From: An Introduction to Computer Simulation Methods Third Edition (revised)
by Harvey Gould, Jan Tobochnik, and Wolfgang Christian

Ch 11: Numerical and Monte Carlo Methods

*Problem 11.12.** Error estimating by bootstrapping

Suppose that we have made a series of measurements, but do not know the underlying probability distribution of the data. How can we estimate the errors of the quantities of interest in an unbiased way? One way is to use a method known as *bootstrapping*, a method that uses random sampling to estimate the errors.

Consider a set of n measurements such as n values of the pairs (x_i, y_i), and suppose we want to fit this data to the best straight line. If we label the original set of measurements, $M = \{m_1, m_2, \dots m_n\}$, then the kth *resampled* data set M_k consists of n measurements that are randomly chosen from the original set. This procedure means that some of the m_i may not appear in M_k and some may appear more than once. We then compute the quantity G_k from the resampled data set. For example, G_k could be the slope found from a least squares calculation. If we do this resampling n_r times, a measure of the error in the quantity G is given by σ_G^2, where

$$\sigma_G = \frac{1}{n_r - 1} \sum_{k=1}^{n_r} \big[G_k - \langle G_k \rangle\big]^2. \tag{11.26}$$

with

$$\langle G_k \rangle = \frac{1}{n_r} \sum_{k=1}^{n_r} G_k. \tag{11.27}$$

a. To see how this procedure works, consider $n = 15$ pairs of points x_i randomly distributed between 0 and 1, with the corresponding values of y given by $y_i = 2x_i + 3 + s_i$, where s_i is a uniform random number between -1 and $+1$. First compute the slope, m, and the intercept, b, using the least squares method and their corresponding errors using (7.41).

b. Resample the same set of data 200 times, computing the slope and intercept each time using the least squares method. From your results estimate the probable error for the slope and intercept using (11.26). How well do the estimates from bootstrapping compare with the direct error estimates found in part a? Does the average of the bootstrap values for the slope and intercept equal m and b, respectively, from the least squares fits. If not why not? Do your conclusions change if you resample 800 times?