A Bayesian Three-Stage Model in Crystallography

By Simon French*

Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, England

A Bayesian method of statistical analysis is presented, one which provides a more powerful approach than classical least squares in circumstances where hitherto the latter technique has been applied. The general principles of the method are discussed and applications to crystallographic refinement and primary reduction of diffractometer data given.

Introduction

The method of least squares finds many applications in crystallography. A sad comment – for most statisticians agree upon the poorness of this method of parameter estimation (Stein, 1956: Marquardt, 1970; Lindley, 1971; Efron & Morris, 1977). Unfortunately, they do not agree upon what should be its replacement. Generalizing very broadly, present-day statistics is divided between two schools: the Bayesian and the frequentist. As may be gathered from the title, the method presented in this paper belongs to the former.

Although the present controversy in statistics is not the subject of this paper, a few brief words are relevant. The dichotomy between Bayesian and frequentist is sharp. It is one of philosophy. The Bayesian's view of scientific methodology differs from that of the frequentist. The vast majority of applications of statistical techniques to crystallography have used the frequentist approach, as described in, for example, Hamilton (1964). The notable published exception is the work of Mendes & de Polignac (1973). Thus, in this paper I am not only explicitly presenting to you a new statistical technique, but probably also implicitly calling into question your scientific methodology.

What is the Bayesian approach to inference? Clearly I cannot hope to answer this question in less than a book, certainly not in a paper. Anyway, there are some excellent books on Bayesian statistics on which I may draw: Box & Taio (1973), De Groot (1970), Jeffreys (1961), Lindley (1965). Barnett (1973) contrasts and compares the Bayesian and frequentist approaches in a very readable fashion. If you are not familiar with Bayesian ideas, I strongly urge you to read the early chapters of one of the above references before reading further in this paper. True, in the next section I briefly

introduce the basic principles of Bayesian statistics, but I do so only sufficiently to refresh a memory. Also, such a brief introduction is a most suitable vehicle for explaining my notation.

The development in this paper continues with a discussion of the various sources of information that a crystallographer might wish to incorporate into a refinement. This motivates the structure of the Bayesian three-stage model introduced in § 3. In the following section crystallographic refinement is discussed as a Bayesian three-stage model. § 5 is by way of an aside. It illustrates, in the context of the refinement of heavy-atom sites used in isomorphous phasing of macromolecules, some of the problems that extreme non-linearity may bring. In § 6 I show how the approach of the Bayesian three-stage model leads directly to a new method of reducing diffractometer data. § 7 contains my summary.

1. Bayesian statistics

A Bayesian and a frequentist are distinguished above all else by their different attitudes towards probability. For a Bayesian, at least in his role as a statistician, probability is simply a numerical representation of his degree of belief in a proposition about the system which he is observing. On the other hand, a frequentist holds that probability only has meaning as a numerical representation of the variability actually present in the observed system. This subtle distinction makes for the world of difference in their approaches to statistics. A Bayesian can clearly talk about his degree of belief in a scientific hypothesis or the value of a parameter, but a frequentist cannot, for there is no variation here. Nature either always does or always does not follow a hypothesis. Similarly, parameter values do not vary; they are fixed even if they are unknown.

* Present address: Department of Decision Theory, University of Manchester, Manchester M13 9PL, England.

For a Bayesian the primary objective of any statistical investigation is the generation of his degree of belief in some chosen aspect of the system under observation. This distribution of belief is based not only upon the evidence of his data, but also upon any information available from similar past investigations. Once he has developed this distribution, he can set about providing estimates of selected parameters, stating his belief in certain hypotheses or whatever.

A Bayesian then handles all his information and knowledge through probabilities; they are his very language. Thus it is important to have a precise notation for probability statements, even though, alas, this precision is gained at the expense of simplicity. The next three paragraphs introduce a suitable notation. To gather all its subtleties, I suggest you read them rapidly and then read them again rather more slowly.

Throughout the paper I shall repeatedly need to write sentences of the form: 'My belief in the value of quantity X given that I know quantity y is represented by the probability distribution $P_X(.|y)$'. For this I shall use the notation

$$X \sim P_X(.|y). \tag{1.0}$$

This distribution function is defined by:

$P_X(x|y) = \text{prob}(X \le x|y)$

= the probability that X is no greater than x given that I know the quantity y. (1.1)

The purpose of the '.' in $P_X(.|y)$ in (1.0) is to emphasize that there I was referring to the entire distribution function. In (1.1) it is replaced by x because there I am referring to the value of the distribution function at a particular point. Thus (1.0) is a general statement of my knowledge of X in the light of my knowledge of y, whereas (1.1) is a statement of a particular facet of this knowledge.

Corresponding to the distribution function $P_X(.|y)$ is the probability density function $p_X(.|y)$, where, of course,

$$P_X(x|y) = \int_{-\infty}^{x} p_X(t|y) \, dt. \tag{1.2}$$

Note that throughout subscripts give the quantities about which I am expressing my beliefs. Quantities that (I assume) I know are given to the right of the conditional vertical stroke. If my beliefs about the quantity X are not conditioned upon my knowledge of some other quantity y, I write

$$X \sim P_X(.) \tag{1.3}$$

with $p_X(.)$ being the corresponding density function.

Now to return to the subject in hand, what is the Bayesian approach to inference? Suppose that I am interested in some parameter θ. Before I carry out any investigations, I must have some knowledge of θ; otherwise I could not define θ, much less give direction to an investigation. Let this prior knowledge be given by the distribution

$$\theta \sim P_\theta(.) \tag{1.4}$$

with density function $p_\theta(.)$. Next I conduct an experiment that is informative on θ. By this I mean the following. Before the experiment, I ask myself: if I know the parameter value θ, what is my distribution of belief in the as yet unmade observation Y? Suppose my reply

$$Y \sim P_Y(.|\theta) \tag{1.5}$$

depends upon the value I choose for θ, then the experiment is informative on θ. I consider this question for each possible value of θ and so generate my family of conditional distributions for Y.

Suppose that I now observe $Y = y$; what should I believe about θ in the light of this information? Put probabilistically, what is $P_\theta(.|y)$? Bayes's theorem provides the answer. (Hence the name 'Bayesian Statistics'.) In terms of density functions, this states

$$p_\theta(\theta|y) \underset{\theta}{\propto} p_Y(y|\theta).p_\theta(\theta), \tag{1.6}$$

where the notation $\propto_\theta$ means 'is proportional to as a function of θ'. The constant of proportionality is easily found on remembering that a probability density must integrate to unity.

In the barest essentials, I have quoted my joint distribution of belief in the observation Y and the parameter θ prior to the experiment, viz

$$Y, \theta \sim P_{Y,\theta}(.,.), \tag{1.7}$$

where

$$p_{Y,\theta}(y,\theta) = p_Y(y|\theta)p_\theta(\theta). \tag{1.8}$$

I have then derived from this joint distribution the conditional distribution of θ given the observed value $Y = y$. This interpretation will be useful in § 3.

It should be noted that throughout the above I have assumed that my distributions of belief obey sufficient conditions for Bayes's theorem to apply (see, for example, De Groot, 1970, ch. 3).

Although my posterior distribution of belief represents all my knowledge of a parameter, I am often required to state rather less than this, namely an estimate of the parameter or whether I would reject a certain null hypothesis concerning the parameter. For a full discussion of how I, as a Bayesian, should approach such requests, see, for example, De Groot (1970). ch. 11. In this paper, I simply ask you to accept that the mean and standard deviation of my posterior distribution for a parameter are reasonable as an estimate and a measure of its accuracy.

2. Crystallographic refinement (1)

Before discussing the Bayesian three-stage model in its general form, it will be helpful to consider a particular example: crystallographic refinement. Here I am interested in the three-dimensional structure of a molecule as represented by its electron density, $\rho(\mathbf{x})$, or, equivalently, the Fourier transform of this, $F(\rho)$.

The information, which I may wish to incorporate into a crystallographic refinement, breaks down rather naturally into three parts.

Stage I. Observation error

Firstly and rather obviously, I believe that my X-ray intensity observations are related to the structure factors of $\rho(\mathbf{x})$. If $I(\rho)$ is the vector of intensity observations and $F(\rho)$ the corresponding vector of structure factors, then I shall have a distribution of belief:

$$I(\rho) \sim P_{I(\rho)}[.|F(\rho)].\qquad(2.0)$$

As in (1.5), this is my distribution of belief in the observations conditional on my assumed knowledge of $F(\rho)$ prior to the experiment. Later on we shall use Bayes's theorem to invert this probability relation, so giving my belief in the structure factors conditional on my actual observations. Further discussion of the distribution in (2.0) and those in (2.1) and (2.2) below is deferred to § 4.

Stage II: Modelling error

If I believe modern molecular theories then I believe that $\rho(\mathbf{x})$ has an atomic form, i.e. I believe that it may be well approximated by a parametric density, $\rho_m(\mathbf{x},\beta)$, where the only unknowns are contained in the vector of atomic parameters, β. Because of the nature of crystallographic observations, it is convenient to take the Fourier equivalent of this statement, viz: from modern theories of molecular structure I believe that the Fourier transform of the true electron density, $F(\rho)$, may be well approximated by the Fourier transform of the parametric density, $F[\rho_m(.,\beta)]$, in which the only unknowns are the atomic parameters, β. Thus I have a distribution of belief

$$F(\rho) \sim P_{F(\rho)}(.|\beta),\qquad(2.1)$$

which expresses my degree of confidence in fitting the vector of structure factors of the true electron density at the points in reciprocal space at which I make intensity observations with structure factors derived from a parametric model of the molecule.

Stage III: Prior knowledge of the parameters

From previous structural investigations, both practical and theoretical, I have available considerable information about bond lengths and angles, shape and volume parameters of specific elements, and the general vibrational characteristics of molecules; i.e. I know much about β. Thus I have a distribution of belief

$$\beta \sim P_\beta(.),\qquad(2.2)$$

which expresses my feelings a priori as to the relative merits of various conjectured structures.

Thus, all the sources of information that I wish to incorporate into a crystallographic refinement are summarized by the three distributions of belief (2.0), (2.1) and (2.2). The question is: once I have observed actual values for $I(\rho)$, how should I modify my belief about $F(\rho)$ and β?

3. The Bayesian three-stage model

Let us now consider the general situation. Suppose prior to performing an experiment I have the following structure to my beliefs.

Stage I

I believe that the observations $\mathbf{Y}$, which I shall make, are related to the parameter θ_1 by:

$$\mathbf{Y} \sim P_Y(.|\theta_1).\qquad(3.0)$$

Stage II

I believe that in turn the parameter θ_1 is related to the parameter θ_2 by:

$$\theta_1 \sim P_{\theta_1}(.|\theta_2).\qquad(3.1)$$

Stage III

Although I do not know θ_2 exactly, I have a distribution of belief for it which expresses all my prior knowledge:

$$\theta_2 \sim P_{\theta_2}(.).\qquad(3.2)$$

It is helpful to consider these three stages as representing the following information.

Stage I: observation errors, i.e. my beliefs as to the reliability and accuracy of my measurements.

Stage II: errors associated with the poorness of my physical model of the underlying situation.

Stage III: prior knowledge of the parameters that are associated with the physical model that I have chosen.

Combining the information in the three stages (3.0), (3.1) and (3.2), my joint distribution of belief in $(\mathbf{Y},\theta_1,\theta_2)$ is given by

$$p_{Y,\theta_1,\theta_2}(\mathbf{Y}, \theta_1, \theta_2) = p_Y(\mathbf{Y}|\theta_1).p_{\theta_1}(\theta_1|\theta_2)\, p_{\theta_2}(\theta_2).\qquad(3.3)$$

Note that probability density functions are used in the above.

After I make the observation $\mathbf{Y} = \mathbf{y}$, my posterior joint distribution of belief in $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ conditional on $\mathbf{y}$ is given by [cf. (1.6) and (1.8)]

$$p_{\theta_1,\theta_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) \underset{\theta_1,\theta_2}{\propto} p_{\mathbf{Y},\theta_1,\theta_2}(\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \quad (3.4)$$

If $\boldsymbol{\theta}_1$ is the parameter that holds all my interest, then I may integrate the nuisance parameter $\boldsymbol{\theta}_2$ out of the joint density given in (3.4) and thus obtain my marginal posterior belief in $\boldsymbol{\theta}_1$,

$$p_{\theta_1}(\boldsymbol{\theta}_1 | \mathbf{y}) = \int p_{\theta_1,\theta_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y})\, d\boldsymbol{\theta}_2. \quad (3.5)$$

[For some cautionary remarks on considering marginal beliefs alone: see Box & Taio (1973) p. 71.]

It is equally possible that $\boldsymbol{\theta}_2$ alone holds my interest, in which case I need my marginal posterior density:

$$P_{\theta_2}(\boldsymbol{\theta}_2 | \mathbf{y}) = \int p_{\theta_1,\theta_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y})\, d\boldsymbol{\theta}_1. \quad (3.6)$$

Thus after making the observation $\mathbf{Y} = \mathbf{y}$ I may answer all questions concerning $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ by referring to the distributions given by (3.4), (3.5) and (3.6).

Lindley & Smith (1972) and Smith (1973) have examined this three-stage model in considerable detail when the distributions involved are normal and the relations between the parameters are linear. Webb has extended their analysis to the case where some of the relations between the parameters are non-linear (Webb, 1974). Both analyses are summarized in Appendix A. However, since it will be used much in the following, it is worth stating Webb's model here.

The normal distribution will be denoted by

$$\mathbf{X} \sim N(\boldsymbol{\mu}, V), \quad (3.7)$$

i.e. the random vector $\mathbf{X}$ has a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix V.

Webb's model takes the form:

stage I: $\quad \mathbf{Y} \sim N[\mathbf{f}_1(\boldsymbol{\theta}_1), \mathbf{C}_1]$ (3.8)

stage II: $\quad \boldsymbol{\theta}_1 \sim N[\mathbf{f}_2(\boldsymbol{\theta}_2), \mathbf{C}_2]$ (3.9)

stage III: $\boldsymbol{\theta}_2 \sim N[\mathbf{A}_3\boldsymbol{\theta}_3, \mathbf{C}_3]$. (3.10)

Assuming that I know the functions $\mathbf{f}_1(.)$ and $\mathbf{f}_2(.)$, the matrices $\mathbf{A}_3$, $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$, and the vector $\boldsymbol{\theta}_3$, Webb deduces approximations to my marginal posterior distributions for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ after I have observed $\mathbf{Y} = \mathbf{y}$. The approximations, together with algorithms for their calculation, are presented in Appendix A. For the main development of the paper, however, it is sufficient merely to note the three-stage structure in (3.8), (3.9) and (3.10).

When the distributions involved are less tractable than the normal, numerical techniques may be employed unless, as is unfortunately true in crystallographic cases, the dimensions of the vectors involved prevent this.

4. Crystallographic refinement (2)

Unless I am being particularly obscure, it should be apparent that the statistical framework for a crystallographic refinement given by relations (2.0), (2.1) and (2.2) is of the general form given by (3.0), (3.1) and (3.2) and discussed in the last section. However to be explicit: $I(\rho)$ identifies with the observation $\mathbf{Y}$, $F(\rho)$ with the first parameter $\boldsymbol{\theta}_1$, and $\boldsymbol{\beta}$ with the second parameter $\boldsymbol{\theta}_2$. The distributions of belief that are involved at each stage are now considered in greater, though not complete, detail.

I shall reluctantly make the assumption that the distributions involved may be well approximated by multivariate normal distributions. My reluctance derives from my awareness that I make the assumption for reasons of computational tractability. Nonetheless it is the usual assumption and is probably not so unreasonable, particularly in small-molecule investigations. [See French (1975) for a full discussion.]

Having made the assumption of normality, the three-stage model is of the form given by relations (3.8), (3.9) and (3.10) above. Thus I need only discuss the mean and covariance matrix at each stage in order to be able to use the iterative-solution procedure described in Appendix A.

The first stage represents my belief in the relation between the vector of intensity observations, $I(\rho)$, and the vector of structure factors, $F(\rho)$. As such it models not only the uncertainties arising from counting statistics (or photographic film error), but also those introduced in applying corrections for absorption, extinction, radiation damage and interlevel scale differences. In the following I shall assume that the corrected intensities are unbiased observations on the squared structure factor moduli. This may often be an unreasonable assumption, e.g. when ordinate analysis is used to reduce the data (Tickle, 1975; French, 1975) or when negative intensities have been set to zero. The covariance matrix, $\mathbf{C}_1$ in (3.8), may be well approximated by carefully inflating counting statistics according to the methods discussed in McCandlish, Stout & Andrews (1975), Dodson (1976a), and French (1975).

The second stage describes my belief as to how well the parametric approximation to the molecular structure actually corresponds with nature. The covariance structure at this stage is, at present, poorly investigated. Webb (1974) has made the suggestion of approximating $\mathbf{C}_2$ [see (3.9)] with $\sigma^2 I$ where σ^2 is set to (say) 10% of the current squared residual between the fitted structure factor moduli and the observed intensities. The sense of this suggestion can be seen by noting the better the fit in any refinement the smaller the uncertainty that is fed in at the second stage. This is discussed at some length in French (1975). Blight & Ott (1975) discuss the general problem of 'approximation error' in terms of polynomial fitting.

The third stage represents my prior knowledge of bond angles and lengths, molecular symmetry and vibrational characteristics. It is a task far beyond my ability to suggest how this may be defined explicitly, but, fortunately, it is a very easy matter to approximate it at each cycle of the iterative solution of the non-linear three-stage model. Quite simply, I make the third stage express complete ignorance, i.e. $C_3 \to \infty$, and imagine that I can make observations directly upon the bond angles and lengths, symmetry relations and the temperature factors. These pseudo-observations take the values predicted from previous structural investigations, both practical and theoretical, and the covariance structure describes the reliability I put in these predictions. This procedure is entirely analogous to augmenting the least-squares objective function with 'soft constraints' (Waser, 1963; Rollett, 1970; Dodson, Isaacs & Rollett, 1976). The mathematically curious may use the inverse-function theorem (Rudin, 1964, p. 193) to show the equivalence of using pseudo-observations and a vague third stage and of using a fully informative third stage. However, its good sense should be apparent.

I have assumed in the above that I know the covariance matrices at the first and second stages, C_1 and C_2. The former, C_1, is the usual least-squares weighting matrix. A considerable proportion of crystallographic refinements do not use variance weighting schemes; instead they use either unit weights with all their associated horrors or 'sensible' weights, e.g. $(k|F_{obs}|)^{-2}$ (Hamilton & Abrahams, 1970). I am opposed to this as it loses the individual error structure at each reflection that the careful data analysis referred to above can maintain. Furthermore, it becomes difficult to balance the weighting of the crystallographic observations to the weighting of the pseudo-observations (soft constraints). I realize, however, that a careful, complete analysis of the data is very time consuming. Perhaps, therefore, a compromise is possible. The shape of C_1 can be kept from the early data reduction so that at refinement the matrix $C_1 = \sigma_1^2 S$ is used, where S is the known shape matrix and σ_1^2 is an unknown multiplicative constant. σ_1^2 can be estimated during the refinement by the modal estimate method proposed in Lindley & Smith (1972). Webb's arbitrary 10% of the squared residual for C_2 could be improved by a similar modal estimate.

At the end of Appendix A, I make some remarks concerning the robustness of the above procedures to breakdowns in the assumptions. I think it is fair to say that if the covariance matrices at the first and second stages have about the right shape and balance then the approximations to the means of my posterior distributions are robust to fairly large departures from normality and linearity. But the same does not apply to the approximations to my posterior standard deviations. These can be very sensitive to non-

normality in particular (Kendall & Stuart, 1961, ch. 31). The poorness of least-squares estimated standard deviations discovered by Hamilton & Abrahams (1970) can, perhaps, partly be attributed to such factors. Furthermore, as they remark, their least-squares refinement makes no allowance for the modelling error, i.e. it does not include stage II of the Bayesian model. These remarks are not meant to discount their explanation on the basis of the undoubted systematic errors, but to complement it.

To my knowledge, the work of Lindley, Milledge & Webb (1974) on small molecules is the only test of the above theory to data. Their results were extremely promising; the method converged well and produced results that were chemically very sensible.

5. Heavy-atom refinement

Obviously to use the Bayesian three-stage model, or least squares for that matter, to refine a structure, it is necessary to have intensity observations on the crystal. Unfortunately, this clear requirement is not always fulfilled. For instance, in the isomorphous replacement method of phasing macromolecules it is necessary to refine the heavy-atom difference structure between the derivative and native molecules without being able to observe its intensities directly. It is only possible to observe intensities on the derivative and native molecules. However, from these it is possible to calculate 'observations' on the difference structure's intensities. These calculated observations serve as the first stage of the three-stage model and, once their distribution has been defined, the theory of the preceding section may be applied in its entirety.

Before continuing some notation will be necessary. For a given reflection, let I_P, J_P be the observed and true native intensities; I_{PH}^+, J_{PH}^+ be the observed and true derivative intensities; I_{PH}^-, J_{PH}^- be the observed and true derivative intensities at the Friedel-related reflection; J_H be the true intensity of the difference structure.

The problem is to find a function, I_H, of the actual intensity observations such that $E[I_H(I_P, I_{PH}^+, I_{PH}^-)] = J_H$. The literature contains many suggestions for this function (Srinivasan, 1966; Singh & Ramaseshan, 1966; Dodson & Vijayan, 1971; Kartha & Parthasarathy, 1965). Unfortunately, all the suggestions contain a common mistake. Essentially the authors have found a function such that

$$I_H(J_P, J_{PH}^+, J_{PH}^-) = J_H$$

and argued that since

$$E(I_P) = J_P, E(I_{PH}^+) = J_{PH}^+, E(I_{PH}^-) = J_{PH}^-,$$

it 'clearly' follows that

$$E[I_H(I_P, I_{PH}^+, I_{PH}^-)] = J_H. \tag{5.0}$$

Alas, such a conclusion is not only not clear but also false.

To a second-order approximation (see Appendix B)

$$E[I_H(I_P, I_{PH}^+, I_{PH}^-)] = J_H + \tfrac{1}{2} \text{ trace } (\mathbf{HV}) + \ldots, \quad (5.1)$$

where $\mathbf{H}$ is the Hessian matrix of I_H with respect to $(I_P, I_{PH}^+, I_{PH}^-)$ evaluated at $(J_P, J_{PH}^+, J_{PH}^-)$, and $\mathbf{V}$ is the variance matrix of $(I_P, I_{PH}^+, I_{PH}^-)$. Thus if $\tfrac{1}{2}$ trace($\mathbf{HV}$) is significantly large, I_H is a biased observation of J_H.

Dodson, Evans & French (1975) and French (1975) have investigated the size of this bias and found it to be considerable; in some cases it exceeds 200% of J_H. We have suggested subtracting from each of the suggested I_H the quantity $\tfrac{1}{2}$ trace($\mathbf{H}_0\mathbf{V}$), where $\mathbf{H}_0$ is the Hessian matrix of I_H evaluated at the observed native and derivative intensities. In this way we have produced approximately unbiased 'observations' of J_H (Dodson, 1976b). The approximation to the variance provided by (B.1) in Appendix B appears to be adequate for the refinement. With the first stage defined as indicated above, the methods of the previous section may be applied to solve the resulting three-stage model.

It is a salutory exercise to consider the above a little more carefully. Essentially the mistaken assumption (5.0) is the result of expanding the function $I_H(I_P, I_{PH}^+, I_{PH}^-)$ in a Taylor series about the mean values of I_P, I_{PH}^+ and I_{PH}^- and then assuming that the second and higher-order terms are negligible. To obtain the more accurate expression (5.1), one is doing little better − only one more term is included. How many terms need be included depends upon how non-linear the function $I_H(I_P, I_{PH}^+, I_{PH}^-)$ is in the region to which the joint probability distribution of I_P, I_{PH}^+ and I_{PH}^- gives most of its mass. In linearizing any non-linear stage of the Bayesian three-stage model precisely similar considerations apply (see Appendix A). The moral is that, if one or more stages of the Bayesian model are particularly non-linear, the methods of Appendix A may be inappropriate and the user should be awake to this possibility.

6. Diffractometer intensity measurements

Moving rather perversely from refinement to primary data reduction, I now introduce a method of 'integrating the intensity under a diffractometer step-scan profile'. My discussion of refinement was of an interpretative nature, showing how existing refinement techniques can be viewed in terms of a Bayesian three-stage model. Here I intend to show how the Bayesian approach leads to a new method of primary data reduction that is more sensitive to the available information, and so produces more accurate integrated intensities, than previously used methods.

In scanning a reflection, a diffractometer measures a sequence of counts as the detector steps through the diffraction peak:

$$c_i \sim P_{c_i}(.\,|\lambda_i), \quad i = 1, 2, \ldots, N, \quad (6.0)$$

where c_i is the ith observed count, λ_i is the true (mean) count at the ith step, N is the number of steps in the scan. The distributions $P_{c_i}(.\,|\lambda_i)$ for $i = 1, 2, \ldots, N$ are approximately independent Poisson with means λ_i, but are slightly perturbed through 'instrument instability' and counting losses.

Our current physical intuition suggests that the λ_i are related to the reflection's intensity through

$$\lambda_i = J\pi(x_i) + b(x_i), \quad (6.1)$$

where J is the true intensity of the reflection; $\pi(.)$ is the peak-shape function:

$$\int \pi(x)\, \mathrm{d}x = 1,$$

$b(.)$ is the background function; x_i is the position of the ith count in the scan. The peak shape $\pi(.)$ and background $b(.)$ are unknown functions. However, physical theory and experience do give certain guides to their form. For instance, the peak shape will be continuous and, in many cases, unimodal. Similarly, for φ scans the background may be expected to be roughly constant over the width of the scan. Let $\{\pi(.,\alpha)\}_\alpha$ and $\{b(.,\beta)\}_\beta$ be two parametric families amongst which I may hope to find good approximations to $\pi(.)$ and $b(.)$. For example, I have found that transformed normal probability densities (Johnson, 1949) are often suitable for $\pi(.,\alpha)$ and that linear functions serve for $b(.,\beta)$.

By now the first two stages of the Bayesian model should be becoming apparent. The first stage, representing the observation error, will express the relation between the c_i and λ_i as given in (6.0). The second stage, representing the modelling error, will express the relation:

$$\lambda_i \simeq J\pi(x_i, \alpha) + b(x_i, \beta), \quad (6.2)$$

where α and β give suitable approximations to the particular peak shape and background at the reflection. The third stage is, perhaps, not so clear.

The third stage should represent my prior knowledge of the parameters at the second stage, viz my prior knowledge of J, α and β. Of J, the true intensity, I typically know nothing, but of α and β I may have considerable knowledge. Consider first the peak shape parameter α. Once I have fitted a few reflections, I will known much about α because Diamond (1969) has shown that the peak shape varies slowly through reciprocal space. Furthermore, the position within the scan of the peak last fitted and the reliability of the diffractometer in moving between reflections will also contribute to my knowledge of α. For the background parameter β I may have prior knowledge in the form of a smoothed background function (see, for example, Krieger, Chambers, Christoph, Stroud & Trus, 1974).

Hopefully the partial sketch of the three-stage model given above will provide an adequate introduction to the more detailed description of it given below. However, I should state at this point that I have no intention of describing the practical details here. My sole purpose is to demonstrate the statistical modelling involved. Details of the implementation must await a further paper, although they are already presented in French (1975).

Stage I: Observation error

This stage is essentially the relation (6.0). However, for convenience the observations are transformed by taking their square root. This is done because I intend to use Webb's three-stage model (3.8–3.10). For this to be applicable, the distributions at each stage need to be normal and the covariance matrices need to be known and independent of the parameters. Were the observations here left untransformed, these conditions would be unsatisfied. Assuming that only errors from counting statistics are present, I have to a good approximation (Box & Taio, 1973, Fig. 1.3.8)

$$c_i \sim N(\sqrt{\lambda_i}, 0.25). \qquad (6.3)$$

Stage I should also include a contribution from machine instability and counting losses. The latter I shall ignore here, as they are outside my realm of experience in protein crystallography. McCandlish, Stout & Andrews (1975), amongst others, have argued that machine instability gives rise to approximately constant relative errors in the counts. This I can model in (6.3) simply by inflating the variance so that

$$\sqrt{c_i} \sim N[\sqrt{\lambda_i}, 0.25(1 + \sigma_1^2 \lambda_i)]. \qquad (6.4)$$

Admittedly the variance now depends on the parameter λ_i, but the dependence is very slight since $\sigma_2^2 \simeq 0.0001$ usually. Such a slight departure from constant variances causes Webb's procedure no problems. At each non-linearity cycle of the algorithm, the unknown parameter λ_i in the variance can be estimated from the second-stage parametric approximation:

$$v_i = J\pi(x_i, \alpha) + b(x_i, \beta). \qquad (6.5)$$

I indicate below how the relative variance, σ_1^2, may be estimated.

Stage II: Modelling error

At this stage the true count λ_i is modelled with the parametric approximation v_i. It seems reasonable to expect the magnitude of the modelling error to increase with the true count. Indeed, I shall assume more than this. I shall take the modelling error to have constant relative variance. Thus the variance of λ_i conditional on v_i is given by:

$$\text{var}(\lambda_i | v_i) = \sigma_2^2 v_i^2. \qquad (6.6)$$

The correlations between the errors at adjacent steps will depend on the smoothness which I expect the true curves to exhibit and the separation of the steps considered. Blight & Ott (1975) have discussed how these correlations may be modelled. They have shown that there is much advantage to be gained from including them in the three-stage model when the size of the error involved at the second stage is of the same or greater order than that at the first stage. However, that is not the case here; except for the summits of the largest diffraction peaks, counting errors dominate the errors from poor modelling. Consequently, I take advantage of the computational gains afforded by assuming the modelling errors to be uncorrelated at different steps.

In order to make the relation between the first and second stages linear and also to reduce the dependence of the second-stage variance on v_i, square roots are taken. I assume normality:

$$\sqrt{\lambda_i} \sim N(\sqrt{v_i}, 0.25\sigma_2^2 v_i). \qquad (6.7)$$

In solving the three-stage model, it is found that the unknown relative variances σ_1^2 and σ_2^2 enter only through their sum. Thus it is only necessary to obtain an estimate of $(\sigma_1^2 + \sigma_2^2)$. This can be done by iteratively adjusting $(\sigma_1^2 + \sigma_2^2)$ in the fitting of the first few reflections in a data set until the standardized residuals have approximately unit variance. I have found it quite satisfactory to use the value obtained for $(\sigma_1^2 + \sigma_2^2)$ for the remainder of the data set.

Stage III: Prior knowledge

It is here that the information gained from previous fits of diffraction peaks is incorporated. I shall assume that my prior knowledge of J, α and β is independent and that my knowledge of J is vague, i.e. I know nothing about J. Specifically I assume

$$p_{J,\alpha,\beta}(.) \propto p_\alpha(.) p_\beta(.). \qquad (6.8)$$

$p_\alpha(.)$ contains all the information that I have learnt about the peak shape and its position within the scan. Suppose I am fitting the sth reflection. From the fit at the $(s-1)$st reflection, I have the marginal posterior distribution $p_\alpha(. | c_{s-1})$, which contains all my current information about the peak shape and position. In moving to the sth reflection I must weaken my belief according to the reliability of the diffractometer and the variability of peak shape that I believe to exist over reciprocal space. To do this I assume that $p_\alpha(.)$ is normal with mean (expectation) and covariance matrix:

$$E(\alpha) = \mathbf{S}_s E(\alpha | c_{s-1}) + \Delta\alpha_s \qquad (6.9)$$

$$\text{var}(\alpha) = \mathbf{S}_s \text{var}(\alpha | c_{s-1}) \mathbf{S}_s^T + \text{var}(\Delta\alpha_s), \qquad (6.10)$$

where $\mathbf{S}_s$ is a shift matrix that allows for changes in scale (step size) and predictable variation in peak

shape, *e.g.* width; $\Delta\alpha_s$ allows for predictable shifts in the peak parameters, *e.g.* position; var($\Delta\alpha_s$) weakens my distribution of belief according to my trust in the predicted shifts and the reliability of the diffractometer.

In defining $p_\beta(.)$ I can choose to do one of the following. (i) Use a method similar to that for $p_\alpha(.)$ above, *i.e.* I can remember the values of previously fitted backgrounds. This would be a partial smoothing of the background. (ii) Use a vague prior, *i.e.* assume that the background is so variable compared with the accuracy that it can be determined within one scan that it is not worth 'learning' the background from previous peaks. (iii) Use a smoothed background function (see, for example, Krieger, Chambers, Christoph, Stroud & Trus, 1974).

Again I shall assume that the distribution is normal.

The description of the three-stage model is now complete, or, at least, in as great a detail as I intend to give here. In addition to the raw counts themselves, the Bayesian method has incorporated four other sources of information: (i) the local behaviour of the background that may be expected from the collection geometry; (ii) properties expected in the peak shape; (iii) the shapes of the peaks already analysed; (iv) the position within the scan of the peak last analysed and the reliability of the diffractometer in moving between reflections. If a multiple-counter diffractometer is used, there is one further source of information: (v) the relative positions of peaks within simultaneously collected scans are defined by the diffraction geometry. Furthermore, the peak shape and, perhaps, the background will be roughly common to all counters. With the exception of the work of Diamond (1969), none of the standard data-reduction methods take account of all these sources of information, and even Diamond does not utilize them to the full. The Bayesian method incorporates them all.

7. Discussion

My purpose in this paper has been to introduce the Bayesian three-stage model to crystallographers in the hope that they might find it a more useful statistical tool in circumstances where they would usually resort to the principle of least squares. I am convinced that once the Bayesian method is fully understood, it gives one a very simple, natural and powerful method of fitting a scientific model to data.The applications to refinement and diffractometer intensity measurement have, I hope, illustrated this. Clearly there are further applications. For instance, in estimating the absorption correction of Huber & Kopfmann (1968, 1969), one could, through the third stage of the Bayesian method, require the fitted absorption surface to be smooth, thus avoiding their *ad hoc* smoothing method.

APPENDIX A
The Bayesian three-stage model with normal uncertainties

Smith (1973) has summarized the Bayesian three-stage model with normal distributions of belief over linear relations between the parameters as:

$$\text{stage I: } \mathbf{Y} \sim N(\mathbf{A}_1\mathbf{\theta}_1, \mathbf{C}_1) \qquad (A.1)$$

$$\text{stage II: } \mathbf{\theta}_1 \sim N(\mathbf{A}_2\mathbf{\theta}_2, \mathbf{C}_2) \qquad (A.2)$$

$$\text{stage III: } \mathbf{\theta}_2 \sim N(\mathbf{A}_3\mathbf{\theta}_3, \mathbf{C}_3), \qquad (A.3)$$

where the notation $\mathbf{X} \sim N(\mathbf{\mu}, \mathbf{V})$ has been used to indicate that $\mathbf{X}$ has a normal distribution of belief with mean $\mathbf{\mu}$ and covariance matrix $\mathbf{V}$. It is possible to add further stages *ad infinitum* (or *nauseam*), but for our purposes three stages are sufficient. Assuming that I know (*i.e.* my uncertainty is negligible) the matrices $\{\mathbf{A}_i, \mathbf{C}_i; i = 1, 2, 3\}$ and the parameter $\mathbf{\theta}_3$, my marginal posterior distributions for $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$ after an observation $\mathbf{Y} = \mathbf{y}$ are given by [see equations (3.5) and (3.6)]:

$$P_{\theta_1}(.|\mathbf{y}) \text{ is } N(\mathbf{\theta}_1^*, \mathbf{D}_1), \qquad (A.4)$$

where

$$\mathbf{D}_1^{-1}\mathbf{\theta}_1^* = \tilde{\mathbf{D}}_1^{-1}\tilde{\mathbf{\theta}}_1 + (\mathbf{C}_2 + \mathbf{A}_2\mathbf{C}_3\mathbf{A}_2^T)^{-1}\mathbf{A}_2\mathbf{A}_3\mathbf{\theta}_3 \qquad (A.5)$$

$$\mathbf{D}_1^{-1} = \tilde{\mathbf{D}}_1^{-1} + (\mathbf{C}_2 + \mathbf{A}_2\mathbf{C}_3\mathbf{A}_2^T)^{-1} \qquad (A.6)$$

$$\tilde{\mathbf{D}}_1^{-1}\tilde{\mathbf{\theta}}_1 = \mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{y} \qquad (A.7)$$

$$\tilde{\mathbf{D}}_1^{-1} = \mathbf{A}_1^T\mathbf{C}_1^{-1}\mathbf{A}_1; \qquad (A.8)$$

and

$$P_{\theta_2}(.|\mathbf{y}) \text{ is } N(\mathbf{\theta}_2^*, \mathbf{D}_2), \qquad (A.9)$$

where

$$\mathbf{D}_2^{-1}\mathbf{\theta}_2^* = \tilde{\mathbf{D}}_2^{-1}\tilde{\mathbf{\theta}}_2 + \mathbf{C}_3^{-1}\mathbf{A}_3\mathbf{\theta}_3 \qquad (A.10)$$

$$\mathbf{D}_2^{-1} = \tilde{\mathbf{D}}_2^{-1} + \mathbf{C}_3^{-1} \qquad (A.11)$$

$$\tilde{\mathbf{D}}_2^{-1}\tilde{\mathbf{\theta}}_2 = \mathbf{A}_2^T\mathbf{A}_1^T(\mathbf{C}_1 + \mathbf{A}_1\mathbf{C}_2\mathbf{A}_1^T)^{-1}\mathbf{y} \qquad (A.12)$$

$$\tilde{\mathbf{D}}_2^{-2} = \mathbf{A}_2^T\mathbf{A}_1^T(\mathbf{C}_1 + \mathbf{A}_1\mathbf{C}_2\mathbf{A}_1^T)^{-1}\mathbf{A}_1\mathbf{A}_2. \qquad (A.13)$$

The derivation of these results is given in Lindley & Smith (1972) and Smith (1973).

The vectors $\tilde{\mathbf{\theta}}_1$ and $\tilde{\mathbf{\theta}}_2$ are the usual least-squares estimates of the parameters $\mathbf{\theta}_1$ and $\mathbf{\theta}_2$. From ($A.5$) and ($A.10$) it can be seen that the posterior means $\mathbf{\theta}_1^*$ and $\mathbf{\theta}_2^*$

are weighted averages between the least-squares estimates and the values for the parameters predicted by θ_3. Furthermore, the weighting is done in an intuitively sensible way, being based on the sampling variances of the least-squares estimates and the prior variances induced from the third stage.

If I know little about θ_2 prior to the experiment, then I want stage III of the model to express this ignorance. Formally this can be done by letting $C_3^{-1} \to 0$, i.e. by letting my prior spread of belief at the third stage become arbitrarily large. In this case it may be shown that (Smith, 1973):

$$\theta_1^* = (A_1^T C_1^{-1} A_1 + C_2^{-1})^{-1}(A_1^T C_1^{-1} A_1 \tilde{\theta}_1 + C_2^{-1} A_2 \tilde{\theta}_2) \quad (A.14)$$

$$D_1^{-1} = A_1^T C_1^{-1} A_1 + C_2^{-1} - C_2^{-1} A_2(A_2^T C_2^{-1} A_2)^{-1} A_2^T C_2^{-1}$$

and $\qquad\qquad\qquad\qquad\qquad\qquad (A.15)$

$$\theta_2^* = \tilde{\theta}_2 \qquad (A.16)$$

$$D_2^{-1} = \tilde{D}_2^{-1}. \qquad (A.17)$$

Thus, in the case of prior ignorance it can be seen that the mean of my posterior belief for θ_2 is the usual least-squares estimate, whilst that for θ_1 is a weighted average of the least-squares estimate for θ_1 and the value of θ_1 predicted by the least-squares estimate of θ_2. The Bayesian analysis is clearly holding on to the underlying structure rather better than the least-squares analysis.

Unfortunately it seldom happens in practice that the relations between the parameters are linear. In the more usual case at least one of the first two stages is non-linear. Webb (1974) has considered the model:

stage I: $\qquad Y \sim N[f_1(\theta_1), C_1] \qquad (A.18)$

stage II: $\qquad \theta_1 \sim N[f_2(\theta_2), C_2] \qquad (A.19)$

stage III: $\qquad \theta_2 \sim N[A_3\theta_3, C_3]. \qquad (A.20)$

If the functions f_1 and f_2 are approximately linear in the regions in which my posterior densities are concentrated, then the theory of the linear model given above may be used to provide approximations to the posterior distributions for the non-linear model. Suppose that $\theta_{1,0}^*$ and $\theta_{2,0}^*$ lie near θ_1^* and θ_2^* respectively, then expanding to the first two terms in the Taylor series for the functions f_1 and f_2:

$$f_i(\theta_i) \simeq f_i(\theta_{i,0}^*) + \frac{\partial f_i^T}{\partial \theta_i}(\theta_{i,0}^*)(\theta_i - \theta_{i,0}^*). \quad (A.21)$$

Using $(A.21)$ we may rewrite the model given by $(A.18)$, $(A.19)$ and $(A.20)$ approximately as:

stage I: $\qquad X \sim N(A_1\varphi_1, C_1), \qquad (A.22)$

where

$$X = Y - f_1(\theta_{1,0}^*) - A_1 f_2(\theta_{2,0}^*)$$
$$+ A_1\theta_{1,0}^* + A_1 A_2\theta_{2,0}^* \qquad (A.23)$$

$$A_1 = \frac{\partial f_1^T}{\partial \theta_1}(\theta_{1,0}^*); \qquad (A.24)$$

stage II: $\qquad \varphi_1 \sim N(A_2\varphi_2, C_2), \qquad (A.25)$

where

$$\varphi_1 = \theta_1 - f_2(\theta_{2,0}^*) + A_2\theta_{2,0}^* \qquad (A.26)$$

$$A_2 = \frac{\partial f_2^T}{\partial \theta_2}(\theta_{2,0}^*); \qquad (A.27)$$

stage III: $\qquad \varphi_2 \sim N(A_3\varphi_3, C_3), \qquad (A.28)$

where

$$\varphi_2 = \theta_2 \qquad (A.29)$$

$$\varphi_3 = \theta_3. \qquad (A.30)$$

The matrices C_1 and C_2 should be inflated to allow for the errors in the linear approximations in $(A.21)$, although this is seldom done simply because it is not clear how to do so. Webb (1974) makes some very reasonable suggestions for doing this which seem to work well in his particular examples of crystallographic refinement.

Depending on my state of prior knowledge, either $(A.4)$ to $(A.13)$ or $(A.14)$ to $(A.17)$ give my posterior beliefs in φ_1 and φ_2 with means φ_1^* and φ_2^*. Replacing $\theta_{1,0}^*$ and $\theta_{2,0}^*$ with

$$\theta_{1,1}^* = \varphi_1^* + f_2(\theta_{2,0}^*) - A_2\theta_{2,0}^* \qquad (A.31)$$

$$\theta_{2,1}^* = \varphi_2^* \qquad (A.32)$$

the approximate linearization in $(A.21)$ to $(A.30)$ may be repeated and the entire process iterated until the shifts $(\theta_{i,j}^* - \theta_{i,j-1}^*)$ $(i = 1, 2; j =$ current shift) become small compared with the posterior standard deviations given by the current approximations, $D_{1,j}$ and $D_{2,j}$, to the posterior covariance matrices, D_1 and D_2. Supposing that the process converges at the kth cycle, my marginal posterior beliefs may be approximated by

$$\theta_1 \sim N(\theta_{1,k}^*, D_{1,k}) \qquad (A.33)$$

$$\theta_2 \sim N(\theta_{2,k}^*, D_{2,k}). \qquad (A.34)$$

In the above it has been assumed that I know the matrices $\{C_i; i = 1, 2, 3\}$. This is not usually the case. Theoretically I should place a distribution of belief *a priori* over these nuisance parameters and then integrate them out of my posterior distribution of belief. However, in crystallographic cases my uncertainty in the C_i is several orders less than my uncertainty in the parameters of interest and it is, apparently, sufficient to use approximations to the C_i derived from counting statistics and careful data analysis (French, 1975; McCandlish, Stout & Andrews, 1975; Dodson, 1976a). See also Lindley & Smith (1972).

The quality of the approximation to my posterior distribution depends on three factors: (i) the amount of non-linearity exhibited by the functions f_i over the

region of highest posterior density, *i.e.* the quality of the linear approximations to the f_i afforded by (A.21); (ii) the accuracy of my approximations to the C_i; (iii) the degree of non-normality actually present in my uncertainty. [Strictly, under the assumptions in (A.18), (A.19) and (A.20) all the distributions involved are normal, but in practice I have to make the assumption that any departures from normality in my beliefs will not seriously affect my posterior beliefs.]

If any of these assumptions breaks down seriously, then I have no reason to expect the above theory to apply. How much trust may I place in the approximations (A.33) and (A.34) then? As with all important questions, this is to a large extent unanswered. From the literature it seems safe to say that $\theta^*_{1,k}$ and $\theta^*_{2,k}$ are approximations to the means of my posterior distributions that are robust to fairly large breakdowns in my assumptions. Unfortunately the same does not seem to apply to $D_{1,k}$ and $D_{2,k}$ (Kendall & Stuart, 1961, ch. 31). So whilst I may be fairly safe in stating where I think the parameters lie, remarks concerning the precision of my beliefs may be unsound. It is important to realise that these comments, somewhat translated, apply equally well to a least-squares analysis of the model.

APPENDIX B
Approximations to the mean and variance of a real function of a random vector

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a smooth, well behaved function, *i.e.* possess all the properties of continuity and differentiability I need below.

Let $X \in \mathbb{R}^n$ be a random vector such that $\Pr(X \in D) = 1$. Furthermore, let the first two moments of X exist and be finite:

$$E(X) = \mu$$
$$\text{var}(X) = V.$$

Set $Z = (X - \mu)$, then by Taylor's expansion in D:

$$f(X) = f(\mu) + Z^T \nabla f(\mu) + \tfrac{1}{2}(Z^T \nabla)^2 f(\mu) + \dots$$

Setting

$$H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mu) \right)$$

and taking expectations:

$$E[f(X)] = f(\mu) + \tfrac{1}{2}\text{trace}(HV) + \dots \qquad (B.0)$$

Thus, to a second-order approximation, $f(X)$ is a biased observation on $f(\mu)$ by an amount $\tfrac{1}{2}$ trace (HV).

Similarly, to the first order:

$$\text{var}[f(X)] = \text{trace}(GV) + \dots \qquad (B.1)$$

where

$$G = \left(\frac{\partial f}{\partial x_i}(\mu) \ \frac{\partial f}{\partial x_j}(\mu) \right).$$

To calculate a second-order approximation to the variance of $f(X)$, the fourth moments of X are required.

References

BARNETT, V. (1973). *Comparative Statistical Inference.* New York: John Wiley.

BLIGHT, B. J. N. & OTT, L. (1975). *Biometrika,* 63, 79–88.

BOX, G. E. P. & TAIO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass.: Addison-Wesley.

DE GROOT, M. H. (1970). *Optimal Statistical Decisions* New York: McGraw-Hill.

DIAMOND, R. (1969). *Acta Cryst.* A25, 43–55.

DODSON, E. (1976a). *Crystallographic Computing Techniques,* edited by F. R. AHMED, pp. 205–211. Copenhagen: Munksgaard.

DODSON, E. (1976b). *Crystallographic Computing Techniques,* edited by F. R. AHMED, pp. 259–268. Copenhagen: Munksgaard.

DODSON, E., EVANS, P. & FRENCH, S. (1975). *Anomalous Scattering,* edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 423–436. Copenhagen: Munksgaard.

DODSON, E., ISAACS, N. & ROLLETT, J. S. (1976). *Acta Cryst.* A32, 311–314.

DODSON, E. & VIJAYAN, M. (1971). *Acta Cryst.* B27, 2402–2411.

EFRON, B. & MORRIS, C. (1977). *Sci. Am.* 236, 119–127.

FRENCH, S. (1975). D.Phil. Thesis, Univ. of Oxford.

HAMILTON, W. C. (1964). *Statistics in Physical Science.* New York: Ronald.

HAMILTON, W. C. & ABRAHAMS, S. C. (1970). *Acta Cryst.* A26, 18–24.

HUBER, R. & KOPFMANN, G. (1968). *Acta Cryst.* A24, 348–351.

HUBER, R. & KOPFMANN, G. (1969). *Acta Cryst.* A25, 143–152.

JEFFREYS, H. (1961). *Theory of Probability.* Oxford: Clarendon Press.

JOHNSON, N. L. (1949). *Biometrika,* 36, 149–168.

KARTHA, G. & PARTHASARATHY, R. (1965). *Acta Cryst.* 18, 745–753.

KENDALL, M. G. & STUART, A. (1961). *The Advanced Theory of Statistics,* Vol. 2. London: Griffin.

KRIEGER, M., CHAMBERS, J. L., CHRISTOPH, G. C., STROUD, R. M. & TRUS, B. L. (1974). *Acta Cryst.* A30, 740–748.

LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint,* Vols. 1 and 2. Cambridge Univ. Press.

LINDLEY, D. V. (1971). *Bayesian Statistics, A Review.* Philadelphia: Society for Industrial and Applied Mathematics.

LINDLEY, D. V., MILLEDGE, H. J. & WEBB, G. I. (1974). Symposium on Bayesian Methods in Crystallography, Oxford. Unpublished paper.

LINDLEY, D. V. & SMITH, A. F. M. (1972). *J. R. Stat. Soc. B,* 34, 1–41.

McCANDLISH, L. E. STOUT, G. H. & ANDREWS, L. C. (1975). *Acta Cryst.* A31, 245–249.

MARQUARDT, D. W. (1970). *Technometrics,* 12, 591–611.

MENDES, M. & DE POLIGNAC, C. (1973). *Acta Cryst.* A **29**, 1–9.

ROLLETT, J. S. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 167–181. Copenhagen: Munksgaard.

RUDIN, W. (1964). *Principles of Mathematical Analysis*. New York: McGraw-Hill.

SINGH, A. K. & RAMASESHAN, S. (1966). *Acta Cryst.* **21**, 279–280.

SMITH, A. F. M. (1973). *J. R. Stat. Soc. B*, **35**, 67–75.

SRINIVASAN, R. (1966). *Z. Kristallogr.* **126**, 175–181.

STEIN, C. (1956). *Proc. 3rd Berkeley. Symp.*, Vol. 1, pp. 197–206.

TICKLE, I. (1975). *Acta Cryst.* B **31**, 329–331.

WASER, J. (1963). *Acta Cryst.* **16**, 1091–1094.

WEBB, G. I. (1974). Symposium on Bayesian Methods in Crystallography, Oxford. Unpublished paper.

Diffraction by Crystals with Planar Faults.
III. Structure Analyses Using Microtwins

BY J. M. COWLEY AND ANDREW Y. AU

Department of Physics, Arizona State University, Tempe, Arizona 85281, USA

The general theory for kinematical diffraction from crystals having planar faults is applied to the case of microtwinning and the related cases occurring in some minerals where there is a disordered sequence of two types of structure having small differences in composition, unit-cell dimensions and axial orientations. It is shown that, if intensities are measured for unresolved or partially resolved pairs of diffraction spots using conventional techniques, errors may well arise if the measurements are interpreted on the usual assumption that the intensities from the different crystal regions are summed incoherently. Calculations for representative cases suggest that errors, due to the neglect of the spreading of intensity maxima into continuous streaks, may amount to 20 or 30% when overlapping diffraction spots have structure amplitudes of opposite sign, but are usually much smaller, especially if the structure amplitudes are of the same sign.

In the first paper of this series [(Cowley, 1976*a*, hereinafter referred to as (I)] a general theory for kinematical diffraction from crystals having planar faults was presented. Applications to particular types of faults were given there and in the second paper of this series (Cowley, 1976*b*). These applications should be sufficient to illustrate the derivation of expressions appropriate to special cases from the general theory and it is not our intention to multiply examples. However in the course of discussions with Drs Gabrielle and J. D. H. Donnay on the interpretation of electron diffraction patterns from feldspars the question was raised as to the influence of microtwinning on the intensities which would be measured and used in the course of an X-ray diffraction structure analysis. We have therefore considered an idealized case of this kind and attempted to estimate the magnitude of any errors which might result from the application of accepted practices of structure analysis.

There are many cases reported in the literature for which it appears that twin planes occur more or less at random and, on the average only a few unit cells apart. One case, illustrated graphically by high-resolution electron microscopy, is that of monoclinic enstatite (Iijima & Buseck, 1976). In this case the two orientations of the monoclinic *a* axis differ by a sufficiently large angle to allow most pairs of related diffraction spots to be clearly separated, although there is considerable diffuse streaking intensity between them.

For some of the feldspars the separation of axial orientations may be very much smaller. The evidence of variation of axial orientations is not clear from diffraction patterns but may be deduced from the irregular mottling of electron-microscope images (the 'tweed' structure). For bytownite (McLaren & Marshall, 1974) the variation in orientation of the unit-cell axes over distance of the order of 100 Å appears to accompany compositional variations associated with an exsolution process. Similar orientational variations appear to be present in a Himalaya mine orthoclase (Prince, Donnay & Martin, 1973), giving rise to a tweed structure in high-resolution electron micrographs (Iijima, private communication), but it is not clear whether a compositional variation is involved in this case.

For convenience we continue our discussion in terms of 'twins', but the treatment will not exclude such cases involving variations of composition as well as small