and b, we can calculate the deviations Δy_i between each of the observed values y_i and the corresponding calculated or fitted values

$$\Delta y_i = y_i - y(x_i) = y_i - a - bx_i \tag{6.4}$$

With well chosen parameters, these deviations should be relatively small. However, the sum of these deviations is not a good measure of how well our calculated straight line approximates the data because large positive deviations can be balanced by negative ones to yield a small sum even when the fit of the function $y(x)$ to the data is bad. We might consider instead summing the absolute values of the deviations, but this leads to difficulties in obtaining an analytical solution. Instead we sum the squares of the deviations.

There in no correct unique method for optimizing the parameters valid for all problems. There exists, however, a method that can be fairly well justified, that is simple and straightforward, and that is well established experimentally. This is the *method of least squares,* similar to the method discussed in Chapter 4, but extended to include more than one variable. It may be considered as a special case of the more general *method of maximum likelihood.*

Method of Maximum Likelihood

Our data consist of a sample of observations drawn from a parent distribution that determines the probability of making any particular observation. For the particular problem of an expected linear relationship between dependent and independent variables, we define parent parameters a_0 and b_0 such that the actual relationship between y and x is given by

$$y_0(x) = a_0 + b_0 x \tag{6.5}$$

We shall assume that each individual measured value of y_i is itself drawn from a Gaussian distribution with mean $y_0(x_i)$ and standard deviation σ_i. We should be aware that the Gaussian assumption may not always be exactly true. In Example 6.2 the $y_i = C_i$ were obtained in a counting experiment and therefore follow a Poisson distribution. However, for a sufficiently large number of counts y_i the distribution may be considered to be Gaussian. We shall discuss fitting with Poisson statistics in Section 6.6.

With the Gaussian assumption, the probability P_i for making the observed measurement y_i with standard deviation σ_i for the observations about the actual value $y_0(x_i)$ is

$$P_i = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - y_0(x_i)}{\sigma_i}\right]^2\right\} \tag{6.6}$$

The probability for making the observed set of measurements of the N values of y_i is the product of the probabilities for each observation:

$$P(a_0, b_0) = \Pi P_i = \prod\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) \exp\left\{-\frac{1}{2}\sum\left[\frac{y_i - y_0(x_i)}{\sigma_i}\right]^2\right\} \tag{6.7}$$

where the product Π is taken with i ranging from 1 to N and the product of the exponentials has been expressed as the exponential of the sum of the arguments. In these products and sums, the quantities $1/\sigma_i^2$ act as weighting factors.

Similarly, for any *estimated* values of the parameters a and b, we can calculate the probability of obtaining the observed set of measurements

$$P(a, b) = \prod\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}\sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\} \tag{6.8}$$

with $y(x)$ defined by Equation (6.1) and evaluated at each of the values x_i.

We assume that the observed set of measurements is more likely to have come from the parent distribution of Equation (6.5) than from any other similar distribution with different coefficients and, therefore, the probability of Equation (6.7) is the maximum probability attainable with Equation (6.8). Thus, the maximum-likelihood estimates for a and b are those values that maximize the probability of Equation (6.8).

Because the first factor in the product of Equation (6.8) is a constant, independent of the values of a and b, maximizing the probability $P(a, b)$ is equivalent to minimizing the sum in the exponential. We define this sum to be our goodness-of-fit parameter χ^2:

$$\chi^2 = \sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2 = \sum\left[\frac{1}{\sigma_i}(y_i - a - bx_i)\right]^2 \tag{6.9}$$

We use the same symbol χ^2, defined earlier in Equation (4.32), because this is essentially the same definition in a different context.

Our method for finding the optimum fit to the data will be to find values of a and b that minimize this weighted sum of the squares of the deviations χ^2 and hence, to find the fit that produces the smallest sum of the squares or the *least-squares fit*. The magnitude of χ^2 is determined by four factors:

1. Fluctuations in the measured values of the variables y_i, which are random samples from a parent population with expectation values $y_0(x_i)$.
2. The values assigned to the uncertainties σ_i in the measured variables y_i. Incorrect assignment of the uncertainties σ_i will lead to incorrect values of χ^2.
3. The selection of the analytical function $y(x)$ as an approximation to the "true" function $y_0(x)$. It might be necessary to fit several different functions in order to find the appropriate function for a particular set of data.
4. The values of the parameters of the function $y(x)$. Our objective is to find the "best values" of these parameters.

6.3 MINIMIZING χ^2

To find the values of the parameters a and b that yield the minimum value for χ^2, we set to zero the partial derivatives of χ^2 with respect to each of the parameters

$$\begin{aligned}\frac{\partial}{\partial a}\chi^2 &= \frac{\partial}{\partial a}\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx)^2\right]\\ &= -2\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)\right] = 0\\ \frac{\partial}{\partial b}\chi^2 &= \frac{\partial}{\partial b}\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)^2\right]\\ &= -2\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)\right] = 0\end{aligned} \tag{6.10}$$

These equations can be rearranged as a pair of linear simultaneous equations in the unknown parameters a and b:

$$\begin{aligned}\sum\frac{y_i}{\sigma_i^2} &= a\sum\frac{1}{\sigma_i^2} + b\sum\frac{x_i}{\sigma_i^2}\\ \sum\frac{x_i y_i}{\sigma_i^2} &= a\sum\frac{x_i}{\sigma_i^2} + b\sum\frac{x_i^2}{\sigma_i^2}\end{aligned} \tag{6.11}$$

The solutions can be found in any one of a number of different ways, but, for generality we shall use the method of determinants. (See Appendix B.) The solutions are

$$\begin{aligned}a &= \frac{1}{\Delta}\begin{vmatrix}\sum\frac{y_i}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2}\\ \sum\frac{x_i y_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2}\end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2}\right)\\ b &= \frac{1}{\Delta}\begin{vmatrix}\sum\frac{1}{\sigma_i^2} & \sum\frac{y_i}{\sigma_i^2}\\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i y_i}{\sigma_i^2}\end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right)\\ \Delta &= \begin{vmatrix}\sum\frac{1}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2}\\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2}\end{vmatrix} = \sum\frac{1}{\sigma_i^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2\end{aligned} \tag{6.12}$$

For the special case in which all the uncertainties are equal ($\sigma = \sigma_i$), they cancel and the solutions may be written

$$\begin{aligned}a &= \frac{1}{\Delta'}\begin{vmatrix}\Sigma y_i & \Sigma x_i\\ \Sigma x_i y_i & \Sigma x_i^2\end{vmatrix} = \frac{1}{\Delta'}(\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i)\\ b &= \frac{1}{\Delta'}\begin{vmatrix}N & \Sigma y_i\\ \Sigma x_i & \Sigma x_i y_i\end{vmatrix} = \frac{1}{\Delta'}(N\Sigma x_i y_i - \Sigma x_i \Sigma y_i)\\ \Delta' &= \begin{vmatrix}N & \Sigma x_i\\ \Sigma x_i & \Sigma x_i^2\end{vmatrix} = N\Sigma x_i^2 - (\Sigma x_i)^2\end{aligned} \tag{6.13}$$