THE USE OF BAYESIAN STATISTICS IN IMAGE ESTIMATION FROM INTERFEROMETER DATA

T. J. Cornwell
Nuffield Radio Astronomy Laboratories,
Jodrell Bank, Macclesfield, Cheshire.

1. INTRODUCTION

The problem of estimating radio sky brightness distributions from incomplete, and noisy, visibility measurements, such as those collected by a long baseline interferometer, has recently been attacked using non-linear data-adaptive techniques such as the Maximum Entropy method (Ables (1974), Wernecke (1976), Wernecke and D'Addario (1977), Gull and Daniell (1978)) and the Maximum Likelihood method (Papadopoulos (1975)).

In general, there are many brightness distributions consistent with the measured visibilities (Bracewell (1956)). The Maximum entropy method selects, as its estimate, the brightness distribution which has maximum entropy and is consistent with the data. It has been interpreted as the most objective way of assigning the missing values of the visibility (Ables (1973), Ponsonby (1973)), and as selecting the most probable image given the Bose-Einstein statistics of photons (Kikuchi and Soffer (1977)). The second interpretation has the advantage that the two entropy measures, in current use namely log f and -f log f where f is the brightness (Wernecke (1976)), can be derived as limiting cases of a more general expression (Kikuchi and Soffer (1977)).

The Maximum Entropy equations possess no known closed-form solution in two dimensions and hence they must be solved numerically using the techniques of constrained optimization (Wernecke and D'Addario (1977)). The value of the Lagrange multiplier used in this method can only be estimated and hence the solution is approximate.

A method of extending the statistical interpretation of the Maximum Entropy method advanced by Kikuchi and Soffer (1977) is presented here and it is shown that a definite value can be assigned to the Lagrange multiplier introduced by Wernecke and D'Addario (1976). This method involves the use of Bayes' rule connecting conditional probabilities (Hoel (1947)).

We discuss first the Maximum Likelihood method and the CLEAN algorithm

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 227–234.

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press

2. THE MAXIMUM LIKELIHOOD METHOD

We wish to estimate the sky brightness f(x,y) within a region of sky described by orthogonal coordinates x,y. We have M samples of the measured visibility

$$m_i = m(u_i, v_i) \qquad i = 1,2,\ldots M \tag{1}$$

The measured visibility is related to the real visibility $\hat{m}_i$ by the addition of an error term

$$m_i = \hat{m}_i + \varepsilon_i \qquad i = 1,2,\ldots M \tag{2}$$

where

$$\hat{m}_i = \int_D f(x,y) \exp\{-j.\ 2\pi.\ (u_i x + v_i y)\}\ dxdy \tag{3}$$

We will approximate this integral by a sum over N pixels of area $\Delta A = \Delta x.\Delta y$

$$\hat{m}_i = \Delta A \sum_{k=1}^{N} f_k \exp\{-j.2\pi.(u_i x_k + v_i y_k)\} \tag{4}$$

The pixel area ΔA must be chosen to be less than the resolution of the interferometer in order that this approximation can fit the observed visibilities.

For convenience we will use vector notation for all variables.

$$\underline{x} = \{x_i : i=1,\ldots \text{(no. of elements in } \underline{x})\} \tag{5}$$

We assume that the error terms ε_i are uncorrelated, zero mean, complex Gaussian distribution variables with σ_i^2 on the i^{th} sample point. Therefore the probability of the vector of error terms $\underline{\varepsilon}$ lying between $\underline{\varepsilon}'$ and $\underline{\varepsilon} + \underline{d\varepsilon}'$ is

$$P(\underline{\varepsilon}')\underline{d\varepsilon}' = (2\pi)^{-M} \prod_{i=1}^{M} \sigma_i^{-2} \exp \frac{-1}{2} \sum_{i=1}^{M} \frac{|\varepsilon'_i|^2}{\sigma_i^2}\ d\varepsilon' \tag{6}$$

The function $P(\underline{\varepsilon}')$ is the probability density function of the error vector $\underline{\varepsilon}'$. from equation (2) we can see that

$$\underline{\varepsilon}' = \underline{m} - \underline{\hat{m}} \tag{7}$$

and therefore $P(\underline{\varepsilon})\ \underline{d\varepsilon}'$ tells us the probability of measuring the visibility between $\underline{m}$ and $\underline{m} + \underline{dm}$ given that the true visibility is $\underline{\hat{m}}$. Thus

$$P(\underline{\varepsilon}') = P(\underline{m}|\underline{\hat{m}}) \tag{8}$$

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press

is the conditional probability density function of the observed visibility vector $\underline{m}$ given that the true visibility vector is $\underline{\hat{m}}$. Since the true visibility $\underline{\hat{m}}$ can be calculated from the true brightness $\underline{\hat{f}}$ we have that $P(\underline{m}|\underline{\hat{m}}) \equiv P(\underline{m}|\underline{\hat{f}})$.

One obvious estimate of the brightness distribution is that which gives maximum probability of measuring the observed visibility vector $\underline{m}$. Formally, we maximise the Likelihood function given by $L = P(\underline{m}|\underline{\hat{f}})$, in fact this means that the true visibility $\hat{m}$ is fitted to the observed visibility $\underline{m}$ in the weighted least squares sense. Schwarz (1978) has shown that the CLEAN algorithm performs such a fit. However, this Maximum Likelihood estimate may have negative components, though usually only at the noise level.

A simple physical constraint is that the brightness must be positive. We may incorporate this constraint into the Maximum Likelihood estimate by the use of barrier functions (Adby and Dempster (1974). A barrier function can be used in optimization to restrict the range of possible solutions to within a subspace of the original space. For example to maintain positivity we may add a logarithmic term to the function to be maximised; this discriminates against solutions near zero and forbids negative solutions. Using this approach we find the restricted Maximum Likelihood estimate by requiring that

$$J = \alpha \sum_{k=1}^{N} \log_e \hat{f}_k + \log_e L \qquad \alpha << 1 \tag{9}$$

be maximised. For convenience, we have used the logarithm of L in equation (9); this produces the same estimate since the logarithm is a monotonically increasing function of argument. From equation (6) we have that

$$\log_e L = -\tfrac{1}{2} \sum_{i=1}^{M} \frac{|m_i - \hat{m}_i|^2}{\sigma_i^2} + \text{constant} \tag{10}$$

This estimate may be found using the iterative Maximum Entropy algorithm developed by Wernecke and D'Addario (1977) with their Lagrange multiplier set to $\Delta A(2\alpha)^{-1}$.

Therefore using a technique of constrained optimization, namely the use of barrier functions we may introduce a physical constraint, the positivity of brightness distributions.

We will now describe another way of introducing physical constraints.

3. THE BAYESIAN APPROACH

The philosophy of Bayesian statistics (Silvey (1970)) is to introduce a priori knowledge about the parameters to be estimated. We introduce a weighting function $Q(\underline{\hat{f}})$ which describes our degree of belief in the vector $\underline{f}$ prior to measurement. This prior knowledge is transformed to <u>posterior</u> knowledge by the sampling distribution $P(\underline{m}|\hat{\underline{f}})$. Bayes' rule gives the conditional probability density function of the brightness vector $\hat{\underline{f}}$, given the measurements $\underline{m}$, to be

$$R(\hat{\underline{f}}|\underline{m}) = \frac{Q(\hat{\underline{f}})\ \underline{P}(\underline{m}|\hat{\underline{f}})}{N(\underline{m})} \tag{11}$$

where

$$N(\underline{m}) = \int_{\text{all }\hat{\underline{f}}} Q(\hat{\underline{f}})\ P(\underline{m}|\hat{\underline{f}})\ \underline{df} \tag{12}$$

The presence of $N(\underline{m})$ in the denominator means that $Q(\hat{\underline{f}})$ does not have to be normalizable (Silvey 1970), i.e.

$$\int_{\text{all }\hat{\underline{f}}} Q(\hat{\underline{f}})\ \underline{df} \neq \text{finite}$$

This leads to an extension of the Maximum Likelihood method, we choose the brightness vector which is most probable given that the measured visibility is $\underline{m}$.

Therefore if we can find a suitable form for $Q(\hat{\underline{f}})$, which we will call the prior, then we can find an estimate based on the measurements and on the knowledge contained in $Q(\hat{f})$.

Jaynes (1968) has discussed two "objective" methods of defining priors in cases where we have very little knowledge about the parameters to be estimated. One of these, now known as Jaynes' Principle involves the maximisation of the entropy of $Q(\hat{\underline{f}})$ subject to certain observations constraints. This principle forms the basis of the Maximum Entropy method (Ables 1974) which therefore should be viewed as an attempt to find a prior distribution suitable for use in Bayes' rule.

However we have some physical knowledge about the brightness distribution which may be used to select a prior.

4. INTRODUCTION OF PHYSICAL KNOWLEDGE

(i) As stated above, we know that the brightness distribution must be positive. Using this fact only we require

$$\begin{aligned} Q(\hat{\underline{f}}) &= \text{constant if all } f_i > 0 \\ &= 0 \text{ or undefined otherwise} \end{aligned} \tag{14}$$

Since we wish to find the maximum of $R(\hat{\underline{f}}|\underline{m})$ we require that the derivatives of $Q(\hat{\underline{f}})$ be calculable. By analogy with the barrier function mentioned above we can easily select a suitable function

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press

$$Q(\underline{\hat{f}}) = \prod_{i=1}^{N} f_i^{\alpha} \qquad \alpha << 1 \tag{15}$$

This is an example of an unnormalizable function. This prior can be made arbitrarily close to unity for any $\underline{\hat{f}} > \underline{0}$ by selecting α small enough.

We now choose as our estimate the vector $\underline{\hat{f}}_B$ which maximises $P(\underline{\hat{f}}|\underline{m})$, or more conveniently its logarithm.

Therefore

$$\frac{\partial}{\partial \underline{\hat{f}}}\left[(\alpha \sum_{i=1}^{N} \log \hat{f}_i - \tfrac{1}{2} \sum_{k=1}^{M} \frac{|m_k - \hat{m}_k|^2}{\sigma_{k2}}) \right]_{\underline{\hat{f}}=\underline{\hat{f}}_B} = 0 \tag{16}$$

Again this may be solved using the algorithm of Wernecke and D'Addario (1977).

(ii) We should recognize that we observe the brightness distribution by photons and therefore we should use Bose-Einstein statistics to select a prior.

Kikuchi and Soffer (1977) use these statistics to derive a form for $Q(\hat{f})$ valid in the radio-astronomical regime.

$$\log_e Q(\underline{\hat{f}}) = z \sum_{k=1}^{N} \log_e \hat{f}_k + \text{constant} \tag{17}$$

where z is the number of degrees of freedom for each photon. Kikuchi and Soffer argue that z is due to a spatial uncertainty about the origin of the photon within the pixel and to a temporal uncertainty about the time of arrival of the photon. This temporal uncertainty clearly depends upon the noise level, bandwidth and integration time and thus, in our formulation, must be included in the sample distribution $P(\underline{m}|\underline{\hat{f}})$. Then z is given by the number of resolution elements in a pixel, therefore we have that z = 1 since z cannot be fractional and the pixels are, by definition, not resolved.

Note that the prior has the same form as that in (i) and thus, the estimate may also be solved by the algorithm of Wernecke and D'Addario (1977) with their Lagrange multiplier set to $\Delta A/2$.

We have used strictly physical knowledge in cases (i) and (ii) and have obtained the most probable estimates given that knowledge. However another application of Bayes' rule is possible and this we discuss now.

5. INTRODUCTION OF HUMAN JUDGEMENT

We have a set of noisy, incomplete measurements of the visibility which is compatible with many different brightness distributions. In the above formulation we have selected the most probable, in some sense, of these as our estimate.

Another approach would be to reject the selection of one of these distributions as an estimate but instead to try and characterise this set of possible brightness distributions in a smaller set.

Let us select a property, A, of brightness distributions and devise a measure M_A of that property. For example, Baker (1978) has proposed the measure

$$M_s = \sum_{k=1}^{N} \hat{f}_k^{\,2} \tag{18}$$

for the sharpness or contrast of an image. Then we postulate a prior dependent upon this measure; for example

$$Q(\hat{f}) = \exp\,(\beta M_A) \tag{19}$$

Using the above Bayesian formulation and this prior we can define a most probable image $\hat{\underline{f}}_A$ which obeys the condition

$$\frac{\partial}{\partial \hat{\underline{f}}}\left[(\beta M_A - \tfrac{1}{2}\sum_{i=1}^{M}\frac{|m_i - m_i|^2}{\sigma_i{}^2})\right]_{\hat{\underline{f}}=\hat{\underline{f}}_A} = 0$$

Then for β positive we select the distribution with property A dominant and for β negative we select the distribution with the opposite of property A dominant.

If we use sharpness as property A then we find both the sharpest and the least sharp brightness distributions consistent with the measurements.

We could now use these two brightness distributions as the subset characterising the full set. The presence of a feature in both estimates would be strong evidence that such a feature was contained in the true brightness distribution.

Under this philosophy an all purpose map such as that described by Gull and Daniell (1978) would not be satisfactory, instead we would produce a number of maps by different algorithms and study this set instead of one member.

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press

6. CONCLUSIONS

We have used Bayesian statistics in two different modes.

In the first we have extended the interpretation of the Maximum Entropy method made by Kikuchi and Soffer (1978). We acknowledge that we observe the radio source by photons and therefore we can apply Bose-Einstein statistics to select, as the most probable, one of the set of distributions consistent with the data. This Bayesian formulation has the advantage that no arbitrary constraints, such as Lagrange multipliers, are involved.

In the second mode we choose not to select one map as the best in some sense but to use several different maps to characterise the set of brightness distributions consistent with the data.

Although the two modes are different philosophically the map produced by using Bose-Einstein statistics can be regarded as belonging to the characterising subset of the second mode. However this second mode has the advantage that the final judgement is left to the person processing the visibility measurement.

ACKNOWLEDGEMENTS

The author is especially indebted to Dr. R.E. Spencer who introduced him to data adaptive image processing and who shared both theoretical and practical insights into this subject. Many other people at Jodrell Bank discussed aspects of map-making with the author, in particular, Drs. J.E.B. Ponsonby, P.N. Wilkinson, R.G. Conway and R.J. Davis. Finally the author would like to thank Prof. J.G. Davies and Dr. J.E.B. Ponsonby for their supervision, the S.R.C. for financial support and Sir Bernard Lovell for the use of the facilities at Jodrell Bank.

REFERENCES

Ables J. G. 1974. Astron and Astrophys. Suppl. 15, 383.
Adby P.R. and Dempster M.A.H. 1974. "Introduction to Optimization Methods", Chapman and Hall, London.
Baker P.L. 1978. Astron. and Astrophys. to be published.
Bracewell R.N. 1956. Aust.J.Phys. 9, 297.
Gull S. F. and Daniell G.J. 1978. Nature 272, 686-690.
Hoel P.G. 1947. "Introduction to Mathematical Statistics", John Wiley and Sons, London.
Jaynes E.T. 1968. I.E.E.E. Trans. SSC-4 227-241.
Kikuchi R. and Soffer B.H. 1977, J.Opt.Soc.Am., 67, 1656-1657.
Papadopoulos G., 1975, Trans. I.E.E.E. AP-23, 45.
Ponsonby J.E.B., 1973. M.N.R.A.S., 163, 369.
Schwarz J., 1978, Astron. and Astrophys. 65, 345.

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press

Silvey S. D. 1970. "Statistical Inference", Chapman Hall London (1970).
Wernecke S.J., 1976, Ph.D. Thesis, Stanford Univ.
Wernecke S.J., 1977, Rad.Sci., 12, 831.
Wernecke S.J. and D'Addario L.R., 1977, Trans. I.E.E.E., C-26, 351.

https://doi.org/10.1017/S0252921100074911 Published online by Cambridge University Press