 = a^2\sigma_u^2 + b^2\sigma_v^2 + 2ab\sigma_{uv}^2.$$

   Ordinarily only the first two terms will be needed. This is the single most common situation that's encountered.

2. **Standard deviation of the mean** (sometimes called the *standard error* of the mean) This is a special case of a sum, so we apply the treatment above. Assume we take N independent measurements x_i of a randomly varying quantity. Since the measurements are independent and they scatter randomly, the covariance between them is zero. Since they are taken under identical conditions,

they all have the same standard deviation σ_{x_i}. If we then calculate the mean and propagate the uncertainties, we find that,

$$\text{if } \bar{x} = \frac{1}{N}\sum_i x_i, \text{ then } \sigma_{\bar{x}} = \frac{1}{\sqrt{N}}\sigma_{x_i}.$$

Thus, as you might expect, the mean is better-determined than the individual measurements. In most real experiments, we use the population standard deviation *s* to estimate the standard deviations in the individual measurements, so the estimate for the standard deviation of the mean will be given by $s/\sqrt{N}$. In analyzing the results of an experiment, one typically uses the mean values calculated from the measurements to find derived quantities like focal lengths or magnifications, so the uncertainties of these mean values are the ones that should be propagated. *In summary, if you use the mean values to determine some other derived quantity, then the standard deviations you insert into the error propagation formulas for sums, products, and so forth should be the standard deviations of the mean values. They are smaller by a factor* $1/\sqrt{N}$ *than the uncertainties of the individual data points.*

3. **Products (or division):** $x = \pm auv$ or $\pm a\frac{u}{v}$. This gives the result,

$$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + 2\frac{\sigma_{uv}^2}{uv},$$

where the + sign for the covariance term must be changed to a - sign for the *u/v* case.

4. **Powers:** $x = au^{\pm b}$. This general power law gives the error propagation formula,

$$\frac{\sigma_x}{x} = b\frac{\sigma_u}{u}.$$

A particular common case is the square, $x=D^2$. In this case we just obtain $\sigma_x = 2D\sigma_D$.

5. **A real formula.** In optics, the focal length of a lens can be determined from two measured lengths, the object-image distance *L* and the separation *D* between the two positions at which a lens will form a focused image of the image at the object plane. The focal length is given by

$$f = \frac{L^2 - D^2}{4L}.$$

This is a particularly nice example because it has some nuances. If you break it into two terms and attempt to use the formulas 1-4 for differences, squares, and divisions, you quickly notice that both terms depend on *L*, so there's a non-zero covariance between them because they aren't independent. This is typical with a formula of any complexity, and makes a real mess out of something that initially seems simple. You can avoid this problem by instead using the error propagation formula directly, so everything is done in terms of the independently measured quantities *L* and *D*. The necessary derivatives are:

$$\frac{\partial f}{\partial L} = \frac{1}{4} + \frac{D^2}{4L^2} \quad \text{and} \quad \frac{\partial f}{\partial D} = \frac{-D}{2L}.$$

The result for the standard deviation of f is then,

$$\sigma_f = \sqrt{\left(\frac{1}{4} + \frac{D^2}{4L^2}\right)^2 \sigma_L^2 \; + \; \frac{D^2}{4L^2}\sigma_D^2}$$