

CHEM1047 – Week 8 Lecture 2 – Error propagation and numerical accuracy

□ Chapter 15 of Monk and Munro, "Maths for Chemistry", 2nd edition.

Uncertainties are intrinsic to the measurement process – either the quantity itself fluctuates with time (air pressure, Brownian motion, etc.), or the instrument readout is imperfect and the measurement itself is consequently an uncertain process. It is often a requirement that such uncertainties be rigorously quantified and propagated through the calculation process.

1. Errors in linear functions

Consider the following function $f(x_1, x_2, \dots)$ of uncertain variables $\{x_1, x_2, \dots\}$:

$$f(x_1, x_2, \dots) = a_0 + a_1x_1 + a_2x_2 + \dots \quad (1)$$

where the coefficients $\{a_0, a_1, a_2, \dots\}$ are known exactly. From the properties of variance established in the previous lecture:

$$\text{Var}[f(x_1, x_2, \dots)] = \text{Var}[a_0] + \text{Var}[a_1x_1] + \text{Var}[a_2x_2] + \dots \quad (2)$$

The first term on the right hand side is zero, and the exact constants may be taken out:

$$\text{Var}[f(x_1, x_2, \dots)] = a_1^2 \text{Var}[x_1] + a_2^2 \text{Var}[x_2] + \dots \quad (3)$$

The squares appear because the definition of variance involves the square of the variable:

$$\text{Var}[ax] = \langle a^2x^2 \rangle - \langle ax \rangle^2 = \left(\langle x^2 \rangle - \langle x \rangle^2 \right) = a^2 \text{Var}[x] \quad (4)$$

We can now note that the coefficients a_k are actually partial derivatives of $f(x_1, x_2, \dots)$ with respect to the corresponding variables x_k , and therefore:

$$\text{Var}[f(x_1, x_2, \dots)] = \left(\frac{\partial f}{\partial x_1} \right)^2 \text{Var}[x_1] + \left(\frac{\partial f}{\partial x_2} \right)^2 \text{Var}[x_2] + \dots \quad (5)$$

Standard deviation was defined as the square root of the variance, therefore:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_{x_2}^2 + \dots \quad \Rightarrow \quad \sigma_f = \sqrt{\sum_k \left(\frac{\partial f}{\partial x_k} \right)^2 \sigma_{x_k}^2} \quad (6)$$

For linear functions this relation is exact.

2. Linear propagation of errors

The error propagation problem for arbitrary functions has no general solution, but an approximate solution may be obtained when the function is well-behaved and the errors are small. The function may be approximated in the vicinity of some reference point $\{x_0, y_0, z_0, \dots\}$ by a Taylor series up to the linear term with respect to each of its variables:

$$f(x, y, z, \dots) \approx f(x_0, y_0, z_0, \dots) + \left. \frac{\partial f}{\partial x} \right|_{x=x_0} (x - x_0) + \left. \frac{\partial f}{\partial y} \right|_{y=y_0} (y - y_0) + \left. \frac{\partial f}{\partial z} \right|_{z=z_0} (z - z_0) \quad (7)$$

The function itself and the reference point are known exactly. This means that everything except $\{x, y, z, \dots\}$ in Equation (7) is exact. This reduces Equation (7) to the form specified for the linear function in the previous section and allows us to use the expression we have derived for the error propagation through a linear function. The result, for statistically independent variables $\{x, y, z, \dots\}$, is:

$$\sigma_f \approx \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + \dots} \quad (8)$$

This relation may be pre-computed for common functions. All variables must be statistically independent.

Table 1. Error propagation rules for common functions

Function	Formula	Error propagation rule
addition / subtraction	$f(x, y) = x \pm y$	$\sigma_f \approx \sqrt{\sigma_x^2 + \sigma_y^2}$
scaling by an exact constant	$f(x) = kx$	$\sigma_f \approx k \sigma_x$
product	$f(x, y) = xy$	$\sigma_f \approx xy \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
ratio	$f(x, y) = \frac{x}{y}$	$\sigma_f \approx \left \frac{x}{y}\right \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
exact powers	$f(x) = x^k$	$\sigma_f \approx kx^{k-1} \sigma_x $
powers of an exact number	$f(x) = k^x$	$\sigma_f \approx k^x \sigma_x \ln k $
logarithm with an exact base	$f(x) = \log_k x$	$\sigma_f \approx \left \frac{\sigma_x}{x \ln k}\right $

In situations where Taylor series is not a good approximation, an alternative way of determining the standard deviation of a function of uncertain parameters is the *Monte-Carlo method*: the parameters are randomly varied within their statistical distributions and the function is recomputed until sufficient statistics is accumulated to determine the distribution parameters of the function.

Example 1: calculate the pressure and estimate its uncertainty for a $1.00 \pm 0.03 \text{ m}^3$ container holding 0.30 ± 0.02 moles of an ideal gas at the temperature of 298 ± 2 Kelvin. The numbers after the “ \pm ” symbol are standard deviations. Assume that the gas constant R is known exactly.

Solution: the ideal gas law is $PV = nRT$. Solving it for pressure yields $P = nRT/V = 743 \text{ Pa}$. The expression for the error propagation rule in Equation (8) is therefore:

$$\begin{aligned} \sigma_P &\approx \sqrt{\left(\frac{\partial P}{\partial n}\right)^2 \sigma_n^2 + \left(\frac{\partial P}{\partial T}\right)^2 \sigma_T^2 + \left(\frac{\partial P}{\partial V}\right)^2 \sigma_V^2} = \\ &= \sqrt{\left(\frac{RT}{V}\right)^2 \sigma_n^2 + \left(\frac{nR}{V}\right)^2 \sigma_T^2 + \left(\frac{nRT}{V^2}\right)^2 \sigma_V^2} = \dots = 54 \text{ Pa} \end{aligned}$$

3. Catastrophic loss of accuracy

There are important situations when manipulating uncertain quantities can result in the answer of such low accuracy as to render the calculation pointless. There are also rare cases when a calculation involving uncertain quantities may yield a very accurate result. Some cautionary tales are given below.

Example 2: calculate the sum of $x = 0.219 \pm 0.090$ and $y = -0.226 \pm 0.050$, and estimate its uncertainty.

Solution: using the expression from the first line of Table 1 yields $x + y = -0.007 \pm 0.103$, where the uncertainty is an order of magnitude bigger than the result.

Example 3: discuss if it would be wise to rely on the value of $1/(x + y)$ for any practical purposes when $x = 0.219 \pm 0.090$ and $y = -0.226 \pm 0.050$.

Solution: as we found in the previous example $x + y = -0.007 \pm 0.103$. The uncertainty interval includes zero on either side, meaning that the uncertainty interval of $1/(x + y)$ extends from minus infinity to plus infinity. Reporting any value would be dangerously misleading.

Example 4: a friend of yours happily proclaims that they had measured the pH of a solution “to the decimal place”, and reports $\text{pH} = 2.1 \pm 0.1$. What is the actual accuracy of the hydrogen ion concentration?

Solution: from the definition of pH, $[\text{H}^+] = 10^{-\text{pH}}$, meaning that $[\text{H}^+] = 7.9 \text{ mM}$, but $\sigma_{[\text{H}^+]} \approx |10^{-\text{pH}} \sigma_{\text{pH}} \ln 10| = 1.8 \text{ mM}$, meaning the error of nearly 25%.

Example 5: having stuck it to your friend, you went and measured the hydrogen ion concentration by potentiometric titration to be $7.92 \pm 0.03 \text{ mM}$. To what accuracy can you quote the pH in the report you will be sitting and writing alone in the empty library on a Friday night?

Solution: from the definition, $\text{pH} = -\log[\text{H}^+]$. Using the error propagation formula from Table 1 yields $\text{pH} = 2.1013 \pm 0.0016$.