8.2. Other refinement methods

BY E. PRINCE AND D. M. COLLINS

Chapter 8.1 discusses structure refinement by the method of least squares, which has a long history of successful use in data fitting and statistical analysis of results. It is an excellent technique to use in a wide range of practical problems, it is easy to implement, and it usually gives results that are straightforward and unambiguous. If a set of observations, y_i, is an unbiased estimate of the values of model functions, $M_i(\mathbf{x})$, a properly weighted least-squares estimate is the best, linear, unbiased estimate of the parameters, $\mathbf{x}$, provided the variances of the p.d.f.s of the populations from which the observations are drawn are finite. This assumes, however, that the model is correct and complete, an assumption whose validity may not necessarily be easily justified. Furthermore, least squares tends to perform poorly when the distribution of errors in the observations has longer tails than a normal, or Gaussian, distribution. For these reasons, a number of other procedures have been developed that attempt to retain the strengths of least squares but are less sensitive to departures from the ideal conditions that have been implicitly assumed. In this chapter, we discuss several of these methods. Two of them, maximum-likelihood methods and robust/resistant methods, are closely related to least squares. A third one uses a function that is mathematically related to the entropy function of thermodynamics and statistical mechanics, and is therefore referred to as the maximum-entropy method. For a discussion of the particular application of least squares to structure refinement with powder data that has become known as the Rietveld method (Rietveld, l969), see Chapter 8.6.

8.2.1. Maximum-likelihood methods

In Chapter 8.1, structure refinement is presented as finding the answer to the question, 'given a set of observations drawn randomly from populations whose means are given by a model, $\boldsymbol{M}(\mathbf{x})$, for some set of unknown parameters, $\mathbf{x}$, how can we best determine the means, variances and covariances of a joint probability density function that describes the probabilities that the true values of the elements of $\mathbf{x}$ lie in certain ranges?'. For a broad class of density functions for the observations, the linear estimate that is unbiased and has minimum variances for all parameters is given by the properly weighted method of least squares. The problem can also be stated in the slightly different manner, 'given a model and a set of observations, what is the *likelihood* of observing those particular values, and for what values of the parameters of the model is that likelihood a maximum?'. This set of parameters is the *maximum-likelihood estimate*.

Suppose the ith observation is drawn from a population whose p.d.f. is $\Phi_i(\Delta_i)$, where $\Delta_i = [y_i - M_i(\mathbf{x})]/s_i$, $\mathbf{x}$ is the set of 'true' values of the parameters, and s_i is a measure of scale appropriate to that observation. If the observations are independent, their joint p.d.f. is the product of the individual, marginal p.d.f.s:

$$\Phi_J(\Delta) = \prod_{i=1}^{n} \Phi_i(\Delta_i). \qquad (8.2.1.1)$$

The function $\Phi_i(\Delta_i)$ can also be viewed as a conditional p.d.f. for y_i given $M_i(\mathbf{x})$, or, equivalently, as a likelihood function for $\mathbf{x}$ given an observed value of y_i, in which case it is written $l_i(\mathbf{x}|y_i)$. Because a value actually observed logically must have a finite, positive likelihood, the density function in (8.2.1.1) and its logarithm will be maximum for the same values of $\mathbf{x}$:

$$\ln[l(\mathbf{x}|\mathbf{y})] = \sum_{i=1}^{n} \ln[l_i(\mathbf{x}|y_i)]. \qquad (8.2.1.2)$$

In the particular case where the error distribution is normal, and σ_i, the standard uncertainty of the ith observation, is known, then

$$\Phi_i(\Delta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-(1/2)\{[y_i - M_i(\mathbf{x})]/\sigma_i\}^2\right), \qquad (8.2.1.3)$$

and the logarithm of the likelihood function is maximum when

$$S = \sum_{i=1}^{n} \{[y - M_i(\mathbf{x})]/\sigma_i\}^2 \qquad (8.2.1.4)$$

is minimum, and the maximum-likelihood estimate and the least-squares estimate are identical.

For an error distribution that is not normal, the maximum-likelihood estimate will be different from the least-squares estimate, but it will, in general, involve finding a set of parameters for which a sum of terms like those in (8.2.1.2) is a maximum (or the sum of the negatives of such terms is a minimum). It can thus be expressed in the general form: find the minimum of the sum

$$S = \sum_{i=1}^{n} \rho(\Delta_i), \qquad (8.2.1.5)$$

where ρ is defined by $\rho(x) = -\ln[\Phi(x)]$, and $\Phi(x)$ is the p.d.f. of the error distribution appropriate to the observations. If $\rho(x) = x^2/2$, the method is least squares. If the error distribution is the Cauchy distribution, $\Phi(x) = [\pi(1 + x^2)]^{-1}$, $\rho(x) = \ln(1 + x^2)$, which increases much more slowly than x^2 as $|x|$ increases, causing large deviations to have much less influence than they do in least squares.

Although there is no need for $\rho(x)$ to be a symmetric function of x (the error distribution can be skewed), it may be assumed to have a minimum at $x = 0$, so that $\mathrm{d}\rho(x)/\mathrm{d}x = 0$. A series expansion about the origin therefore begins with the quadratic term, and

$$\rho(x) = (x^2/2)\left(1 + \sum_{k=1}^{\infty} a_k x^k\right). \qquad (8.2.1.6)$$

This procedure is thus equivalent to a variant of least squares in which the weights are functions of the deviation.

8.2.2. Robust/resistant methods

Properly weighted least squares gives the best *linear* estimate for a very broad range of distributions of random errors in the data and the maximum-likelihood estimate if that error distribution is normal or Gaussian. But the best linear estimator may nevertheless not be a very good one, and the error distribution may not be well known. It is therefore important to address the question of how good an estimation procedure may be when the conditions for which it is designed may not be satisfied. Refinement procedures may be classified according to the extent that they possess two properties known as robustness and resistance. A procedure is said to be *robust* if it works well for a broad range of error distributions and *resistant* if its results are not strongly affected by fluctuations in any small subset of the data. Because least squares is a linear estimator, the influence of any single data point on the parameter estimates increases without limit as the difference between the observation and the model increases. It therefore works poorly if the actual error

references