electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898

Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering

Steen Hansen

Journal of
Applied Crystallography
ISSN 0021-8898

Received 3 July 2000
Accepted 21 September 2000

Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering

Steen Hansen

Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 FRB C, Denmark. Correspondence e-mail: slh@kvl.dk

Bayesian analysis is applied to the problem of estimation of hyperparameters, which are necessary for indirect Fourier transformation of small-angle scattering data. The hyperparameters most frequently needed are the overall noise level of the experiment and the maximum dimension of the scatterer. Bayesian methods allow the posterior probability distribution for the hyperparameters to be determined, making it possible to calculate the distance distribution function of interest as the weighted mean of all possible solutions to the indirect transformation problem. Consequently no choice of hyperparameters has to be made. The applicability of the method is demonstrated using simulated as well as real experimental data.

1. Introduction

For interpretation of the experimental results from small-angle scattering (SAS), it is often relevant to represent the scattering data in direct space, which requires a Fourier transformation of the data. This transformation preserves the full information content of the experimental data, but a direct Fourier transform is of limited use because of noise, smearing and truncation. Attempts to take these effects into account by indirect Fourier transformation (IFT) have been described by *e.g.* Glatter (1977), Moore (1980), Müller & Glatter (1982), Svergun *et al.* (1988), Hansen & Pedersen (1991) and Svergun (1992). IFT in SAS is an underdetermined problem, several solutions of which may fit the data adequately. Consequently, some principle for choosing one of the many possible solutions must be used if a single representation is wanted. The most frequently used principle (Glatter, 1977) imposes a smoothness criterion upon the distribution to be estimated, giving higher preference to smoother solutions. This is in good agreement with the prior knowledge that most SAS experiments have very low resolution. Consequently, this method has demonstrated its usefulness for analysis of SAS data for more than two decades.

For application of the original method of Glatter (1977), it is necessary to choose (i) a number of basis functions (usually of the order of 20–40), (ii) a regularizing or 'smoothing' parameter, as well as (iii) a maximum length in direct space. Various *ad hoc* guidelines as to how to make these choices have been provided.

As a result of improved computing facilities, with respect to both hardware and software, the restriction on the number of basis functions must now be considered obsolete. Using the algorithm described by Steenstrup (1985), as performed for the calculations presented in this paper, it is easily possible to use 1000 points for the estimation of the distance distribution function, requiring just a few seconds of CPU time. This number of points corresponds to a resolution that is much better than that obtained by small-angle scattering. However, the choices of smoothing parameter and maximum length for the calculation still present occasional problems for the applicability of the indirect transformation method. Other methods for IFT in SAS face similar problems with regard to the determination of hyperparameters (see below). As the method of Glatter (1977) is by far the one most frequently used for IFT in SAS, it seems most relevant to use Glatter's method as the main example when demonstrating the applicability of the Bayesian method for the selection of hyperparameters.

Estimation of the smoothing parameter corresponds to estimation of the noise level of the experiment. This may not be of great interest and consequently this parameter may be integrated out of the analysis. However, the maximum diameter of the scatterer is clearly very often a structural parameter of interest. As such, methods other than Glatter's IFT have been suggested for the estimation of the largest particle dimension (*e.g.* Müller *et al.*, 1980). By the Bayesian method suggested here, a two-dimensional probability distribution for the hyperparameters is calculated from the basic rules of probability theory. From this distribution it is possible to make unique choices for the hyperparameters as well as reliable error estimates. However, an even more appealing approach, and the method which has been used for the examples presented in the present paper, is to use all the calculated distance distribution functions weighted by their appropriate probabilities. This method provides a mean distribution as well as a reliable error estimate, including contributions from the uncertainty in the hyperparameters.

Using the Bayesian method, the only choice left to make is which method of regularization to use. But this too can be determined by integration over all the hyperparameters of

each method of regularization, allowing the posterior probability for each method to be evaluated. If prior knowledge about the scatterer is available, this should be included in the calculation.

2. Theory

2.1. Small-angle scattering

In small-angle scattering, the intensity I is measured as a function of the length of the scattering vector $q = 4\pi\sin(\theta)/\lambda$, where λ is the wavelength of the radiation and θ is half the scattering angle. For scattering from a dilute solution of monodisperse molecules of maximum dimension d, the intensity can be written in terms of the distance distribution function $p(r)$ (see Glatter, 1982):

$$I(q) = 4\pi \int_0^d p(r)\sin(qr)/qr\,\mathrm{d}r. \tag{1}$$

Approximating the distance distribution function $p(r)$ by $\mathbf{p} = (p_1, \ldots, p_N)$ and measuring the intensity at

$$I(q_i) = \sum_{j=1}^{N} A_{ij}p_j + e_i, \tag{2}$$

where e_i is the noise at data point i and the matrix A is given by $A_{ij} = 4\pi\Delta r\sin(q_i r_j)/(q_i r_j)$, where $\Delta r = r_j - r_{j-1}$, the aim of the indirect Fourier transformation is to restore $\mathbf{p}$, which, by virtue of the Fourier transform, contains the full information present in the scattering profile. The distance distribution function is related to the density–density correlation $\gamma(r)$ of the scattering length density $\rho(\mathbf{r})$ by

$$p(r) = r^2\gamma(r) = r^2\left\langle \int_V \rho(\mathbf{r}')\rho(\mathbf{r}+\mathbf{r}')\,\mathrm{d}\mathbf{r}'\right\rangle, \tag{3}$$

where $\rho(\mathbf{r})$ is the scattering contrast, given by the difference in scattering density between the scatterer $\rho_{\rm sc}(\mathbf{r})$ and the solvent $\rho_{\rm so}$, *i.e.* $\rho(\mathbf{r}) = \rho_{\rm sc}(\mathbf{r}) - \rho_{\rm sc}$; the enclosing angle brackets indicate averaging over all orientations of the molecule and V is the volume of the molecule. For uniform scattering density of the molecule, the distance distribution function is proportional to the probability distribution for the distance between two arbitrary scattering points within the molecule. For non-uniform scattering density, the distance distribution may have negative regions (if the scattering density of some region of the scatterer is less than the scattering density of the solvent). High concentrations may also give negative regions in the distance distribution function, around the maximum size of the scatterer (see *e.g.* Glatter, 1982).

In the case of polydispersivity, the size distribution for the molecules can be calculated from the scattering profile by a similar IFT if the shape of the molecule is known (usually spheres are assumed). The mathematics of this problem is completely analogous to the problem of estimation of the distance distribution function for monodisperse systems (but at least the physics here does not allow the distribution to become negative). The actual calculations may cause additional difficulties as the system of equations to be solved for calculation of a size distribution is often more ill-conditioned than for the determination of the distance distribution function (as will be evident from a singular value decomposition of the transformation matrix). However, the numerical problems encountered are similar.

2.2. IFT in SAS

The most frequently used method for IFT in SAS is that of Glatter (1977). Described briefly below, this method is very similar to the method of Tikhonov & Arsenin (1977), who used a more general expression for the regularization of a distribution. Tikhonov & Arsenin estimated a distribution $\mathbf{f} = (f_1, \ldots, f_N)$ by minimizing the expression

$$\alpha\Omega(\mathbf{f}, \mathbf{m}, \rho) + \chi^2, \tag{4}$$

where α is a Lagrange multiplier, which is found by allowing the χ^2 to reach a predetermined value, and where the regularizer is given by the general expression

$$\Omega(\mathbf{f}, \mathbf{m}, \rho) = \|\,\mathbf{f} - \mathbf{m}\,\|^2 + \rho\,\|\,\mathbf{f}'\,\|^2, \tag{5}$$

where the prime indicates a derivative (first and/or higher) of $\mathbf{f}$ and ρ determines the relative weight of the two terms. The first term minimizes the deviation of $\mathbf{f}$ from the prior estimate $\mathbf{m} = (m_1, \ldots, m_N)$ with respect to a given norm, and the second term imposes a smoothness constraint on the distribution to be estimated. The method of Glatter only uses the last term in (5). Using only the first term in (5) will give a regularization similar to that of the maximum-entropy method. The norm is then to be replaced by the Shannon entropy, which measures the distance between two distributions $\mathbf{f}$ and $\mathbf{m}$ (Kullback, 1959). It may be noted that a second-order Taylor expansion of the expression for the entropy will give the quadratic constraint $\sum_{j=1}^{N}(f_j - m_j)^2/2m_j$.

The χ^2 is defined in the conventional manner, *i.e.*

$$\chi^2 = \sum_{i=1}^{M}[I_m(q_i) - I(q_i)]^2/\sigma_i^2, \tag{6}$$

where $I_m(q_i)$ is the measured intensity and σ_i is the standard deviation of the Gaussian noise at data point i. The smoothness constraint can be expressed by writing the distance distribution function as a sum of smooth basis functions, *e.g.* cubic B-splines: $p(r) = \sum_{j=1}^{K} a_j B_j(r)$, where the sum $A = \sum_{j=1}^{K-1}(a_{j+1} - a_j)^2$ is then to be minimized subject to the constraint that the χ^2 takes some sensible value (Glatter, 1977). This problem leaves two parameters to be determined: the maximum diameter used, d, and the Lagrange multiplier, α, which determines the relative weighting of the constraints from the data and the smoothness, respectively.

Using this method, the number of basis functions K is chosen as sufficiently large to accommodate the structure in the data.

The smoothing parameter (or noise level) is found by the 'point of inflexion' method by plotting A and the χ^2 as a function of the Lagrange multiplier α. Using this method, a plateau in A has to be found when χ^2 has reached a low value; this region determines the correct noise level. A problem with

the point of inflexion method is that the aforementioned plateaus may not exist. Furthermore, when they do exist, the point on the plateau may not be uniquely determined.

2.3. Bayesian analysis

To incorporate IFT within a Bayesian framework, the regularization constraints ('smoothness', 'maxent', *etc.*) are considered to be 'models' and the parameters of the models determine the distribution of interest. The hyperparameters can be considered to be part of the model. Using Gaussian approximations around the maximum probability for the parameters in each model, the total probability of each model can be calculated by integration over the parameters, as shown below. For an introduction to Bayesian methods see *e.g.* MacKay (1992). When applied to the problem of inferring which model or hypothesis H_i is most plausible after data d have been measured, for this posterior probability $P(H_i|D)$, Bayes' theorem gives

$$P(H_i \mid D) = P(D \mid H_i)\, P(H_i)/P(D), \tag{7}$$

where $P(D|H_i)$ is the probability of the data d assuming that the model H_i is correct; $P(H_i)$ denotes the prior probability for the model H_i, which is assumed constant for 'reasonable' hypotheses (*i.e.* different reasonable hypotheses should not be assigned different prior probabilities), and $P(D)$ is the probability for measuring the data, but this just gives rise to a renormalization constant after the data have been measured. The *evidence* $P(D|H_i)$ for a hypothesis can be found by integrating over parameters $\mathbf{f} = (f_1, \ldots, f_N)$ in the model H_i, and

$$P(H_i \mid D) \propto P(D \mid H_i) = \int P(D, \mathbf{f} \mid H_i)\, \mathrm{d}^N f, \tag{8}$$

N being the number of parameters in the model. Again using Bayes' theorem,

$$P(D \mid H_i) = \int P(D \mid H_i, \mathbf{f})\, P(\mathbf{f} \mid H_i)\, \mathrm{d}^N f, \tag{9}$$

where the likelihood is written

$$P(D \mid H_i, \mathbf{f}) = \exp(-L)/Z_L \tag{10}$$

with

$$Z_L = \int \exp(-L)\, \mathrm{d}^M D, \tag{11}$$

M being the number of data points. For the usual case of Gaussian errors, $L = \chi^2/2$ and $Z_L = \prod (2\pi\sigma_i^2)^{-1/2}$, where σ_i is the standard deviation of the Gaussian noise at data point i.

Correspondingly, it is now assumed that the prior probability for the distribution $\mathbf{f}$ can be expressed through some functional S (to be chosen) and written

$$P(\mathbf{f} \mid H_i) = \exp(\alpha S)/Z_S \tag{12}$$

with

$$Z_S = \int \exp(\alpha S)\, \mathrm{d}^N f. \tag{13}$$

According to this expression, the model H_i is the hypothesis that the prior probability for the distribution of interest $\mathbf{f}$ can be written as above with some (regularizing) functional S and a parameter α that determines the 'strength' of the prior (through S) relative to the data (through χ^2). Both the functional form of S as well as the value for the parameter α are then part of the hypothesis and are to be determined subsequently .

Inserting (10) and (12) into (9) and writing $Q = \alpha S - \chi^2/2$, the evidence is given by

$$\begin{aligned} P(D \mid H_i) &= \left[\int \exp(\alpha S)\, \exp(-\chi^2/2)\, \mathrm{d}^N f \right] / Z_S Z_L \\ &= \left[\int \exp(Q)\, \mathrm{d}^N f \right] / Z_S Z_L. \end{aligned} \tag{14}$$

Using Gaussian approximations for the integrals, expanding Q around the maximum in $\mathbf{f}$, and writing $\mathbf{A} = -\nabla\nabla S$, $\mathbf{B} = \nabla\nabla\chi^2/2$ evaluated at the maximum value of $Q(\mathbf{f}_0) = Q_0$, where $\nabla Q = 0$, we obtain

$$P(D \mid H_i) = \frac{(2\pi)^{-N/2} \exp(Q_0) \det^{-1/2}(\alpha\mathbf{A} + \mathbf{B})}{(2\pi)^{-N/2} \exp(S_{\max}) \det^{-1/2}(\alpha\mathbf{A}) \prod (2\pi\sigma_i^2)^{-1/2}}, \tag{15}$$

where $S_{\max}$ is the maximum for the functional S. Usually $S_{\max} = 0$, which will be assumed in the following (otherwise just a renormalizing constant is left). Furthermore, the term $\prod (2\pi\sigma_i^2)^{-1/2}$ from the experimental error is redundant for comparison of different hypotheses and is also left out in the following. The probability for different hypotheses each being equally probable *a priori* can then be calculated from the expressions

$$P(D \mid H_i) = \det^{1/2}(\mathbf{A}) \exp(Q_0) \det^{-1/2}(\mathbf{A} + \alpha^{-1}\mathbf{B}) \tag{16}$$

and

$$\begin{aligned} \log P(D \mid H_i) = {} & \tfrac{1}{2} \log\det(\mathbf{A}) + \alpha S_0 - \chi_0^2/2 \\ & - \tfrac{1}{2} \log[\det(\mathbf{A} + \alpha^{-1}\mathbf{B})] \end{aligned} \tag{17}$$

Previously, any hyperparameters were implicitly included in the models or hypotheses. Now writing H_i for the model *without* the hyperparameters α and d, we obtain from Bayes' theorem that the posterior probability is determined by

$$P(D, \alpha, d \mid H_i) = P(D \mid \alpha, d, H_i)\, P(\alpha)\, P(d), \tag{18}$$

where $P(\alpha)$ is the prior probability for α and $P(d)$ is the prior probability for d. Assuming α to be a scale parameter (see Jaynes, 1983), it is found that $P(\alpha) = \alpha^{-1}$ should be used for the prior probability of α. For a parameter d which is usually known *a priori* within an order of magnitude, the prior probability should be uniform within the allowed interval.

2.3.1. Application to SAS. For application of the Bayesian method to small-angle scattering, the regularizing functional has to be chosen. For smoothness regularization, several forms for S exist. For the present calculations, the expression $S = \int f''(x)^2\, \mathrm{d}x$ has been selected, which takes the discrete form $S = \sum (f_{j+1} - 2f_j + f_{j-1})^2 (\Delta x)^{-3}$. With the regularization functional

$$S = \sum_{j=2}^{N-1} [f_j - (f_{j-1} + f_{j+1})/2]^2 + \tfrac{1}{2} f_1^2 + \tfrac{1}{2} f_N^2, \tag{19}$$

a constraint similar to Glatter's original smoothness constraint is obtained. Consequently (19) and Glatter's constraint often give almost identical results. Note that (19) can also be written $S = \sum (f_j - m_j)^2$, with $m_j = (f_{j-1} + f_{j+1})/2$. Adding the terms $\frac{1}{2} f_1^2$

and $\frac{1}{2}f_N^2$ to the sum (corresponding to the condition that the distribution has fixed end points $f_0 = f_{N+1} = 0$) leads to the matrix elements for **A**: $a_{ij} = 1$ for $i = j$, $-\frac{1}{2}$ for $|i - j| = 1$ and 0 elsewhere. As $\det\mathbf{A} = (1/2)^N(N + 1)$, for smoothness regularization we obtain

$$P(D \mid H_i, \alpha, d) = [(1/2)^N(N+1)]^{1/2} \frac{\exp(\alpha S_0 - \chi_0^2/2)}{\det^{1/2}(\mathbf{A} + \alpha^{-1}\mathbf{B})}. \quad (20)$$

Regularizing by the maximum-entropy method, S takes the form

$$S = \sum_{j=1}^{N} -f_j \ln(f_j/m_j) + f_j - m_j, \quad (21)$$

where **m** is a prior estimate of **f** (Skilling, 1988). Using the expression for the entropy metric (Skilling, 1988), $ZS = \int\int\prod f_j^{-1/2} \exp(\alpha S)\, \mathrm{d}^N f$ and the corresponding expressions for the matrices **A** and **B** are $\mathbf{A} = [f^{1/2}](-\nabla\nabla S)[f^{1/2}]$ (writing $[f] = \mathrm{diag}\{f\}$) and $\mathbf{B} = [f^{1/2}](\nabla\nabla\chi^2/2)[f^{1/2}]$. As $\nabla\nabla S = -f^{-1}$, $\mathbf{A} = \mathbf{I}$, where **I** is the identity matrix, and $\det^{-1/2}(\mathbf{A}) = 1$.

It is apparent from a second-order expansion of the expression for the entropy that the maximum-entropy method in a flat metric using the 'prior' $m_j = (f_{j-1} + f_{j+1})/2$ is equivalent to the smoothness regularization.

For an application of Bayesian methods to the analysis of dynamic light scattering data using maximum-entropy regularization see work by Bryan (1990).

3. Results

The first simulated example, shown in Fig. 1, was taken from May & Nowotny (1989) and has been analysed previously with the purpose of comparing methods for IFT in SAS (Hansen & Pedersen, 1991; Hansen & Müller, 1996). The original distance distribution function as well as the distance distribution functions estimated by the methods described above [regularizing using (19) and (21), respectively] are shown in Fig. 2. The maximum-entropy calculation was performed using a spherical prior. The estimated distributions shown have been calculated as means, weighted according to their posterior probabilities (shown in Fig. 4) using (18).

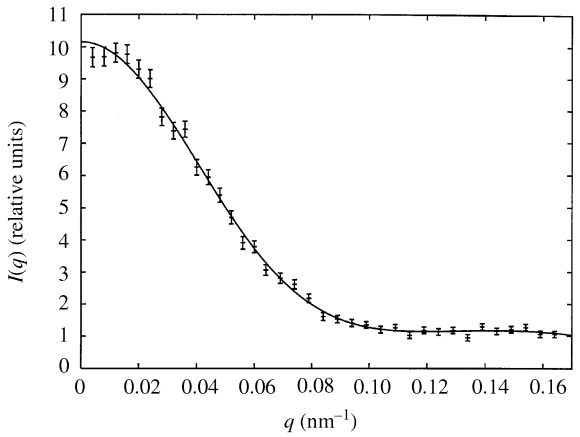

Figure 1
Simulated data and fit to data.

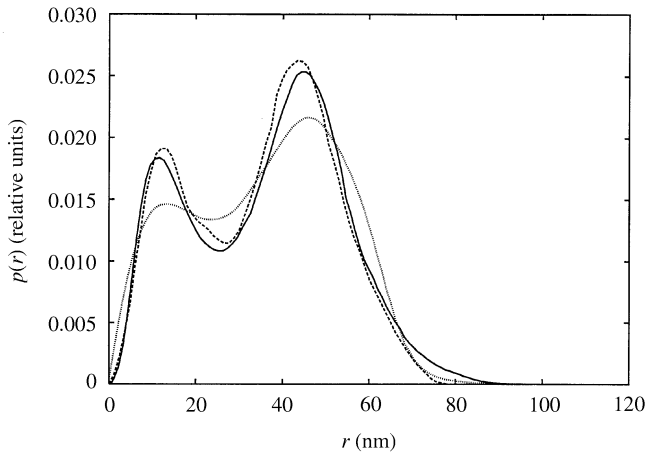

Figure 2
Distance distribution functions corresponding to data in Fig. 1. Full line: maximum-entropy calculation. Dashes: original distribution. Dots: calculation using smoothness regularization.

It may be noted that the result presented in Fig. 2 is slightly better than that reported by Hansen & Pedersen (1991), probably as a result of lifting the restriction of having to construct the distance distribution function from a relatively small number of cubic *B*-splines.

For the calculations represented in Fig. 2, the maximum diameter d was varied between 56 and 104 nm. Extending the calculations beyond these limits did not change the result.

Fig. 3 is the conventional plot used to find the 'point of inflexion', which is traditionally used to determine the noise level of the experiment. The curve showing χ^2 as a function of α has a large plateau for decreasing α. The correct value is $\chi^2 = 1.0$. The curve showing S has a negative slope for all α values, leaving a range of possible α to be chosen for the 'best' solution.

Using the Bayesian method for estimation of α and d, the probability $P(D, \alpha, d|H_i)$ (the evidence), shown in Fig. 4, is found. Fig. 4 displays a clear maximum, making it possible to select the most likely set of hyperparameters (α, d).

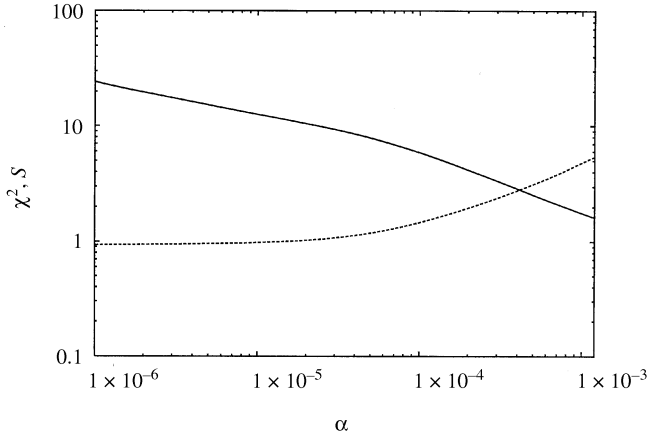

Figure 3
χ^2 (dashed) and S (full line, rescaled for clarity) calculated as a function of the Lagrange multiplier α using the correct maximum diameter.

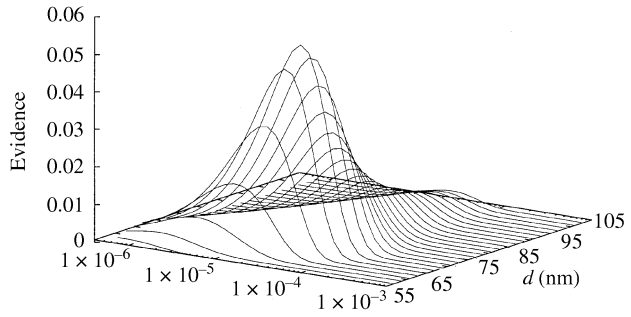

Figure 4
Evidence for α and d calculated from the simulated data shown in Fig. 1 using smoothness regularization. The maximum is at $\alpha = 3 \times 10^{-5}$ and d = 68 nm.

For comparison with Fig. 3, one-dimensional distributions for α and d are shown in Fig. 5. Correspondingly, the one-dimensional distribution for d is shown in Fig. 6. Fig. 7 represents a comparison of χ^2 and α.

Using the maximum-entropy method and a spherical prior gives the evidence for α and d shown in Fig. 8.

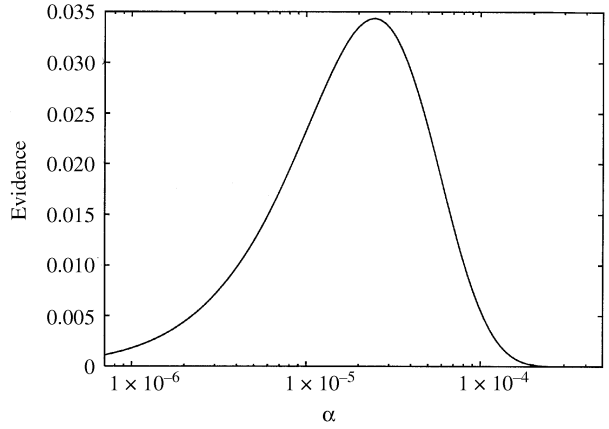

Figure 5
The evidence as a function of α using smoothness regularization and d = 74 nm.

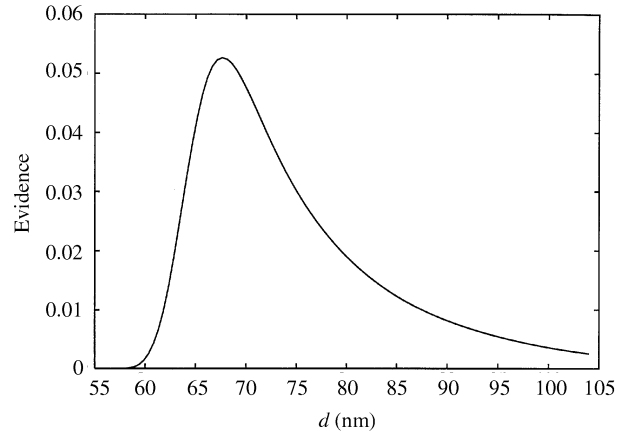

Figure 6
The evidence as a function of d using smoothness regularization and α = 2×10^{-5}.

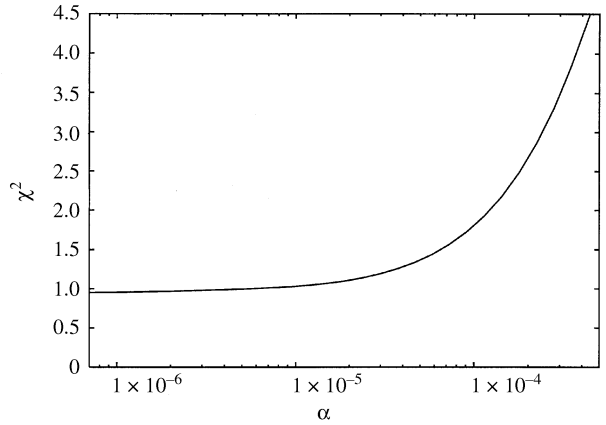

Figure 7
χ^2 as a function of α calculated for d = 74 nm.

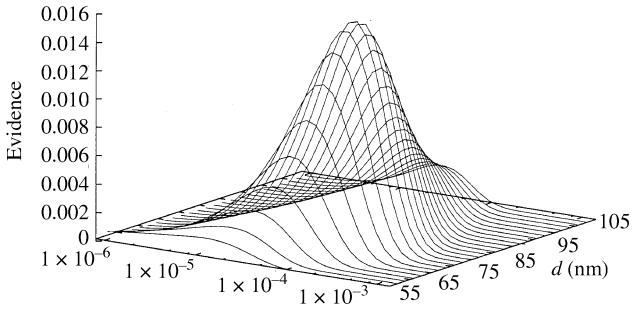

Figure 8
Evidence for α and d calculated from the simulated data shown in Fig. 1 using maximum entropy. The maximum is at $\alpha = 8 \times 10^{-5}$ and d = 75 nm.

For this example, 36 data points were used, with approximately 100 points for the estimation of $p(r)$. For each value of (α, d), $p(r)$ was calculated, requiring less than 1 s of CPU time on an ordinary PC (400 MHz Pentium processor). The speed of the algorithm can be increased significantly if necessary; however, the determination of the two-dimensional probability distribution for α and d described here only took a few minutes of CPU time.

The results obtained for a typical experimental example, using SANS measurements on casein micelles, are represented in Figs. 9–11 [scattering data were taken from Hansen *et al.* (1996)]. Fig. 9 shows the experimental data, including error bars determined by counting statistics (error bars are so small that they hardly show up in the figure). The fit corresponding to the average distance distribution function is also shown. In Fig. 10, the distance distribution function calculated using the smoothness constraint (19) is shown, while Fig. 11 presents the evidence as a function of α and d.

4. Discussion

From the result shown in Fig. 2, it is obvious that the smoothness constraint has the effect of broadening the peaks in the estimated distribution, as should be expected. In spite of this deviation, the maximum dimension for the scatterer as well as the noise level are determined with good accuracy.

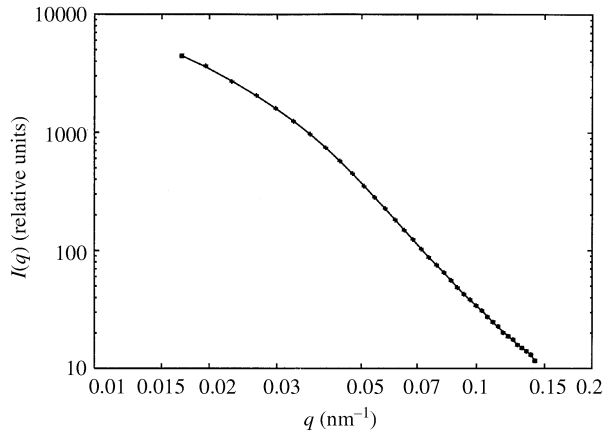

Figure 9
SANS on casein micelles and fit to data.

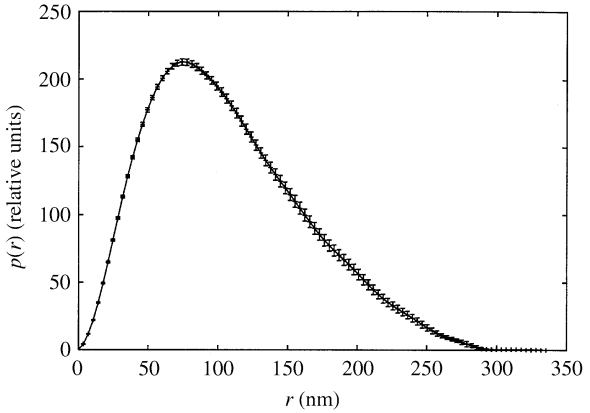

Figure 10
Estimate of the distance distribution function for the casein data.

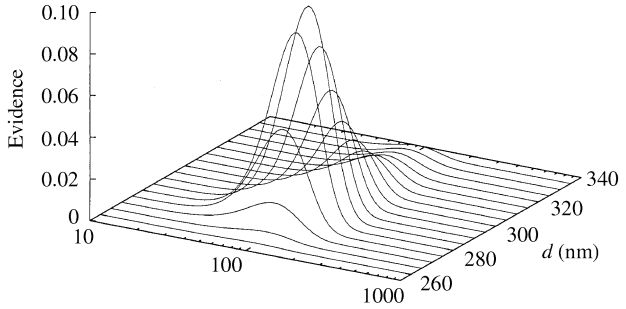

Figure 11
Evidence for α and d from the data in Fig. 10.

From the simulated distance distribution function, the maximum diameter of the scatterer is about 77 nm. For the smoothness constraint, the maximum probability is found for a diameter of 68 nm, but the tail of the probability distribution leads to an average of 74 nm for the maximum diameter. At the maximum probability, the value for the regularization parameter α is 3×10^{-5}. The corresponding numbers using the maximum-entropy regularization are 75 and 78 nm for the diameter and 8×10^{-5} for the regularization parameter, which is also in good agreement with the original distance distribution function. The estimated distance distribution functions have slightly higher values for the maximum diameter, but the distributions only approach zero very slowly in this region, which makes it difficult to point out which values should be quoted.

For the determination of the noise level, the smoothness constraint gives an estimate that is about 10–15% larger than the correct value. This is to be expected as the distance distribution function contains relatively sharp features, which do not comply with the smoothness constraint. The maximum-entropy method leads to an estimate that is about 2–3% below the correct level. However, as mentioned previously, the noise level is rarely of interest on its own and for the present calculations all noise levels are included in the analysis and contained in the final result.

From the probability distributions it is possible to give error estimates for the hyperparameters. The probability distribution for the maximum diameter is skew with a full width at half-maximum (FWHM) value of 13 nm. For the corresponding maximum-entropy estimate, the FWHM is 19 nm.

Note that in spite of the relatively large variation in diameters and noise levels used, the posterior probabilities ensure that small and large values for the hyperparameters are assigned relatively small weights, making the average $p(r)$ well behaved.

The situation for the experimental data is similar to that for the simulated data. The posterior distribution for α and d in Fig. 11 has a clear and well defined maximum and the average distance distribution in Fig. 10 appears to be free from artifacts, although the data are fitted quite closely. The fact that this example is polydisperse makes the determination of the maximum dimension even more difficult. However, the maximum dimension calculated from Figs. 10 and 11 is in very good agreement with that obtained from model fitting using the knowledge that the casein micelles are polydisperse spherical particles. From the data, it may be noted that $q_{\min}/\pi \simeq 200$ nm, which means that the shape of the distance distribution function beyond 200 nm is increasingly influenced by the regularization term.

The error bars on $p(r)$ have been calculated in the conventional manner from the posterior probability distributions evaluated at each r.

By comparison of the different methods for selection of hyperparameters, it is evident that the Bayesian method provides clearer and correct answers to the questions posed. The use of the original *ad hoc* method for the determination of the noise level and the maximum diameter of the scatterer will leave the user to choose values from relatively wide plateaus. The Bayesian method facilitates the use of IFT in SAS as the task of selecting the relevant parameters to be used can now be left completely to probability theory (and hence to a computer). Furthermore, it is not necessary to restrict analysis to the most likely distance distribution; deduced parameters, such as the radius of gyration, can be found by integration over all distributions weighted by their probabilities.

After integration of the posterior probabilities over all the hyperparameters, a 'total' probability for each method for IFT

can be obtained. This allows different methods for IFT to be compared from a probabilistic point of view. At this point it might be taken into consideration that various experimental situations can and usually will correspond to different states of prior knowledge, and as a consequence different prior probabilities for the hypotheses $P(H_i)$ in (7) should be introduced. In a situation where information about the scatterer is available from electron micrographs, it would probably be more sensible to use the maximum-entropy method with a prior constructed from the micrographs than simply to use a smoothness constraint without regard to the prior information. However, it should be noted that calculations have indicated that in general, without any form of prior knowledge about the scatterer, a smoothness constraint is more likely to give a reliable result than the (conventional) maximum-entropy method using *e.g.* a simple spherical prior. This agrees well with the intuitive notion that the more restrictive a model is, the worse it is. Having to choose between two equally 'complicated' models (*e.g.* in the sense of having equally many parameters and each model fitting the data equally well at the most likely choice of parameters, *i.e.* having identical likelihoods) the one should be chosen which has the largest range of parameters fitting the data adequately (largest accessible volume of hypothesis space). In other words, the model to be preferred is characterized by the property that the measurement of the data leads to the least reduction in the accessible volume of hypothesis space.

The program used for the calculations presented here is available from the author. It is written in Fortran and includes corrections for experimental smearing as described by Pedersen *et al.* (1990).

5. Conclusions

Bayesian methods have been applied to the problem of estimation of hyperparameters used for IFT in SAS. It has been demonstrated that the distance distribution of interest can be found by integrating out the hyperparameters using the posterior probability distribution for the hyperparameters.

References

Bryan, R. K. (1990). *Eur. Biophys. J.* **18**, 165–174.

Hansen, S., Bauer, R., Lomholt, S. B., Quist, K. B., Pedersen, J. S. & Mortensen, K. (1996). *Eur. Biophys. J.* **24**, 143–147.

Hansen, S. & Müller, J. J. (1996). *Maximum Entropy and Bayesian Methods, Cambridge, 1994,* edited by J. Skilling, pp. 69–78. Dordrecht: Kluwer Academic Publishers.

Hansen, S. & Pedersen, J. S. (1991). *J. Appl. Cryst.* **24**, 541–548.

Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.

Glatter, O. (1982). *Small Angle X-ray Scattering*, edited by O. Glatter & O. Kratky. London: Academic Press.

Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz. Dordrecht: Kluwer Academic Publishers.

Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.

MacKay, D. J. C. (1992). *Maximum Entropy and Bayesian Methods, Seattle, 1991,* edited by C. R. Smith, G. J. Erickson & P. O. Neudorfer, pp. 39–66. Dordrect: Kluwer Academic Publishers.

May, R. P. & Nowotny, V. (1989). *J. Appl. Cryst.* **22**, 231–237.

Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.

Müller, J. J., Schmidt, P. W. & Damaschun, G. (1980). *J. Appl. Cryst.* **27**, 257–270.

Müller, K. & Glatter, O. (1982). *Makromol. Chem.* **183**, 465–479.

Pedersen, J. S., Posselt, D. & Mortensen, K. (1990). *J. Appl. Cryst.* **23**, 321–333.

Skilling, J. (1988). *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, edited by G. J. Erickson & C. R. Smith, pp. 173–187. Dordrecht: Kluwer Academic Publishers.

Steenstrup, S. (1985). *Aust J. Phys.* **38**, 319–327.

Svergun, D. I. (1992). *J. Appl. Cryst*. **25**, 495–503.

Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (1988). *Acta Cryst.* A**44**, 244–250.

Tikhonov, A. N. & Arsenin, V. Ya. (1977). *Solution of Ill-Posed Problems*. New York: Wiley.