ELSEVIER

Physica B 248 (1998) 327–337

PHYSICA B

The Bayesian approach to reflectivity data

D.S. Sivia*, J.R.P. Webster

CLRC, ISIS Facility, Rutherford Appelton Laboratory, Chilton, Oxon, OX11 0QX, UK

Abstract

We outline the basic principles of Bayesian probability theory and illustrate its use with reflectivity data. This approach provides a unified rationale for data analysis, which both justifies many of the commonly used analysis procedures and indicates some natural extensions that enhance their potency. Thus, for example, we find that the ubiquitous least-squares apparatus of parameter estimation is easily adapted to tackle the more intriguing question of model selection. A free-form solution application is also presented, as is a discussion of the difficult but important question of optimal experimental design. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Bayesian probability theory; Model selection; Maximum entropy; Experimental design

1. Introduction

The analysis of reflectivity data, in common with all experimental investigation, ultimately entails the drawing of suitable inferences about an object of interest given a set of incomplete and noisy measurements. While we may strive to optimise the collection of the latter, any rationale method of analysis must take the uncertainties into account and provide a corresponding measure of reliability for the conclusions reached.

About three hundred years ago, people like Bernoullis [1], Bayes [2] and Laplace [3] started to think seriously about the question of how to reason in situations where we cannot argue with certainty. Although the origins and context of their deliberations were far removed from our modern experimental facilities, the probabilistic apparatus they developed to tackle the problem is still highly relevant today. Indeed, more recently, Cox [4] has shown that any method of plausible reasoning (or scientific inference) that satisfies simple rules of logical consistency must be equivalent to the use of ordinary probability theory.

There are just two basic rules in probability theory; namely, the sum rule and product rule

$$\text{prob}(x|I) + \text{prob}(\bar{x}|I) = 1, \tag{1}$$

$$\text{prob}(x, y|I) = \text{prob}(x|y, I) \times \text{prob}(y|I), \tag{2}$$

where $\bar{x}$ represents the proposition that x is false, the vertical bar "|" means "given" and the comma is read as the conjunction "and". Many other relationships can be derived from Eqs. (1) and (2), with

*Corresponding author. Fax: +44 0 1235 44 5720; e-mail: dss@isise.rl.ac.uk.

PII: S0921-4526(98)00259-2

Bayes' theorem and marginalisation being two of the most useful corollaries

$$\text{prob}(x|y,I) = \frac{\text{prob}(y|x,I) \times \text{prob}(x|I)}{\text{prob}(y|I)}, \tag{3}$$

$$\text{prob}(x|I) = \int \text{prob}(x,y|I)\,\mathrm{d}y. \tag{4}$$

While these equations are well known, they become a particularly potent force (in terms of their wider applicability) when interpreted from the Bayesian standpoint. That is to say a probability encodes a state of knowledge, and all probabilities are conditional. The second point is emphasised above by the inclusion of the symbol I on every right-hand side, which denotes the relevant background information and assumptions.

In this paper, we illustrate the (direct) use of probability theory for the analysis of neutron reflectivity data; a more general tutorial introduction to this Bayesian approach can be found in Sivia [5]. We begin with elementary parameter estimation in Section 2, showing both how some familiar statistical procedures can easily be justified within our unified framework and how they can be generalised. In Sections 3–5 we move on to increasingly demanding examples involving model selection, free-form solutions and optimal experimental design. We conclude with Section 6.

2. Parameter estimation

Let us begin with the simplest type of data analysis problem, namely the estimation of the values of M parameters $\{X_i\}$ of a certain model given N data $\{\boldsymbol{D}_k\}$. A two-parameter example relevant to reflectivity data from a bare substrate could be $X_1 = \beta$ and $X_2 = W$, referring to the average scattering–length density and (Gaussian) surface roughness of the substrate, respectively. The probability density function (pdf) $\text{prob}(\boldsymbol{X}|\boldsymbol{D},\boldsymbol{I})$, where $\boldsymbol{X} = \{X_i\}$ and $\boldsymbol{D} = \{\boldsymbol{D}_k\}$, then encapsulates what we can say about the quantities of interest in the light of the empirical evidence and our background information (such as a knowledge of the experimental setup): the values of the parameters that maximise this pdf represent a best estimate of $\boldsymbol{X}$, while its spread about this optimal point gives a measure of the reliability.

In order to calculate this so-called posterior pdf, we can use Bayes' theorem to relate it to two others that are easier to assign:

$$\text{prob}(\boldsymbol{X}|\boldsymbol{D},\boldsymbol{I}) \propto \text{prob}(\boldsymbol{D}|\boldsymbol{X},\boldsymbol{I}) \times \text{prob}(\boldsymbol{X}|\boldsymbol{I}). \tag{5}$$

The term on the far right is known as the prior pdf, and represents our state of knowledge (or ignorance) about $\boldsymbol{X}$ before the analysis of the current data. This is modified by the (new) measurements through the other pdf on the right, which is called the likelihood function. We should note that proportionality in Eq. (5) arises from the omission of $\text{prob}(\boldsymbol{D}|\boldsymbol{I})$ in the denominator; although this is fine for the parameter estimation problem, because its just a normalisation constant (not explicitly involve $\boldsymbol{X}$), it will play a crucial role when we come to model selection.

Some of the most widely used statistical procedures follow immediately from Eq. (5) once we make some suitable simplifying approximations. First of all, suppose we assign a uniform prior, $\text{prob}(\boldsymbol{X}|I) = \text{constant}$, to express gross initial ignorance; then the posterior pdf becomes proportional to the likelihood function, and we have a justification for the use of the maximum likelihood estimate. If the noise in the data is assumed to be (roughly) Gaussian, independent and additive, then the likelihood function takes the form

$$\text{prob}(\boldsymbol{D}|\boldsymbol{X},\boldsymbol{I}) \propto \exp(-\chi^2/2), \tag{6}$$

where χ^2 is given by the familiar sum of the squares of the "datum $\boldsymbol{D}_k$ minus the fit F_k over the errorbar σ_k".

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{D_k - F_k}{\sigma_k} \right)^2. \tag{7}$$

Thus, the maximum likelihood estimate is given by the ubiquitous least-squares solution. An obvious extension to simply minimising χ^2 is then to include additional terms, such as the sum of $[(X_i - a_i)/\varepsilon_i]^2$, to encode some cogent prior knowledge (like $X_i = a_i \pm \varepsilon_i$).

As an elementary, but concrete, example of parameter estimation, Fig. 1 shows the results of the analysis of neutron reflectivity data from a

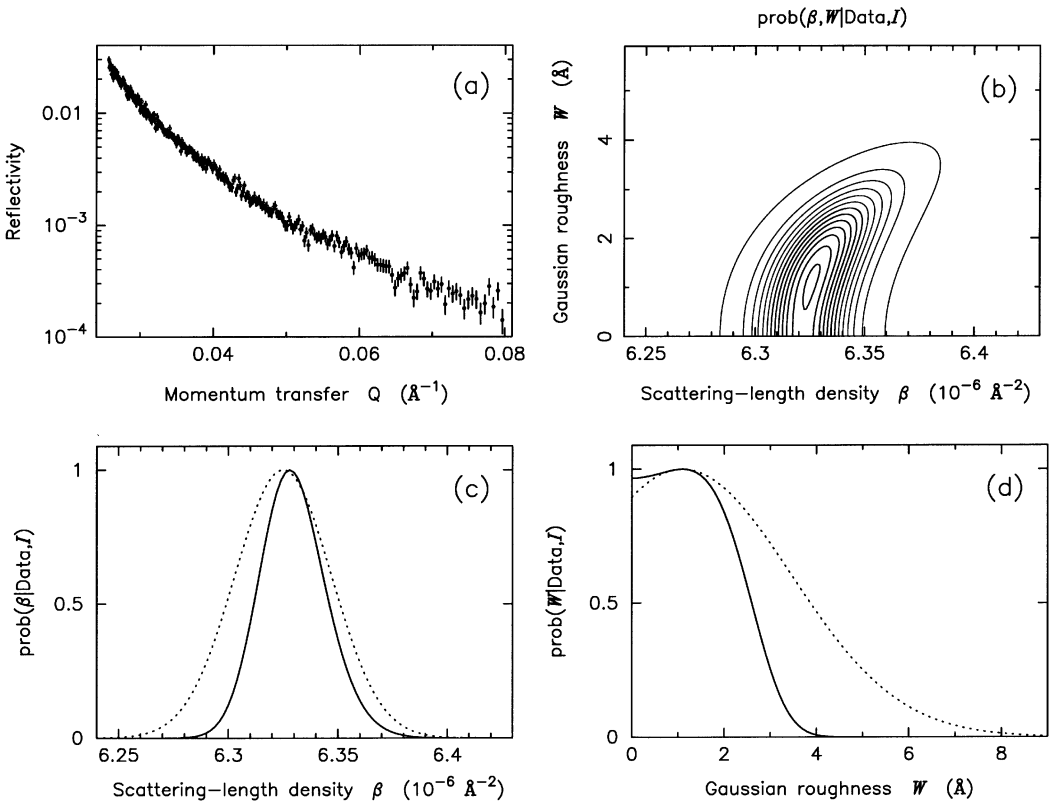

Fig. 1. (a) Neutron reflectivity data from a bare substrate ($\mathbf{D}_2\mathrm{O}$); (b) the posterior pdf for the scattering-length density β and the interfacial roughness W of the substrate; (c) and (d) The marginal posterior pdfs for β and W (solid line), and the approximations to them (dotted lines) given by a quadratic Gaussian expansion around the maximum of the joint pdf in (b).

bare substrate (which happens to be heavy water D_2O). The contour plot of the posterior pdf prob($\beta, W|\boldsymbol{D,I}$) in Fig. 1b defines all that we can infer about the average scattering-length density β and roughness W of the substrate; it was generated by calculating the values of $\exp(-\chi^2/2)$ on a uniform grid of points in the two-dimensional (β, W)-space. We often summarise the conclusions of such an analysis by providing the best estimates, $1-\sigma$ error-bars and a correlation for the parameters: $\beta = \beta_0 \pm \sigma_\beta$, $W = W_0 \pm \sigma_W$ and $\langle(\beta-\beta_0)(W-W_0)\rangle$. This is equivalent to approximating the posterior pdf by a multivariate Gaussian, usually chosen by matching up the location and curvature of the maximum. While this may be the most practical thing that can be done, Fig. 1b highlights its potential inadequacies: not only does it fail to capture some of the important characteristics of the shape of prob($\beta, W|\boldsymbol{D,I}$), it does not take into account the sharp cut-off $W \geqslant 0$ imposed by the prior.

If we were only interested in the surface roughness of the substrate, so that its scattering–length density was nothing more than a nuisance parameter (required to compute the fit to the data but of no intrinsic value), then our inference is defined by the marginal posterior pdf prob($W|\boldsymbol{D,I}$). This is obtained by integrating the two-dimensional pdf of Fig. 1b along the horizontal β-axis, and the result is shown in Fig. 1c. It tells us that the value of W most probably lies between 0 Å and 2 Å, but is very unlikely to be bigger than 4 Å; by contrast, the Gaussian approximation around the best-fit parameters for β and W would have suggested that $W = 1.1 \pm 2.3$ Å so that W might even be as large as 8 Å (about $3-\sigma$). If, on the other hand, our sole

interest lay in the average scattering-length density, then we would integrate the two-dimensional pdf of Fig. 1b along the vertical W-axis to obtain the marginal posterior pdf prob($\beta|\boldsymbol{D},\boldsymbol{I}$) of Fig. 1d.

Finally, before moving on to model selection, let us consider how systematic uncertainties can be handled. For example, given that no critical edge is apparent in the data of Fig. 1d, there might be an open question with regard to the (absolute, vertical) scaling of the measurements. Well, this situation can be dealt with in a straightforward manner, at least in principle, through marginalisation. That is to say, if ϕ represents a set of calibration parameters (i.e. essential for the calculations but intrinsically uninteresting) then they can be eliminated from the inference problem by integrating them out

$$\begin{aligned} \text{prob}(\boldsymbol{X}|\boldsymbol{D},\boldsymbol{I}) &= \int \text{prob}(\boldsymbol{X},\phi|\boldsymbol{D},\boldsymbol{I})\,\mathrm{d}\phi \\ &\propto \int \text{prob}(\phi|\boldsymbol{I})\exp(-\chi^2/2)\,\mathrm{d}\phi, \end{aligned} \tag{8}$$

where we have used Bayes' theorem, and assigned a uniform prior prob($\boldsymbol{X}|\boldsymbol{I}$) and a least-squares likelihood prob($\boldsymbol{D}|\boldsymbol{X},\phi,\boldsymbol{I}$), in the second line. Thus, for our bare substrate, we simply compute the values of $\exp(-\chi^2/2)$ on a three-dimensional grid of points in (β, W, A)-space and sum-up along the vertical scale-factor A-axis. This integral must be weighted by prob($A|\boldsymbol{I}$), so that due account is taken of any previous calibration measurements; the assignment of $\delta(A-A_0)$ indicates certainty, where as gross ignorance may be expressed with a uniform pdf. The marginal posterior pdf for β and W under the latter conditions is shown in Fig. 2; as expected, it is much broader than the case of Fig. 1b where the scaling was assumed to be known. The difference would not have been as striking, however, if the reflectivity data had been measured to lower Q's so that a critical edge was evident. If we were also not sure about the magnitude of the resolution blurring $\Delta Q/Q$, then we would need to compute $\exp(-\chi^2/2)$ on a four-dimensional grid of points in $(\beta, W, A, \Delta Q/Q)$-space and integrate out the two nuisance parameter weighted by prob($A, \Delta Q/Q|\boldsymbol{I}$); in our present case, the resultant marginal posterior pdf for β and W would look very similar to that of Fig. 2 because there are no sharp fringes predicted or seen in the data. For problems of higher dimensionality, the marginalisation of Eq. (8) is best done through a Monte Carlo algorithm [6] or estimated analytically with a multivariate Gaussian approximation [5].

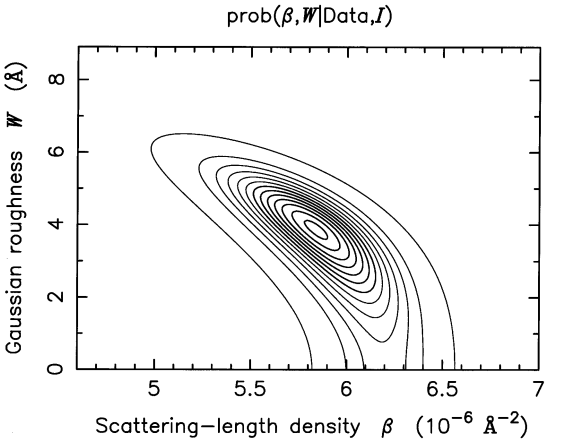

Fig. 2. The posterior pdf for β and W resulting from the data of Fig. 1a when the absolute (vertical) scaling of the measurements is assumed to be unknown.

3. Model selection

In the previous section, the M components of the vector $\boldsymbol{X}$ referred to the parameters of a particular model, or hypothesis, H; thus all the pdfs were implicitly conditional on H, so that prob($\boldsymbol{X}|\boldsymbol{D},\boldsymbol{I}$) was really shorthand for prob($\boldsymbol{X}|\boldsymbol{D},\mathrm{H},\boldsymbol{I}$) and so on. Now, suppose that we have competing hypotheses H_1 and H_2, each generally having a different number of associated variables; which one should be preferred in the light of data $\boldsymbol{D}$?

Well, to answer this sort of model selection problem, we need to consider the ratio of the posterior probabilities for the alternative hypotheses

$$\frac{\text{prob}(\mathrm{H}_1|\boldsymbol{D},\boldsymbol{I})}{\text{prob}(\mathrm{H}_2|\boldsymbol{D},\boldsymbol{I})} = \frac{\text{prob}(\boldsymbol{D}|\mathrm{H}_1,\boldsymbol{I})}{\text{prob}(\boldsymbol{D}|\mathrm{H}_2,\boldsymbol{I})} \times \frac{\text{prob}(\mathrm{H}_1|\boldsymbol{I})}{\text{prob}(\mathrm{H}_2|\boldsymbol{I})}, \tag{9}$$

where we have used Bayes' theorem, top and bottom, on the right-hand side. If this ratio is very much greater than one (say $\geqslant 100$), then H_1 is to be

preferred; if its a lot smaller than one (say $\leqslant 0.01$), then H_2 is better; and if the ratio is of order unity, then the measurements are not sufficient to make an informed judgement. The term on the far right of Eq. (9) is just the ratio of our prior probabilities for the two models; unless we have good initial reasons for preferring one over the other, this contribution can be set to unity. The data-dependent term looks rather awkward by comparison, as it involves an assessment of the probability of the measurements given the truth of the hypotheses but without knowing the values of their associated parameters. A little thought, however, reveals that the so-called evidence term prob($\boldsymbol{D}$|H, $\boldsymbol{I}$) is just the omitted normalisation constant in the Bayes' theorem expression of the posterior pdf in Eq. (5). In other words,

$$\text{prob}(\text{D}|\text{H},\boldsymbol{I}) \propto \int\int \cdots \int \text{prob}(\boldsymbol{X}|\text{H},\boldsymbol{I})\exp(-\chi^2/2)\text{d}^{\text{M}}\boldsymbol{X}, \tag{10}$$

where we have made the least-squares likelihood assignment of Eq. (6) for prob($\boldsymbol{D}|\boldsymbol{X}$, H, $\boldsymbol{I}$), and implicitly assumed that the error-bars for the data $\{\sigma_k\}$ are known. Thus, the only real difference between parameter estimation and model selection is that we must now give some consideration to a suitable prior-range for the parameters $\boldsymbol{X}$, even if we are going to assign a uniform pdf initially, because prob($\boldsymbol{X}$|H, $\boldsymbol{I}$) needs to be normalised properly.

A very simple example of this type of analysis concerns the question of whether or not the data in Fig. 1a have been scaled correctly in terms of absolute reflectivity; if they have not, then we must allow from the possibility of a multiplicative factor A (between 0.01 and 100, say). The evidence for the hypothesis that the scaling is correct ($A = 1$) is then given by the average value of $\exp(-\chi^2/2)$ computed over the two-dimensional grid of points in (β, W)-space allowed by the prior probability. The evidence for the proposition of an unknown scale factor is given by the correponding average of $\exp(-\chi^2/2)$ computed over the three-dimensional grid of points in (β, W, A)-space. Carrying out this calculation, we find in favour of Fig. 1b by a factor of about a hundred relative to Fig. 2.

Fig. 3a shows neutron reflectivity data from a mixture of two polymers taken by Geoghagen et al. [7]. Such measurements are often analysed in terms of an average scattering-length density depth profile $\beta(z)$ consisting of a few uniform layers of material with some interfacial roughness; a five-component profile that gives good agreement with the data is shown in Fig. 3b. A question that arises naturally concerns the significance of the features, particularly the subtle ones to the right of the large enhancement at the air interface. One way to tackle this issue is to construct a series of models of increasing complexity and see which one is most strongly supported by the data in terms of its probabilistic evidence. Carrying out such an analysis for the data of Fig. 3a, along the lines outlined in Sivia et al. [8], we obtain the results shown in Fig. 4. Thus, we find very strong evidence for the four-component model, which shows both a depletion region and slight enhancement bump following the

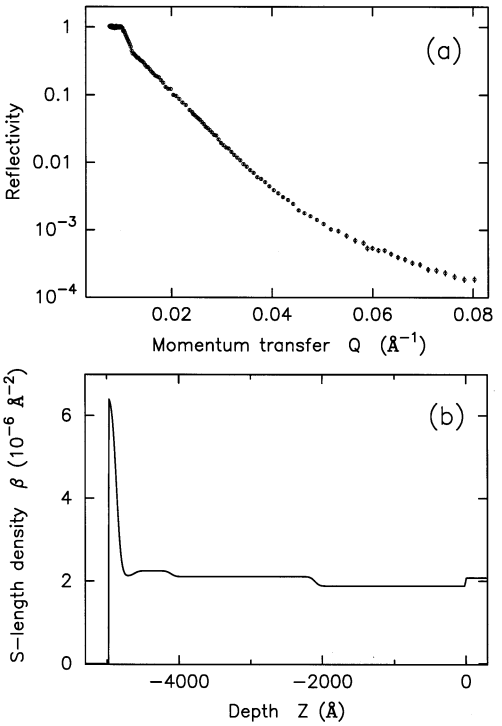

Fig. 3. (a) Neutron reflecivity data from a mixture of two polymers; (b) A scattering-length density profile $\beta(Z)$, consisting of five uniform layers with some interfacial roughness, that gives reasonable agreement with the measurements; the substrate (silicon) is at $Z > 0$, where as the air interface ($\beta = 0$) is at about $Z < -5000$ Å.

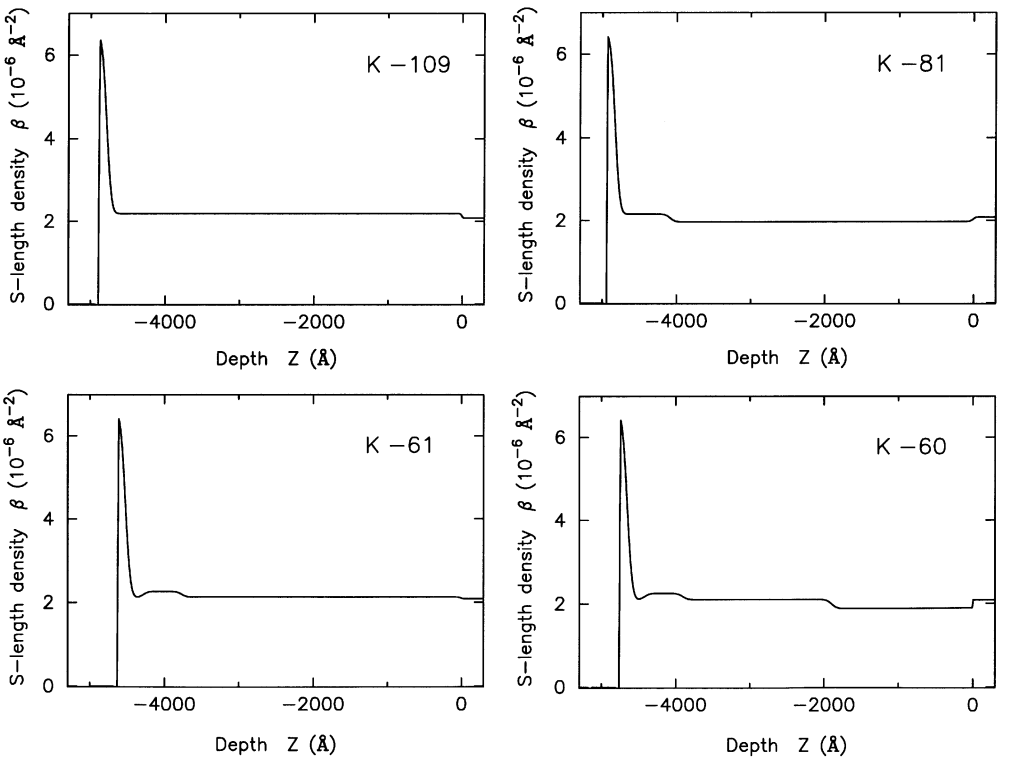

Fig. 4. Four best-fit $\beta(Z)$ profiles, corresponding to models with different numbers of uniform layers, and their probabilistic evidence; the numbers in the top right-hand corners are the (base-10) logarithms of the probabilities for the models, given the data of Fig. 3a, and K is a normalisation constant common to all.

huge enrichment at the air surface, but would be reluctant to put too much faith in any greater structure. Indeed, we would expect the formal probabilities to start falling if more complicated models were considered; this will be so despite the better agreement with the data because, in accordance with Ockham's razor [9], the value of χ^2 achieved is not the only thing that matters in the integral of Eq. (10).

4. Free-form solutions

One of the most important things that the Bayesian view of probability teaches us is that the results of all analyses are conditional on the assumptions that underlie them. The highly significant evidence values displayed in Fig. 4 are meaningful in an absolute sense, therefore, only in the context of the specific models that were considered. It can be useful, then, to check whether our conclusions change very much when the strong assumptions implicit in a parametric model are relaxed. Thus we may be led to think about "free-form" solutions for the depth profile $\beta(z)$.

For the problem of Fig. 3, which involves the mixture of two known polymers with scattering length densities β_1 and β_2, it is the volume fraction profile $f(z)$, rather than $\beta(z)$, that is of greater interest; the two are related simply by

$$\beta(z) = \beta_1 + f(z)[\beta_2 - \beta_1]. \tag{11}$$

To obtain a free-form solution, we must define $f(z)$ by a large number of parameters to allow for a lot of flexibility. This in turn causes difficulties (such as having more variables than data) that require the use of "weak" prior information about $f(z)$ to ensure stability (and a unique best estimate); for example,

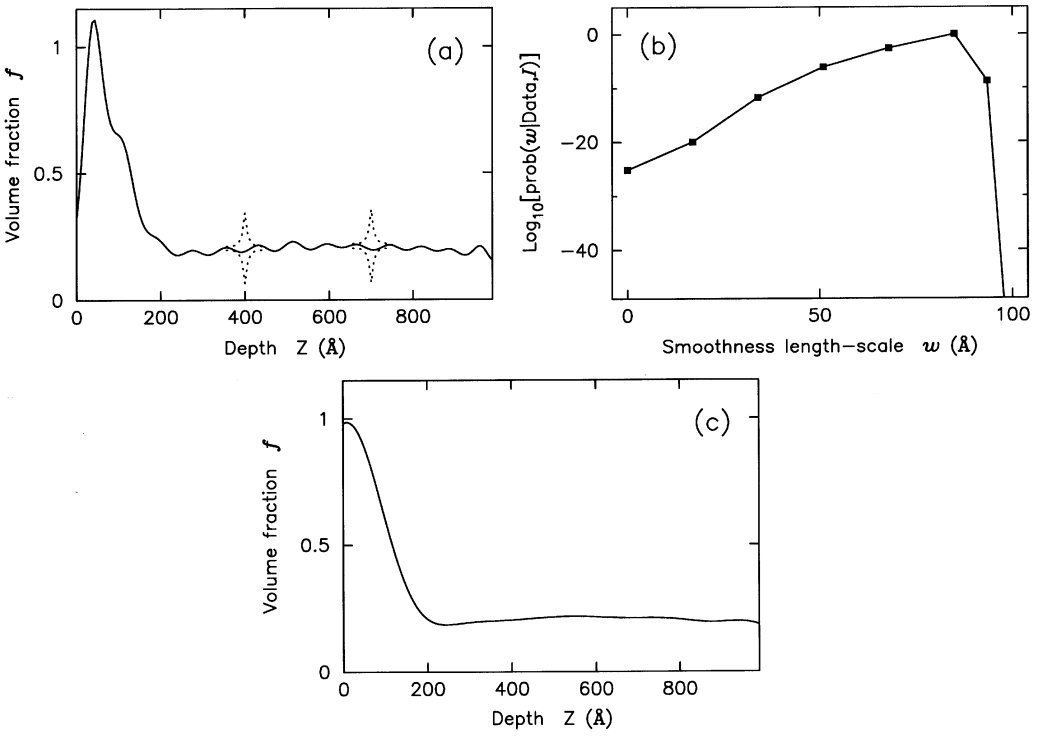

Fig. 5. The maxent estimate of the volume fraction profile $f(Z)$, given the data of Fig. 3a; the air interface is at $Z = 0$, and the silicon substrate is (way off) on the far right. The dotted lines show two error-stars, which indicate that the rippling structure is not reliable; (b) the probabilistic evidence for the smoothness length-scale w for $f(Z)$; (c) the maxent estimate of $f(Z)$ using the optimal value of $w = 85$ Å.

positivity, smoothness, bounds, and so on. A prior pdf that is appropriate for positive, and additive, distribution is

$$\text{prob}(\boldsymbol{f}|\boldsymbol{I}) \propto \exp(\alpha S), \tag{12}$$

here S is the entropy of the distribution $f(z)$ and α is a Lagrange multiplier [10]; its use with the data of Fig. 3a leads to maximum entropy (MaxEnt) estimate of Fig. 5a. While this might be the "best" solution under the circumstances, how much better is it than the alternatives? In other words, how reliable are the features in Fig. 3a? A good way of answering this question, for our non-parametric formulation, is to consider the average value of $f(z)$ over some region

$$\Phi = \frac{1}{z_2 - z_1}\int_{z_1}^{z_2} f(z)\,\mathrm{d}z = \Phi_0 \pm \sigma, \tag{13}$$

and see how the error-bar σ changes with the limits of the integration. The uncertainty in Φ will be large when z_2–z_1 is small, and vice versa, indicating that fine detail in $f(z)$ cannot be inferred reliably but gross properties can; this information is conveyed pictorially by the two error-stars in Fig. 5a. Although they correctly warn us not to take the rippling feature in the volume fraction profile at face value, can we do anything to discourage it; after all, intermixing polymers are unlikely to exhibit rapidly varying structure.

A simple extension to the traditional MaxEnt formulation that allows us to combine our prior knowledge about the positivity of $f(z)$ with a preference for its local smoothness is through the use of "fuzzy pixels" [10]. While we leave the details to Ref. [5], it will suffice to say that we introduce an additional nuisance parameter w pertaining to the length-scale of the spatial correlations. This can be

dealt with in the usual manner by marginalisation, but reduces to an optimisation for w when the pdf prob($w|\boldsymbol{D},\boldsymbol{I}$) is sharply peaked as is in Fig. 5b. Using a value of $w = 85$ Å yields the MaxEnt solution shown in Fig. 5c; not only is it aesthetically more pleasing than Fig. 5a, which is equivalent to using $w = 0$, it is overwhelmingly favoured in terms of its formal probabilistic evidence. Also of note is the fact the volume fraction is now less than unity everywhere, a known bound that was not encoded in the entropic prior. The error-stars for Fig. 5c have not been shown but are fairly small due to the enforced local smoothness. An examination of the free form solution for $f(z)$ leads to essentially the same conclusions as from the model selection analysis of the previous section.

5. Experimental design

So far, we have considered the analysis of a given set of data. If we have some choice in the matter, can we say anything about the optimal design of the experiment itself? The answer is yes: all we have to do, in essence, is to make the data most sensitive to the object of interest. In practice, however, the implementation of this simple guiding principle is far from straightforward. The problem is not just one of messy algebra, but of a rather more fundamental nature: what's best depends on the precise question being asked and, even worse, what's best often depends on the answer! To illustrate the principles, and some of the difficulties, let's outline a specific case; namely, the optimal choice of neutron scattering-length contrasts for obtaining partial structure factors [11].

Suppose that a sample is known to consist of several distinct components, with scattering lengths $b_1, b_2, \ldots, b_M$; for example, the solvent, chain and head group of a surfactant layer on a liquid substrate. Then, the depth profile $\beta(z)$ can be written as

$$\beta(z) = b_1 n_1(z) + b_2 n_2(z) + \cdots, \tag{14}$$

where the $n_j(z)$, for $j = 1$ to M, are the number-density profiles for the various components. Working within the kinematic approximation, the reflectivity data $R(Q)$ are given by

$$\frac{Q^2}{16\pi^2} R(Q) = b_1^2 h_{11}(Q) + b_2^2 h_{22}(Q) + \cdots + 2b_1 b_2 h_{12}(Q) + \cdots, \tag{15}$$

where the partial structure factors $h_{ij}(Q)$ are related to the Fourier transforms of the number–density profiles $\{\hat{n}_j(Q)\}$ by

$$h_{ij}(Q) = \mathrm{Re}\{\hat{n}_i(Q)\hat{n}_j^*(Q)\}. \tag{16}$$

The diagonal elements ($i = j$) tell us something about the size and spread of the individual components, where as the cross terms ($i \neq j$) convey information about their separations.

For any given value of the momentum transfer Q, Eq. (15) states that the partial structure factors of interest are related linearly to the data. With a suitable redefinition of the variables, therefore, the problem can be recast in the general form of a noisy set of simultaneous equations

$$\sum_{i=1}^{\mu} \boldsymbol{B}_{ji} X_i = D_j \pm \sigma_j. \tag{17}$$

Here the $\{X_i\}$ are the $\mu = M(M+1)/2$ parameters about which inferences are to be drawn, and the matrix $\boldsymbol{B}$ defines the experimental set-up for the ν measurements $\{D_j\}$. Elementary linear algebra tells us that we require at least $\nu = \mu$ data to obtain a unique solution to Eq. (17); this means that we must ascertain the reflectivity for several different sets of b's in Eq. (14), achieved for neutrons through isotopic substitution, with each choice of M values determining a row of the $\boldsymbol{B}$ matrix. The design question we are trying to answer here is which combinations of scattering-lengths are best?

Well, according to Bayes' theorem of Eq. (5), the information content of the data is encapsulated in the likelihood function; with the least-squares assignment of Eqs. (6) and (7), its logarithm $L = \log_e[\mathrm{prob}(\boldsymbol{D}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{I})]$ is given by

$$L = \mathrm{Constant} - \frac{1}{2}\sum_{j=1}^{\nu} \frac{1}{\sigma_j^2}\left(D_j - \sum_{i=1}^{N} \boldsymbol{B}_{ji} X_i\right)^2. \tag{18}$$

The congenial properties of Eqs. (17) and (18) ensure that prob($\boldsymbol{D}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{I}$) is a unimodal multivariate Gaussian, whose orientation and spread about the maximum (at $\nabla L = 0$) is governed by the

eigen-characteristics of the second-derivative Hessian matrix

$$\frac{\partial^2 L}{\partial X_i \partial X_{i'}} = -\sum_{j=1}^{v} \frac{\boldsymbol{B}_{ji}\boldsymbol{B}_{ji'}}{\sigma_j^2}, \tag{19}$$

the situation being shown graphically in Fig. 6. In a more general context, the likelihood function for $\beta(z)$ from a single set of reflectivity measurements could well be multimodal; an optimal design might then entail the collection of several data-sets to eliminate the source of the ambiguities (examples of such experiments can be found in Refs. [12–14]). For our present case, however, increasing the sensitivity of the measurements to the object of interest simply means making prob($\boldsymbol{D}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{I}$) as sharply peaked as possible.

Although the above objective is easy to state, two difficulties immediately confront us when we consider its implementation. First of all, how should we assess a potential improvement if a proposed design change makes the likelihood function narrower in one direction but wider in another? There is no real answer to this except to say that we must think more deeply about exactly what we are after, and use the eigenvalues $\{\lambda_i\}$ (and the eigenvectors) appropriately in coming to a decision. One straightforward criterion is to maximise the (square-root of the modulus of the) determinant of the $\nabla\nabla L$ matrix in Eq. (19), which is equal to the product of the λ's, since this will minimise the hyper-volume in parameter-space that gives reasonable agreement with the data. The second problem is that the error-bars $\{\sigma_j\}$ in the summation of Eq. (19) depend on the actual values of the partial structure factors that we are trying to infer! This is because for Poisson counts, such as the number of neutrons reflected, the variance of the measurements is equal to the data $\{D_j\}$ and this, in turn, is determined by the answer we seek. Short of conducting a brief experiment, or carrying out computer simulations, we must make some gross simplifying approximations to continue further analytically. The easiest assumption is to presume that all the error-bars will be roughly the same magnitude, for a given value of Q, and remove them from Eq. (19) as a proportionality constant; we could even set $\sigma_j \propto 1/\sqrt{t_j}$ if we wanted to optimise the times for the different experimental runs (with the constraint that $\Sigma t_j = T$), but omit this level of complexity here.

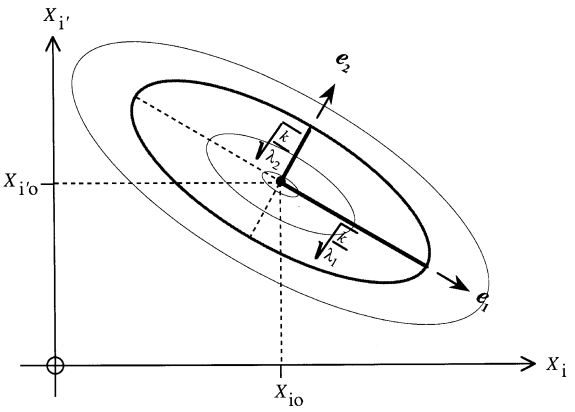

Fig. 6. A schematic illustration of the likelihood function for a linear least-squares inference problem. The principle axes and spread of the ellipsoid are given by the eigenvectors $\{\boldsymbol{e}_i\}$ and eigenvalues $\{\lambda_i\}$ of (minus) the second-derivative Hessian matrix $\nabla\nabla L$.

Finally, let us consider the case of a two-component system as the simplest non-trivial example. Suppose that there is a monolayer of a surfactant with a dominant chain, and a negligible head group, on top of a solvent substrate. Then, depending on whether the chain is hydrogenated (H) or deuterated (D), the neutron scattering-length c (or b_1) can be either -1.0 or $24.0\,\text{Å}$ ($\times 10^{-4}$); with suitable H/D mixing, the solvent contribution s (or b_2) can lie anywhere in the range -0.17–$1.91\,\text{Å}$ ($\times 10^{-4}$). Since $M = 2$, there will be $\mu = 3$ $\{X_i\}$ of interest: namely, the partial structure factors h_{cc}, h_{ss} and h_{cs}. Using the minimum number $v = 3$ of data-sets required for a solution to Eq. (17), our quality factor of the "square-root of the modulus of the determinant of the $\nabla\nabla L$ matrix in Eq. (19)" reduces to $|\det(\boldsymbol{B})|$, where

$$\boldsymbol{B} \propto \begin{pmatrix} c_1^2 & s_1^2 & c_1 s_1 \\ c_2^2 & s_2^2 & c_2 s_2 \\ c_3^2 & s_3^2 & c_3 s_3 \end{pmatrix}. \tag{20}$$

With only the H or D options for the chain, there are just four possibilities for the three c's to be

Table 1
The four combinations of hydrogenated (H) and deuterated (D) surfactant chains, for the three partial structure-factor data sets, and the corresponding optimal scattering-lengths of the solvent (achieved through H/D mixing)

Chain	Solvent			Det(***B***)
H H H	−0.17	0.87	1.91	2
D H H	1.91	1.91	0.93	1134
D D H	−0.17	1.91	1.91	108 860
D D D	−0.17	0.87	1.91	31 099

investigated: (H, H, H), (D, H, H), (D, D, H) and (D, D, D). For each of these cases, the optimal choice of the solvent combination (s_1, s_2, s_3) , and resultant maximum value of $|\det(\boldsymbol{B})|$, is given in Table 1. This indicates that it is well worth the effort of deuterating the chain, with the best triplet of c and s being (D, −0.17), (D, 1.91) and (H, 1.91). We must remember, of course, that this prediction is based on at least one gross approximation ($\sigma_j = \sigma$), so that the reality may be a little different; nevertheless, we do not expect to be too far wrong. To test our proposed optimal set of scattering-length contrasts, we simulated one triplet of reflectivity data under our preferred option and another using only a deuterated chain (the second best alternative); these are shown in Fig. 7a. The analysis of these computer-generated measurements, to obtain the three partial structure factors, is also shown in Fig. 7 and confirms that our prediction of the optimal choice was correct.

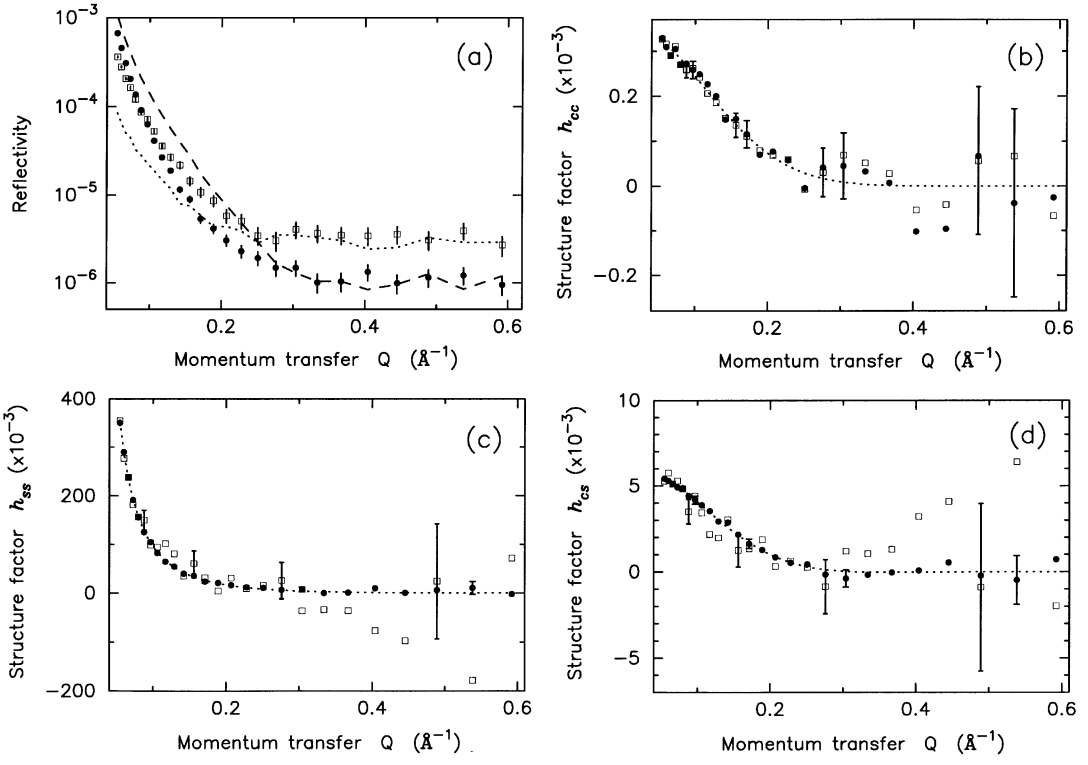

Fig. 7. (a) Simulated neutron reflectivity data: the dotted and dashed lines are both for deuterated surfactant chains, but on H_2O and D_2O respectively; the solid circles (●) are from a hydrogenated chain on D_2O, and the open squares (□) are from a deuterated chain on a 50 : 50 mixture of H_2O&D_2O; (b) and (c) the inferred partial structure-factors h_{cc}, h_{ss} and h_{cs}, and their $1-\sigma$ error-bars (open-square on the left and solid circle on the right); the dotted lines show the answer used to generate the simulated data sets.

6. Conclusions

Bayesian probability theory gives a logical and unified approach to data analysis: it provides the justification for many conventional statistical procedures, and gives improved prescriptions when they fail. We have tried to illustrate this point with a broad spectrum of examples from reflectivity data analysis, ranging from elementary parameter estimation to optimal experimental design, and hope that the reader will come to share Laplace's view that "probability theory is nothing but common sense reduced to calculation".

References

[1] J. Bernoulli, Ars conjectandi (Thurnisiorum, Basel, 1713).
[2] T. Bayes, Phil. Trans. Roy. Soc. London 53 (1763) 370.
[3] P.S. Laplace, Theorie Analytique des Probabilités, Courcier, Paris, 1812.
[4] R.T. Cox, Am. J. Phys. 14 (1946) 1.
[5] D.S. Sivia, Data Analysis: a Bayesian tutorial, Oxford Univ. Press, Oxford, 1996.
[6] J.J.K. O. Ruanaidh, W.J. Fitzgerald, Numerical Bayesian Methods Applied to Signal Processing, Springer, London, 1996.
[7] M. Geoghegan, R.A.L. Jones, D.S. Sivia, J. Penfold, A.S. Clough, Phys. Rev. E 53 (1996) 825.
[8] D.S. Sivia, W.A. Hamilton, G.S. Smith, Physica B 173 (1991) 121.
[9] W.M. Thorburn, Mind 27 (1918) 345.
[10] J. Skilling, Maximum Entropy, In: B. Buck, V.A. Macaulay (Eds.), Action, Oxford University Press, Oxford, 1991.
[11] E.A. Simister, E.M. Lee, R.K. Thomas, J. Penfold, J. Phys. Chem. 96 (1992) 1373. Also see Thomas et al. in these Proceedings (SXNS-5), Physica B 248 (1998).
[12] D.S. Sivia, W.A. Hamilton, G.S. Smith, T.P. Rieker, R. Pynn, J. Appl. Phys. 70 (1991) 732.
[13] V.O. deHaan, A.A. vanWell, S. Adenwalla, G.P. Felcher, Phys. Rev. B 15 (1995) 10830.
[14] C.F. Majkrzak, N.F. Berk, Phys. Rev. B 15 (1995) 10 827. Also see their paper in these Proceedings (SXNS-5), Physica B 248 (1998).