Quadratic fitting to find extrema

Neil D. Drummond

January 28, 2009

Suppose we have N estimates of a quantity y at different values of a variable x upon which y depends, i.e. we have $\{y_1 \pm \sigma_{y_1}, \ldots, y_N \pm \sigma_{y_N}\}$ at $\{x_1, \ldots, x_N\}$. Suppose y has a local extremum as a function of x that we wish to determine. We can attempt to find the extremum by fitting a quadratic $y(x) = a_1 x^2 + a_2 x + a_3$ to our data, provided the $\{x_i\}$ are sufficiently close to the extremum.

We determine the parameters a_1, a_2 and a_3 in the quadratic by minimizing

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - y(x_i)}{\sigma_{y_i}} \right)^2 = \sum_{i=1}^{N} \left(\frac{y_i - a_1 x_i^2 - a_2 x_i - a_3}{\sigma_{y_i}} \right)^2. \tag{1}$$

At the minimum, $\partial\chi^2/\partial a_1 = \partial\chi^2/\partial a_2 = \partial\chi^2/\partial a_3 = 0$. The resulting three equations can be written in matrix form as $M\mathbf{a} = \mathbf{c}$, where

$$M = \sum_{i=1}^{N} \frac{1}{\sigma_{y_i}^2} \begin{pmatrix} x_i^4 & x_i^3 & x_i^2 \\ x_i^3 & x_i^2 & x_i \\ x_i^2 & x_i & 1 \end{pmatrix}, \tag{2}$$

and

$$\mathbf{c} = \sum_{i=1}^{N} \frac{y_i}{\sigma_{y_i}^2} \begin{pmatrix} x_i^2 \\ x_i \\ 1 \end{pmatrix}. \tag{3}$$

Hence we can evaluate the three fitting parameters as $\mathbf{a} = M^{-1}\mathbf{c}$. As usual, the fit should only be regarded as good if $\chi^2 \approx N$ at the minimum.

Suppose we wish to evaluate the variance of some function $g(\mathbf{a})$ of the fitted parameters. Let $p_j = \partial g/\partial a_j$ and let

$$\mathbf{d}_i \equiv \sigma_{y_i} \frac{\partial \mathbf{c}}{\partial y_i} = \frac{1}{\sigma_{y_i}} \begin{pmatrix} x_i^2 \\ x_i \\ 1 \end{pmatrix}. \tag{4}$$

Note that

$$\frac{\partial \mathbf{a}}{\partial y_i} = M^{-1} \frac{\partial \mathbf{c}}{\partial y_i} = \sigma_{y_i}^{-1} M^{-1} \mathbf{d}_i. \tag{5}$$

Hence

$$\frac{\partial g}{\partial y_i} = \sum_{j=1}^{3} \frac{\partial g}{\partial a_j} \frac{\partial a_j}{\partial y_i} = \sigma_{y_i}^{-1} \mathbf{p}^T M^{-1} \mathbf{d}_i. \tag{6}$$

Assuming that the random errors in the data points are uncorrelated and are normally distributed with variance $\sigma_{y_i}^2$, the variance in $g(\mathbf{a})$ is approximately given by

$$\sigma_g^2 \approx \sum_{i=1}^{N} \left(\sigma_{y_i} \frac{\partial g}{\partial y_i} \right)^2 = \sum_{i=1}^{N} \mathbf{p}^T M^{-1} \mathbf{d}_i \mathbf{d}_i^T M^{-1} \mathbf{p} \tag{7}$$

where we have made use of the fact that M is symmetric. But it can easily be verified that $\sum_i \mathbf{d}_i\mathbf{d}_i^T = M$. Hence

$$\sigma_g^2 \approx \mathbf{p}^T M^{-1}\mathbf{p}. \tag{8}$$

- If $g(\mathbf{a}) = a_j$, i.e., we compute the variance in parameter a_j, then $\mathbf{p} = \mathbf{e}_j$ and so $\sigma_{a_j}^2 = M_{jj}^{-1}$. So the error bar on parameter a_j is

$$\sigma_{a_j} = \sqrt{M_{jj}^{-1}}. \tag{9}$$

- The fitted function $y(x)$ is extremized at $x = -a_2/(2a_1) \equiv z$. Taking $g(\mathbf{a}) = z$ gives $\mathbf{p} = (a_2/(2a_1^2), -1/(2a_1), 0)$; Eq. (8) can then be used to calculate the variance in z. We find that the error in z is

$$\sigma_z = \sqrt{\frac{M_{11}^{-1}a_2^2}{4a_1^4} - \frac{M_{12}^{-1}a_2}{2a_1^3} + \frac{M_{22}^{-1}}{4a_1^2}}. \tag{10}$$

- The variance in $y = a_1x^2 + a_2x + a_3 \equiv g(\mathbf{a})$ at a particular point x is given by Eq. (8) with $\mathbf{p} = (x^2, x, 1)$. So the error bar on the fitted quadratic at x is

$$\sigma_y = \sqrt{M_{11}^{-1}x^4 + 2M_{12}^{-1}x^3 + (M_{22}^{-1} + 2M_{13}^{-1})x^2 + 2M_{23}^{-1}x + M_{33}^{-1}}. \tag{11}$$

- The variance in the extremal function value $y_0 \equiv y(z) = -a_2^2/(4a_1) + a_3 \equiv g(\mathbf{a})$ is given by Eq. (8) with $\mathbf{p} = (a_2^2/(4a_1^2), -a_2/(2a_1), 1) = (z^2, z, 1)$, which gives the same result as would be obtained by inserting $x = z$ in Eq. (11).

Suppose the error bars on the data are unknown, but it is believed that (i) the data really are normally distributed about the least-squares-fitted quadratic function, and (ii) all the error bars σ_{y_i} are the same. Then the error bars must be such that $\chi^2 \approx N$. If the error bars are all the same then the χ^2 fit is equivalent to a least-squares fit. Let us first choose the error bars to be an arbitrary constant, and let C^2 be the corresponding minimum value of the χ^2 function. We may then rescale the error bars by $C/\sqrt{N}$, so that the new χ^2 value is N. We now have suitable error bars σ_{y_i} on our data, and hence we can use the formulae given above to compute the error in the extremum, etc.