

# Introduction to Error Analysis for the Physical Chemistry Laboratory

January 22, 2007

## 1 Introduction

In the physical chemistry laboratory you will make a variety of measurements, and then manipulate them to arrive at a numerical value for a physical property. However, without an estimate of the error of these numbers they are largely useless. Some published numbers in physics and chemistry are accurate to 10 significant figures or more, while others are only accurate to an order of magnitude (*no* significant figures!). The estimated error of a published number is a crucial piece of information that must be calculated.

This handout provides a practical introduction to the error analysis required in a typical physical chemistry lab. Error in an experiment is classified into two types: **random error** and **systematic error**. Random error arises from uncertainty in measurement devices, and is the subject of this handout. Systematic error includes all other sources of error, including simplifications in your model, biased instrumentation, impure reagents, etc. Random error measures the **precision** of the experiment, or the reproducibility of a given result. Systematic error measures the **accuracy** of a result, or how close a result is to the true value. The distinction between accuracy and precision is illustrated schematically in figure 1.

## 2 What is uncertainty?

Most chemists have an intuitive idea of what uncertainty is, but it is instructive to give a more rigorous definition. Suppose you perform an experiment to determine the boiling point of a liquid, and you measure 32 degrees. How confident are you in this number? This question could in principle be answered by repeating the experiment many times and collecting the results. If you took this large collection of results and counted all of the values that lie within specified intervals (e.g. between 30 and 30.1, 30.1 and 30.2, etc.) you could make a histogram plot (figure 2). You can see that the results are spread over a range of values. The width of this spread is a measure of the uncertainty in

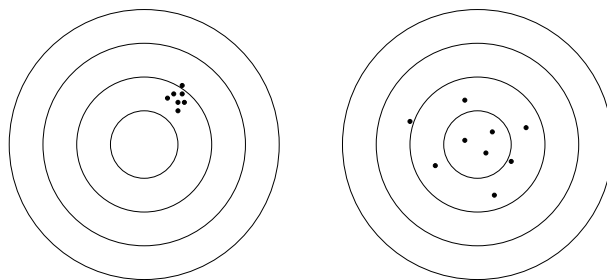


Figure 1: Schematic illustration of accuracy and precision. The left-hand target represents a high precision but low accuracy experiment. The right-hand target represents a low precision but high accuracy experiment

your initial measurement. The scatter can be quantified using the **standard deviation** ( $\sigma$ ) of the distribution which is defined as

$$\sigma = \sqrt{\sum_{i=1}^N \frac{(f_i - \bar{f})^2}{N - 1}} \quad (1)$$

where  $f_i$  are the results of your individual experiments,  $\bar{f}$  is the average of your results, and  $N$  is the number of trials performed. The standard deviation gives limits above and below a measured value in which subsequent experimental results will probably lie (with 70% certainty).

### 3 Error propagation

Most often in the physical chemistry lab you will only perform an experiment one or two times. In this case it is not possible to calculate the standard deviation of a large number of trials directly. Instead, it must be estimated by *propagating* the errors in the individual measurements that lead to your final result.

First of all, we need to discuss the uncertainties in individual measurements. Sometimes an instrument such as a volumetric flask will state its uncertainty in its technical specifications. If not, you can make an educated estimate of the uncertainty. If an instrument gives a digital reading, you can generally take the uncertainty to be half of the last decimal place. For example, if a digital thermometer reads 25.4 degrees, the uncertainty is 0.05 degrees. For analog instruments, first read the measurement to as many significant figures as there are marks on the gauge, and then estimate one more significant figure. The uncertainty should then be estimated based on how confident you are in the estimated significant figure. For example, if a mercury thermometer has marks at every degree, you would read the number of degrees, and estimate the tenths of a degree. The uncertainty might be plus or minus 0.2 degrees (e.g.  $25.4 \pm 0.2$  degrees).

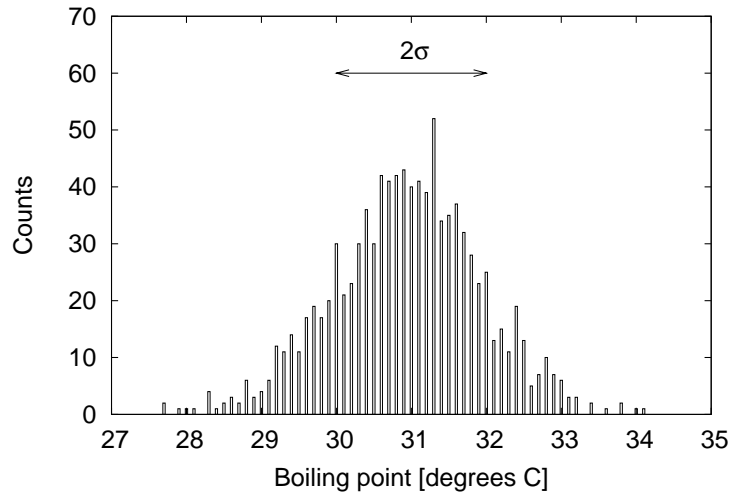


Figure 2: Distribution of a series of 1000 boiling point experiment results. The uncertainty in a single trial is related to the width of the distribution, and is called the standard deviation of the distribution ( $\sigma$ ).

To begin our discussion of error propagation, consider an experiment that measures some quantity  $x$ . The result we are looking for is some function  $f(x)$ . The measurement of  $x$  is subject to some uncertainty bounds, and the most general case does not assume symmetric uncertainties above and below  $x$ . In this case, the measured value  $x_0$  is within the range

$$x_0 - \sigma_- < x_0 < x_0 + \sigma_+ \quad (2)$$

where  $x_0$  is the measured value of  $x$ ,  $\sigma_+$  is the uncertainty above  $x_0$ , and  $\sigma_-$  is the uncertainty below  $x_0$ . The desired property  $f$  is then within the range

$$f(x_0 - \sigma_-) < f(x_0) < f(x_0 + \sigma_+) \quad (3)$$

If the uncertainty in  $x$  is assumed to be small, the uncertainty in  $f$  becomes

$$\sigma_f = \left. \frac{df}{dx} \right|_{x_0} \sigma_x \quad (4)$$

where  $\sigma_x$  is the uncertainty in  $x$ , which we now take to be symmetric. If we have a function of many variables, and if the errors are both small and *independent*, the uncertainty is

$$\sigma_f = \sqrt{\left( \left. \frac{\partial f}{\partial x} \right|_{x_0} \right)^2 \sigma_x^2 + \left( \left. \frac{\partial f}{\partial y} \right|_{y_0} \right)^2 \sigma_y^2 + \dots} \quad (5)$$

We can derive some special cases from equation 5. If a function only contains addition and subtraction operations, the uncertainty is

$$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2 + \dots} \quad (6)$$

If a function only contains multiplication and division operations, the uncertainty is

$$\sigma_f = |f| \times \sqrt{\left(\frac{\sigma_x}{x_0}\right)^2 + \left(\frac{\sigma_y}{y_0}\right)^2 + \dots} \quad (7)$$

Now that we are equipped with these formulas, we can proceed to propagate our individual uncertainties. We will illustrate this procedure with an example. Suppose you want to measure the molar heat of solvation of LiCl in water. This involves (1) weighing an amount of LiCl, (2) measuring a volume of water, and (3) measuring the temperature change when the reagent is dissolved. We will first calculate the heat of solvation ( $H$ ) itself, which is expressed in terms of our three measurements:

$$H = -\frac{C \times V \times T}{m/M} \quad (8)$$

Here  $m$  is the mass of LiCl,  $M$  is the molecular weight,  $C$  is the heat capacity of water per unit volume,  $V$  is the volume of water, and  $T$  is the temperature change. Note that the function  $H$  only contains multiplication/division operations, so we can use the error propagation rule for multiplication and division (equation 7). The variables we need to consider are  $m$ ,  $V$ , and  $T$ . We do not include  $M$  and  $C$  in our list of variables, because they are assumed to be known to much higher (relative) precision than  $m$ ,  $V$  and  $T$ . If they weren't, we would have to include them in our error analysis, even if we didn't measure them. Plugging our variables into equation 7, we have

$$\sigma_H = |H| \sqrt{\left(\frac{\sigma_m}{m}\right)^2 + \left(\frac{\sigma_V}{V}\right)^2 + \left(\frac{\sigma_T}{T}\right)^2} \quad (9)$$

**Problem 1** Calculate the heat of solvation of LiCl and its associated uncertainty as discussed above, if the mass is  $2.1 \pm 0.05$  g, the molecular weight is  $42.394 \pm 0.0005$  g/mol, the volume is  $0.1 \pm 0.02$  L, the temperature change is  $4 \pm 0.5$  K, and the heat capacity per volume is exactly 4.184 kJ/LK.

(Answer:  $H = -33.79$  kJ/mol,  $\sigma_H = 8.01$  kJ/mol, so you would report the heat of solvation as  $-33.8 \pm 8$  kJ/mol.)

**Problem 2** Perform the same calculation as in the last problem, but include the molecular weight in your list of error propagation variables.

(Answer: You should get the same answer as before, to at least 6 significant figures! This is why it is often possible to ignore variables known to high precision in your error propagation.)

Sometimes a function may contain both addition/subtraction and multiplication/division, in which case the two rules can be combined. The easiest way to do this is to break the calculation into steps. Going back to the heat of solvation experiment, suppose that two separate masses were weighed, and then both masses were added to the solvent. Now the total mass is  $m = m_1 + m_2$ , and the equation for the heat of solvation becomes

$$H = -\frac{C \times V \times T}{(m_1 + m_2)/M} \quad (10)$$

The first step is to calculate the uncertainty in  $m = m_1 + m_2$  using the error propagation rule for addition/subtraction, i.e.

$$\sigma_m = \sqrt{\sigma_{m_1}^2 + \sigma_{m_2}^2} \quad (11)$$

Then simply use the total mass  $m$  and its calculated uncertainty, and proceed as in Problem 1.

**Problem 3** Calculate the uncertainty for the previous example if the two weights were  $0.5 \pm 0.05$  g and  $1.1 \pm 0.05$  g, the volume is again  $0.1 \pm 0.02$  L, and the temperature change is  $3 \pm 0.5$  K.

(Answer:  $\sigma_m = 0.07$  g,  $H = 33.26$  kJ/mol, and  $\sigma_H = 8.78$  kJ/mol.)

Almost all of the error propagation you will do in the physical chemistry lab will only require the rules for addition/subtraction and multiplication/division. However, occasionally you might come across a more complicated function, in which case we need to use equation 5. For example, suppose you have determined  $\Delta G$  for a reaction and are interested in calculating the equilibrium constant:

$$K = \exp\left(-\frac{\Delta G}{RT}\right) \quad (12)$$

Assuming that  $R$  and  $T$  are known to high precision, we only need to calculate the partial derivative of  $K$  with respect to  $\Delta G$ :

$$\frac{\partial K}{\partial \Delta G} = \frac{-1}{RT} \exp\left(-\frac{\Delta G}{RT}\right) \quad (13)$$

and using equation 5, the uncertainty in  $K$  is

$$\sigma_K = \frac{-1}{RT} \exp\left(-\frac{\Delta G}{RT}\right) \sigma_{\Delta G} \quad (14)$$

In many experiments you will be required to calculate the slope or intercept of a linear function. For example, the rate of a unimolecular reaction obeys an exponential rate law

$$c(t) = A \exp(-kt) \quad (15)$$

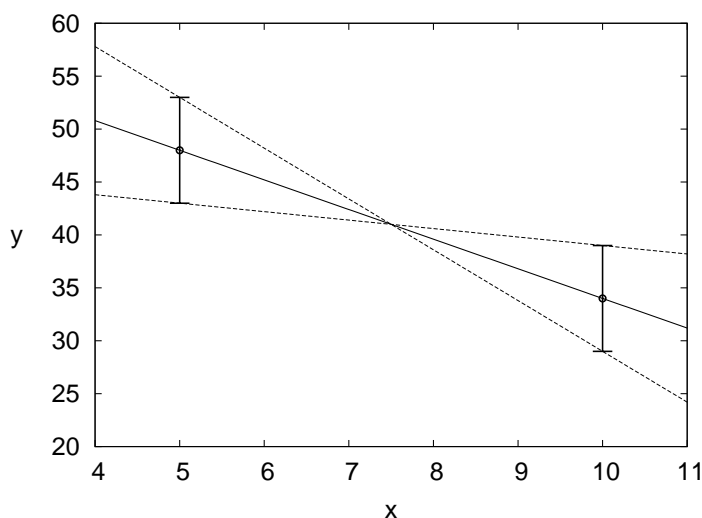


Figure 3: How to determine the uncertainty in the slope and intercept of two data points. The bars indicate the uncertainty in the  $y$  variable, and the dashed lines give upper and lower bounds for the line.

where  $c$  is the concentration of reagent,  $k$  is the rate constant,  $t$  is time, and  $A$  is a constant. If you want to calculate  $k$ , you will need values for  $c$  at different times. Taking the natural logarithm of equation 15 gives

$$\ln c = -kt + \ln A \quad (16)$$

which is a linear equation with the familiar form  $y = mx + b$ . In this case, what interests us is the slope as a function of  $t$ , which is our desired rate constant.

Using two data points,  $(x_1, y_1)$  and  $(x_2, y_2)$ , the slope and intercept can be calculated directly:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (17)$$

$$b = y_1 - mx_1 \quad (18)$$

and the uncertainties can be calculated using error propagation. If the error in  $x$  is negligible compared to the error in  $y$ , you can calculate the maximum and minimum values for the slope and intercept as illustrated in figure 3. Drawing a line through the upper bound of  $y_1$  and the lower bound for  $y_2$  gives a lower limit to the slope and an upper limit to the intercept. Similarly, drawing a line through the lower bound of  $y_1$  and the upper bound of  $y_2$  gives an upper limit to the slope and a lower limit for the intercept.

If you have many data points, the “average” slope and intercept can be calculated using **linear regression**. In practice this is always calculated using

a computer program, so the details of the procedure do not concern us. You simply need to become acquainted with a program capable of performing a linear regression with error analysis. The result of the calculation will give a value and an uncertainty for the slope and intercept.

## Further reading

- [1] P. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill, 2002.
- [2] E. B. Wilson Jr., *An Introduction to Scientific Research*, Dover, 1990.
- [3] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*; University Science Books, 1997.
- [4] D. P. Shoemaker, C. W. Garland, and J. W. Nibler, *Experiments in Physical Chemistry*, McGraw-Hill, 1996.