International Tables for Crystallography (2006). Vol. C, Section 8.2.3.1, p. 691.

might seem to have the effect of making the weights dependent on the calculated values, so that the right-hand side of (8.2.2.6) is no longer zero, but this applies only if the weights are changed during the refinement. There is thus no conflict with the result in (8.1.2.9). In practice, in any case, many other sources of uncertainty are much more important than any possible bias that could be introduced by this effect.

8.2.3. Entropy maximization

8.2.3.1. *Introduction*

Entropy maximization, like least squares, is of interest primarily as a framework within which to find or adjust parameters of a model. Rationalization of the name 'entropy maximization' by analogy to thermodynamics is controversial, but there is formal proof (Shore & Johnson, 1980) supporting entropy maximization as the unique method of inference that satisfies basic consistency requirements (Livesey & Skilling, 1985). The proof consists of discovering the consequences of four consistency axioms, which may be stated informally as follows:

(1) the result of the inference should be unique;
(2) the result of the inference should be invariant to any transformations of coordinate system;
(3) it should not matter whether independent information is accounted for independently or jointly;
(4) it should not matter whether independent subsystems are treated separately in conditional problems or collected and treated jointly.

The term 'entropy' is used in this chapter as a name only, the name for variation functions that include the form $\varphi \ln \varphi$, where φ may represent probability or, more generally, a positive proportion. Any positive measure, either observed or derived, of the relative apportionment of a characteristic quantity among observations can serve as the proportion.

The method of entropy maximization may be formulated as follows: given a set of n observations, y_i, that are measurements of quantities that can be described by model functions, $M_i(\mathbf{x})$, where $\mathbf{x}$ is a vector of parameters, find the prior, positive proportions, $\mu_i = f(y_i)$, and the values of the parameters for which the positive proportions $\varphi = f[M_i(\mathbf{x})]$ make the sum

$$S = -\sum_{i=1}^{n} \varphi_i' \ln(\varphi_i'/\mu_i'), \qquad (8.2.3.1)$$

where $\varphi_i' = \varphi_i / \sum \varphi_j$ and $\mu_i' = \mu_i / \sum \mu_j$, a maximum. S is called the *Shannon–Jaynes entropy*. For some applications (Collins, 1982), it is desirable to include in the variation function additional terms or restraints that give S the form

$$S = -\sum_{i=1}^{n} \varphi_i' \ln(\varphi_i'/\mu_i') + \lambda_1 \xi_1(\mathbf{x}, \mathbf{y}) + \lambda_2 \xi_2(\mathbf{x}, \mathbf{y}) + \ldots, \qquad (8.2.3.2)$$

where the λs are undetermined multipliers, but we shall discuss here only applications where $\lambda_i = 0$ for all i, and an unrestrained entropy is maximized. A necessary condition for S to be a maximum is for the gradient to vanish. Using

$$\frac{\partial S}{\partial x_j} = \sum_{i=1}^{n} \left(\frac{\partial S}{\partial \varphi_i}\right)\left(\frac{\partial \varphi_i}{\partial x_j}\right) \qquad (8.2.3.3)$$

and

$$\frac{\partial S}{\partial \varphi_i} = \sum_{k=1}^{n} \left(\frac{\partial S}{\partial \varphi_k'}\right)\left(\frac{\partial \varphi_k'}{\partial \varphi_i}\right), \qquad (8.2.3.4)$$

straightforward algebraic manipulation gives equations of the form

$$\sum_{i=1}^{n} \left\{ \frac{\partial \varphi_i}{\partial x_j} - \varphi_i' \left(\sum_{k=1}^{n} \frac{\partial \varphi_k}{\partial x_j} \right) \right\} \ln\left(\frac{\varphi_i'}{\mu_i'}\right) = 0. \qquad (8.2.3.5)$$

It should be noted that, although the entropy function should, in principle, have a unique stationary point corresponding to the global maximum, there are occasional circumstances, particularly with restrained problems where the undetermined multipliers are not all zero, where it may be necessary to verify that a stationary solution actually maximizes entropy.

8.2.3.2. *Some examples*

For an example of the application of the maximum-entropy method, consider (Collins, 1984) a collection of diffraction intensities in which various subsets have been measured under different conditions, such as on different films or with different crystals. All systematic corrections have been made, but it is necessary to put the different subsets onto a common scale. Assume that every subset has measurements in common with some other subset, and that no collection of subsets is isolated from the others. Let the measurement of intensity I_h in subset i be J_{hi}, and let the scale factor that puts intensity I_h on the scale of subset i be k_i. Equation (8.2.3.1) becomes

$$S = -\sum_{h=1}^{n} \sum_{i=1}^{m} (k_i I_h)' \ln\left[\frac{(k_i I_h)'}{J_{hi}'}\right], \qquad (8.2.3.6)$$

where the term is zero if I_h does not appear in subset i. Because k_i and I_h are parameters of the model, equations (8.2.3.5) become

$$\sum_{i=1}^{m} k_i \ln\left[\frac{(k_i I_h)'}{J_{hi}'}\right] - \sum_{h=1}^{n} \sum_{i=1}^{m} (k_i I_h)' \left(\sum_{l=1}^{m} k_l\right) \ln\left[\frac{(k_i I_h)'}{J_{hi}'}\right] = 0, \qquad (8.2.3.7a)$$

and

$$\sum_{h=1}^{n} I_h \ln\left[\frac{(k_i I_h)'}{J_{hi}'}\right] - \sum_{h=1}^{n} \sum_{i=1}^{m} (k_i I_h)' \left(\sum_{l=1}^{n} I_l\right) \ln\left[\frac{(k_i I_h)'}{J_{hi}'}\right] = 0. \qquad (8.2.3.7b)$$

These simplify to

$$\ln I_h = Q - \sum_{i=1}^{m} k_i' \ln(k_i/J_{hi}) \qquad (8.2.3.8a)$$

and

$$\ln k_i = Q - \sum_{h=1}^{n} I_h' \ln(I_h/J_{hi}), \qquad (8.2.3.8b)$$

where

$$Q = \sum_{h=1}^{n} \sum_{i=1}^{m} (k_i I_h)' \ln[(k_i I_h)/J_{hi}]. \qquad (8.2.3.8c)$$

Equations (8.2.3.8) may be solved iteratively, starting with the approximations $k_i = \sum_{h=1}^{n} J_{hi}$ and $Q = 0$.

The standard uncertainties of scale factors and intensities are not used in the solution of equations (8.2.3.8), and must be computed separately. They may be estimated on a fractional basis from the variances of estimated population means $\langle J_{hi}/I_h \rangle$ for a scale factor and $\langle J_{hi}/k_i \rangle$ for an intensity, respectively. The maximum-entropy scale factors and scaled intensities are relative, and either set may be multiplied by an arbitrary, positive constant without affecting the solution.