[3] Bretthorst, G. Larry, (1988), "Excerpts from Bayesian Spectrum Analysis and Parameter Estimation," in *Maximum-Entropy and Bayesian Methods in Science and Engineering,* **1,** G. J. Erickson and C. R. Smith *eds.,* Kluwer Academic Publishers, Dordrecht, Holland, pp. 75-145.

[4] Levine, M. A., (1987), "Root Sum Square Approximation of Source Convolved with Decollimation," in *Infrared Systems and Components,* Robert L. Caswell, Editor, Proc. SPIE **750,** pp. 79-89.

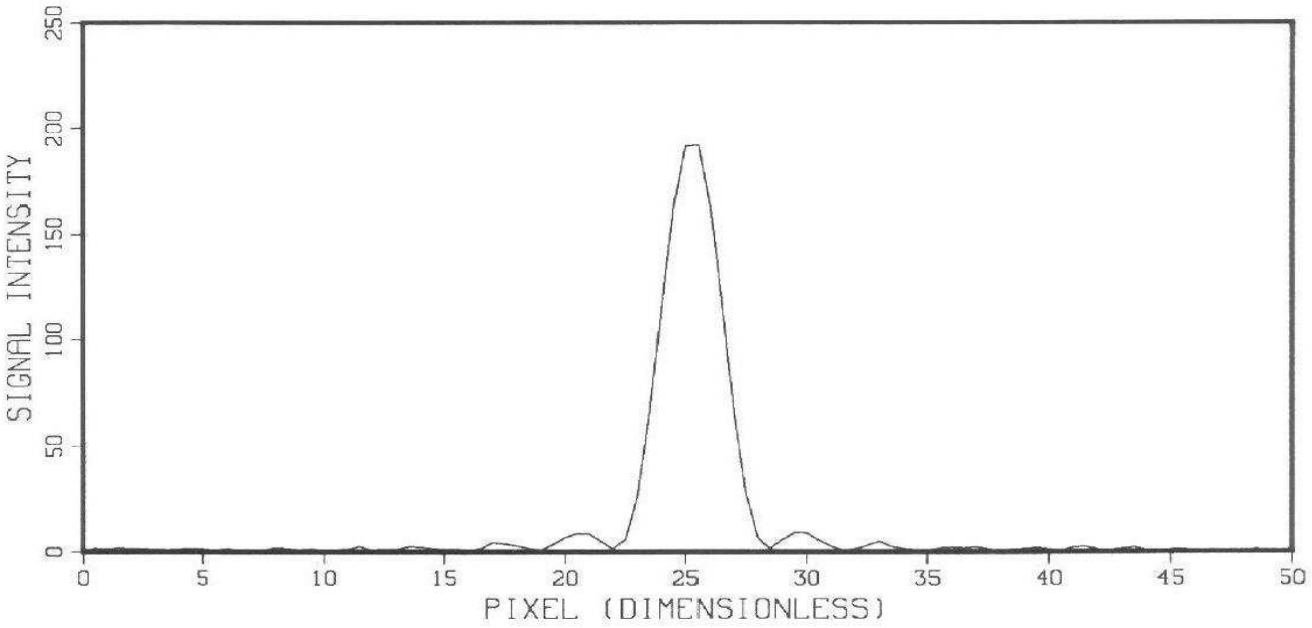
[5] Zellner, A., (1971), *An Introduction to Bayesian Inference in Econometrics,* John Wiley and Sons, New York. Second edition, (1987).

[6] Jaynes, E. T., (1988), "Bayesian Spectrum and Chirp Analysis," in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems,* C. Ray Smith and G. J. Erickson, eds., D. Reidel, Dordrecht, Holland, pp. 1-37.

[7] Jeffreys, H., (1939), *Theory of Probability,* Oxford University Press, London, (Later editions, 1948, 1961).

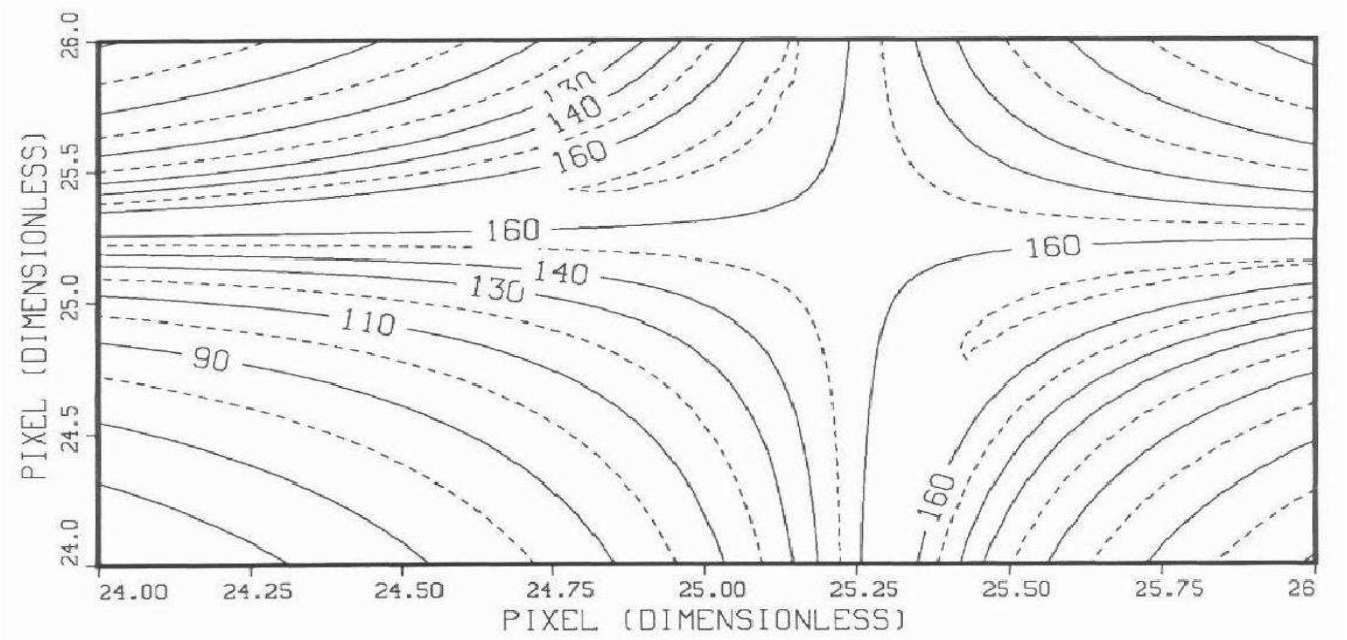
[8] Jaynes, E. T., (1988), "Detection of Extra-Solar System Planets," in *Maximum-Entropy and Bayesian Methods in Science and Engineering,* **1,** G. J. Erickson and C. R. Smith, *eds.,* Kluwer Academic Publishers, Dordrecht, Holland, pp. 147-160.

Figure 7: Computer Simulated Data – Two Point Sources With Sinc-Function Smearing



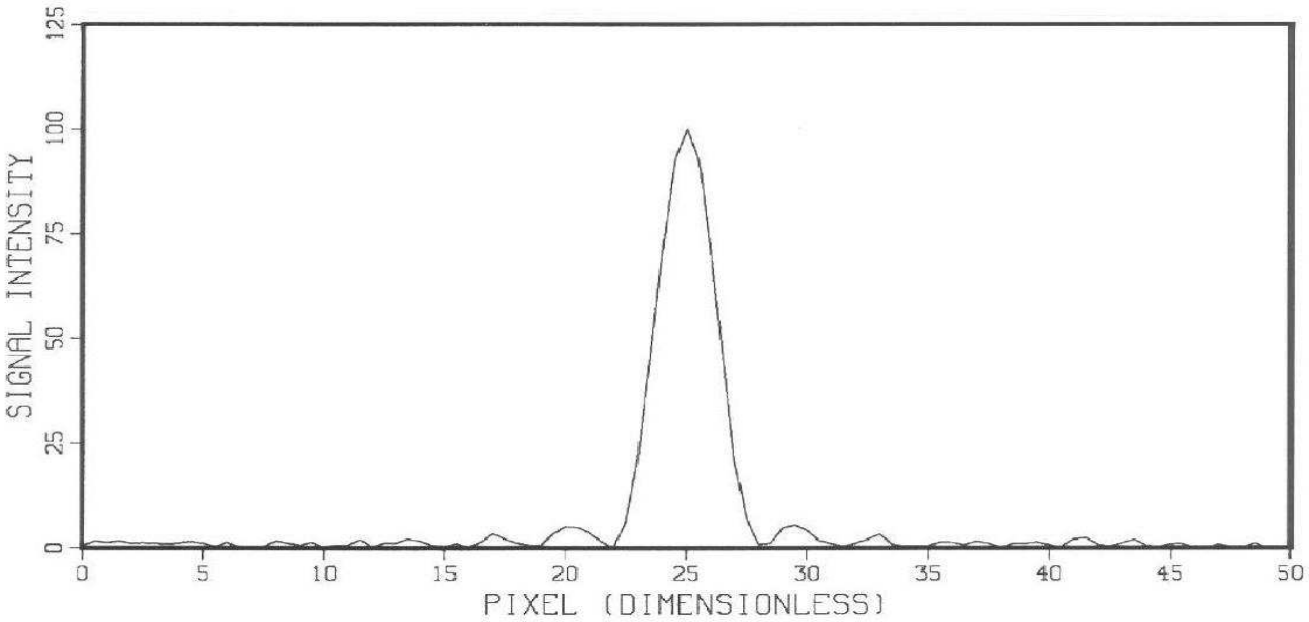
These computer simulated data contains two point sources, one at 25.0001 and the other 25.5001 (we again displaced them by 1/10000 to avoid possible numerical difficulties with the sinc functions). The pixel values are again the half integers $(0.5, 1.0, \cdots, 50.0)$; thus, the two point sources are separated by one pixel. There are $N = 100$ data values, and there are approximately 6 data values in the vicinity of the peak.

Figure 8: Log_{10} Posterior Probability of Two Point Sources With Sinc-Function Smearing



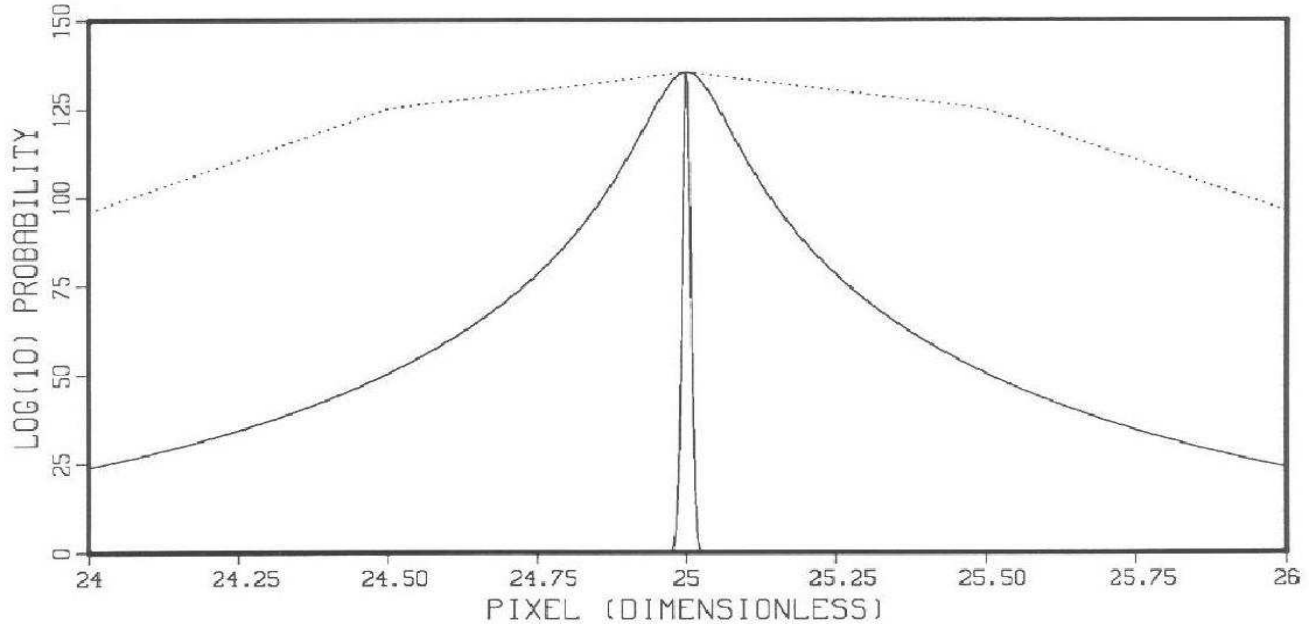
This contour plot is the base 10 logarithm of the posterior probability of two point sources with sinc-function smearing in the data. Notice that the contours clearly resolve the two point sources even though there are very few actual data values. Around the region of the maximum, a drop of one corresponds to the region containing 90% of the posterior probability, a drop of 2 corresponds to a region containing 99%, etc. Thus the region around the maximum corresponds to a very sharp peak, indicating good resolution of the point sources.

Figure 5: Computer Simulated Data – One Point Source With Sinc-Function Smearing



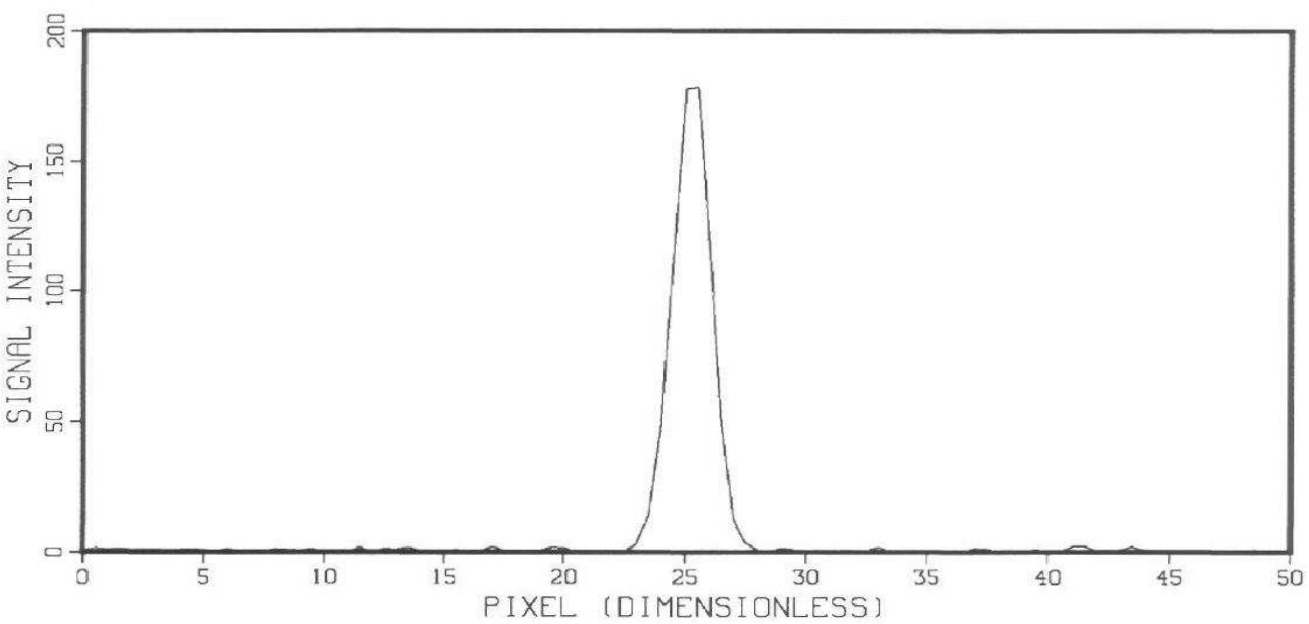
These computer simulated data contain one point source at 25.0001 (we displaced it by 1/10000 to avoid possible numerical difficulties with the sinc functions). The signal-to-noise ratio of the data is 100. There are $N = 100$ data values, with the pixel values corresponding to the half integers. The amplitude of the point source is 100 and the variance of the noise is 1.

Figure 6: Log_{10} Posterior Probability of One Point Source – Versus the Data



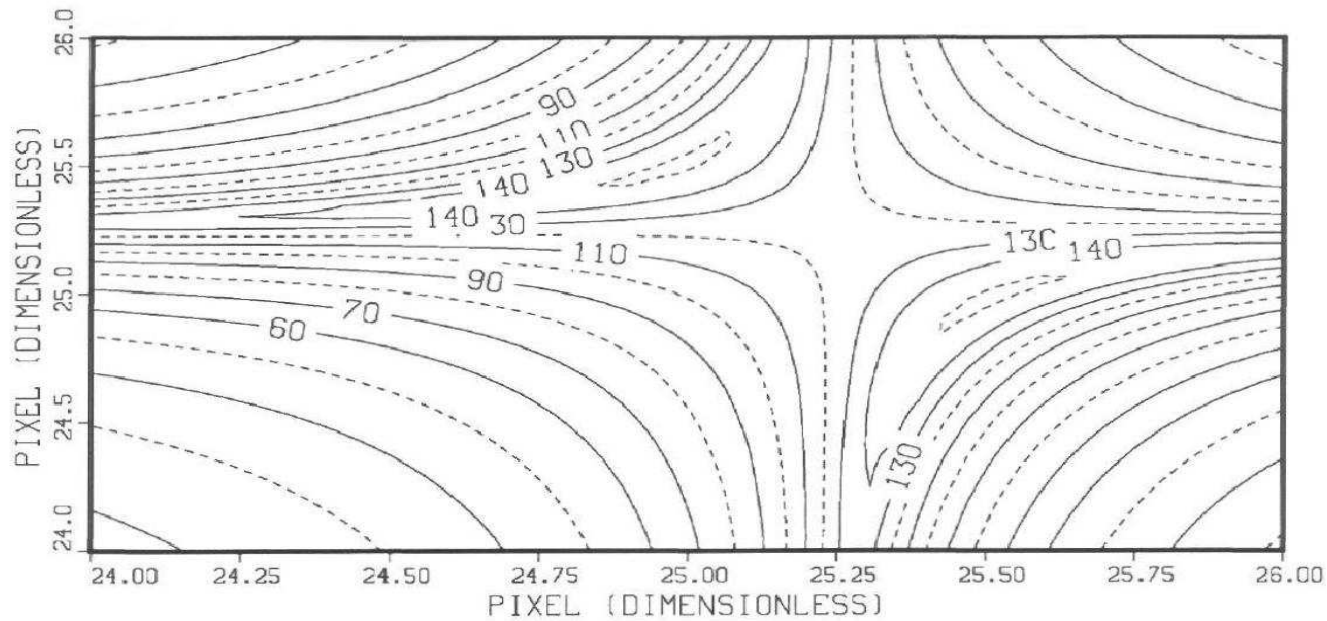
The dotted line is a close-up of the data. There are 5 data values in this region. The solid line inside the data is the base 10 logarithm of the posterior probability of the center position of one point source with sinc-function smearing. Notice how sharply peaked this is compared with the data. The line in the center that looks like we have marked the position of the point source is the posterior probability of the center position (normalized to the maximum of the log posterior). The width of this line is indicative of how accurately the point source has been located.

Figure 3: Computer Simulated Data – Two Point Sources With Gaussian Smearing



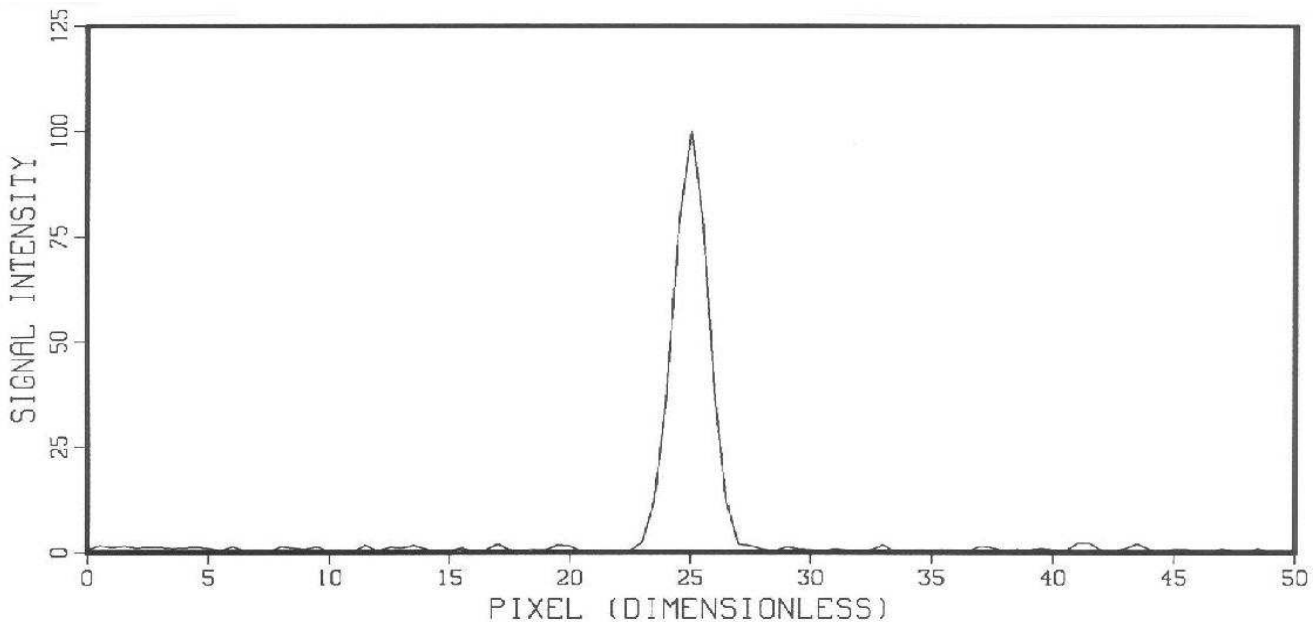
These computer simulated data contain two point sources, one at 25 and the other at 25.5. The pixel values correspond to the half integers; thus these two point sources are separated by one pixel. There are $N = 100$ data values, and there are approximately 6 data values in the vicinity of the peak.

Figure 4: Log_{10} Posterior Probability of Two Point Sources With Gaussian Smearing



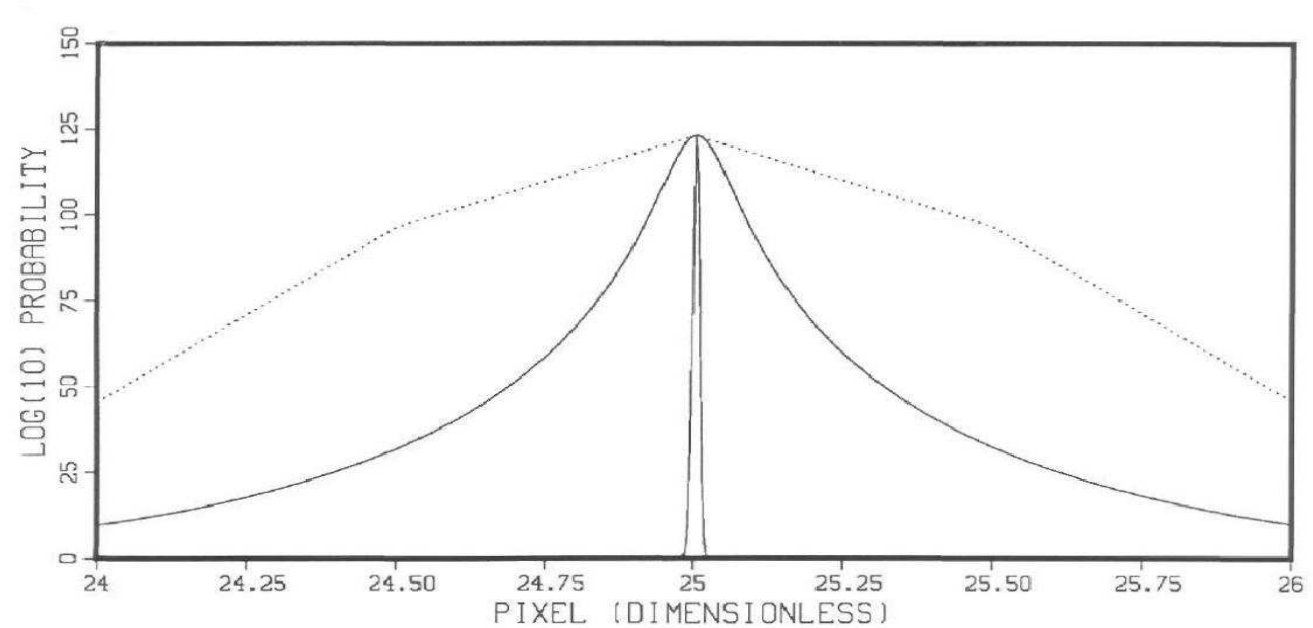
This contour plot is the base 10 logarithm of the posterior probability of two Gaussians in the data. Notice that the contours clearly resolve the two Gaussians even though there are very few actual data values. Around the region of the maximum, a drop of one corresponds to a region containing 90% of the posterior probability, a drop of 2 corresponds to a region containing 99%, etc. Thus the region around the maximum corresponds to a very sharp peak, indicating good resolution of these two center positions.

Figure 1: Computer Simulated Data – One Point Source With Gaussian Smearing



These computer-simulated data contain one point source with Gaussian smearing; the point source is at 25. The signal-to-noise ratio of the data is approximately 18. There are $N = 100$ data values, and there are approximately 5 data values in the vicinity of the peak, with two more located near the bottom of the Gaussian. The amplitude of the Gaussian is 100 and the variance of the noise is 1. Pixel values, in this example, correspond to half integers.

Figure 2: Log_{10} Posterior Probability of One Point Source – Versus the Data



The dotted line is a close-up of the data (scaled to the maximum Log posterior). There are only five data values in this region. The sharply peaked solid line inside the data is the base 10 logarithm of the posterior probability of the point source with Gaussian smearing. Notice how sharply peaked this is compared to the data. The line in the center that looks like we have marked the position of the point source is really the posterior probability of the point source scaled to the maximum Log posterior. The width of this line is indicative of how accurately probability theory has determined the center position.

This problem has been considered in some detail by Jaynes,[8] and we refer the reader to this analysis for analytic details. Here we will simply use this model and compute the posterior probability of one or more point sources using the general theory given by Bretthorst.[1,2,3] For this third example we will again use dimensionless units and run the pixel values from one to 50 by half integers, so that there are $N = 100$ data values. We generated the data, Fig. 5, from

$$d_i = 100 \left[\frac{\sin(25.0001 - x_i)}{25.0001 - x_i}\right]^2 + e_i, \tag{35}$$

where there is one point source located at 25.0001 (we displaced the location off the exact pixel value to avoid numerical problems with the sinc functions). Although the point source is located in nearly the same place and has the same amplitude as the single point source with Gaussian smearing, the signal-to-noise in these data is 100. Next we computed the posterior probability of the location of a single point source with sinc-function smearing, obtaining the "student t-distribution" – see Bretthorst.[1] We plot the base 10 logarithm of the "student t-distribution," Fig. 6. We have displayed the data (dotted line) as a reference. The sharp spike around 25 is the "student t-distribution" normalized to the maximum of the log "student t-distribution." Like with the Gaussian smearing, the position of a single point source with Airy smearing has been determined to ± 0.01 pixels.

In the last example, we will consider the case where the data contain two point sources with Airy smearing. We generated the data, Fig. 7, from

$$d_i = 100 \left[\frac{\sin(25.0001 - x_i)}{25.0001 - x_i}\right]^2 + 100 \left[\frac{\sin(25.5001 - x_i)}{25.5001 - x_i}\right]^2 + e_i. \tag{36}$$

The noise was again generated in the same manner as the Gaussian case. Also, these two point sources are separated by one real pixel. The signal-to-noise is again approximately double that of the one point source with sinc-function smearing, or approximately 200. We computed the posterior probability of the center position of two one point sources with sinc-function smearing independent of the amplitudes and variance of the noise, obtaining a "student t-distribution". The two parameters of interest are the positions of the two point sources. We again plot the base 10 logarithm of the posterior probability of the position of the point sources independent of the amplitudes and variance of the noise, the "student t-distribution," Fig. 8. Notice the contour lines are dropping by 10 orders of magnitude in probability density, indicating we have determined the positions of the point sources accurately. The elongated form of these contour lines is indicative of the difference in the center positions being better determined than the sum of the positions.

5 SUMMARY – CONCLUSIONS

In this paper we have demonstrated how the use of Bayesian probability theory can be used to improve the resolution of multiple closely-spaced objects. This analysis indicates that using probability theory, multiple closely-spaced point sources are easily resolvable at separations of less than one pixel, provided one has reasonable signal-to-noise, order 10, and enough data in the region of the peak to get a good determination of the shape. But Bayesian probability theory is much more than a parameter estimation procedure. Indeed when little prior information is incorporated into a Bayesian calculation, the estimates one obtains cannot differ significantly from the maximum likelihood estimates, and if one examines the results presented here one will find that we have indeed obtained the maximum likelihood results. The real power of Bayesian probability theory comes from its ability to carry the problem forward and to answer questions like "How precise are our estimates?" or "Has a signal been detected?" or "How many point sources are present?"

References

[1] Bretthorst, G. Larry, (1988), "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics* **48,** Springer-Verlag, New York, New York.

[2] Bretthorst, G. Larry, (1987), *Bayesian Spectrum Analysis and Parameter Estimation,* Ph.D. thesis, Washington University, St. Louis, MO.; available from University Microfilms Inc., Ann Arbor, Mich.

these two point sources are estimated to be

$$\begin{aligned}(\theta_1)_{\text{est-Bayes}} &= \hat{x}_1 \pm \sqrt{\frac{16[(700+300).1+100+100]2}{(100+100)^3(1.7+3)}} \text{ pixels} \\ &\approx \hat{x}_1 \pm 0.016 \text{ pixels} \\ (\theta_2)_{\text{est-Bayes}} &\approx \hat{x}_2 \pm 0.016 \text{ pixels.}\end{aligned} \tag{31}$$

This is only slightly worse than the one point source estimate of ± 0.014 pixels. Thus when two point sources are as close as one fifth of a pixel, they are almost noninterfering and are still resolvable at better than 6 standard deviations!

4 EXAMPLES

In these examples we will generate data containing one or more point sources and then apply the full Bayesian solution to these computer-generated data. We will specify carefully how the data were generated so that these demonstrations may be repeated easily. In the first example, Fig. 1, we have generated data from

$$d_i = 100 \exp\left\{-(25.0 - x_i)^2\right\} + e_i, \tag{32}$$

where $x_i = \{0.5, 1.0, \cdots 50.0\}$, corresponding to $N = 100$ data values. The noise term e_i was generated from the absolute value of a unit normal random number generator. The signal-to-noise ratio (defined as the peak square signal value divided by the mean-square noise) is 100. This is the example used in deriving the accuracy estimates for the one point source with Gaussian smearing, Eq. (20). Notice that on this scale the digitization of the Gaussian is apparent. We analyzed the data using a one-point-source model with Gaussian smearing and computed the posterior probability of the position of the point source independent of the amplitude and variance of the noise,[1] the so-called "student t-distribution." We have plotted the base 10 logarithm of the "student t-distribution" as the solid line in Fig. 2. The dotted line is the data, after subtracting a constant so that the peak data value just touches the maximum log "student t-distribution." The spike around 25 is, up to a normalization constant, the posterior probability of the position of a single point source with Gaussian smearing. The posterior probability has determined the location of the point source very accurately, within approximately ± 0.01 pixels. This is in good agreement with the theoretical calculations.

For this second demonstration we use data which contain two point sources with Gaussian smearing. The smearing is the same as that used in the last example. We generated data from the following equation:

$$d_i = 100 \exp\left\{-(25.0 - x_i)^2\right\} + 100 \exp\left\{-(25.5 - x_i)^2\right\} + e_i. \tag{33}$$

The noise values were generated in the same manner as for the one point source with Gaussian smearing. Here we have two point sources separated by one pixel. The data, Fig. 3, look like one Gaussian. The total signal-to-noise ratio of these data is just double that of the one point source. To perform this analysis we considered a two-point-source model with Gaussian smearing and computed the "student t-distribution." There are now two parameters of interest to be determined from the data. We computed the base 10 logarithm of the "student t-distribution" and have displayed it as a contour plot, Fig. 4. Notice that there are two well-defined maxima in this contour plot. Had we plotted the "student t-distribution," instead of the logarithm of the "student t-distribution," the two maxima would look like two very sharp spikes located at 25 and 25.5. The stretched out appearance of these contours indicates that the "student t-distribution" has determined the difference in position of the two point sources better than the sum.

In the next two examples we again use computer-generated data. However, we now consider the case of a one-dimensional Airy distribution as the smearing function. In one dimension, the Airy distribution reduces to the so-called sinc function. We take our signal to be of the form

$$f(x) = \sum_{j=1}^{m} B_j \left[\frac{\sin(\theta_j - x)}{\theta_j - x}\right]^2. \tag{34}$$

where $\overline{h^2}$ is now given by

$$\overline{h^2} = \frac{h_1^2 + h_2^2}{2} \tag{23}$$

with

$$h_1 \equiv \sum_{i=1}^{N} d_i H_1(x_i), \qquad h_2 \equiv \sum_{i=1}^{N} d_i H_2(x_i), \tag{24}$$

and

$$H_1(x) \approx \frac{(2/\pi)^{\frac{1}{4}}}{\sqrt{2(1+\exp\left\{-\frac{(\theta_1-\theta_2)^2}{2}\right\})}} \left[\exp\left\{-(\theta_1 - x)^2\right\} + \exp\left\{-(\theta_2 - x)^2\right\}\right], \tag{25}$$

$$H_2(x) \approx \frac{(2/\pi)^{\frac{1}{4}}}{\sqrt{2(1-\exp\left\{-\frac{(\theta_1-\theta_2)^2}{2}\right\})}} \left[\exp\left\{-(\theta_1 - x)^2\right\} - \exp\left\{-(\theta_2 - x)^2\right\}\right]. \tag{26}$$

The quantity $(\theta_1 - \theta_2)^2$ in the square root in the above equations tells us how the estimation of two point sources interfere with each other. As long as $(\theta_1 - \theta_2)^2$ is large, the exponential term will be small and the center position of each Gaussian can be estimated as if the other were not present. It is only when the estimated separation distance is less than or approximately equal to 2 that any significant interference can occur.

We can derive the accuracy estimates for two point sources with Gaussian smearing in a way analogous to what was done for the one point source. We postulate a functional form of the data with high signal-to-noise ratio; here, we take

$$d_i = (2/\pi)^{\frac{1}{4}} \left[\hat{B}_1 \exp\left\{-(\hat{x}_1 - x_i)^2\right\} + \hat{B}_2 \exp\left\{-(\hat{x}_2 - x_i)^2\right\}\right] + e_i \tag{27}$$

as the data, where $\hat{x}_1$ is the estimated position of the first point source, $\hat{x}_2$ is the estimated position of the second point source, and $\hat{B}_1$ and $\hat{B}_2$ are the estimated amplitudes of the sources. Then we substitute the data, Eq. (27), into the posterior probability, Eq. (22), and Taylor expand about the maximum posterior probability to obtain the (mean $\pm$ standard deviation) estimates. The calculation is long and tedious and we do not repeat the details here. When the two point sources are well separated, the estimates reduce to the one point source estimates. When the two point sources are so close that they overlap almost completely, the accuracy estimates reduce to

$$(\theta_1)_{\text{est-Bayes}} = \hat{x}_1 \pm \sqrt{\frac{16[(7\hat{B}_1 + 3\hat{B}_2)\Delta + \hat{B}_1 + \hat{B}_2]\sigma^2\alpha^2}{(\hat{B}_1 + \hat{B}_2)^3(17\Delta + 3)}} \text{ pixels} \tag{28}$$

for θ_1 and

$$(\theta_2)_{\text{est-Bayes}} = \hat{x}_2 \pm \sqrt{\frac{16[(3\hat{B}_1 + 7\hat{B}_2)\Delta + \hat{B}_1 + \hat{B}_2]\sigma^2\alpha^2}{(\hat{B}_1 + \hat{B}_2)^3(17\Delta + 3)}} \text{ pixels} \tag{29}$$

for θ_2, where Δ, the estimated separation, is defined as

$$\Delta \equiv |\hat{x}_1 - \hat{x}_2| \tag{30}$$

and $\hat{B}_1$ is the estimated amplitude If Δ is very small, these estimates will confound the two point sources; i.e., they will be indistinguishable from one point source, and the estimates will be determined to an accuracy which depends on the sum of the amplitudes of the two point sources, as intuition might have suggested. However, these estimates will be a little worse than those obtained from the one-point-source model. Again this is what one would have expected. After all, in the limit $\Delta \to 0$ we have estimated the position of a single point source with a probability density for two: probability theory is hedging its bets, by making the estimates less certain.

We take an example similar to the previous one and set $\sigma = 1$, $\hat{B}_1 = \hat{B}_2 = 100$, $\alpha = \sqrt{2}$, and position the two point sources 0.2 pixels apart ($\Delta = .1$ in dimensionless units); then according to probability theory the positions of

This is the posterior probability density for θ_1, the position of the point source, and will tell us everything the data have to say about it independent of the amplitude B_1 and independent of the noise variance σ^2. This probability density is referred to as a "student t-distribution" in the literature, and it is the generalization of this distribution to arbitrary models that we apply in Section 4.

To gain a better analytic understanding of these equations, suppose we have data with very high signal-to-noise ratio of the form

$$d_i = (2/\pi)^{\frac{1}{4}} \hat{B} \exp\left\{-(x_i - \hat{x})^2\right\} + e_i, \tag{15}$$

where $\hat{x}$ is the estimated location of the point source and $\hat{B}$ is the estimated amplitude. Then, if we make the approximation that sums may be replaced by integrals in Eq. (11), the sufficient statistic $\overline{h^2}$ is given approximately by

$$\overline{h^2} \approx \hat{B}^2 \exp\left\{-(\hat{x} - \theta_1)^2\right\}. \tag{16}$$

If we assume the variance of the noise σ^2 to be known, we can Taylor expand the sufficient statistic $\overline{h^2}$ about $\hat{x}$ and make the (mean $\pm$ standard deviation) estimates of the center position. Performing the indicated calculation the posterior probability of the center position is given approximately by

$$P(\theta_1|\sigma, D, I) \approx \sqrt{\frac{\hat{B}^2}{\pi\sigma^2}} \exp\left\{-\frac{\hat{B}^2(\hat{x} - \theta_1)^2}{\sigma^2}\right\} \tag{17}$$

from which we estimate the position to be

$$(\theta_1)_{\text{est-Bayes}} = \hat{x} \pm \frac{\sigma}{\sqrt{2}\hat{B}} \tag{18}$$

where, due to the approximation in Eq. (16), the estimated value $\hat{x}$ is the "true" location of the point source. We remind the reader that we are using dimensionless units. To convert to real physical pixels, let α represent the standard deviation of the Gaussian smearing function; then the location of the point source is estimated to be

$$(\theta_1)_{\text{est-Bayes}} = \hat{x} \pm \frac{\sigma\alpha}{\hat{B}}, \tag{19}$$

where θ_1 and $\hat{x}$ are now the dimensional quantities of interest. This indicates, as one's intuition might have suggested, that the accuracy of the center position depends inversely on the signal-to-noise ratio of the data and linearly on how broadly the instrument smears the data. To be more specific, suppose $\sigma = 1$, $\hat{B} = 100$, and $\alpha = \sqrt{2}$, i.e., the smearing is such that we have significant data over only approximately 5 pixels; then we estimate the center position of the Gaussian to be

$$\begin{aligned} (\theta_1)_{\text{est-Bayes}} &= \hat{x} \pm \frac{\sqrt{2}}{100} \\ &\approx \hat{x} \pm 0.014 \text{ pixels.} \end{aligned} \tag{20}$$

If we had used the standard deviation of the Gaussian smearing function as an indication of how accurately we determine the location of the point source, we would have had

$$(\theta_1)_{\text{est-Gaussian}} = \hat{x} \pm 1.4 \text{ pixels.} \tag{21}$$

Thus the use of probability theory has improved the resolution by a factor of 100. That is, had the point source been displaced from $\hat{x}$ by more than $\alpha/100$, it is very unlikely that we would have obtained data pointing to the "true" position.

This same calculation may be repeated for the case of two point sources. The calculation proceeds by postulating a two-point-source model function, Eq. (2), with $m = 2$. The posterior probability of the location of the two point sources is then computed in an analogous way to the one point source problem. From this we find the posterior probability of two point sources located at θ_1 and θ_2 with Gaussian smearing to be

$$P(\theta_1, \theta_2|\sigma, D, I) \propto \exp\left\{\frac{\overline{h^2}}{\sigma^2}\right\}, \tag{22}$$

Substituting Eqs. (2) Eq. (3) into Eq. (6) and taking the simplest case of one Gaussian, we have

$$P(B_1, \theta_1|\sigma, D, I) \propto \sigma^{-N} \exp\left\{-\sum_{i=1}^{N} \frac{[d_i - B_1(2/\pi)^{\frac{1}{4}} \exp\{-(x_i-\theta_1)^2\}]^2}{2\sigma^2}\right\}. \tag{7}$$

But this is the joint posterior probability of both the amplitude B_1 and the center position θ_1 given the variance of the noise σ^2, the data D, and the prior information I. We would like to compute $P(\theta_1|\sigma, D, I)$ to better understand how to estimate the important position of the point source. The rules of probability theory uniquely determine how this is to be done. One computes

$$P(\theta_1|\sigma, D, I) = \int dB_1 P(B_1, \theta_1|\sigma, D, I). \tag{8}$$

This equation essentially says the probability of a series of mutually exclusive propositions is the sum of the probabilities of the individual propositions; when probability densities are used the sums are replaced by integrals and one obtains Eq. (8). Using Eq. (7) and performing the indicated integral, one obtains

$$P(\theta_1|\sigma, D, I) \propto \sigma^{-N} \exp\left\{-\frac{N\overline{d^2} - \overline{h^2}}{2\sigma^2}\right\}, \tag{9}$$

where

$$\overline{d^2} \equiv \frac{1}{N}\sum_{i=1}^{N} d_i^2 \tag{10}$$

is the mean-square observed data value, and

$$\overline{h^2} \equiv \sqrt{2/\pi}\left[\sum_{i=1}^{N} d_i \exp\left\{-(x_i-\theta_1)^2\right\}\right]^2 \tag{11}$$

where we have made the approximation that

$$\begin{aligned}\sqrt{2/\pi}\sum_{i=1}^{N}\exp\left\{-2(x_i-\theta_1)^2\right\} &\approx \sqrt{2/\pi}\int_{-\infty}^{\infty} dx \exp\left\{-2(x-\theta_1)^2\right\} \\ &\approx 1\end{aligned} \tag{12}$$

which is valid when the smearing function is wide compared to a pixel. The approximation made in the above calculation is, essentially, harmless and is removed in the general formulation of the problem.[1] The Gaussian model functions, Eq. (3), are normalized in the sense that their mean-square length is one.

If the variance of the noise σ^2 is known, then we may simplify Eq. (9) to obtain

$$P(\theta_1|\sigma, D, I) \propto \exp\left\{\frac{\overline{h^2}}{2\sigma^2}\right\}. \tag{13}$$

This is the posterior probability of the possible positions of the point source and will tell us everything the data have to say about it, but it assumes we know σ. In most real problems this is not the actual case. However, this presents little difficulty because we can always multiply Eq. (9) by a prior probability for the variance and integrate with respect to σ. This will give us the posterior probability of the center position independent of the amplitude and noise variance. The variance of the noise is a scale parameter, and Jeffreys[7] has shown that the appropriate uninformative prior probability for a scale parameter is the Jeffreys prior $1/\sigma$. Multiplying Eq. (9) by a Jeffreys prior and integrating with respect to σ, one obtains

$$P(\theta_1|D, I) \propto \left[1 - \frac{\overline{h^2}}{N\overline{d^2}}\right]^{-\frac{N-1}{2}}. \tag{14}$$

where we define $\mathbf{B} \equiv \{B_1, \cdots, B_m\}$ as the collection of amplitudes and $\mathbf{\Theta} \equiv \{\theta_1, \cdots, \theta_r\}$ is a set of nonlinear parameters. Both $\mathbf{B}$ and $\mathbf{\Theta}$ are to be determined from the data. However, in this problem we are interested in the location of the point sources, and not nearly so interested in their magnitude. For this demonstration we will take the signal functions to be either an Airy or Gaussian functions. The optical smearing function can, in many cases, be closely approximated by a Gaussian.[4] We will assume, for now, that the signal functions G_j are of the form

$$G_j(x, \theta_j) \equiv (2/\pi)^{\frac{1}{4}} \exp\left\{-(x - \theta_j)^2\right\}, \tag{3}$$

where we are using dimensionless units, i.e., we have absorbed the spread of the Gaussian into the pixel value x and θ_j (in these units the pixels are not integers). The reason for the unusual normalization will become apparent as we proceed. The $\mathbf{\Theta}$ parameters are the locations of the point sources. In this calculation we will assume there is only a single point source, and later we will present analytic results for the two-point source problem. We restrict the calculation to one dimension, primarily because there is no convenient way to exhibit the results for higher dimensional calculations. Indeed the generalization to any number of dimensions is straightforward.

We will solve this problem using Bayesian probability theory, assigning uninformative priors; thus, from a probability standpoint we will make very conservative estimates of the parameters and their accuracy. We will compute the posterior probability of the position of the point sources independent of the amplitudes. We do this to see what probability theory can tell us about the position of the point source without the nuisance amplitude confusing what probability theory is telling us.

3 THE POSTERIOR PROBABILITY OF THE PARAMETERS

Bayes' theorem tells us that, given the data D and prior information I, the joint probability density of all for the parameters is

$$P(\mathbf{B}, \mathbf{\Theta}|D, I) = \frac{P(\mathbf{B}, \mathbf{\Theta}|I)P(D|\mathbf{B}, \mathbf{\Theta}, I)}{P(D|I)}. \tag{4}$$

Here, $P(\mathbf{B}, \mathbf{\Theta}|D, I)$ is the joint posterior probability density of the amplitudes $\mathbf{B}$ and the positions of the point sources $\mathbf{\Theta}$ given the data D and the prior information I. It is this probability that we would like to compute. $P(\mathbf{B}, \mathbf{\Theta}|I)$ is the joint probability of the amplitude and the nonlinear parameters given only our prior information I. This term is called a prior probability and represents what was known about these parameters before we took the data. $P(D|\mathbf{B}, \mathbf{\Theta}, I)$ is the probability of the data D given the amplitude, the positions of the point sources and the prior information I. This term is often called a likelihood or sampling distribution. And $P(D|I)$ is the probability of the data D given only our prior information I. This term is a normalization constant and will be ignored in this calculation.

To proceed, we must supply both $P(\mathbf{B}, \mathbf{\Theta}|I)$ and $P(D|\mathbf{B}, \mathbf{\Theta}, I)$. For this problem we will assume we have little prior information; the prior probability will be effectively a constant over the range of values where the likelihood is sharply peaked. Consequently, the prior we use will make almost no difference. For this calculation we will take $P(\mathbf{B}, \mathbf{\Theta}|I)$ to be a uniform prior – for a more extensive discussion of uninformative priors see Zellner[5] and Bretthorst.[1] Thus to proceed we must compute the likelihood $P(D|\mathbf{B}, \mathbf{\Theta}, I)$. We can compute the likelihood of the data if we can assign a prior probability to the noise. We can assign a noise prior using maximum entropy once we identify what prior information we have. For this calculation we will assign a noise prior that is least informative for a given second moment of the noise, or noise power. Thus any other maximum entropy prior based on more information will always make more precise estimates than those that follow. The maximum entropy calculation is straightforward and gives

$$P(e_1, \cdots, e_N|\sigma, I) = \Pi_{i=1}^{N} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{e_i^2}{2\sigma^2}\right\}, \tag{5}$$

as the prior probability of the noise, where σ^2 is the given variance of the noise. For now we will assume σ^2 to be known; at the end of the calculation, if σ^2 is unknown, we can eliminate it from consideration or even estimate it. For a discussion of this important point, see Jaynes[6] and Bretthorst.[1,2,3] From the noise prior probability, Eq. (5), we can compute the joint posterior probability of all of the parameters:

$$P(\mathbf{B}, \mathbf{\Theta}|\sigma, D, I) \propto \sigma^{-N} \exp\left\{-\sum_{i=1}^{N} \frac{(d_i - f(x_i))^2}{2\sigma^2}\right\}. \tag{6}$$

in *Infrared Systems and Components III,* pp 93.104,
Robert L. Caswell *ed., SPIE* Vol. **1050,** 1989

Bayesian Analysis of Signals from Closely-Spaced Objects

G. Larry Bretthorst
Washington University, Department of Chemistry
1 Brookings Drive, St. Louis, Missouri 63130
and
C. Ray Smith
Radar Technology Branch, Advanced Sensors Directorate, Research, Development and Engineering Center
U. S. Army Missile Command, Redstone Arsenal, Alabama 35898-5253

Abstract

Bayesian probability theory is applied to the problem of the resolution of closely-spaced objects. The conditions assumed are: point sources, observed through a known smearing function (i.e., point-spread function). For this demonstration we use a Gaussian smearing function so that we can obtain analytic results; however, we present graphical results for both the Gaussian and the Airy smearing functions. The generalizations to arbitrary smearing functions may be found in other works by Bretthorst.[1,2,3] The results obtained for one and two point sources indicate explicitly the dependence of resolution on signal-to-noise and on the smearing function.

1 INTRODUCTION

In many important problems, we know the functional form of the signal received by the sensor; when this is so we say that we know the model for the signal. But even if we do not know the functional form of the signal, Bayesian probability theory is such that an adequate representation of the functional form can be determined from measurements. Thus the procedures we will be describing can be applied to any real devices that sample data at discrete intervals. Here we use probability theory to investigate how accurately one can resolve an isolated point source with Gaussian smearing in a simple one-dimensional telescope and then generalize to two point sources with Gaussian smearing. Additionally, numerical results are presented for Airy smearing.

2 THE MODEL

One advantage of a Bayesian probability calculation is that it requires one to state precisely what assumptions are being made in the calculation. Thus one always knows what problem is being solved. We assume the data may be separated into a systematic part $f(x)$, which we will call the signal, and a random part e_i, which we will call noise. We assume the signal and the noise are additive to make the data D. The data are assumed to be sampled at discrete positions which we will refer to as pixels. The model of the experiment may then be written as

$$d_i = f(x_i) + e_i \qquad (1 \leq i \leq N) \tag{1}$$

where $D \equiv \{d_1, \cdots, d_N\}$ is the discrete values of the data, $f(x_i)$ is the signal evaluated at the discrete pixel values, and e_i is noise. The pixel values x_i are not assumed uniformly spaced, nor do we assume these values to be integers. Indeed, the analysis does not depend upon what these values are; for other models they could equally well be time values. The signal $f(x)$ is assumed to be a sum over a set of signal functions. We will designate these signal functions as $G_j(x, \mathbf{\Theta})$ in such a way that the signal may be written as

$$f(x) = \sum_{j=1}^{m} B_j G_j(x, \mathbf{\Theta}), \tag{2}$$