Chi-Square Minimization With Correlated Errors
John Fowler
31 July, 2002

The familiar definition of a chi-square random variable with N degrees of freedom is

$$\chi_N^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2}$$

where x_i is a sample of a Gaussian random variable whose mean is indicated by the bar and whose standard deviation is σ_i (if the samples are all from the same population, and if the mean is computed from the sample, then the number of degrees of freedom drops to *N-1*). What is somewhat less familiar is the fact that this expression is really a special case of a more general expression. We can define a vector **u** whose components are the terms squared in the numerators in the sum above, and we can define an error covariance matrix **Ω** whose diagonal contains the corresponding variances:

$$u \equiv (u_1, u_2, u_3, \ldots, u_N)$$

$$u_i \equiv x_i - \bar{x}_i$$

$$\Omega \equiv \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & . & . \\ 0 & 0 & . & \sigma_N^2 \end{pmatrix}$$

$$W \equiv \Omega^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ 0 & 0 & . & . \\ 0 & 0 & . & \frac{1}{\sigma_N^2} \end{pmatrix}$$

Now we can write the chi-square as

$$\chi_N^2 = u \; W \; u^T$$

where the superscript *T* indicates the transpose of the vector (whether the **u** vector is defined as a row or column vector is arbitrary, as is whether the superscript *T* is applied to the **u** on the left or the one on the right). This is the general expression for chi-squares. When the error covariance matrix is diagonal, it is algebraically equivalent to the expression at the top of the page, but it also applies when the error covariance matrix is *not* diagonal, in which case ***W***, the inverse matrix, is not simply a diagonal matrix containing inverse variances.

This can be applied to curve fitting based on chi-square minimization as follows. Let the regression equation be indicated as

$$y = f(x)$$

where x is the independent variable, y is the dependent variable, $f(x)$ is any desired functional form involving coefficients a_n, (linear or nonlinear with regard to x and a_n), and we have a set of data pairs (x_i, y_i), $i = 1, N$, to which we wish to fit the given functional form. In the case of interest, the errors in the y samples are generally correlated, and we have the error covariance matrix $\mathbf{\Omega}$ whose inverse is $\boldsymbol{W}$. We define

$$u_i \equiv y_i - f(x_i)$$

and treat the u_i as components of a vector. Then we carry out the vector-matrix-vector multiplications in the general expression for the chi-square to obtain

$$\chi^2 = \sum_i \sum_j w_{ij} \, u_i \, u_j$$

where the sums are from one to N, and we suppress the subscript N on the chi-square to avoid clutter in what follows. The functional form to be fit contains coefficients a_n, and the system of equations that must be solved in order to evaluate these coefficients is

$$\frac{\partial \chi^2}{\partial a_n} = 0$$

and the error covariance matrix $\boldsymbol{C}$ expressing the uncertainties in these coefficients is obtained from

$$\alpha_{mn} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_m \, \partial a_n}$$
$$C = \alpha^{-1}$$

The second-derivative matrix is called the *Hessian* matrix; it is multiplied by ½ and inverted to obtain the error covariance matrix for the coefficients. The error model that depends on this covariance matrix may be derived by writing the regression equation in terms of *true* values plus *noise* as follows. Let y_T denote the true value of y, and let ε_y denote the error in y. Similarly, let a_{Tn} denote the true value of a_n, and let ε_n denote the error in a_n. Then

$$y = y_T + \varepsilon_y$$
$$a_n = a_{Tn} + \varepsilon_n$$

We substitute these in the regression equation and subtract $y_T = f_T(x)$, where the subscript T on $f(x)$ indicates the functional form using the true values of the coefficients, a_{Tn}.

$$\varepsilon_y = f(x) - f_T(x)$$

$$\sigma_y^2 = \left\langle \varepsilon_y^2 \right\rangle = \left\langle \left[f(x) - f_T(x) \right]^2 \right\rangle$$

Expanding the last term on the right will generally contain terms in $<\varepsilon_n^2>$ and $<\varepsilon_n \varepsilon_m>$. These correspond to the diagonal and off-diagonal elements of the coefficient error covariance matrix. By plugging in these error covariance terms and the value of *x* at which *y* and its uncertainty are desired, the error estimate σ_y is obtained. The details depend on the exact form of *f(x)*.