Homework 1; due on Thursday, September 15

PY 502, Computational Physics (Fall 2011)

Department of Physics, Boston University

Instructor: Anders Sandvik

FITTING A STRAIGHT LINE TO DATA

In experimental and computational science one frequently encounters the problem of fitting a function with a number of parameters to a set of measured (or simulated) data points. Here we will consider the simplest case of fitting a straight line. You will derive the expression for the line parameters and their statistical errors, and then write a program that reads data from a file and gives as output the line parameters and errors.

We have N measured data pairs (x_i, y_i), with y_i affected by a statistical error σ_i corresponding to one standard deviation of the distribution of y_i (a normal distribution is assumed). The values x_i are assumed to have no errors. We want to draw the "best" line, $y = a + bx$, through the data points, i.e., we want to minimize

$$\chi^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2/\sigma_i^2. \tag{1}$$

The line parameters a and b are obtained by setting the corresponding derivatives to zero;

$$\frac{\partial \chi^2}{\partial a} = 0, \qquad \frac{\partial \chi^2}{\partial b} = 0. \tag{2}$$

The statistical errors of the parameters are obtained from their derivatives with respect to y_i: The leading-order changes in a and b if y_i is changed by an amount σ_i is

$$\delta a_i = \sigma_i \frac{\partial a}{\partial y_i}, \qquad \delta b_i = \sigma_i \frac{\partial b}{\partial y_i}. \tag{3}$$

With the errors σ_i independent normally distributed stochastic variables, the variance of the parameters a, b are given by the sums of the individual variances $\delta a_i, \delta b_i$. The statistical errors (the standard deviations) of the parameters are thus

$$\sigma_a = \sqrt{\sum_i (\delta a_i)^2}, \qquad \sigma_b = \sqrt{\sum_i (\delta b_i)^2}. \tag{4}$$

The quality of the fit is quantified by χ^2. If N is large, one would expect χ^2 per degree of freedom, here $\chi^2/(N-2)$, to be close to 1. The distribution of χ^2 for given N (assuming that the data indeed is described by a straight line and the errors are normally distributed) can be found in statistical tables.

Assignment: Derive expressions for a, b, σ_a, σ_b in terms of the data $\{x_i, y_i, \sigma_i\}$, using the approach outlined above. Write a Fortran 90 program that reads data from a file. The program should ask the user to give the name of the file. The data is assumed to be in the format of three columns, corresponding to x_i, y_i, σ_i (as in the example file `xy.dat` given on the course web site). The user should not have to give the number of data, N; it should be obtained by the program from the length (the number of rows) of the input file. As output, written on the screen, the program should give the line parameters and errors, the number N, as well as the goodness of the fit; $\chi^2/(N-2)$.

You should test the program using the data set given on the course web site and produce a graph showing the data with error bars and the fitted line. Turn in this graph, showing also the output values produced by your program in this case, along with your derivations of the line parameters. Submit your program as an e-mail attachment to `py502@buphy.bu.edu`.

You have to submit your solution by the due date; for each late day 10% of the maximum will be deducted from your score.