
Propagation of Error

25 January 2019

I. The General Error Equation

A. Overview

Say you want to find the volume of an irregular object. You measure it by immersion in water in five separate trials and find the average volume \bar{V} with a standard deviation of s_V , which means that there is a 68.3% chance* that the actual volume of the object is between $\bar{V} - s_V$ and $\bar{V} + s_V$.

Now, what if you really want the density of the object? You could measure the mass in a number of trials to find \bar{m} and s_m , but then how can we find the values of \bar{d} and s_d given \bar{V} , \bar{m} , s_V , and s_m ? Can one say that $\bar{d} = \bar{m}/\bar{V}$ (in which case, how would s_d be calculated?), or does one have to calculate the density from each trial and then average the densities to find \bar{d} and s_d ? These questions are the subject of the current lecture.

B. The Average of a Calculated Value

Let $f(x, y)$ be some quantity that can be calculated using the measured values of x and y . It can be shown that the most *probable* result for the *expectation value* (a.k.a. the *population mean*; the average you would get if you did an infinite number of trials) for measured quantities is given by the arithmetic mean of a finite number of trials. The arithmetic mean is the usual definition for the mean: for example, if you measure the value of x n times and get the results x_1, x_2, \dots, x_n , then the arithmetic mean (i.e., average) of x is given by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Similarly, the average of y is calculated using the formula

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Note that n trials have been performed for both x and y . Using those n pairs of numbers, we can find n values of $f(x_i, y_i)$, where the average is expressed in

*If you'd like to know where this number comes from, please see Dr. Freitag.

an analogous way:

$$\bar{f} = \frac{1}{n} \sum_{i=1}^n f(x_i, y_i)$$

If we assume that the errors in measurement are relatively small, i.e., $x_i - \bar{x} \ll 1$ and $y_i - \bar{y} \ll 1$, then we can expand $f(x_i, y_i)$ in a multi-dimensional Taylor expansion about the mean values of x and y and drop out the second-order and higher terms. This gives us the following:

$$f(x_i, y_i) = f(\bar{x}, \bar{y}) + (x_i - \bar{x}) \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} + (y_i - \bar{y}) \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}}$$

Substituting this into the above mean for $f(x, y)$, we have

$$\begin{aligned} \bar{f} &= \frac{1}{n} \sum_{i=1}^n f(x_i, y_i) \\ &= \frac{1}{n} \sum_{i=1}^n f(\bar{x}, \bar{y}) + \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} + \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \\ &= f(\bar{x}, \bar{y}) + \frac{1}{n} \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \sum_{i=1}^n (x_i - \bar{x}) + \frac{1}{n} \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \sum_{i=1}^n (y_i - \bar{y}) \end{aligned}$$

Consider the sum in the middle term:

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= (x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) \\ &= \sum_{i=1}^n x_i - n\bar{x} \\ &= n\bar{x} - n\bar{x} \\ &= 0 \end{aligned}$$

... where we have used Eq. (1) to introduce \bar{x} in the penultimate step. The last term in \bar{f} above evaluates to zero in the same way. Therefore, we have the intuitive result (which is really only approximately true since we've dropped the higher-order terms in the Taylor expansion),

$$\bar{f} = f(\bar{x}, \bar{y})$$

In other words, the mean value of f is the value that is calculated using the mean values of x and y , so long as the errors are small.

For example, to find the average density from a set of data points for mass and volume, you don't have to calculate the density n times; you can just find the average of the mass and the average of the volume and then write $\bar{d} = \bar{m}/\bar{V}$.

C. The Sample Variance of a Calculated Value

Now let's consider the sample *variance* (i.e., the standard deviation squared) of f . By definition, the sample variances of x and y are given by

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{and} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

By analogy, the sample variance of f is given by

$$s_f^2 = \frac{1}{n-1} \sum_{i=1}^n [f(x_i, y_i) - \bar{f}]^2$$

This can be evaluated using the same Taylor expansion used above:

$$\begin{aligned} s_f^2 &= \frac{1}{n-1} \sum_{i=1}^n [f(x_i, y_i) - \bar{f}]^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(f(\bar{x}, \bar{y}) + (x_i - \bar{x}) \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} + (y_i - \bar{y}) \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} + \dots - f(\bar{x}, \bar{y}) \right)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left((x_i - \bar{x}) \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} + (y_i - \bar{y}) \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \right)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left[(x_i - \bar{x})^2 \left(\left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \right)^2 + 2(x_i - \bar{x})(y_i - \bar{y}) \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} + (y_i - \bar{y})^2 \left(\left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \right)^2 \right] \end{aligned}$$

... using the definitions for s_x^2 and s_y^2 , we have...

$$\begin{aligned} &= \left(\left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \right)^2 s_x^2 + 2 \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \left(\left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \right)^2 s_y^2 \\ &= \left(\left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \right)^2 s_x^2 + 2 \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} s_{xy} + \left(\left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \right)^2 s_y^2 \end{aligned}$$

... where we have defined the *sample covariance* in the last step:

$$s_{xy} \equiv \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

The sample covariance tells us something about the correlation between the values of x_i and the corresponding values of y_i . If there is a correlation, i.e., if x_i and y_i are both less (or more) than their averages for most values of i , then s_{xy} will be large and positive. On the other hand, if x_i is larger than its average and y_i is lower than its average (or vice versa) for most values of i , then s_{xy} will be large and negative. If the two variables x and y are independent, so that there is no correlation between values of x_i and y_i relative to their averages, then we will have $s_{xy} = 0$. In that case, the variance in f is

$$s_f^2 = \left(\left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{y}} \right)^2 s_x^2 + \left(\left. \frac{\partial f}{\partial y} \right|_{\bar{x}, \bar{y}} \right)^2 s_y^2$$

This procedure can easily be extended to more than two independent variables, such that if $f = f(x, y, z, \dots)$, the variance in f is

$$s_f^2 = \left(\frac{\partial f}{\partial x} \right)^2 s_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 s_y^2 + \left(\frac{\partial f}{\partial z} \right)^2 s_z^2 + \dots = \sum_{i=1}^k \left(\frac{\partial f}{\partial \xi_i} \right)^2 s_{\xi_i}^2 \quad (2)$$

where it is understood that the derivatives are evaluated at the mean values for the independent variables. This is the general equation for the error (variance) of a calculated value; it is the same as Eq. (2) on p. 11 of your lab manual.

II. Specific Formulas

A. Addition and Subtraction

For example, if we have $f(x, y) = x \pm y$, the variance in f is given by Eq. (2):

$$\begin{aligned} s_f^2 &= \left(\frac{\partial f}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 s_y^2 \\ &= \left(\frac{\partial}{\partial x}(x \pm y)\right)^2 s_x^2 + \left(\frac{\partial}{\partial y}(x \pm y)\right)^2 s_y^2 \\ &= s_x^2 + s_y^2 \end{aligned}$$

or simply

$$s_f = \sqrt{s_x^2 + s_y^2} \quad (3)$$

For less precise work, one can take a single standard deviation to be a reasonable estimate of the error,[†] such that the value of f is reported as:

$$f = (x \pm y) \pm \sqrt{s_x^2 + s_y^2}$$

Note that it is understood that the values of x and y in these formulas will be the mean values, and so the macron is usually omitted.

B. Multiplication and Division

Consider another example: $f(x, y) = xy^{\pm 1}$. In this case, Eq. (2) says the variance in f is:

$$\begin{aligned} s_f^2 &= \left(\frac{\partial f}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 s_y^2 \\ &= \left(\frac{\partial}{\partial x}xy^{\pm 1}\right)^2 s_x^2 + \left(\frac{\partial}{\partial y}xy^{\pm 1}\right)^2 s_y^2 \\ &= y^{\pm 2}s_x^2 + x^2y^{\pm 2-2}s_y^2 \end{aligned}$$

Dividing by $f^2 = x^2y^{\pm 2}$, we have:

$$\begin{aligned} \frac{s_f^2}{f^2} &= \frac{y^{\pm 2}}{x^2y^{\pm 2}}s_x^2 + \frac{x^2y^{\pm 2-2}}{x^2y^{\pm 2}}s_y^2 \\ &= \frac{s_x^2}{x^2} + \frac{s_y^2}{y^2} \end{aligned}$$

[†]As mentioned earlier, there is a 68.3% chance that the actual mean is within one standard deviation of the calculated average if the error is due to random fluctuations.

or

$$s_f = f \sqrt{\frac{s_x^2}{x^2} + \frac{s_y^2}{y^2}} \quad (4)$$

Note that we have the same formula for both multiplication and division. Again, remember that x and y are the mean values for these quantities.

C. Exponentiation

As a final example of a general formula, consider $f(x) = e^x$. Using the formula for the variance, Eq. (2), we have

$$s_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 s_x^2 = \left(\frac{\partial}{\partial x} e^x\right)^2 s_x^2 = (e^x)^2 s_x^2 = f^2 s_x^2$$

or simply

$$s_f = f s_x \quad (5)$$

Specific examples of error analysis using these formulas are given starting on p. 13 of the lab manual.

III. An Illustrative Example

Here we will consider another, more involved example and give some practical suggestions for performing the calculation.

A. The Clausius-Clapeyron Equation

In general chemistry you learned that the Clausius-Clapeyron equation can be used to calculate the vapor pressure of a substance at one temperature if the vapor pressure is known at another temperature, so long as the enthalpy of vaporization of the substance is also known, i.e.,

$$p_1 = p_2 \exp \left[\frac{\Delta_{\text{vap}}H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \right]$$

... where p_1 is the vapor pressure at temperature T_1 , p_2 is the vapor pressure at temperature T_2 , and R is the gas constant. Say we have the following measured values,

$$\begin{aligned} p_2 &= 101.32 \pm 0.05 \text{ kPa} \\ T_2 &= 373.15 \pm 0.12 \text{ K} \\ \Delta_{\text{vap}}H &= 40.67 \pm 0.02 \text{ kJ mol}^{-1} \\ T_1 &= 364.75 \pm 0.12 \text{ K} \\ R &= 8.314472 \pm 0.000005 \text{ J mol}^{-1}\text{K}^{-1} \end{aligned}$$

and we want to calculate p_1 and s_{p_1} . We *could* apply Eq. (2) directly and write

$$s_{p_1}^2 = \left(\frac{\partial p_1}{\partial p_2}\right)^2 s_{p_2}^2 + \left(\frac{\partial p_1}{\partial T_2}\right)^2 s_{T_2}^2 + \left(\frac{\partial p_1}{\partial \Delta_{\text{vap}}H}\right)^2 s_{\Delta_{\text{vap}}H}^2 + \left(\frac{\partial p_1}{\partial T_1}\right)^2 s_{T_1}^2 + \left(\frac{\partial p_1}{\partial R}\right)^2 s_R^2$$

However, it would be tedious to work out the derivatives and plug into the resulting equation. Instead, we can perform the calculation stepwise using the general formulas we have already derived for subtraction, multiplication and division, and exponentiation. This is made much easier with the help of a spreadsheet.

To begin, enter the known quantities and their uncertainties in the spreadsheet:

	A	B	C	D	E
1	Quantity			Uncertainty	
2	p ₂	101.32 ±		0.05 kPa	
3	T ₂	373.15 ±		0.12 K	
4	Δ _{vap} H	40.67 ±		0.02 kJ mol ⁻¹	
5		40670 ±		20 J mol ⁻¹	
6	T ₁	364.75 ±		0.12 K	
7	R	8.314472 ±		0.000005 J mol ⁻¹ K ⁻¹	

Now, we can do the calculations one at a time. Applying Eq. (4), the uncertainty in 1/T₂ is

$$s_{1/T_2} = \frac{1}{T_2} \sqrt{\frac{s_{T_2}^2}{T_2^2}} = \frac{s_{T_2}}{T_2^2}$$

There is an analogous equation for the error in 1/T₁. To perform the calculation, enter '=1/B3' in cell B9 to find the value of 1/T₂, and then calculate the uncertainty separately:

	A	B	C	D	E
1	Quantity			Uncertainty	
2	p ₂	101.32 ±		0.05 kPa	
3	T ₂	373.15 ±		0.12 K	
4	Δ _{vap} H	40.67 ±		0.02 kJ mol ⁻¹	
5		40670 ±		20 J mol ⁻¹	
6	T ₁	364.75 ±		0.12 K	
7	R	8.314472 ±		0.000005 J mol ⁻¹ K ⁻¹	
8					
9	1/T ₂	0.0026799 ±		=D3/POWER(B3,2)	
10	1/T ₁	0.0027416 ±		9.02E-07 K ⁻¹	

(Create your own spreadsheet for this example and make sure you're getting the same results. Note that 'POWER(B3,2)' means you are taking the value in cell B3 and raising it to the second power.) Once we have 1/T₂ and 1/T₁, we can subtract them and calculate the error using Eq. (3):

	A	B	C	D	E	F
1	Quantity			Uncertainty		
2	p ₂	101.32 ±		0.05 kPa		
3	T ₂	373.15 ±		0.12 K		
4	Δ _{vap} H	40.67 ±		0.02 kJ mol ⁻¹		
5		40670 ±		20 J mol ⁻¹		
6	T ₁	364.75 ±		0.12 K		
7	R	8.314472 ±		0.000005 J mol ⁻¹ K ⁻¹		
8						
9	1/T ₂	0.0026799 ±		8.618E-07 K ⁻¹		
10	1/T ₁	0.0027416 ±		9.02E-07 K ⁻¹		
11	1/T ₂ - 1/T ₁	-6.17E-05 ±		=SQRT(POWER(D9,2)+POWER(D10,2))		
12	Δ _{vap} H/R	4891.4712 ±		2.4054462 K		

Then, the quantity $\Delta_{\text{vap}}H/R$ can be calculated using Eq. (4):

	A	B	C	D	E	F	G
1	Quantity			Uncertainty			
2	p_2	101.32	±	0.05 kPa			
3	T_2	373.15	±	0.12 K			
4	$\Delta_{\text{vap}}H$	40.67	±	0.02 kJ mol ⁻¹			
5		40670	±	20 J mol ⁻¹			
6	T_1	364.75	±	0.12 K			
7	R	8.314472	±	0.000005 J mol ⁻¹ K ⁻¹			
8							
9	$1/T_2$	0.0026799	±	8.618E-07 K ⁻¹			
10	$1/T_1$	0.0027416	±	9.02E-07 K ⁻¹			
11	$1/T_2 - 1/T_1$	-6.17E-05	±	1.248E-06 K ⁻¹			
12	$\Delta_{\text{vap}}H/R$	4891.4712	±	=B12*SQRT(POWER(D5/B5,2) + POWER(D7/B7,2))			

Eq. (4) can be similarly used to find the error of the entire exponent, which will be labelled x for simplicity:

$$x = \frac{\Delta_{\text{vap}}H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right)$$

Once the exponent is known, we can apply Eq. (5) to find the error in the exponential:

	A	B	C	D	E
1	Quantity			Uncertainty	
2	p_2	101.32	±	0.05 kPa	
3	T_2	373.15	±	0.12 K	
4	$\Delta_{\text{vap}}H$	40.67	±	0.02 kJ mol ⁻¹	
5		40670	±	20 J mol ⁻¹	
6	T_1	364.75	±	0.12 K	
7	R	8.314472	±	0.000005 J mol ⁻¹ K ⁻¹	
8					
9	$1/T_2$	0.0026799	±	8.618E-07 K ⁻¹	
10	$1/T_1$	0.0027416	±	9.02E-07 K ⁻¹	
11	$1/T_2 - 1/T_1$	-6.17E-05	±	1.248E-06 K ⁻¹	
12	$\Delta_{\text{vap}}H/R$	4891.4712	±	2.4054462 K	
13	exponent (x)	-0.301884	±	-0.006104	
14	e^x	0.7394239	±	=B14*D13	
15	p_1	74.918426	±	0.4587899 kPa	

Finally, Eq. (4) can be used one last time to find the error in the product of p_2 and e^x . We can report the final result as

$$p_1 = 74.92 \pm 0.46 \text{ kPa}$$

B. Reducing the Uncertainty

Another benefit of doing the calculation like this is the following: you can easily change the uncertainties of various quantities to see how it affects the uncertainty of the final calculated quantity. For example, is it worth the time, effort, and expense to try and measure p_2 to ± 0.005 kPa rather than ± 0.05 kPa? Try it on your spreadsheet (the answer is no; the uncertainty in the final result changes in the thousandths place). What if we could measure the temperatures

	A	B	C	D	E
1	Quantity			Uncertainty	
2	p_2	101.32	±	0.01 kPa	
3	T_2	373.15	±	0.01 K	
4	$\Delta_{\text{vap}}H$	40.67	±	0.01 kJ mol ⁻¹	
5		40670	±	10 J mol ⁻¹	
6	T_1	364.75	±	0.01 K	
7	R	8.314472	±	0.000001 J mol ⁻¹ K ⁻¹	
8					
9	$1/T_2$	0.0026799	±	7.182E-08 K ⁻¹	
10	$1/T_1$	0.0027416	±	7.516E-08 K ⁻¹	
11	$1/T_2 - 1/T_1$	-6.17E-05	±	1.04E-07 K ⁻¹	
12	$\Delta_{\text{vap}}H/R$	4891.4712	±	1.2027223 K	
13	exponent (x)	-0.301884	±	-0.000514	
14	e^x	0.7394239	±	-0.00038	
15	p_1	74.918426	±	0.0392042 kPa	

Figure 1. The error calculation with uncertainties according to the significant figures assumption.

to the nearest hundredth of a degree? Yes, this will make a significant difference, reducing the uncertainty to ± 0.054 kPa. Therefore, we should focus our experimental efforts on better thermometers.

C. Comparison with the Significant Figures Approximation

Finally, it is interesting to compare this result with what you would have estimated the error to be using the rules for significant figures. Recall that the fundamental assumption in significant figures is that the last reported digit is uncertain by ± 1 . Therefore, we would have had the uncertainties shown in Fig. 1 instead. Note that since the uncertainties in each quantity have all decreased, the overall uncertainty decreases as well, and so now we know the pressure to the nearest four hundredths. However, using the ‘rules’ for significant figures, we would say that the result is 74.9 kPa, or

$$74.9 \pm 0.1 \text{ kPa}$$

whereas the actual error bar is

$$74.92 \pm 0.04 \text{ kPa}$$

So, while significant figures give you a fairly good estimate of the error (assuming the uncertainties are really ± 1 in the last digit, which they typically are not), for more precise work the full error calculation is essential.