The MaxEnt Method Applied to Spectral Analysis and Image Reconstruction

Suzette C. Lizamore
Institute of Statistics and Operations Research
Victoria University of Wellington
suzette@isor.vuw.ac.nz

Abstract

In this paper the Maximum Entropy technique is applied to two problems, the analysis of an unknown spectrum and the reconstruction of an image from tomographic data. In the spectral analysis case we seek to determine the frequencies present in a series of additive sine waves. The power of the MaxEnt technique to identify these frequencies accurately in the presence of noise and in the case where the frequencies are very close is shown in examples and compared to traditional Fourier Analysis.

In the image reconstruction case we seek to identify an underlying image from data based on randomly placed cross-sections of the image. The method is developed for situations analogous to tomography and some results of using the method are shown.

1 Introduction to Maximum Entropy

There are many situations in which we cannot measure the quantity (or image) we are interested in directly. Instead we obtain data on some transform of this image and then seek to infer the image from the data. This is known as an inverse problem. Unfortunately, the inverse transform may be intractable and in the case of incomplete or imperfect data, there is often not enough information in the data set to allow for the use of an inverse transform, or the inverse, if it exists, may be badly conditioned or not uniquely defined. According to Skilling [7], the maximum entropy technique (MaxEnt) is not just a good technique for such problems, it is the technique.

MaxEnt has proved to be a valuable tool in a diverse range of fields including astronomy, structural molecular biology, medical tomography, spectral analysis, theoretical physics (Skilling and Bryan [6]) and in restoring pictures taken by the flawed Hubble telescope (Alexander [1]).

Bayesian probability theory provides a unique mathematical language for dealing with inference and reasoning under uncertainty. Bayes theorem lies at the core of maximum entropy theory and takes the following form

\[
P(image|data) = \frac{P(image) \times P(data|image)}{P(data)} \propto P(image) \times P(data|image) \quad (1)
\]
The prior probability, \( P(\text{image}) \), encodes what is known about the image before the data is taken into account. The likelihood, \( P(\text{data} | \text{image}) \), represents the probability of obtaining the data given a certain image. The evidence, \( P(\text{data}) \), is the probability of obtaining the data no matter what the underlying image is. The posterior probability, \( P(\text{image} | \text{data}) \), is the revised probability of the image. \( P(\text{data}) \) is a normalisation constant and is often excluded, leaving Bayes theorem in the “proportional” format.

To keep the theory objective it is necessary that people with the same information about a problem assign the same probabilities. It is possible to construct a numerical measure of the uncertainty that is implicit in a probability assignment to a set of propositions. It takes a very simple form when the propositions are exhaustive and exclusive.

\[
S = -\sum p_n \log p_n \quad p_n = P(A_n | B)
\]

This is the entropy of a probability distribution. It is maximised with the value \( \log N \) when all probabilities are equal and minimized with the value 0 when one proposition is certainly true (Macaulay [5]).

The configurational entropy of the image — the entropy function applied to the image strengths \( f_n \) rather than to a probability distribution \( p_n \) is used to define the prior distribution of the image. At the moment no links exist to connect the entropy of a probability distribution (a measure of uncertainty) to the entropy of an image (a measure of probability).

### 2 Monkeys and MaxEnt

The proverbial monkey argument, described by Daniell [3], provides a good illustration of the theory of MaxEnt. An image can be regarded as divided into \( N \) cells. A team of monkeys throw \( M \) balls at the cells of the image randomly, with Poisson expectations \( \mu_i \) in each cell; the image intensity in cell \( i \), \( f_i \), is given by the number of balls in the cell \( (n_i) \). The Poisson expectations \( \mu_i \) form the underlying model, \( m_i \), of the image. The probability of the image relative to the underlying model is given by

\[
P(f_1, f_2, ... | m_1, m_2, ...) = \prod \frac{m_i^f_i e^{-m_i}}{f_i!}
\]

Using Stirling’s approximation \((f_i! = \sqrt{2\pi f_i f_i^e e^{-f_i}}\), the prior probability of the image is

\[
P(\text{image}) = P(f_1, f_2, ...) \propto \exp(\alpha S(f, m)), \quad (4)
\]

where \( \alpha \) is a constant and \( S(f, m) \) is the configurational entropy of the image.

\[
S(f, m) = \sum f_i - m_i \log \left( \frac{f_i}{m_i} \right) \quad (5)
\]

Given an image, the predicted values of the observations can be calculated using the response function of the data. From these predicted values, the residuals, \( r_i \) (differences between the predicted and true values of the data) are calculated. If the \( r_i \) are not zero, it is because of experimental errors in the data. If these errors have a Gaussian distribution, the likelihood distribution has the following form

\[
P(\text{data} | \text{image}) \propto \prod \exp(-r_i^2/2\sigma^2) = \exp(-\chi^2/2) \quad (6)
\]
Using Bayes theorem, the posterior distribution is therefore given by

\[ P(\text{image}|\text{data}) \propto \exp \left( \alpha S - \frac{1}{2} \lambda^2 \right) \] (7)

\( \alpha \) is a regularising parameter which controls the competition between the entropy and the likelihood, the larger \( \alpha \) is, the smoother the reconstruction. There are a number of ways of setting \( \alpha \) (see Gull [4]). The maximum entropy image is obtained by maximising \( P(\text{image}|\text{data}) \). This is a numerical analysis problem solved by an algorithm developed by Skilling and Bryan [6] that works on images of up to one million points. For the examples referred to in the paper I have used a modified version of this algorithm developed by Macaulay [5].

### 3 Spectral Analysis

Spectral estimation consists of fitting harmonic functions (sines and cosines) to a data set to determine the frequencies present, their amplitudes and their phases. The problem is formulated by Buck and Macaulay [2] as follows:

Measurements of the signal, \( S_t \), are made at discrete times \( t_p \) at each of which the datum might be modeled as containing an additive noise component. Let \( D_p \) represent the \( p^{th} \) datum and \( n_p \) the noise associated with \( D_p \). Then

\[ D_p = S(t_p) + n_p \] (8)

\( S_t \) can be expressed by the Hartley transform as

\[ S(t) = T \int_{-f_0}^{f_0} H(f) \cos(2\pi ft) df \] (9)

where \( H(f) \) is the Hartley Spectrum, \( T \) is the total time span and \( \cos(x) = \cos(x) + \sin(x) \)

The image of interest is the amplitude function of the spectrum defined by

\[ A(f_i) = \sqrt{H^2(f_i) + H^2(-f_i)} \quad f_i > 0 \] (10)

The prior is derived from the probability of the amplitude function (see Macaulay [5]) and is given by

\[ P(\text{image}) \propto \exp \left( \alpha \sum \sqrt{A_i^2 + m_i^2} - m_i - A_i \sinh^{-1} \frac{A_i}{m_i} \right) \] (11)

Knowledge of \( H(f > 0) \) and \( H(f < 0) \) is equivalent to determining the usual Fourier spectrum, the relative sizes of which are determined from a Gaussian likelihood function (6), equivalent to assuming a uniform prior on the phases. The posterior distribution takes the form of (7). The Hartley transform (9) is the response function for transition between image and data space.

The method has been tested on a number of series with multiple component harmonics of differing phases, amplitudes and wavelengths with additive Gaussian noise. A program provided by Simon McAuliffe (private communication) was used to generate the test data. The results presented are those for two different series. The first is a sum of three sine
wave series with amplitude ratios (AR) 2:3:1, wavelengths 19.5, 20 and 20.5 and phases of 30°, 270° and 180°. The second series is a sum of five sine wave series with amplitude ratios 1:2:1:2:1 and phases of 0°, 45°, 300°, 120° and 230° respectively.

The signal to noise ratios (SN) correspond to adding N(0,1) noise to a signal in the range of ±1. 2000 data points were used though good results can be obtained with as few as 300 points. The MaxEnt results are presented in the form of the Power density of the Hartley spectrum and are compared with the unsmoothed Fourier periodogram in figure 1.

Figure 1: MaxEnt and Periodogram spectral analysis of two time series with multiple component harmonics and Gaussian noise.
Figure 1 shows that in the first case, MaxEnt has identified the three frequencies clearly in the appropriate ratio of the square of the amplitudes (4:9:1). The periodogram identifies the high amplitude signal of wavelength 20, but does not give a good indication of what the other two frequencies are or what their amplitudes might be.

In the second case, MaxEnt has identified the 5 frequencies present, giving a fairly good idea of the amplitudes involved. The periodogram has also identified the frequencies involved, but presents a less comprehensive view of the situation.

The MaxEnt method works very well under test conditions of high noise, very close frequencies and varying amplitudes and phases. It has also been successfully applied to "real" data such as the Wolf sunspot index (see Buck and Macaulay [2]).

The model may be recovered by integrating over the peaks in the spectrum to give the overall amplitude and phase of the component frequencies. This is described by Macaulay [5].

4 Image Reconstruction

In this section, a method for the tomographic reconstruction of data is developed specifically for the case where the tomography is performed by fish trawling operations. The area of the sea (or image) is divided into $n$ smaller areas or blocks and $p$ trawls of random directions and lengths are run over the area of interest. The image or the density of fish in block $i$ is $f_i$, the total catch on trawl $p$ is $D_p$ and $n_p$ is the error associated with $D_p$. $R$ is the $p \times n$ response matrix where $R_{ij}$ represents the length of trawl $i$ in block $j$ of the image.

Theoretically we can calculate our expected catch (data) by multiplying our trawl distance matrix by the image, so that $D_p$ is given by the sum of the length of the trawl multiplied by the image density of the trawl in each block. We can therefore move from image space to data space in the following way

$$D_p = \sum_i R_{pi} f_i + n_p \quad (12)$$

A 10x10 trial image was used to test the effectiveness of the algorithm, the prior, likelihood and posterior distributions being those of (5), (6) & (7). MaxEnt produces an estimate of the image with very few tomography lines (or trawls). Running ten horizontal or vertical trawls results in a one dimensional view of the image — the relative densities of the rows or columns of the image. Sixty trawls produce a reasonable estimation of the 100 block image. One hundred trawls produce a very good estimation and with one hundred and twenty trawls, the image produced is near to perfect. The algorithm holds up reasonably well when noise is introduced, the basic shape of the image is still visible even under high levels of Gaussian noise as is seen in figure 2, where SN represents the signal to noise ratio.

Image reconstruction is more sensitive to noise than spectral analysis since more parameters require estimation, one hundred as opposed to three. We are now trying to apply the method to determine fish densities using Hoki trawling data from the 1992 season.
Figure 2: MaxEnt image reconstruction for data with decreasing signal to noise ratios.

References


