UNCERTAINTY IN MEASUREMENT:
NOISE AND HOW TO DEAL WITH IT

On a full-grown Coast Live Oak there are, by rough estimate, over a million leaves, in general all the same, but in detail, all different. Like fingerprints, no one leaf is exactly like any other. Such variation of pattern is *noise*.

A mountain stream flows with an identifiable pattern of waves and ripples, but with no pattern repeating itself exactly, either in space or in time. That too is *noise*.

Clouds form, which we may classify as cirrus, or cumulus or nimbus, but no one cloud is exactly like any other. More noise.

Noise, that delightfully random bit of disorder that is present everywhere, is an essential ingredient of our physical universe, to be understood, appreciated and revered.

One has only to imagine a world without noise: the leaves of a plant without variation of pattern, a stream without random gurglings, a campfire without random flickerings. It's a world without butterflies as we know them, a world with both predictable weather and a predictable stock market.

It is not a world we would want to achieve.

It's more fun to ponder the unpredictable. From noise comes spontaneity, creativity and perhaps even life itself. One's sense of humor may even be a manifestation of noise—a kind of noise in the brain, causing the eruption of an unexpected thought or phrase, a joke.

Now this lab course is not designed to show why jokes are humorous—at least not intentionally. However in the lab there will be lots of opportunity to observe noise and to understand it—from the point of view of the physicist.

Because of noise, every measurement of any physical quantity is *uncertain*. For example, here is a recorder trace of the output voltage from an ohmmeter:

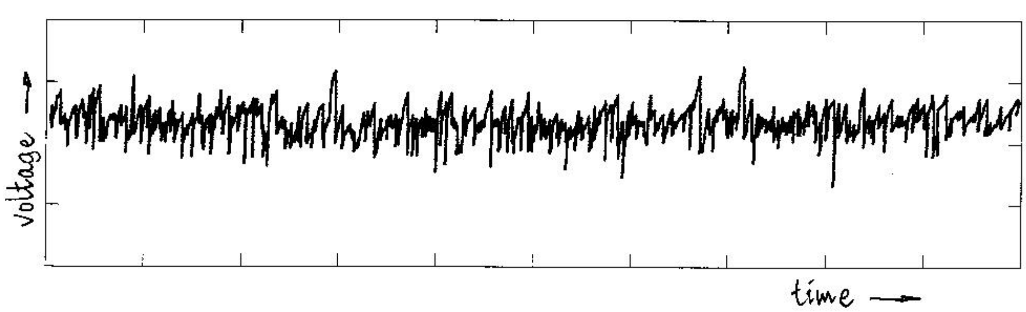


Figure 1 — Noise in a meter reading. The voltage *fluctuates* because of noise.

In another example, taken directly from the Radioactivity experiment, the intensity of a radioactive source is monitored with a Geiger counter. The counter is used to count the number of pulses in each of a sequence of one-second intervals, producing this graph of counting rate versus time:

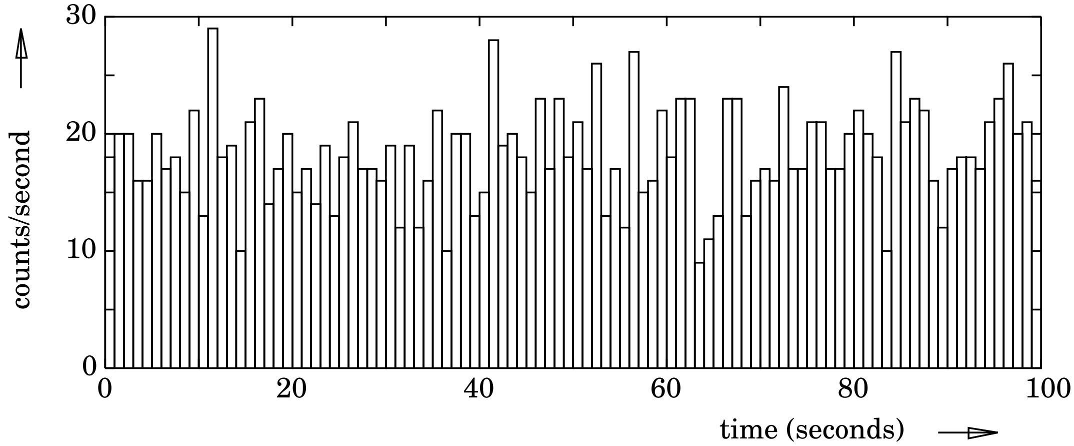


Figure 2 — Noise in a pulse counter.

The number of counts recorded in each interval will *fluctuate* from one interval to the next. We use the term *noise* to describe such fluctuations. It is our aim in the following paragraphs to understand noise as a source of uncertainty, to describe techniques for quantifying it, and to give meaning to the concept of *precision*.

Noise is also called *random error*, or *statistical uncertainty*. It is to be distinguished from *systematic error*. Systematic error, which is an error in measurement arising from a defect, such as the mis-calibration of a meter or some physical effect not taken into account in the measurement, can in principle be checked and corrected for.[1] Noise, on the other hand, is more basic. It arises, as in the first example (Fig. 1), from the thermal motion of individual atoms, or, as in the second example (Fig. 2), from the quantum-mechanical uncertainty associated with the radioactive emission of particles.[2]

In this second example, the question arises: How accurately may we estimate the "true" intensity of the radioactive source (*i.e.*, the "true" counting rate), when we measure for only a finite number of time intervals? Such a finite number of measurements, which in the above example is 100 (in general we'll call it n) is called a "sample", or more precisely, a "random sample", of the *total* population of such measurements. In this example, the

---

[1] The event depicted on the cover of John Taylor's monograph, *An Introduction to Error Analysis*, might have arisen from a *systematic* error in engineering design—or perhaps just a colossal blunder by the train operator.

[2] Noise can also arise from what has more recently been described as *deterministic chaos*—see, for example, James Gleick's book entitled *Chaos—Making a New Science* (Penguin Books, 1988). Connections may exist between such deterministic chaos and the thermal fluctuations or quantum fluctuations on the atomic scale; such connections are the object of recent research.

total population is infinite.[3] If we could make an infinite number of measurements, we could, in principle, reduce the statistical uncertainty to an infinitesimal value. We cannot make an infinite number of measurements, so we are stuck with a finite sample of n measurements, and hence with a finite statistical uncertainty in the determination of the counting rate.

For any such sample of n measurements, a few key statistical parameters may be calculated that serve the purpose of describing the measurement sample in the context of its associated noise. There are three parameters that are particularly useful:

1. The sample *mean* $\overline{x}$:

$$\overline{x} \equiv \frac{1}{n}\sum_{k=1}^{n} x_k \tag{1}$$

Here x_k is the *kth* measurement.

2. The sample *variance* s^2:

$$s^2 \equiv \frac{1}{n-1}\sum_{k=1}^{n}(x_k-\overline{x})^2 \tag{2}$$

The square root of the sample variance is s, and is called the sample *standard deviation.*

3. The *variance of the mean* $\sigma_{\overline{x}}^2$:

$$\sigma_{\overline{x}}^2 \approx \frac{s^2}{n} \tag{3}$$

Note the distinction between the *sample* variance and the variance of the *mean.* The square root of the variance of the mean is $\sigma_{\overline{x}}$, and is called the *standard deviation of the mean.* The meaning of the approximation sign in Eq. 3 is that the quantity s^2/n is an *estimate* of the variance of the mean.

An experimental result, *i.e.*, the best estimate we can make of the "true" value of x, is conveniently expressed in the form

$$\text{“RESULT”} = \overline{x} \pm \sigma_{\overline{x}} \approx \overline{x} \pm \frac{s}{\sqrt{n}} \tag{4}$$

As we shall see in the discussion contained in the following paragraphs, the meaning of this statement is that we expect the "true" value of x, taking into account only the random effects of noise or random error, to have about a 68 per cent chance, or *level of confidence*, of lying between $\overline{x}-\sigma_{\overline{x}}$ and $\overline{x}+\sigma_{\overline{x}}$.[4] These two values of x are the approximate *confidence limits.* They delimit a range of x-values called the *confidence interval.*

[3] We make the assumption that our source of radioactive particles is inexhaustible, which of course cannot be strictly true. This has no bearing on the point of our discussion, however. We'll just take "infinite" to mean "very very large".

[4] Equation 4 is not quite correct. See footnote 8 on Page 2–11 regarding further discussion of Eqs. 4 and 7.

There is one further point that we shall discuss later in more detail. It frequently happens that we wish to determine the mean, and the variance of the mean, for a quantity u that is a *function* $f(x, y, \ldots)$ of a number of experimentally measured, *independent* quantities $x, y, \ldots$. That is, $u = f(x, y, \ldots)$.

The value of $\overline{u}$ (the mean of u), and the best estimate for $\sigma_{\overline{u}}^2$ (the variance of the mean of u), can be calculated using the following formulas:

$$\overline{u} = f(\overline{x}, \overline{y}, \ldots) \tag{5}$$

and

$$\sigma_{\overline{u}}^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_{\overline{x}}^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_{\overline{y}}^2 + \cdots \tag{6}$$

Each of the variances on the right side of Eq. 6 may be estimated using an expression like that of Eq. 3. Hence a result for the derived measurement of u should be expressed in the form

$$\text{“RESULT”} = \overline{u} \pm \sigma_{\overline{u}} \tag{7}$$

The process of doing the calculations described by Eqs. 5 and 6 is called the *propagation of uncertainty through functional relationships*. These formulas, which are valid if $\sigma_{\overline{x}}$, $\sigma_{\overline{y}}$, ... are not too large, are quite general.

In what follows, we discuss the details of each of these points. Further references are cited at the end of this chapter.

1. The sample mean $\overline{x}$

The sample mean $\overline{x}$ is simply the average of the n individual measurements:

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{8}$$

Consider our second example shown graphically in Fig. 2. The number of counts in each of the first 25 one-second intervals is

18, 20, 20, 16, 16, 20, 17, 18, 15, 22, 13, 29, 18, 19, 10, 21, 23, 14, 17, 20, 15, 17, 14, 19, 13

then $n = 25$ and

$$\overline{x} = \frac{1}{25}(18 + 20 + 20 + 16 + 16 + 20 + 17 + \cdots) = \frac{444}{25} = 17.76$$

For this particular sample, certain numbers appear more than once. 13, 14, 15, 16 and 19 each appear twice, 17 and 18 appear three times, and 20 appears four times. In general, the value x_k might appear $g(x_k)$ times; $g(x_k)$ is called the *frequency* of the value x_k. Thus, an expression equivalent to Eq. 8 may be written as

$$\overline{x} = \frac{1}{n}\sum_{x_k} x_k g(x_k) \tag{9}$$

Note that while the sum in Eq. 8 is over k (the interval number), the sum in Eq. 9 is over the values of x_k.

For our example, $g(18) = 3$, $g(20) = 4$, $g(16) = 2$, etc., and Eq. 9 looks like this:

$$\overline{x} = \frac{1}{25}(10 \cdot 1 + 13 \cdot 2 + 14 \cdot 2 + \cdots + 20 \cdot 4 + 21 \cdot 1 + 22 \cdot 1 + 23 \cdot 1 + 29 \cdot 1) \; = \; 17.76$$

Now the total number of intervals n is just the sum of the interval frequencies $g(x_k)$, that is, $n = \sum g(x_k)$, so that

$$\overline{x} \; = \; \frac{\sum\limits_{x_k} x_k g(x_k)}{n} \; = \; \frac{\sum\limits_{x_k} x_k g(x_k)}{\sum\limits_{x_k} g(x_k)}$$

Furthermore, we expect that as n becomes very large, the quantity $g(x_k)/n$ will approach the *probability* $p(x_k)$ that the value x_k will appear. This defines $p(x_k)$:

$$p(x_k) \; \equiv \; \lim_{n \to \infty} \frac{g(x_k)}{n} \tag{10}$$

The introduction of the probability $p(x_k)$ now leads us to a diversion—a brief discussion about ways of thinking about population distributions, and about the commonly encountered *normal* distribution.

Background: Properties of the total population

The probability $p(x_k)$ is descriptive of the total (in our case, infinite) population of all possible measurements. The *total* population is also called the *parent* population. In general, we expect that $p(x_k)$ will be *normalized*:[5]

$$\sum_{x_k} p(x_k) \; = \; 1$$

Although for infinitely large populations such as the one we are considering, $p(x_k)$ is not accessible to us (we can only *estimate* it through the measurement of large samples), it is conceptually well-defined, and with it we can define the *mean* μ and the *variance* σ^2 of the

[5] In the following discussion we assume that x is limited to only the discrete values indicated by x_k. If x is in fact a continuous variable, sums over x_k should be replaced by integrals over x. Thus, for example

$$\sum_{x_k} p(x_k) = 1 \qquad \text{becomes} \qquad \int p(x)\, dx = 1$$

total population:[6]

$$\mu \;\equiv\; \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k \;=\; \sum_{x_k} x_k p(x_k) \tag{11}$$

and

$$\sigma^2 \;\equiv\; \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} (x_k-\mu)^2 \;=\; \sum_{x_k} (x_k-\mu)^2 p(x_k) \tag{12}$$

Note that these definitions are similar to Eqs. 1 and 2 defining the mean and variance for a particular finite sample of measurements; the difference is that we are here considering the *total* population.

In general, the *mean* value, also called the *average* value, or *expectation* value of *any* function $f(x_k)$ is given by

$$E[f(x_k)] \;=\; \text{ave}[f(x_k)] \;=\; \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) \;=\; \sum_{x_k} f(x_k)p(x_k) \tag{13}$$

where $E[f]$ stands for the *expectation* value of f.

Note that μ and σ^2 are the expectation values, or mean values, of particular functions of x_k. Thus $\mu = E[x_k]$ and $\sigma^2 = E[(x_k-\mu)^2]$.

The square root of the population variance σ^2 is σ, the *standard deviation* for the total population. σ is a statistical parameter describing the *dispersion* of the (infinite) number of measured values about the population mean μ. It describes how closely the measured values are clustered about the mean, and thus gives a measure of the width of the distribution of the values of x_k.

The Normal distribution

The interpretation of the parameter σ is easily envisaged if the measured quantities x_k are distributed according to the commonly encountered *Normal*, or *Gaussian* distribution. The probability distribution function for a normally distributed continuous random variable x is given by[7]

$$p(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{14}$$

Here is a graph of $p(x)$ vs. x:

[6] Greek letters are often used to denote parameters that are descriptive of the *parent* population.

[7] See Taylor, Section 5.3.

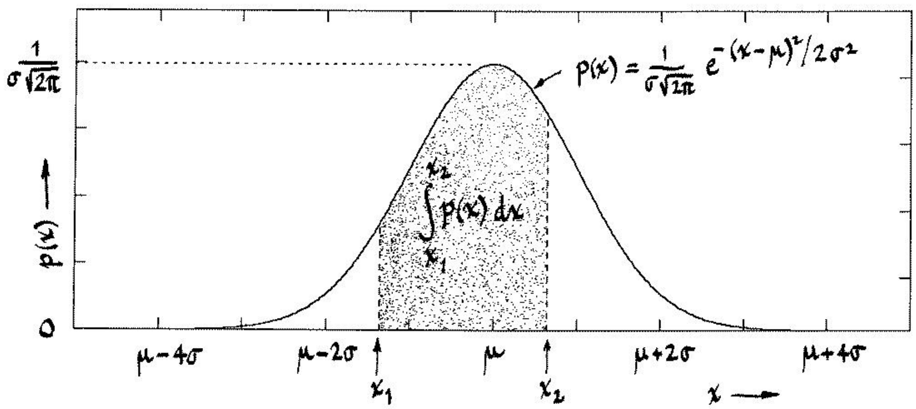


Figure 3 — The Normal distribution.

$p(x)\,dx$ is the probability that any particular value of x falls between x and $x+dx$, and $\int_{x_1}^{x_2} p(x)\,dx$ is the probability that any particular x falls between x_1 and x_2. This integral is represented by the shaded area in Fig. 3.

If $x_1 = -\infty$ and $x_2 = +\infty$, then it is certain that any particular x falls in this interval, and $\int_{x_1}^{x_2} p(x)\,dx = 1$. The *normalization factor* $1/\sigma\sqrt{2\pi}$ ensures that this is the case. If σ is reduced, $p(x)$ becomes more sharply peaked.

If $x_1 = \mu - \sigma$ and $x_2 = \mu + \sigma$, the shaded area is approximately 0.6827. That is, for a normal distribution, there is approximately a 68 per cent chance that any particular x falls within one standard deviation of the mean. Furthermore, the chance that an x will fall within two standard deviations of the mean is approximately 0.9545, and within three standard deviations, approximately 0.9973. It is striking that *for measurements that are normally distributed about some mean value*, almost all of them (over 99 per cent) will lie within three standard deviations of the mean.

From a random sample of n measurements one may form a frequency distribution that may be compared with any particular probability distribution function $p(x)$. Here is a bar graph, or histogram, formed from the data shown in Fig. 2:

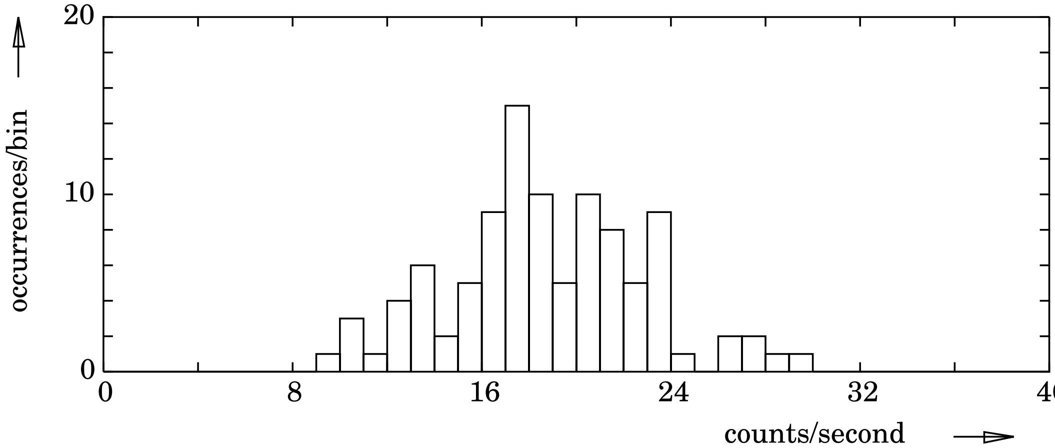


Figure 4 — A sample distribution.

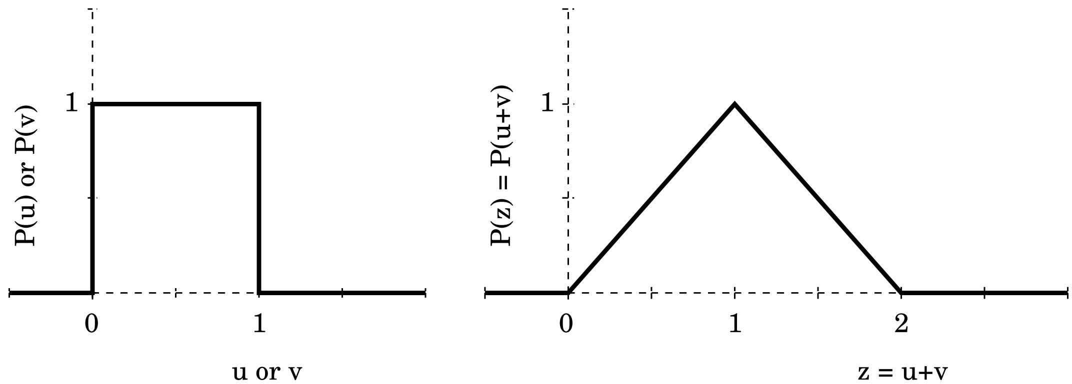
Note that it looks qualitatively similar to the Normal distribution shown in Fig. 3. A quantitative comparison may be made using *Pearson's Chi-square Test*, as described in Chapter 4 of this manual.

Why the Normal distribution is so commonly encountered

We have mentioned that the fluctuations in measured quantities are commonly found to be approximately described by a Normal distribution. Why? The answer is related to a powerful theorem, much beloved by physicists, called the *Central Limit Theorem.*

This theorem states that if we have a number of random variables, say $u, v, w, \ldots$, and that if we form a new variable z that is the *sum* of these ($z = u + v + w + \cdots$), then as the number of such variables becomes large, z will be distributed *normally*, *i.e.*, described by a Normal distribution, *regardless* of how the individual variables $u, v, w, \ldots$ are distributed.

While we won't prove the Central Limit Theorem here (it's not an easy proof), we can present a "physicist's proof"—an example that is easily tested: Let each of $u, v, w, \ldots$ be real numbers randomly and uniformly distributed between 0 and 1. That is, each is drawn from a *flat* distribution—clearly *not* a Normal distribution. Then let $z = u+v+w+\cdots$. It is not hard to show, using a simple computer program, that for even as few as four or five such terms in the sum, z will be nearly normally distributed. In fact if there are only two terms, we can already see the peaking near the center, with the result being a *triangular* distribution, like this:



A simpler example involves dice. Throw *one* die, and the probability that any number between 1 and 6 shows is 1/6—a uniform distribution. Throw *two* dice, however, and the distribution, for x-values between 2 and 12, is triangular. As an exercise, try plotting out the distribution for *three* dice. The x-values range between 3 and 18. Does the distribution look bell-shaped?

Now a typical quantity measured in a physics experiment results from the sum of a large number of random processes, and so is likely to be distributed normally. For example, the pressure of a gas results from summing the random motions of a very large number of molecules, so we expect measured fluctuations in gas pressure to be normally distributed.

Nevertheless, we must be careful to not put too much faith in the results of the Central Limit Theorem. One frequently sees measured values that are obviously non-normal—too far away from the mean—that could arise, say, from some voltage spike or from some vibration caused by a truck passing by. Not every data point can be expected to fall within this classic bell curve.

This ends our diversion. We continue now with our discussion of the sample mean, the sample variance, the variance of the mean, and how uncertainties are propagated through functional relationships.

How is $\overline{x}$ related to μ?

In general, our desire is to determine, from a finite sample of measurements, best estimates of parameters, such as μ and σ^2, that are descriptive of the total population. The simplest relationship is that between $\overline{x}$ and μ: $\overline{x}$ is the best estimate of μ. This is equivalent to saying that the expectation value of $\overline{x}$ is μ, or $E[\overline{x}] = \mu$. While this statement may seem intuitively obvious, here is a proof:

$$E\left[\overline{x}\right] \;=\; E[\frac{1}{n}\sum_k x_k] \;=\; \frac{1}{n}\sum_k E[x_k] \;=\; \frac{1}{n}\sum_k \mu \;=\; \frac{1}{n}\cdot n\mu \;=\; \mu$$

2. The sample variance s² and the sample standard deviation s

The sample variance s^2 is defined by:

$$s^2 \;\equiv\; \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2 \tag{15}$$

By substituting $\overline{x} \equiv \frac{1}{n}\sum x_k$ we obtain

$$s^2 \;=\; \frac{n\sum x_k^2 - (\sum x_k)^2}{n(n-1)} \tag{16}$$

which is an expression useful for numerical calculation, in that it involves only $\sum x_k$, $\sum x_k^2$ and n, which are easily computed.

For the complete measurement sample shown in Fig. 2, $\sum x_k = 1825$, $\sum x_k^2 = 355013$, and $n = 100$, which yields $s^2 = 17.240$ as the sample variance, and $s = (17.240)^{\frac{1}{2}} = 4.152$ as the sample standard deviation.

How is s^2 related to σ^2?

The sample variance s^2 is the best estimate of the variance σ^2 for the total population. This is equivalent to the statement that the expectation value of s^2 is equal to σ^2. The proof of this statement runs as follows. We start by taking expectation values of both sides of Eq. 15:

$$E[s^2] \;=\; \frac{1}{n-1}E\Big[\sum(x_k-\overline{x})^2\Big] \;=\; \frac{1}{n-1}E\Big[\sum x_k^2 - 2\overline{x}\sum x_k + \sum \overline{x}^2\Big]$$

$$= \frac{1}{n-1}E\Big[\sum x_k^2 - 2n\overline{x}^2 + n\overline{x}^2\Big] \;=\; \frac{1}{n-1}\Big\{\sum E[x_k^2] - nE[\overline{x}^2]\Big\} \;=\; \frac{n}{n-1}\Big\{E[x_k^2] - E[\overline{x}^2]\Big\}$$

To evaluate $E[x_k^2]$, we note from Eq. 12 that

$$\sigma^2 \;=\; E[(x_k-\mu)^2] \;=\; E[x_k^2] - 2\mu E[x_k] + E[\mu^2] \;=\; E[x_k^2] - \mu^2 \tag{17}$$

so that

$$E[x_k^2] \;=\; \mu^2 + \sigma^2$$

To evaluate $E[\overline{x}^2]$, we expand to find

$$E[\overline{x}^2] \;=\; E\Big[\Big(\frac{x_1+x_2+\cdots+x_n}{n}\Big)^2\Big] \;=\; \frac{1}{n^2}\Big\{E\Big[\sum x_k^2\Big] + E\Big[\sum_{k\neq j} x_k x_j\Big]\Big\}$$

where the quantity $\sum\limits_{k\neq j} x_k x_j$ represents all cross-products of two different measurements of x. Since x_k and x_j are independent for $k \neq j$, we have

$$E[x_k x_j] \;=\; E[x_k]E[x_j] \;=\; \mu\cdot\mu \;=\; \mu^2$$

Since there are n terms of the form x_k^2 and $n(n-1)$ cross-product terms of the form $x_k x_j$, we have

$$E[\overline{x}^2] \;=\; \frac{1}{n^2}\{n(\sigma^2+\mu^2) + n(n-1)\mu^2\} \;=\; \frac{\sigma^2}{n} + \mu^2 \tag{18}$$

Hence we find (finally!)

$$E[s^2] \;=\; \frac{n}{n-1}\{(\sigma^2+\mu^2) - (\frac{\sigma^2}{n}+\mu^2)\} \;=\; \sigma^2$$

and the assertion is proved.

This proof also provides the justification for dividing by $n-1$, rather than n, when we calculate the sample variance. Qualitatively, when we calculate the sample variance s^2 using Eq. 15, the use of $\overline{x}$ as an estimate of μ in that expression will tend to reduce the magnitude of $\sum(x_k-\overline{x})^2$ somewhat. That is,

$$\sum(x_k-\overline{x})^2 \;<\; \sum(x_k-\mu)^2$$

Division by $n-1$ rather than n serves to compensate for this slight reduction.

3. The variance of the sample mean $\sigma_{\bar{x}}^2$ and its associated standard deviation $\sigma_{\bar{x}}$

For a sample of n measurements x_k we have seen that $\overline{x}$ is the best estimate of the population mean μ. If the x_k are *normally* distributed, an additional single measurement will fall within $\overline{x} \pm s$ at approximately the 68 per cent level of confidence. This is the interpretation of the standard deviation s for the sample of n measurements.

If we take additional samples, of n measurements each, we expect to gather a collection of sample *means* that will be clustered about the population mean μ, but with a distribution that is *narrower* than the distribution of the individual measurements x_k. That is, we expect the variance of the sample *means* to be *less* than the population variance. For a sample of n measurements, it turns out (see proof below) that the variance of the *mean* is just the population variance divided by n:

$$\sigma_{\overline{x}}^2 = \frac{\sigma^2}{n} \tag{19}$$

or, since we may estimate the value of σ^2 by calculating s^2,

$$\sigma_{\overline{x}}^2 \approx \frac{s^2}{n} \tag{20}$$

The quantity $s/\sqrt{n}$ thus provides us with an estimate of the *standard deviation of the mean.* An experimental result is conventionally stated in the form shown in Eq. 4, namely[8]

$$\text{“RESULT”} = \overline{x} \pm \frac{s}{\sqrt{n}} \tag{21}$$

As an example, we look once again at the data sample of 100 measurements of the counting rate shown in Fig. 2. Since for that sample we have $\overline{x} = 18.25$ and $s = 4.152$, we

[8] Equations 4, 7 and 21 are not quite correct. Because of the non-normal distribution of the sample variance, it should be written

$$\text{“RESULT”} = \overline{x} \pm t_{n-1} \frac{s}{\sqrt{n}} \tag{21a}$$

where t_{n-1} is a constant called the “Student” t-factor. In the general case, t_{n-1} depends on the level of confidence chosen and the sample size n. If, as usual, we chose a confidence level of 68.27 per cent, t_{n-1} approaches 1.0 for large n, and is not much larger than 1.0 even for small n. A table at the end of this chapter displays commonly used values of the “Student” t-factor t_ν. (ν, here $n-1$, is the number of “degrees of freedom”.) In this course, our interest in t_ν is largely academic, and frequently (as in Eq. 21) we omit it. With more conservative confidence intervals such as 95 or 99 per cent, its use becomes more meaningful. Its use also arises in the fitting of data to a mathematical model, where confidence intervals on the estimates of parameters are desired. The computer programs we use for such data modeling (see Chapter 5) include “Student” t-factors in the estimation of confidence intervals. For a complete discussion of the “Student” t (and a little story about who “Student” was), see the book by Bennett and Franklin.

may express our measured counting rate in the form

$$\text{Counting rate} \;=\; 18.25 \pm \frac{4.152}{\sqrt{100}} \;=\; 18.25 \pm 0.42 \quad \text{counts/second}$$

This result implies that if one were to take an additional sample of 100 measurements, there would be about a 68 per cent chance that this new sample mean would lie between 17.83 and 18.67 counts/second. Note that we rounded off the uncertainty to two significant figures, since a third significant figure makes no sense. We also did not include any more significant figures in the value of the result (here "18.25") than are implied by the uncertainty. Thus to have stated our result as 18.250 ± 0.42 counts/second would have been incorrect.

> It is meaningless to include more than two significant figures in the uncertainty. It is also meaningless to include more significant figures in the result than are implied by the uncertainty.

A result so expressed thus allows us to compare our own experimental result with those of others. If the result stated in the form of Eq. 21 brackets, or overlaps a similar result obtained elsewhere, we say that the two experimental results are in agreement. We have ignored, of course, any *systematic errors* that may be present in either measurement.

Equation 19 may be easily proved:

$$\sigma_{\overline{x}}^2 \;=\; E\left[(\overline{x}-\mu)^2\right] \;=\; E\left[\overline{x}^2 - 2\mu\overline{x} + \mu^2\right] \;=\; E\left[\overline{x}^2\right] - E\left[\mu^2\right] \;=\; E\left[\overline{x}^2\right] - \mu^2$$

From Eq. 18 we have

$$E\left[\overline{x}^2\right] \;=\; \frac{\sigma^2}{n} + \mu^2$$

from which it follows that $\sigma_{\overline{x}}^2 = \sigma^2/n$.

Finally, there is one additional point to discuss: Suppose we measure a quantity u several times, or by several different methods, and for each measurement u_i we estimate its uncertainty σ_i. The σ_i are not necessarily equal; some of the measurements will be better than others, because of larger sample sizes (more repetitions), or because of other factors—like better apparatus. How do we determine our best estimate of u, and how do we find the uncertainty in that estimate?

For example, suppose a length x is measured by one person n_1 times and by another person n_2 times, so that the first person finds

$$u_1 \equiv \overline{x}_1 = \frac{1}{n_1} \sum_k x_k \qquad \text{with} \qquad \sigma_1^2 = \frac{1}{n_1} \sigma^2$$

while the second person finds

$$u_2 \equiv \overline{x}_2 = \frac{1}{n_2} \sum_j x_j \qquad \text{with} \qquad \sigma_2^2 = \frac{1}{n_2} \sigma^2$$

Here σ_1 is the uncertainty in u_1, σ_2 is the uncertainty in u_2, and σ^2 is the population variance of the x-values. How should u_1 and u_2 be combined to yield an overall $\overline{u}$, and what is the uncertainty in this final $\overline{u}$? Since $n_1 = \sigma^2/\sigma_1^2$ and $n_2 = \sigma^2/\sigma_2^2$

$$\overline{u} \;=\; \frac{1}{n_1 + n_2} \left(\sum_k x_k + \sum_j x_j \right) \;=\; \frac{1}{n_1 + n_2} \left(n_1 u_1 + n_2 u_2 \right) \;=\; \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \left(\frac{u_1}{\sigma_1^2} + \frac{u_2}{\sigma_2^2} \right)$$

with

$$\sigma_{\overline{u}}^2 \;=\; \frac{\sigma^2}{n_1 + n_2} \;=\; \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

In general, if there are n values of u, here is the generalized result, in a form that depends *only* on each u_k and its uncertainty σ_k:

$$\overline{u} \;=\; \frac{\sum_{k=1}^{n} u_k/\sigma_k^2}{\sum_{k=1}^{n} 1/\sigma_k^2}; \qquad \sigma_{\overline{x}}^2 \;=\; \frac{1}{\sum_{k=1}^{n} 1/\sigma_k^2} \tag{22}$$

Note how more measurements, or more accurate measurements, *reduce* the uncertainty by *increasing* its reciprocal. The results expressed in Eq. 22 may also be derived from a principle of maximum likelihood or a principle of least squares. An explicit example of the application of the formulas in Eq. 22 appears on page 4–5 of this manual.

4. The propagation of uncertainty through functional relationships

It frequently occurs that one wishes to determine the uncertainty in a quantity that is a function of one or more (independent) random variables. As we have seen, if we measure a counting rate x, we may express our result as $\overline{x} \pm \sigma_{\overline{x}}$. Suppose, however, we are interested in a quantity u that is proportional to the square of x, that is, $u = ax^2$, where a is some constant. What is the resulting uncertainty in u?

Using the concepts of differential calculus, one expects that if x fluctuates by an amount dx, then u will fluctuate by an amount $du = (\partial u/\partial x)\, dx = 2ax\, dx$. In statistical terms, where the sign of the fluctuation is irrelevant, and if the fluctuations are not too large, one expects that

$$\sigma_u = \left| \frac{\partial u}{\partial x} \right| \sigma_x$$

and also, for the standard deviation of the mean,

$$\sigma_{\overline{u}} = \left| \frac{\partial u}{\partial x} \right| \sigma_{\overline{x}}$$

In each case, the derivative should be evaluated at the point $x = \overline{x}$. We may generalize this idea to include situations where u depends on more than one random variable: Suppose $u = f(x, y, \ldots)$, where $x, y, \ldots$ are random independent variables. Then

$$\sigma_u^2 = \left(\frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \sigma_y^2 + \cdots \tag{23}$$

and also, for the variance of the mean,

$$\sigma_{\overline{u}}^2 = \left(\frac{\partial f}{\partial x} \right)^2 \sigma_{\overline{x}}^2 + \left(\frac{\partial f}{\partial y} \right)^2 \sigma_{\overline{y}}^2 + \cdots \tag{24}$$

Note that we sum the squares of the individual terms; this is appropriate when the variables $x, y, \ldots$ are *statistically independent.*

We illustrate the idea with an example taken from the Radioactivity experiment—one of the experiments in the Intermediate lab. In that experiment, a counter dead time τ may be estimated from a measurement of three counting rates, x, y and z. Here x is the counting rate from one source, y is the counting rate from a second source, and z is the counting rate from both sources simultaneously. For this illustration, we use a simple (albeit inaccurate) formula for the dead time:

$$\tau \approx \frac{x+y-z}{2xy} \tag{25}$$

From this expression, we can calculate

$$\frac{\partial\tau}{\partial x} = \frac{z-y}{2x^2y}; \quad \frac{\partial\tau}{\partial y} = \frac{z-x}{2xy^2}; \quad \frac{\partial\tau}{\partial z} = -\frac{1}{2xy} \tag{26}$$

In a particular experiment, the number of counts in one minute were measured for each of the three configurations, yielding

$$x = 55319; \quad \sigma_x = 235$$

$$y = 54938; \quad \sigma_y = 234$$

$$z = 86365; \quad \sigma_z = 294$$

where the units of all quantities are counts per minute.

At the point (x, y, z):

$$\frac{\partial\tau}{\partial x} = 9.35\times10^{-11}; \quad \frac{\partial\tau}{\partial y} = 9.30\times10^{-11}; \quad \frac{\partial\tau}{\partial z} = -16.5\times10^{-11}$$

with the units being mins^2/count. Inserting these values into Eq. 23 yields

$$\sigma_\tau^2 = (9.35\times10^{-11})^2(235)^2+(9.3\times10^{-11})^2(234)^2+(-16.5\times10^{-11})^2(294)^2 = 3.31\times10^{-15}\ \text{min}^2$$

from which

$$\sigma_\tau = (3.31\times10^{-15})^{\frac{1}{2}} = 5.75\times10^{-8}\ \text{minutes} = 3.5\ \text{microseconds}$$

Using Eq. 25 to evaluate τ at the point (x, y, z) we find

$$\tau = \frac{x+y-z}{2xy} = 3.93\times10^{-6}\ \text{minutes} = 236\ \text{microseconds}$$

Hence we may express the final result of this measurement of the dead time as

$$\tau = 236 \pm 3.5\ \text{microseconds}$$

It turns out that if we used the more accurate formula for the dead time we would have obtained 300 microseconds instead of 236. These values may now be compared with the value obtained by measuring the dead time of the counter directly from the oscilloscope screen, which in this particular experiment was found to be about 220 microseconds, with an error of several tens of microseconds. The two results are thus found to be in rough agreement.

Table of "Student" t-factors t_ν

Degrees of Freedom (ν)	Level of Confidence in per cent			
	68.269	95.0	95.45	99.0
1	1.8373	12.7062	13.9678	63.6567
2	1.3213	4.3027	4.5266	9.9248
3	1.1969	3.1824	3.3068	5.8409
4	1.1416	2.7764	2.8693	4.6041
5	1.1105	2.5706	2.6487	4.0321
6	1.0906	2.4469	2.5165	3.7074
7	1.0767	2.3646	2.4288	3.4995
8	1.0665	2.3060	2.3664	3.3554
9	1.0587	2.2622	2.3198	3.2498
10	1.0526	2.2281	2.2837	3.1693
11	1.0476	2.2010	2.2549	3.1058
12	1.0434	2.1788	2.2314	3.0545
13	1.0400	2.1604	2.2118	3.0123
14	1.0370	2.1448	2.1953	2.9768
15	1.0345	2.1315	2.1812	2.9467
16	1.0322	2.1199	2.1689	2.9208
17	1.0303	2.1098	2.1583	2.8982
18	1.0286	2.1009	2.1489	2.8784
19	1.0270	2.0930	2.1405	2.8609
20	1.0256	2.0860	2.1330	2.8453
21	1.0244	2.0796	2.1263	2.8314
22	1.0233	2.0739	2.1202	2.8188
23	1.0222	2.0687	2.1147	2.8073
24	1.0213	2.0639	2.1097	2.7969
25	1.0204	2.0595	2.1051	2.7874
26	1.0196	2.0555	2.1009	2.7787
27	1.0189	2.0518	2.0969	2.7707
28	1.0182	2.0484	2.0933	2.7633
29	1.0175	2.0452	2.0900	2.7564
30	1.0169	2.0423	2.0868	2.7500
35	1.0145	2.0301	2.0740	2.7238
40	1.0127	2.0211	2.0645	2.7045
45	1.0112	2.0141	2.0571	2.6896
50	1.0101	2.0086	2.0513	2.6778
55	1.0092	2.0040	2.0465	2.6682
60	1.0084	2.0003	2.0425	2.6603
120	1.0042	1.9799	2.0211	2.6174
∞	1.0000	1.9600	2.0000	2.5758

References

1. Taylor, John R., *An Introduction to Error Analysis*, 2nd Ed. (University Science Books, 1997). This book is a good place to start. It includes almost all the material set forth in this chapter, but without some of the derivations and proofs.

2. Bevington, Philip R., and Robinson, D. Keith, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd Ed. (McGraw-Hill, 1992). Bevington's book has long been a standard reference for physicists. It is, however, a little tedious.

3. Evans, Robley D., *The Atomic Nucleus* (McGraw-Hill, 1969). Here one may find clearly written sections relating to much of the preceding material. It also contains a good description of *Pearson's Chi-square Test*, as noted later in Chapter 4.

4. Bennett, Carl A., and Franklin, Norman L., *Statistical Analysis in Chemistry and the Chemical Industry* (Wiley, 1954). Material for this chapter was gleaned from the first 60 or so pages of this tome. The book is excellent, although it contains much more information than we need.