## Notes - Errors and Noise

PHYS 3600, Northeastern University, Don Heiman

## 1. Accuracy versus Precision

1.1 Precision - how exact is a measurement, or how "fine" is the scale (\# of significant figures).

Suppose you measure a resistor with a digital ohmmeter. The ohmmeter reads $1.53483 \Omega$. This number has a high precision. However, it may not represent the "true" resistance as the wires connecting the resistor and ohmmeter have some small resistance that contributes to the measurement.
1.2 Accuracy - how close is the measurement to the "true" value.

Accuracy is a measure of the correctness of the measurement. To determine a more accurate value for the resistor's resistance, you could measure the resistance of the wires and subtract it.

## 2. Errors

We often want to know how close we are to the truth. The error is simply the quantifiable difference between the value obtained in real life and the "true" value. There are three main categories of errors:
blunders, systematic errors, and random errors.
2.1 Blunders - These can usually be avoided by examining the results as you proceed with measurements. Examples are: reading and recording the wrong scale on the instrument, such as millivolts instead of volts; reading milliseconds instead of microseconds on the oscilloscope time base; forgetting to convert frequency $f$ in cycles/second (Hz) to $\omega$ in radians/second; or mixing up centimeters and inches.
2.2 Systematic Errors - these are "reproducible" errors from faulty calibration or biased observation.

They can often be estimated from analysis of the experimental techniques. If you are measuring a distance with an incorrect (shortened or lengthened) ruler, you can calibrate ruler and make the correction by multiplying the original distances by a calibration factor.
2.3 Random Errors - these are due to fluctuations in measurements from repeated experiments.

Random errors are due to statistical fluctuations, often referred to as noise, or limited precision of an instrument. They can be minimized by repeating the experiment many times. Note that random errors which are small produce high precision, but not necessarily high accuracy.

Imagine you are timing the vibration of a mass on a spring with a stopwatch. You could try to measure the period of a single oscillation. One measurement might produce a value of 0.8 sec . By repeating the experiment you get values of $1.1,1.0,0.9$, and 1.2 sec . You can average these five numbers to get a more precise value than from a single measurement. The average of the five measurements is 1.0 sec .

In addition, there could also be systematic errors in which your reaction time with the stopwatch has added a few tenths of a second to the timed interval. The systematic error can be reduced by simply timing many oscillations instead of a single oscillation. In this way the timing error is reduced (divided) by the number of oscillations.

## 3. Significant Digits and Round-off

Calculators and digital meters produce a much larger number of significant digits than is usually justified. The answer from a calculator has very high precision, typically to 8 or 10 digits. In experiments, the number of significant digits is usually much less than this. Suppose you want to divide two values obtained from an experiment - one value has two significant digits and the other value has three. Although the calculator gives 8 significant digits, the answer is only significant to the smallest number of significant digits, only two. You must round-off the calculator number to two significant digits while the third significant figure is dropped. When the third significant figure is greater than 5 the second significant figure is incremented one unit. Thus, 6.23 becomes 6.2 , whereas 6.25 becomes 6.3 .

## - Significant Digits -

## 1. UNCERTAINTIES HAVE LOW PRECISION

Standard deviations or uncertainties are rounded off to one significant digit (or sometimes two digits if the most significant digit is between " 1 " and " 1.4 ").
2. The computed number is then rounded off at the same digit as the uncertainty,

$$
\begin{aligned}
& \text { e.g. } 4.537 \pm 0.250 \mathrm{~cm}^{2} \text { should be written } 4.6 \pm 0.3 \mathrm{~cm}^{2}, \\
& \text { or }(4.6 \pm 0.3) \times 10^{-2} \mathbf{~} \mathbf{~} .
\end{aligned}
$$

It is often helpful to write numbers in scientific notation. For example, the number 0.0000325 would be expressed as $3.25 \times 10^{-5}$. When measuring quantities such as voltage, it is best express the values in engineering notation which has powers of 10 in increments of 3 . Thus, 0.0000325 volts becomes $32.5 \times 10^{-6}$ V , or $32.5 \mu \mathrm{~V}$. With an uncertainty, the result should then be written as: $32.5 \pm 0.4 \mu \mathrm{~V}$.

## 4. Mean and Standard Deviation

Suppose that the only errors in a set of experiments are random in nature. Then by repeating the experiment an infinite number of times we expect the average to be equal to the "true" value. For a finite number $N$ of experiments, the mean value of the set of experimental values, $y_{i}=y_{1}, y_{2}, \ldots, y_{N}$, is given by

$$
\langle y\rangle=\Sigma y_{i} / N .
$$

The deviation of individual experiments from the mean is $\delta y_{i}=y_{i}-\langle y\rangle$. The standard deviation $\sigma$ of the set of experiments is then given by

$$
\begin{aligned}
\sigma^{2} & =\Sigma\left(\delta y_{i}\right)^{2} /(N-1) \\
& =\left[\Sigma y_{i}^{2}-\langle y\rangle^{2}\right] /(N-1) .
\end{aligned}
$$

The $\mathrm{N}-1$ in the denominator is used when $\langle\mathrm{y}\rangle$ is the average of the data set. If $\langle\mathrm{y}\rangle$ is already the "true" value found by another method, then the denominator is $N$ instead of $N-1$. If $\langle y\rangle$ is given by a polynomial of order $n$, then the denominator is $N-1-n$. This can be generalized to a denominator $N-v$, where $v$ is the number of degrees of freedom. When N is large, it makes little difference in the value of the standard deviation.

## 5. Propagation of Errors

In general, for any function $F(x, y, \ldots)$ of a set of variables $x, y, \ldots$, having standard deviations $\sigma_{x}, \sigma_{y}, \ldots$, the standard deviation for the function $F$ is given by

$$
\sigma_{F}{ }^{2}=(\partial F / \partial x)^{2} \sigma_{x}{ }^{2}+(\partial F / \partial y)^{2} \sigma_{y}{ }^{2}+2(\partial F / \partial x)(\partial F / \partial y) \sigma_{x y}{ }^{2}+\ldots .
$$

When the deviations in the variables ( $x, y, \ldots$ ) are not correlated, the covariances vanish, $\sigma_{x y}=\ldots=0$, the equation reduces to

$$
\sigma_{F}{ }^{2}=(\partial F / \partial x)^{2} \sigma_{x}{ }^{2}+(\partial F / \partial y)^{2} \sigma_{y}{ }^{2}+\ldots .
$$

The following recipes are used for finding the cumulative error for various mathematical operations.

### 5.1 Addition and Subtraction

Given $\mathbf{z}=\mathbf{a x} \pm \mathbf{b y}$, and their standard deviations, $\sigma_{\mathbf{x}}$ and $\sigma_{y}$, are not correlated, the standard deviation in $z$ is given by

$$
\sigma_{z}^{2}=a^{2} \sigma_{x}^{2}+b^{2} \sigma_{y}^{2}
$$

For simple addition of two numbers the standard deviation are added in quadrature as

$$
\sigma_{7}{ }^{2}=\sigma_{x}{ }^{2}+\sigma_{v}{ }^{2} .
$$

(Note that for $z=x+x, \sigma_{z}{ }^{2}=\sigma_{x}{ }^{2}+\sigma_{x}{ }^{2}+2 \sigma_{x x}{ }^{2}=2^{2} \sigma_{x}{ }^{2}$, since $\sigma_{x x}=\sigma_{x}$.)
For example, compute the perimeter of a rectangle with sides $\mathrm{h}=2.0 \pm 0.2 \mathrm{~cm}$ and $\mathrm{w}=3.5 \pm 0.4 \mathrm{~cm}$. The perimeter is $\mathrm{P}=\mathrm{h}+\mathrm{h}+\mathrm{w}+\mathrm{w}=2 \mathrm{~h}+2 \mathrm{w}$. The standard deviation in the perimeter is
$\sigma_{P}{ }^{2}=2^{2 *} 0.2^{2}+2^{2} 0.4^{2}=0.9^{2}$. Thus, the resulting perimeter is $\mathbf{P}=\mathbf{1 1 . 0} \pm \mathbf{0 . 9} \mathbf{~ c m}$.

### 5.2 Multiplication and Division

Given that $\mathbf{z}=\mathbf{a x}^{\mathrm{n}} \mathbf{y}$ or $\mathbf{z = a x} \mathbf{n} / \mathbf{y}$, and their standard deviations, $\sigma_{\mathrm{x}}$ and $\sigma_{y}$, are not correlated. Then the standard deviation in z is given by the fractional uncertainties

$$
\left(\frac{\sigma_{Z}}{z}\right)^{2}=n^{2}\left(\frac{\sigma_{X}}{x}\right)^{2}+\left(\frac{\sigma_{Y}}{y}\right)^{2}
$$

For example, compute the area of the rectangle with sides $\mathrm{h}=2.0 \pm 0.2 \mathrm{~cm}$ and $\mathrm{w}=3.5 \pm 0.4 \mathrm{~cm}$. The area is $A=h w$. The standard deviation in the area is now $\sigma_{A}{ }^{2} / A^{2}=0.2^{2} / 2.0^{2}+0.4^{2} / 3.5^{2}=0.15^{2}$. Thus, the resulting area is $\mathbf{A}=\mathbf{7} \pm \mathbf{~ c m}^{\mathbf{2}}$.

### 5.3 Exponents

$$
\begin{array}{ll}
\text { For } z=a x^{ \pm b}, & \sigma_{z} / z=b \sigma_{x} / x . \\
\text { For } z=a e^{ \pm b x}, & \sigma_{z} / z=b \sigma_{x} . \\
\text { For } z=a \ln ( \pm b x), & \sigma_{z}=a \sigma_{x} / x
\end{array}
$$

## 6. Interpolation

Suppose you have a precise set of $y_{i}\left(x_{i}\right)$ data points and want to find a $y$-value that lies between a pair of $x$ values. By linear interpolation, the value $y(x)$ between $y_{1}\left(x_{1}\right)$ and $y_{2}\left(x_{2}\right)$ is given by

$$
y=y_{1}+\left(x-x_{1}\right)\left[\left(y_{2}-y_{1}\right) /\left(x_{2}-x_{1}\right)\right]
$$

In cases where the $y(x)$ curve is significantly nonlinear between successive data points, a more precise method would be to replace the linear approximation by a polynomial which better represents the real curvature of $\mathrm{y}(\mathrm{x})$. For a polynomial interpolation, more than $2 \mathrm{y}_{\mathrm{i}}\left(\mathrm{x}_{\mathrm{i}}\right)$ data points are required.

## 7. Smoothing $y(x)$ Curves

Smoothing noisy data is a convenient way to visually see what the general behavior looks like. For example, smoothing can be used to estimate the midpoint and width of a very noisy peaked curve. Note that when curve fitting the data to a function, identical results will be achieved by fitting the smoothed or unsmoothed data. EasyPlot uses the following sliding window and bucket smoothing procedures.

### 7.1 Sliding Window Smoothing

Smoothing data by the sliding window method replaces each $y_{i}$ value by an average of $y_{i}$ and several adjacent $y$-values. The $y$-values adjacent to $y_{i}$ are weighted less than $y_{i}$, by an amount which decreases for increasing $x$-distance from the $y_{i}$ point. In most cases, a Gaussian weighting function $\left\{\exp \left[-\left(x_{i}-x\right)^{2} / b^{2}\right]\right\}$ is appropriate, however, a triangle function yields nearly identical results. For example, using a triangle smoothing function and a window of $M=5$ points, each $y$-value is replaced by

$$
y_{i}=\left(1^{*} y_{i-2}+2^{*} y_{i-1}+3^{*} y_{i}+2^{*} y_{i+1}+1^{*} y_{i+2}\right) / 9
$$

Notice that the points within $(\mathrm{M}-1) / 2$ of the end of the curve are not smoothed properly.

### 7.2 Bucket Smoothing

Bucket smoothing of a $y(x)$ data set averages over buckets having equal intervals of $x$ and replaces the $y$ values in each bucket by their average. The x-value is usually taken as the midpoint of the bucket interval rather than the average $x$-value. This smoothing function in plotting software can be used to generate an effective set of data with a significantly smaller number of points. It is also useful for converting a data set into an effective data set with equally spaced, monotonically increasing $x$-values. This method is useful for generating equally spaced points for Fourier transforming a data set.

## Rules for Eliminating Negligible Uncertainties

Addition/Subtraction - eliminate uncertainties that are at least 3-times smaller.

Multiplication/Division - eliminate fractional uncertainties that are at least 3-times smaller, even if the magnitude of the uncertainty is larger.

