SHORT COMMUNICATIONS

Acta Cryst. (1994). A**50**, 547–550

On uncertainty in maximum-entropy maps and the generalization of 'classic MaxEnt'.

By STEEN HANSEN, *Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Thorvaldsensvej* 40, 1871 *FRB C, Denmark*, and STEPHEN W. WILKINS, *CSIRO, Division of Materials Science and Technology, Private Bag* 33, *Rosebank MDC, Clayton, Victoria* 3169, *Australia*

(*Received* 19 *January* 1994; *accepted* 22 *April* 1994)

Abstract

A Bayesian approach to the treatment of uncertainty in maximum-entropy maps is developed based on a deeper hypothesis space that includes the prior map. The new approach is able to take into account uncertainty in the knowledge of the prior and provides a generalization of 'classic MaxEnt'. Some examples of the new approach are presented for simulated SAXS data for cytochrome c.

1. Introduction

Maximum entropy has become a powerful tool for statistically inverting scattering data, *e.g.* in analysing X-ray scattering data from large molecules in crystalline form (Bricogne, 1993) and also in solution (*e.g.* Müller & Hansen, 1994). Typically, an estimate for the 'best map' based on the given data and prior information is presented. The question naturally arises as to the degree of belief to be attached to different points in the map. A partial treatment of this problem has been presented by Skilling (1989) in connnection with the development of 'classic MaxEnt' based on a Bayesian argument. However, this treatment can lead to unrealistically small estimates for the uncertainty in a map relative to a prior even when the prior map in the given region is not known with a high degree of confidence. In the present note, we explore this matter further and also consider the effect that uncertainty in the prior map has on the uncertainty in the 'best map'. It transpires that, given some simple and reasonable assumptions, there is an elegant symmetry between the influence of uncertainty in the prior map and uncertainty in the data leading to a generalization of 'classic MaxEnt'. The resulting expression should be useful in assessing the reliability of different regions of a map in a wide class of scattering problems.

2. Generalized MaxEnt

Let us assume that we have a map described by the continuous variable $\mathbf{f}$ over the N pixels of the map, *i.e.* $\mathbf{f} = (f_1, f_2, ..., f_N)$ and that we also have a prior map $\mathbf{m} = (m_1, m_2, ..., m_N)$. Then, for data $\mathbf{d} = (d_1, d_2, ..., d_M)$, Bayes's theorem leads to the expression

$$P(\mathbf{f}|\mathbf{d}, \mathbf{m}) = P(\mathbf{d}|\mathbf{f})P(\mathbf{f}|\mathbf{m})/P(\mathbf{d}) \qquad (1)$$

for the probability P of $\mathbf{f}$ conditional on $\mathbf{d}$ and $\mathbf{m}$ with the assumption that $\mathbf{m}$ and $\mathbf{d}$ are independent.

Regarding $\mathbf{m}$ as nuisance parameters (see *e.g.* Jaynes, 1986) which can (at least in principle) be integrated out, the probability of $\mathbf{f}$ conditional only on the measurement $\mathbf{d}$ is given by

$$P(\mathbf{f}|\mathbf{d}) \propto \int_{\text{prior}} P(\mathbf{d}|\mathbf{f})P(\mathbf{f}|\mathbf{m})P(\mathbf{m})\mathrm{d}\mathbf{m} \qquad (2)$$

assuming that the data have been measured and consequently omitting $P(\mathbf{d})$. For the probabilities on the right-hand side of (2) we have

$$P(\mathbf{d}|\mathbf{f}) \propto \exp(-\chi_d^2/2) \qquad (3)$$

with

$$\chi_d^2 = \sum_{i=1}^{M} \left(d_i - \sum_{j=1}^{N} a_{ij} f_j \right)^2 \Big/ \sigma_{d,i}^2, \qquad (4)$$

where the a_{ij} are the elements of the reponse matrix and we assume that the d_i are Gaussian variates.

For the case of small-angle scattering from solutions, the response matrix is given by the Fourier transformation $a_{ij} = \sin(q_i r_j)/(q_i r_j)$, where q_i is the length of the scattering vector at data point i and f_j is the value of the distance distribution function at the distance r_j.

For crystallography and the case of measured structure values with assumed known phases, (4) takes the form

$$\chi_d^2 = \sum_{\mathbf{k}} \left| D_{\mathbf{k}} - \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{k}) \right|^2 \Big/ \sigma_{\mathbf{k}}^2, \qquad (4a)$$

where $D_{\mathbf{k}}$ is the measured structure factor and $\mathbf{r}_j$ is the position of the jth pixel.

If we now invoke the general expression for the probability of a map $\mathbf{f}$ about an arbitrary prior map $\mathbf{m}$, which is based on the maximum-entropy principle and the assumption of the stochastic model for maps known as the 'monkeys and balls' model then [see equations (5) and (7) in Wilkins, Steenstrup & Varghese (1985)]

$$P(\mathbf{f}|\mathbf{m}) \propto \exp(\alpha S) \qquad (5)$$

with

$$S = -\sum_{j=1}^{N} f_j \ln(f_j/m_j) \qquad (6)$$

and α a Lagrange multiplier determining the relative weighting of the prior information (as given by the entropy) and the 'new' information from the data (as given by the χ_d^2). Various methods for the choice of this Lagrange multiplier have been suggested (see *e.g.* Titterington, 1985).

If the distribution of the prior is assumed to be Gaussian with a mean $\bar{\mathbf{m}}$ and standard deviation σ_m we have

$$P(m) \propto \exp(-\chi_m^2/2) \tag{7}$$

with

$$\chi_m^2 = \sum_{j=1}^{N} (m_j - \bar{m}_j)^2 \Big/ \sigma_{m,j}^2 \tag{8}$$

and

$$P(\mathbf{f}|\mathbf{d},\bar{\mathbf{m}}) \propto \int_{\text{priors},\alpha} \exp(\alpha S - \chi_d^2/2 - \chi_m^2/2)\mathrm{d}\mathbf{m}\mathrm{d}\alpha. \tag{9}$$

The error bars in a MaxEnt map may be obtained from (9) *via*

$$P(\mathbf{f} \in V|\mathbf{d},\bar{\mathbf{m}}) \propto \int_V \int_{\text{priors},\alpha} \exp(\alpha S - \chi_d^2/2 - \chi_m^2/2)\mathrm{d}\mathbf{m}\mathrm{d}\alpha\mathrm{d}\mathbf{f}, \tag{10}$$

where $\mathbf{m}$ (and α which is to be determined for each $\mathbf{m}$) acts as a set of nuisance parameters to be integrated over.

If the Gaussian distribution of the prior is very narrow we have in the limit of $\sigma_m = \mathbf{0}$ the usual expression for the error determination

$$P(\mathbf{f} \in V|\mathbf{d},\bar{\mathbf{m}}) \propto \int_V \exp(\alpha S - \chi_d^2/2)\mathrm{d}\mathbf{f}, \tag{11}$$

assuming α to be known. This leads to the error matrix $[\nabla\nabla(\alpha S - \chi_d^2/2)]^{-1}$ taken in the maximum-entropy solution $\mathbf{f}^{\max}$, as usually quoted in the literature.

Furthermore, if $\mathbf{f}^{\max} \simeq \mathbf{m}$, the value for α (given by $\alpha = \frac{1}{2}|\nabla\chi_d^2|/|\nabla S|$) will be large (as $\nabla S \simeq 0$), so that $\alpha\nabla\nabla S$, which is diagonal with elements $-\alpha/f_j$, will dominate the error matrix. In consequence, the error estimate for the solution will be determined only by the entropy metric leading to the possibility of unrealistically small estimates for the uncertainty in the map as mentioned in the *Introduction*.

In the general case [(10)], for heuristic purposes the quadratic expansion for the entropy term is here employed:

$$\alpha S \simeq -\sum_{j=1}^{N} (f_j - m_j)^2/(2m_j/\alpha) = -\chi_s^2/2 \tag{12}$$

(for $\mathbf{f} \simeq \mathbf{m}$), in which case (10) can be approximated by

$$P(\mathbf{f} \in V|\mathbf{d},\bar{\mathbf{m}}) \propto \int_V \int_{\text{priors}} \exp(-\chi_s^2/2 - \chi_d^2/2 - \chi_m^2/2)\mathrm{d}\mathbf{m}\mathrm{d}\mathbf{f}, \tag{13}$$

demonstrating the symmetry in the roles of the variable $\mathbf{f}$ and the prior $\mathbf{m}$. This symmetry is displayed schematically in Fig. 1(*a*), where $\mathbf{f}$ and $\mathbf{m}$ are connected to each other, to the measured data $\mathbf{d}$ and to the estimated prior $\bar{\mathbf{m}}$ by constraints corresponding to 'rubber bands' (with an energy proportional to the square of elongation) and the problem of finding the 'best' $\mathbf{m}$ and $\mathbf{f}$ is now reduced to a minimization of the total energy of the bands. In this picture, the strengths of the three bands are determined by the belief in the prior (through σ_m), the quality of the data (σ_d) and the assumption of Poisson statistics ('monkeys and balls' model) for the quantity determining the form of the entropy constraint (Wilkins, Steenstrup & Varghese, 1985). In analogy with three rubber bands, the solution (the minimum energy) is found as shown in Fig. 1(*b*) where the linking points are aligned.

As is evident from Fig. 1(*b*), the estimate of $\mathbf{f}$ is influenced by the strength of the binding of $\mathbf{m}$ to the prior model $\bar{\mathbf{m}}$, as a strengthening of this band will lead to a value for $\mathbf{f}$, which is moved closer to $\bar{\mathbf{m}}$. The estimate of $\mathbf{f}$ can be found by combining the two rubber bands χ_s^2 and χ_m^2 into a single effective band, leading to the new constraint

$$\chi_{s,m}^2 = \sum_{j=1}^{N} (\bar{m}_j - f_j)^2 \Big/ \left[\sigma_{m,j} + (\bar{m}_j/\alpha)^{1/2}\right]^2 \tag{14}$$

and the approximation

$$P(\mathbf{f} \in V|\mathbf{d},\bar{\mathbf{m}}) \propto \int_V \exp(-\chi_{s,m}^2/2 - \chi_d^2/2)\mathrm{d}\mathbf{f}. \tag{15}$$

Equation (15) is solved in the conventional manner and the error matrix can be found as described above but using the new effective prior described by (14). The new error matrix is now dependent on the uncertainty of the Gaussian prior.

For the purpose of illustration, consider the special case that the reponse matrix is equal to unity. Consequently, the error matrix is diagonal and given by

$$e_{jj} = \left\{1\Big/\left[\sigma_{m,j} + (\bar{m}_j/\alpha)^{1/2}\right]^2 + 1/\sigma_{d,j}^2\right\}^{-1}, \tag{16}$$

where the usual expression (11) will lead to an error matrix with elements

$$e_{jj} = (\alpha/f_j + 1/\sigma_{d,j}^2)^{-1}. \tag{17}$$

Comparison of (16) and (17) shows that the main effect of the introduced uncertainty or spread in the prior is to impose a lower limit upon the elements of the error matrix. The exact estimate of this uncertainty (σ_m) will usually not be very important as its main purpose is to avoid dominance of the α/f_j terms in the error matrix if, perhaps more by coincidence than firm belief, a prior close to the solution $\mathbf{f}^{\max}$ is used.

With the errors estimated from (17), the α/f_j terms in the error matrix will be dominant. A similar scenario where the density of distributions defined by the entropy metric leads to absurdities has been described by Jaynes (1986), elaborating on the kangaroo problem invented by Gull & Skilling (1984). In this problem, the density of states around the maximum-entropy

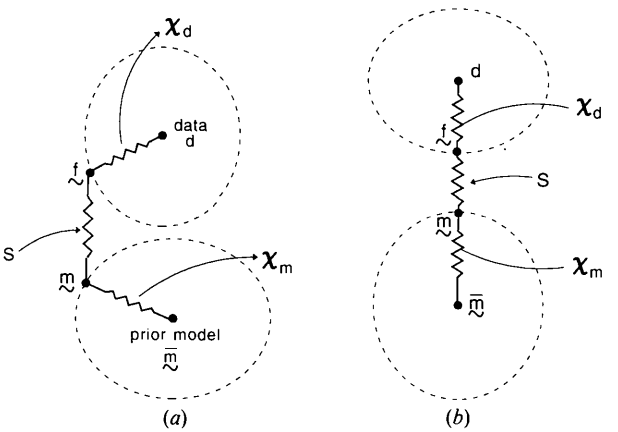

Fig. 1. Schematic illustration of combined constraint and regularization relations for generalized MaxEnt in (*a*) the general case and (*b*) the case where the MaxEnt condition is satisfied.

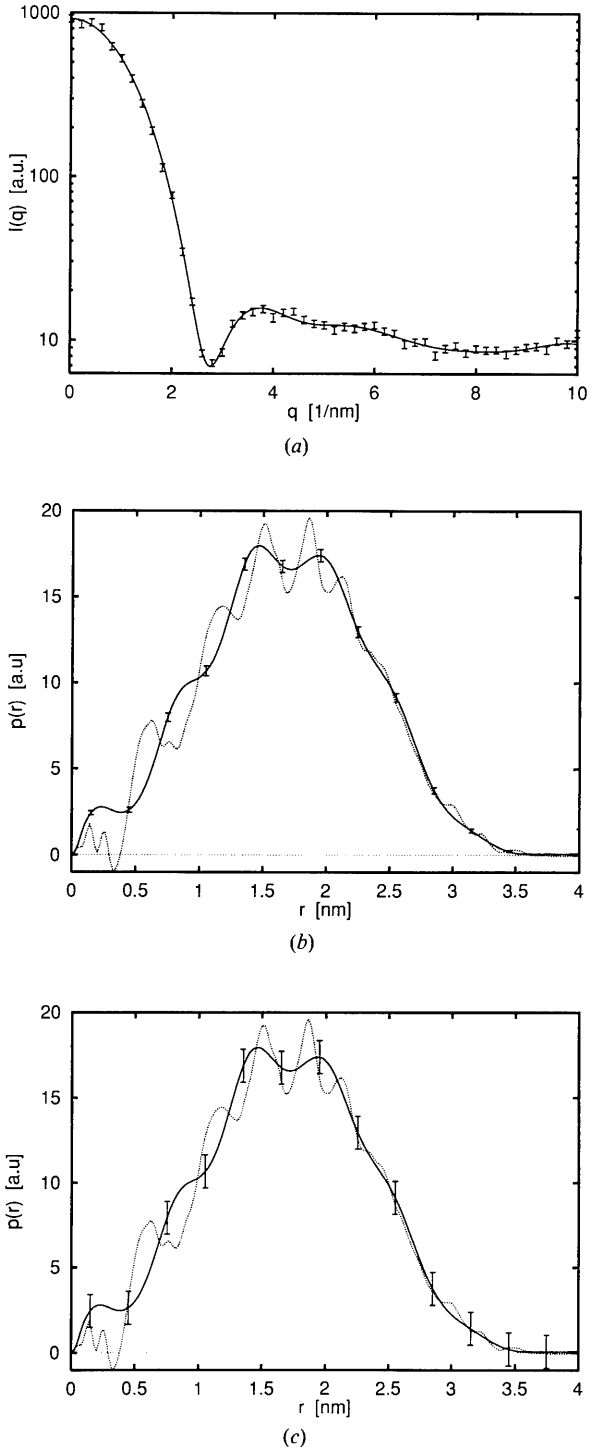

Fig. 2. (*a*) Error bars: simulated data for cytochrome c. Full line: MaxEnt fit to data. (*b*): Full line and error bars: classic MaxEnt estimate of distance distribution function corresponding to (*a*). Error bars are shown at distances corresponding to the resolution of the simulated data. Dotted line: the distance distribution function for cytochrome c. (*c*) Full line and error bars: generalized MaxEnt estimate of distance distribution function corresponding to (*a*). Error bars are shown at distances corresponding to the resolution of the simulated data. Dotted line: the distance distribution function for cytochrome c.

solution increases as the number of kangaroos increase (the number of kangaroos or the density resolution or the number of grey-scale levels of a picture corresponds to the Lagrange multiplier α in the above treatment) leading to an error estimate that is obviously too small. In analogy with the kangaroo problem where the resolution of the problem is obtained by the introduction of a variability of the parameters describing the prior information (the generating function for $\mathbf{m}$, the form of which is determined by some simple requirements), we have here attempted to treat the general problem of error estimation in MaxEnt by the introduction of a similar variability of $\mathbf{m}$.

In a similar attempt to 'soften' the implicit assumptions of the prior, Gull (1989) introduces 'hidden variables' and a 'pre-blur' to account for the strength of the binding of the solution $\mathbf{f}$ to the prior $\bar{\mathbf{m}}$. Both Gull's hidden variables and Jaynes's generating function could be argued to be the consequence of an uncertainty in the prior and an attempt to avoid the penalty for using a 'wrong' prior in the calculations. ('wrong' in the sense of $\bar{\mathbf{m}}$ being far from $\mathbf{f}^{\max}$). From the above it seems that the problem of error calculation is more fundamentally approached if the concept of an uncertainty in the prior is introduced at the outset and the estimate of this uncertainty is allowed to propagate through the further calculations.

To illustrate the problems mentioned above, simulated small-angle X-ray scattering data on cytochrome c is shown in Fig. 2(*a*) (from Müller & Hansen, 1994). The result of a classic MaxEnt calculation of the distance distribution function is shown in Fig. 2(*b*) using a prior close to the final estimate (a prior that could have been obtained, for example, from measurements at lower resolution). The size of the error bars is clearly underestimated especially at short distances. The inclusion of uncertainty in the prior as described above gives the result shown in Fig. 2(*c*). For this case, the uncertainty in the prior was simply taken as a constant to illustrate our point. If the prior is taken from previous measurements, the errors from these measurements could be used to estimate the errors on the prior. However, in spite of this very simplistic example, it is evident from Fig. 2(*c*) that the size of the error bars now takes a more realistic value demonstrating the validity of the approach outlined above.

Finally, it should be noted that the dominance of the regularizing term is frequently a problem in regularization (not just when the entropy happens to be the regularizer). A Bayesian argument similar to the one given above could be carried through for the cases of these and other regularizing functions.

3 Concluding remarks

In the present paper, our main purpose has been to provide a broad conceptual framework for treating the problem of error estimation in MaxEnt. Furthermore, we have attempted to give a simple and transparent approach to the problem of error estimation in maps by suggesting the substitution of (14) for (12) in the calculation of the error matrix. In the case of the important problem of crystal structure determination from diffraction data, the present formulation offers a framework in which local structural information together with a degree of belief can be introduced into a crystal-structure determination and so provide an additional driving force toward the solution, *e.g.* in the map-determination cycle of the combined maximum-entropy and maximum-likelihood approach to this problem developed by Bricogne (1993, and references cited therein). As particular regions of a map become more accurately determined and identified, the constraint on $\mathbf{m}$ can be tightened up and

this in turn will tend to make the Hessian more strongly convex, thus aiding the structure-determination process [see (14)]. By this means, the information contained in a molecular envelope (solvent flattening) and molecular fragments could, for example, be incorporated in the map.

References

BRICOGNE, G. (1993). *Acta Cryst.* D**49**, 37-60.

GULL, S. F. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. SKILLING, pp. 53-71. Dordrecht: Kluwer Academic Publishers.

GULL, S. F. & SKILLING, J. (1984). *Indirect Imaging*, edited by J. A. ROBERTS, pp. 267-279. Cambridge Univ. Press.

JAYNES, E. T. (1986). *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. JUSTICE, pp. 1-58. Cambridge Univ. Press.

MÜLLER, J. J. & HANSEN, S. (1994). *J. Appl. Cryst.* In the press.

SKILLING, J. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. SKILLING, pp. 42-52. Dordrecht: Kluwer Academic Publishers.

TITTERINGTON, D. M. (1985). *Astron. Astrophys.* **144**, 381-387.

WILKINS, S. W., STEENSTRUP, S. & VARGHESE, J. N. (1985). *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 113-123. Adenine Press.

Acta Cryst. (1994). A**50**, 550

Determination of quasicrystalline structures: a refinement program using symmetry-adapted parameters. Erratum.

By L. ELCORO, J. M. PEREZ-MATO and G. MADARIAGA, *Departamento de Física de la Materia Condensada, Facultad de Ciencias, Universidad del País Vasco, Apartado* 644, *Bilbao, Spain*

(*Received* 1 *May* 1994)

Abstract

A typesetting error in equation (28) of Elcoro, Perez-Mato & Madariaga [*Acta Cryst.* (1994), A**50**, 182–193] is corrected. The correct equation is

$$F(\mathbf{H}) = [|\mathbf{A}|/V(\mathbf{a}_i)] \sum_{\mu,m} p_m(\mu) f_m(\mathbf{H}) \times \sum_R \exp(-\widetilde{\mathbf{RH}}\boldsymbol{\beta}_\mu^m \mathbf{RH}) \times \exp[2\pi i \hat{\mathbf{h}} \cdot (\hat{\mathbf{R}}\hat{\boldsymbol{\theta}}_\mu + \hat{\mathbf{t}})] \int_\Omega d\varphi_1 \ldots d\varphi_{n-4} \times \int_a^b dr J(r,\varphi_1,\ldots,\varphi_{n-4}) \exp[2\pi i (\tilde{\mathbf{T}}_\mu^{-1} \tilde{\mathbf{R}}_I \mathbf{h}_I) \cdot \mathbf{x}_I], \quad (28)$$

where

$$a = \sum_i a_i^{\mu,\text{in}} Z_i(\varphi_1,\ldots,\varphi_{n-4})$$

and

$$b = \sum_i a_i^{\mu,\text{ex}} Z_i(\varphi_1,\ldots,\varphi_{n-4}).$$

All information is given in the *Abstract*.

Acta Cryst. (1994). A**50**, 550

Dynamical X-ray diffraction from imperfect crystals: a solution based on the Fokker–Planck equation. Erratum.

By T. J. DAVIS, *CSIRO Division of Materials Science and Technology, Private Bag* 33, *Rosebank MDC, Clayton, Victoria* 3169, *Australia*

(*Received* 25 *May* 1994)

Abstract

The following mathematical expressions were incorrectly printed in the paper by Davis [*Acta Cryst.* (1994), A**50**, 224–231].

Page 225: the correct expression for χ_{-h} is

$$\chi_{-h} = -C(\gamma_h/\gamma_o)\chi'_{-h}.$$

Page 228: equations (22) and (23) should read

$$R(t) = R_o + [R(t') - R_o] \exp[2i\alpha\omega(t-t')] \times (1 + [R(t') - R_o](\chi_{-h}/2\omega) \times \{1 - \exp[2i\alpha\omega(t-t')]\})^{-1}, \quad (22)$$

$$\omega = \pm(\beta^2 - \chi_h \chi_{-h})^{1/2}. \quad (23)$$

All information is given in the *Abstract*.