

Simple uncertainty propagation and linear fits to uncertain measurements

J. L. Lanfranchi

April 2, 2014

1 Propagation of uncertainty

1.1 Basics

If you have some function of the variables x_1, x_2, x_3 , etc.: $f(x_1, x_2, x_3, \dots)$, and you want to find the uncertainty in f given the measured values $\hat{x}_1, \hat{x}_2, \hat{x}_3, \dots$ and uncertainties $\sigma_1, \sigma_2, \sigma_3, \dots$, respectively, which are all *statistically independent* of one another (more on this below), then use

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1} \Big|_{x_1=\hat{x}_1} \right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2} \Big|_{x_2=\hat{x}_2} \right)^2 \sigma_2^2 + \left(\frac{\partial f}{\partial x_3} \Big|_{x_3=\hat{x}_3} \right)^2 \sigma_3^2 + \dots \quad (1)$$

1.2 Example

It's easy to find the rule now for simple situations, such as $f(x, y) = xy$, as follows

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \sigma_y^2} \quad (2)$$

$$= \sqrt{y^2 \sigma_x^2 + x^2 \sigma_y^2} \quad (3)$$

$$= \sqrt{x^2 y^2 \left(\frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} \right)} \quad (4)$$

$$= xy \sqrt{\frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}}. \quad (5)$$

Now you just have to plug in the measured values for each of the dependent variables (the measured value of x is called \hat{x} and the measured value of y , \hat{y}) and your estimates for the uncertainties in these measured values, σ_x and σ_y , respectively. Of course you can use Eqn. 3, 4, or 5, whichever form is most convenient, since they're all equivalent.

1.3 Exercises

Find the expression for the uncertainty in f given uncertainties σ_x and σ_y . (These simple examples should cover many of the scenarios you'll encounter.)

a) $f(x, y) = 2x - 3y$

b) $f(x) = 3x^2$

c) $f(x) = 1/x$

d) $f(x, y) = x/y$

e) $f(x) = \log(x)$

1.4 Assumption of statistical independence

As mentioned above, Eqn. 1 only applies to statistically-independent errors. There are situations where the uncertainty in one variable is dependent upon that of another to some non-zero degree. This requires some modification of the equations, which I'll not discuss here in this "simple" treatment.

1.5 Bias due to nonlinearity

Note that there is bias introduced by using Eqn. 1 for non-linear functions of the variables (this bias lessens the smaller the uncertainty, but can be quite significant nonetheless), but don't worry about that for this class. Just don't forget about this! You'll deal with it more at a later stage in your education and in your working life, whatever scientific, engineering, or data-analytic discipline (economics, finance, etc.) you land in.

2 Linear fit to measurements with uncertainty

First, a disclaimer: Once again, the "real world" presents many common situations that add complexity to what follows. For example, if you want to fit a line to *transformed* data (e.g., the logarithm of the measurements, or the reciprocal of the measurements), then what follows isn't strictly true.¹ Non-normally-distributed errors further complicate the analysis.

For now, just use what follows; just remember that there's more to the story. (And that's what's *fun* about the world: there's always more to discover about it!)

2.1 Uncertainty in only y -values

I'm going to start with fitting a line to data with uncertainties only in the y measurements, which I'm now using as the dependent variable (i.e., $y = f(x)$). I'm switching

¹ See <http://arXiv.org/abs/0706.1062v2> for how to deal with fitting a line to a $\log(y)$ vs. $\log(x)$ plot of data gathered from a power-law process, i.e., $f(x) = 1/x^n$. Power-law processes are *extremely* common to find in practice.

to this notation because you're used to thinking about the "x-axis" and the "y-axis," and the equation for a line is commonly written $y = mx + b$.

Note that you generally have uncertainty in *both* independent and dependent variables, although occasionally you don't have uncertainty (or it's relatively insignificant) in the independent variable. One example of where it is possible to see no error in the independent variable is in numerical simulations that you might perform on a computer; the input you give is an exact, known value (note that the *output* of a computer simulation is almost never exact due to things like round-off error). It is also possible that the relative error in the independent variable is so small that it can be ignored for simple analyses.

For N measurements numbered from $i = 1$ to $i = N$ with errors in the measured dependent variable σ_i , define the following intermediate quantities

$$S \equiv \sum_{i=1}^N \frac{1}{\sigma_i^2} \quad (6)$$

$$S_x \equiv \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \quad (7)$$

$$S_y \equiv \sum_{i=1}^N \frac{y_i}{\sigma_i^2} \quad (8)$$

$$t_i \equiv \frac{1}{\sigma_i} \left(x_i - \frac{S_x}{S} \right) \quad (9)$$

$$S_{tt} \equiv \sum_{i=1}^N t_i^2 \quad (10)$$

then the parameters of slope, m , and y -intercept, b , in the equation $y = mx + b$ are given by

$$m = \frac{1}{S_{tt}} \sum_{i=1}^N \frac{t_i y_i}{\sigma_i} \quad (11)$$

$$b = \frac{S_y - S_x m}{S} \quad (12)$$

with uncertainties

$$\sigma_m^2 = \frac{1}{S_{tt}} \quad (13)$$

$$\sigma_b^2 = \frac{1}{S} \left(1 + \frac{S_x^2}{S S_{tt}} \right). \quad (14)$$

2.2 Uncertainty in x - and y -values

Now I'm going to treat a linear fit to data with uncertainties in both y , the dependent variable, *and* x , the independent variable (i.e., $y = f(x)$). This will most likely be the case for the data you're measuring. Also particular to the following analysis, which you have to be careful about, is that it's only valid if the uncertainties are the same for each measurement. Uncertainties in x (σ_x) and y (σ_y) *can* be different from one another.

2.3 Strategy for using linear fits

In order to take advantage of your linear fit, re-write the equation that relates the y -variable to the x -variable, and group the parameter(s) such that they look like

$$y = (\text{slope}) \cdot x + (\text{y-intercept}). \quad (15)$$

If the parameter appears anywhere where the slope goes, either multiplying or dividing x , then you can set the expression multiplying/dividing x equal to the slope of your linear fit, and then solve for the parameter you're looking for. Similarly, if the parameter appears in the part of the expression that *adds to* the x -term, then set the that expression equal to the y -intercept of your linear fit and solve for the parameter you're interested in.

For example, let's look at calculating the resistivity of a wire from the Ohm's Law lab. Resistance in a wire with constant cross-section is given by

$$R = \rho L / A \quad (16)$$

where ρ is the resistivity, L is the length of wire, and A is the cross-sectional area of the wire. You took measurements of R at various lengths of wire L and you'll plot R vs. L , so re-write Eqn. 16 in terms of a linear function $R(L)$

$$R = R(L) = (\text{slope}) \cdot L + (\text{y-intercept}) \quad (17)$$

$$= \left(\frac{\rho}{A} \right) \cdot L \quad (18)$$

Recall that you had a non-zero y -intercept for that lab *not* in the resistance vs. wire-length model, but due to a systematic offset to R due to how you measured it, which included the internal resistance of the digital multimeter (DMM) and the resistance of the hook-up wires. This can be modeled by modifying the equation above to include this as an additive offset to the resistance measurements:

$$R = \left(\frac{\rho}{A} \right) \cdot L + (R_{\text{DMM}} + R_{\text{hook-up wires}}). \quad (19)$$

Now to find the resistivity, just set the quantity in parentheses multiplying L equal to the slope from your linear fit, and solve for ρ :

$$\frac{\rho}{A} = \text{slope} \quad (20)$$

$$\rho = A \cdot \text{slope}. \quad (21)$$

Save for finding the cross-sectional area, A , of the wire, you now have an estimate of resistivity, ρ , that takes into account *every data point* you used for your linear fit.