Introduction to Uncertainty in Measurement

Mark Chilenski

February 19, 2014

1 What is uncertainty?

- It is impossible to ever measure a quantity exactly: there will *always* be some level of uncertainty associated with a measurement.
- **Uncertainty** is exactly what it sounds like: it is how unsure you are about the value of a quantity.
- A statement of a measurement is incomplete if it does not state the uncertainty in the measurement along with the measured value.
- The uncertainty typically comes from a variety of sources:
  - Uncertainties that come from a lack of information are called **epistemic uncertainties**.
    - This category of uncertainty usually includes what you might be familiar with as **systematic errors** or **systematic effects**.
    - Epistemic uncertainties can be removed by performing calibrations or other auxiliary measurements.
    - Example: You want to measure the current in a circuit without taking it apart to insert an ammeter. So, you connect a voltmeter across a $1\,\mathrm{k\Omega}$ resistor that forms part of the circuit. You can then get the current in that part of the circuit using Ohm's law: $I = V/R$. But, the resistor only has a tolerance of 5% – its actual value is uncertain, and can be anywhere from $0.95\,\mathrm{k\Omega}$ to $1.05\,\mathrm{k\Omega}$. This means that if you measured $V = 1\,\mathrm{V}$ across the resistor with your voltmeter,[1] the current could be anywhere from $1.05\,\mathrm{mA}$ to $0.95\,\mathrm{mA}$. You could in principle remove the resistor from the circuit and measure the actual value, so the uncertainty is epistemic.

[1] There will also be uncertainty in the voltage measurement itself, we will see how to include that later in this lecture.

- Uncertainties that come from a fundamental random fluctuation in the quantity being measured are called **aleatory uncertainties**.
  - This category of uncertainty usually includes what you might be familiar with as **random errors** or **random effects**.
  - Aleatory uncertainties are a reflection of an intrinsic property of the thing being measured and hence cannot ever be completely eliminated from a measurement.
  - Example 1: You have perhaps heard the phrase "Poisson statistics" or "counting statistics" in association with radiation detectors. This phrase refers to the fact that, given a source that emits particles randomly at some *average* rate λ particles per second, there will be random variation in the number of particles emitted during a time interval Δt. *On average* there will be $N = \lambda \Delta t$ particles emitted in the time interval Δt, but the typical amount of variation can be shown to be $\pm\sqrt{\lambda \Delta t}$ particles. Hence, if you are trying to estimate $\lambda = N/\Delta t$, there will be an uncertainty of $\pm\sqrt{N}/\Delta t$ in your estimate for λ. You can reduce this uncertainty by collecting data for a longer period of time, but it will never go away completely.
  - Example 2: If you were to take very careful measurements of a power supply like you charge your phone with, you will find that the output isn't perfectly constant – there are variations from how the voltage is produced, as well as noise that is picked up by the circuit. Thus, even if you had a perfect meter,[2] you still would have to state the output of the supply with some uncertainty that comes from the fluctuations in the output.

2 Basics of probability and statistics

2.1 Random variables and probability density functions

- A **random variable** is a number that depends on the outcome of an experiment.[3] The variable is random in the sense that the outcome of the experiment is random – it is determined by some underlying probability law.

- A simple example of a random variable is as follows: let the experiment consist of flipping a coin. Take the random variable X to be 1 if the coin comes up heads, 0 otherwise.[4] The experiment in this case is clearly random – about half of the time the coin will be heads, about half of the time it will be tails, but we don't know which until we flip the coin.

[2] "Perfect" is used here both in the sense of being "ideal" as discussed earlier in the class, as well as in the sense of delivering an infinitely precise reading – such a thing does not exist!

[3] In our case, the "experiment" will be when you take a measurement.

[4] This is called a **Bernoulli random variable**.

- The **probability density function** (PDF) describes the likelihood that the random variable takes on a given value.
- Specifically, given the PDF $f_X(x)$ for random variable X, the probability that X lies between x_1 and x_2 is

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\,\mathrm{d}x \tag{1}$$

2.2 Summary quantities

- The PDF describes everything there is to know about the associated random variable, but can often contain more information than is useful.
- One way of summarizing the distribution is its **mean**:

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \tag{2}$$

  This is essentially the "center of mass" of the PDF.
- Depending on the application, a different way of summarizing the center of the distribution may be justified – for instance, the **mode** is the point for which $f_X(x)$ is maximum.
- The mean corresponds to the expected "typical" value – this is what we would quote for the quantity itself. When discussing uncertainty we are also interested in how much the distribution is spread out.
- The most common way to summarize this is the **variance**:

$$\sigma^2 = \mathrm{var}[X] = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,\mathrm{d}x \tag{3}$$

- You can re-write this in a more convenient form:

$$\sigma^2 = \mathrm{var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{4}$$

- The variance is akin to the "moment of inertia" of the PDF, taken about the mean.
- For our purposes, we will be very concerned with the **standard deviation**:

$$\sigma = \sqrt{\mathrm{var}[X]} \tag{5}$$

- We can discuss the relationship between two random variables in terms of their **covariance**:

$$\sigma_{XY} = \mathrm{cov}[X,Y] = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \tag{6}$$

- Note that people will sometimes write the covariance as σ^2_{XY} – you need to check the context and the units to make sure you understand the expression!

2.3 Important distributions

2.3.1 Normal (Gaussian) distribution

- The normal distribution is quite ubiquitous – it describes, for example, the bell-shaped curve of exam scores and the result of adding up many independent sources of noise.
- Its PDF is determined by the mean μ and standard deviation σ:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{7}$$

- A shorthand you will often see to say that a random variable X follows a normal distribution with mean μ and variance σ^2 is $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Because the normal distribution is the limiting case of many other distributions, it is often a very good model for the noise in a measurement.

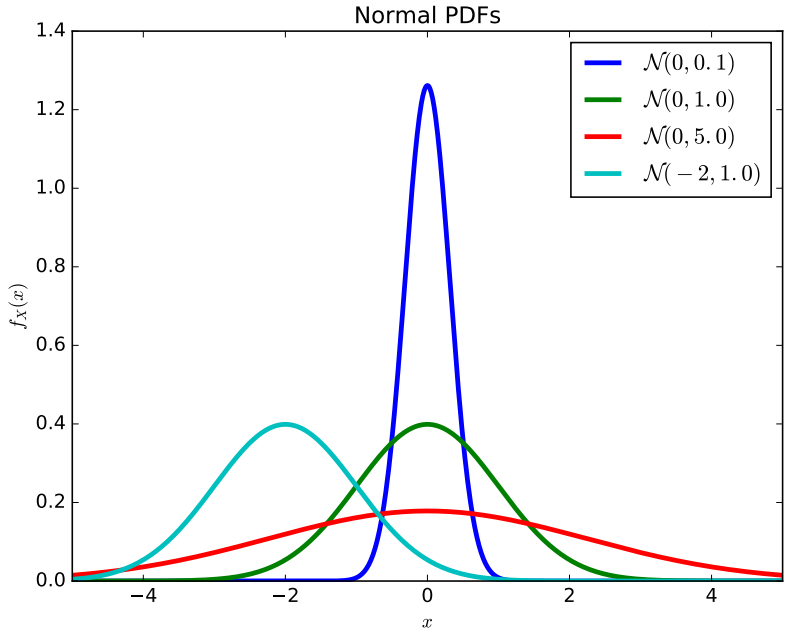

Figure 1: Several examples of the normal distribution for various values of μ, σ.

2.3.2 Uniform distribution

- A random variable that follows the uniform distribution has equal probability of lying anywhere in a given interval, and zero probability of lying outside the interval:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases} \tag{8}$$

- You should be able to convince yourself pretty easily that the mean of the uniform distribution is

$$\mu = \frac{a+b}{2} \tag{9}$$

- The variance of the uniform distribution is

$$\sigma^2 = \frac{(b-a)^2}{12} \tag{10}$$

- A shorthand to indicate that a random variable X follows the uniform distribution between a and b is $X \sim \mathcal{U}(a, b)$.

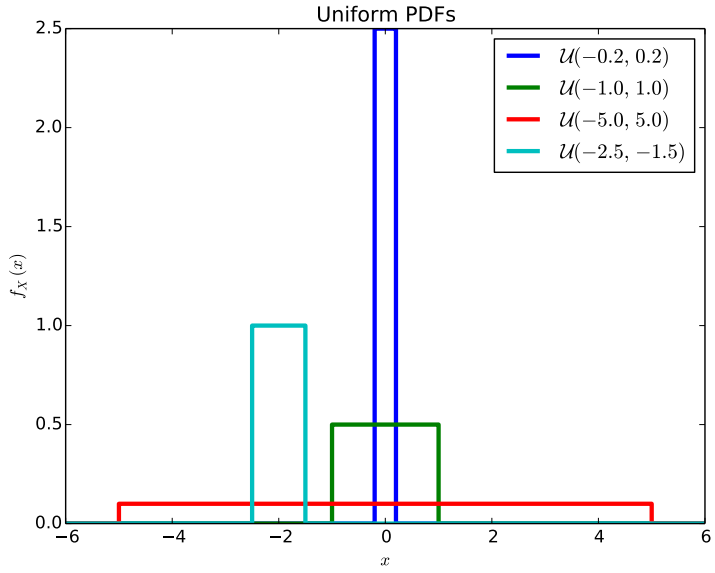

Figure 2: Several examples of the uniform distribution for various a, b.

2.4 Sample statistics

- One of the ways to reduce certain types of (aleatory) uncertainty is to take repeated measurements and compute the measured value and its uncertainty from the samples.
- The **sample mean** gives an estimate of the value of the quantity:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{11}$$

- The **sample standard deviation** gives an estimate of the width of the distribution, and is what will typically be used to quantitatively express the uncertainty:[5]

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{12}$$

- If each individual measurement x_i has the same uncertainty σ, then you can show that the mean computed from the samples has uncertainty

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \approx \frac{s}{\sqrt{N}}, \tag{13}$$

  where in the last step the sample standard deviation has been used as an approximation for σ. From this expression it is clear why taking a large number of samples is desirable: by taking many samples you can reduce the random components of uncertainty to the point that the epistemic uncertainties dominate. Note that this is only true when you are discussing the uncertainty in a mean: considering the power supply example from before, no amount of measurements will change the fact that the output is fluctuating and that this fluctuation represents an uncertainty when the value of the voltage is used elsewhere. But, additional measurements will let you pin down what the *mean value* of the output is.

[5] The factor is $1/(N-1)$ instead of $1/N$ as you might have expected for subtle statistical reasons. It is related to the fact that the expression involves $\overline{x}$, which was itself computed from the set of samples. This will inevitably lead to an underestimation of the variance, which using $1/(N-1)$ corrects for. Look up **Bessel's correction** if you would like to know more.

3 Expressing uncertainty

- There is no universal standard for how to express uncertainty in measurements.[6]
- You will typically encounter expressions of the form "1.0 ± 0.2," or sometimes the equivalent but less clear "1.0(2)."
  - This means that the best estimate of the quantity is 1.0, and it has an "uncertainty" of ± 0.2.
  - The exact interpretation of this uncertainty depends on context, and *should* be defined somewhere in the document it is used in.
  - The most common convention is that the uncertainty (here, 0.2) is the standard deviation of the probability distribution for the quantity.
- You might also see uncertainties written as intervals, such as "the 95% **confidence interval** for the quantity is $[0.6, 1.4]$."
  - Loosely speaking,[7] this means there is a 95% chance that the interval from 0.6 to 1.4 contains the true quantity.
  - 95% is a very popular confidence interval to use, with other common choices being 50% and 99% depending on the application.
  - *If* you are justified in treating the quantity being measured as normally-distributed, then σ is nicely related to the confidence interval:
    - $\pm 1\sigma$ is a 68.3% confidence interval.
    - $\pm 2\sigma$ is a 95.4% confidence interval.
    - $\pm 3\sigma$ is a 99.7% confidence interval.

4 Propagating uncertainty

- Aleatory uncertainties can usually be modeled as random variables that obey a probability distribution that can be determined from your measurements and other knowledge of the experiment.
  - This permits the use of the standard deviation as an uncertainty estimate, as described above.

[6] BIPM JCGM 100:2008 (a URL is provided at the end of these notes) is the closest thing to a standard that has widespread acceptance today, but it does not have universal adoption.

[7] The actual statistical meaning of the phrase "confidence interval" is a little more complex than this, but these nuances are beyond our scope. You may also see the phrase "**credible interval**," which *does* have the more obvious interpretation that there is a 95% (for instance) likelihood that the interval contains the correct value of the quantity, but forming credible intervals usually requires more assumptions than forming confidence intervals.

– Given some quantity x which is a function of variables u, v, and so on, the **uncertainty propagation equation** gives the uncertainty in x as

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \cdots + 2\sigma_{uv} \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + \ldots, \tag{14}$$

where the second set of terms includes all possible pairings of the variables that x depends on.

– For independent errors (so that $\sigma_{uv} = 0$) some specific cases of interest are:

$$x = au + bv \quad \rightarrow \quad \sigma_x^2 = a^2\sigma_u^2 + b^2\sigma_v^2 \tag{15}$$

$$x = auv \quad \rightarrow \quad \frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2}, \tag{16}$$

where a and b are constants (i.e., they have no uncertainty).

- It is usually much harder to justify treating epistemic uncertainties as random variables.

  – To see this, consider again the example of the resistor to measure current: the dominant uncertainty is the epistemic one arising from the fact that we only know the value of the resistor to within the $\pm 5\%$ as given by its tolerance band. Consider what would happen if you were to repeat the experiment many times: for each trial, the (unknown) value of the resistor stays basically the same, so there will be no fluctuation in the data you collect resulting from the unknown value of the resistor. If you tried to compute the uncertainty in the current based solely on the standard deviation of this set of trials, you would almost certainly get an absurdly narrow confidence interval on the current.[8]

  – One of the simpler correct ways to propagate epistemic uncertainty is what is called an "interval analysis" where you use the range of values the uncertain input (in this case the value of the resistor) can take to determine the range of values the uncertain output (in this case the current) can take.

  – Combining epistemic and aleatory uncertainty is, in general, a difficult problem. It is always preferred to reduce your epistemic uncertainties as much as possible!

[8] You could, however, use a *different* resistor of the same type for each trial, in which case the variability in the resistor would properly enter your measurements – this would permit a correct statistical interpretation of the variability due to the uncertainty on the resistance. But if you are going to do that much work, you should just measure the value of the one resistor properly from the start!

5 Linear regression

5.1 Example

Suppose you have a precision current source and a cheap voltmeter and you want to know the Thévenin equivalent for a "black box" two terminal circuit. You could measure the open-circuit voltage to get V_{th} then measure the voltage across the terminals for a given current and solve for R_{th}. But, you will reduce the impact of noise if you instead measure at a number of currents and find the line that best goes through the data:

$$V = a + bI \tag{17}$$

This also has the advantage that you can take a good look at the data to make sure there aren't any nonlinear circuit elements hiding in the black box. Example data are given in table 1 and the results are shown graphically in figure 3. The current is assumed to be very precisely known compared to the voltage. The voltage measurements are given with the number of figures that would be reported on the face of the meter. The uncertainties were estimated with reference to the manufacturer's data sheet:[9]

$$\sigma_V = (1\% \text{ of the reported value}) + (\text{least significant digit}[= 0.1]) \tag{18}$$

Using the analysis below, you can find

$$a = (0.08 \pm 0.15)\,\text{V}, \quad b = (21.8 \pm 0.5)\,\text{k}\Omega \tag{19}$$

The data were produced starting from $V_{th} = 0\,\text{V}$ and $R_{th} = 22.5\,\text{k}\Omega$, so the analysis was rather successful.

Table 1: Measurements for linear regression example

I [mA]	V [V]	σ_V [V]
0.1	2.2	0.122
0.2	4.6	0.146
0.3	6.6	0.166
0.4	8.6	0.186
0.5	11.1	0.211

[9] The data shown here were simulated, but this prescription is similar to that given in the data sheet for Fluke Series III multimeters. The Fluke instruments use 0.3% – 1% was used here to make the example more interesting.

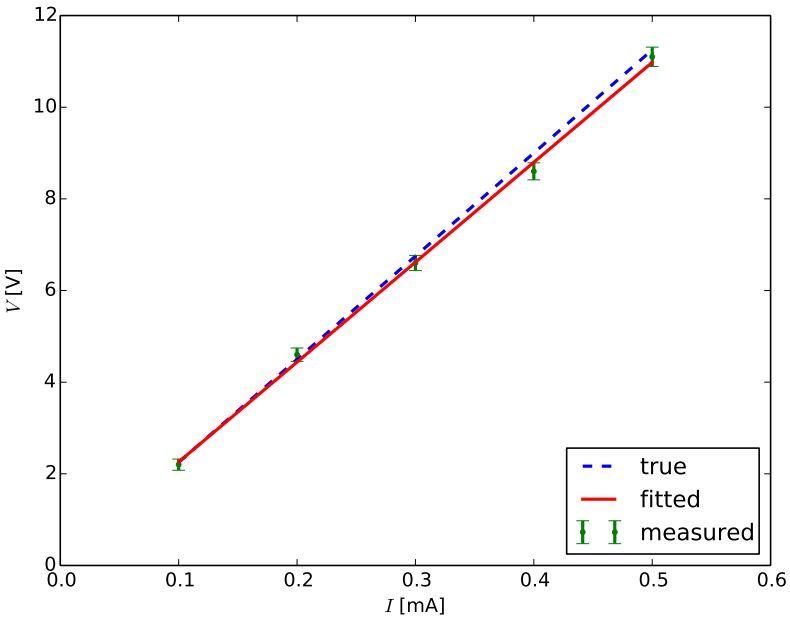

Figure 3: Example of linear regression. The vertical bars are **error bars**, which are a way of graphically representing the width of the distribution describing the value being measured – typically they are $\pm 1\sigma$, but this should always be spelled out in context. Because error bars are linked to the probability distribution, you should generally only include uncertainties that are modeled as such: typically error bars should only include the aleatory components of the uncertainty. Here, the error bars are $\pm 1\sigma$ and were computed according to equation 18.

5.2 Details of linear regression

- Consider the situation where you have measured a quantity y at several locations x: your observations consist of N pairs of (x_i, y_i).
- x is the **independent variable** and y is the **dependent variable**.
- It is often of interest to fit a model of the form[10]

$$y = a + bx \tag{20}$$

- Assume that the x_i have negligible uncertainty[11] and that the uncertainty

[10] The method given here is easily applied to more complicated models – see the book by Bevington and Robinson in the references for more details.

[11] When this assumption does not hold, a more sophisticated approach called **errors-in-variables regression** must be used.

in each of the y_i follows a normal distribution with mean $y(x_i) = a + bx_i$ and standard deviation σ_i.

- The probability of observing a given value of y_i at location x_i is then

$$P_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(y_i - y(x_i))^2}{2\sigma_i^2}\right] \tag{21}$$

- The probability of observing the entire set of N points is then

$$P = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - y(x_i))^2}{\sigma_i^2}\right] \tag{22}$$

- The **method of maximum likelihood** takes the values for the coefficients a and b that maximize this probability to be the best estimates.

- Maximizing the probability is the same as minimizing the (weighted) sum of the **squared residuals**:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - y(x_i))^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - a - bx_i)^2}{\sigma_i^2} \tag{23}$$

- Setting the derivatives with respect to a and b to zero and doing a bit of algebra gives the solution:

$$a = \mu_y^* - b\mu_x^* \tag{24}$$

$$b = \frac{\mathrm{cov}_*[x, y]}{\mathrm{var}_*[x]} \tag{25}$$

where the **weighted mean**, **weighted variance** and **weighted covariance** are

$$\mu_x^* = \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}, \quad \mu_y^* = \frac{\sum \frac{y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} \tag{26}$$

$$\mathrm{var}_*[x] = \frac{\sum \frac{(x_i - \mu_x^*)^2}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}, \quad \mathrm{cov}_*[x, y] = \frac{\sum \frac{(x_i - \mu_x^*)(y_i - \mu_y^*)}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} \tag{27}$$

- For the special case where all of the uncertainties are the same[12] ($\sigma_i = \sigma$), the weighted quantities from above are replaced with their unweighted counterparts:

$$a = \overline{y} - b\overline{x} \tag{28}$$

$$b = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{\mathrm{cov}[x, y]}{\mathrm{var}[x]} \tag{29}$$

$$\overline{x} = \frac{1}{N}\sum x_i, \quad \overline{y} = \frac{1}{N}\sum y_i \tag{30}$$

[12] A dataset having this property is said to be **homoscedastic**, as opposed to **heteroscedastic**.

- The value of χ^2 tells you how good the fit is: $\chi^2/(N-2)$ should be approximately one if the fitted line goes through the data nicely and the uncertainties were estimated properly.

- Using the uncertainty propagation equation, you can find the following uncertainties in the fitted intercept and slope, as well as their covariance:

$$\sigma_a^2 = \sigma_b^2 \frac{\sum \frac{x_i^2}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}, \quad \sigma_b^2 = \frac{1}{\mathrm{var}_*[x] \sum \frac{1}{\sigma_i^2}}, \quad \sigma_{ab} = -\frac{\mu_x^*}{\mathrm{var}_*[x] \sum \frac{1}{\sigma_i^2}} \tag{31}$$

- For homoscedastic data, these simplify to

$$\sigma_a^2 = \sigma_b^2 \frac{\sum x_i^2}{N}, \quad \sigma_b^2 = \frac{\sigma^2}{\mathrm{var}[x] N}, \quad \sigma_{ab} = -\frac{\overline{x}\sigma^2}{\mathrm{var}[x] N} \tag{32}$$

- Using these expressions, you can therefore obtain the slope b and an estimate of its uncertainty σ_b.

- Many software packages are available that will carry out all of these operations for you, but it is important to have this background in order to properly interpret what the software is telling you.

6 References and resources

- Brief and complete (but often somewhat terse) introduction to classical methods of data analysis: Bevington and Robinson, *Data Reduction and Error Analysis*, third edition. McGraw-Hill, 2003.

- Clear introduction to probability theory (but with only a few extensions to experimental statistics given): Bertsekas and Tsitsiklis, *Introduction to Probability*, second edition. Athena Scientific, 2008.

- Introduction to data analysis using the Bayesian framework: Sivia and Skilling, *Data Analysis: A Bayesian Tutorial*, second edition. Oxford, 2006.

- Nice summary of user-oriented recommendations for dealing with uncertainty in measurements: http://physics.nist.gov/cgi-bin/cuu/Info/Uncertainty/index.html

- International guidelines for expressing uncertainty in measurement: BIPM JCGM 100:2008. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

Assignment: Due 2/28

Required Problem

Write a program using the language of your choice that takes data (x_i, y_i) with heteroscedastic uncertainties σ_i and computes:

- a
- b
- σ_a
- σ_b
- χ^2

It is perfectly acceptable to write your program using an existing linear regression routine such as is present in most programming languages, just make sure you can specify heteroscedastic uncertainties and obtain the uncertainties on the fitted parameters.

Perform the following tests:

1. Run your code on the sample data from the lecture notes and make sure the results match.
2. Plot the data with error bars and the fitted line to visually verify the fit.
3. Check the value of $\chi^2/(N-2)$ as described in the lecture notes.

There is an additional dataset on page 100 of the book by Bevington and Robinson if you want to double-check your code (or you can try making your own test data set).

You will use this code for Extensive Lab 1, and your work will be checked on a new dataset in lab on 2/28. Make sure you have what you need to demonstrate your code with you that day (either test it on your lab computer ahead of time or bring your laptop).

Bonus Problem

Using the uncertainty propagation equation it is quite straightforward to obtain uncertainties on the result of evaluating $y = a + bx$. Obtain an expression for this uncertainty as a function of x, and plot the "uncertainty envelope" $y \pm \sigma$ on the figure asked for above.