"A desision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible concequences, if it was unreasonable to expect those consequences."

Herodotus, 500 B.C.

The Maximum Entropy Method (MaxEnt):

— used in making inferences about a positive and additive distribution (*Scene*).

Bayes' Theorem:

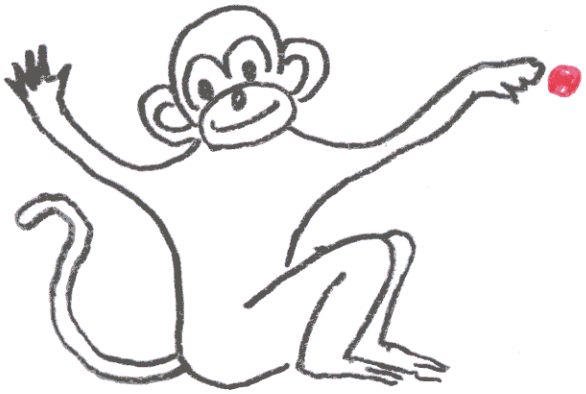
$$Pr(Scene \mid Data,I) = Pr(Scene \mid I)\ Pr(Data \mid Scene,I) \,/\, Pr(Data \mid I)$$

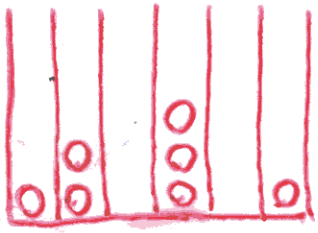
I is prior information.

Given some information about this distribution we should choose such distribution that is:
1) consistent with the given information (I);
2) most probable.

Monkey Example

n cells; N spheres (total); N_i spheres in i-th cell ($N = \Sigma_{i=1}^{n} N_i$).





$\mathbf{m} = (m_1, m_2, \ldots, m_n)$ are the prior probabilities for the spheres to get into corresponding cells.

Then the probability of a certain distribution $(N_1, N_2, \ldots, N_n)$ is:

$$Pr(N_1, N_2, \ldots, N_n|\mathbf{m}) = \frac{N!}{N_1!N_2!\ldots N_n!} m_1^{N_1} m_2^{N_2} \cdots m_n^{N_n}$$

For $x \to \infty$: $\log x! \approx x \log x + O(x)$ (Stirling's Formula). Then for $N, N_i \to \infty$:

$$\log Pr(N_1, N_2, \ldots, N_n|\mathbf{m}) \approx - N \sum_{i=1}^{n} (N_i/N) \log\left((N_i/N)/m_i\right)$$

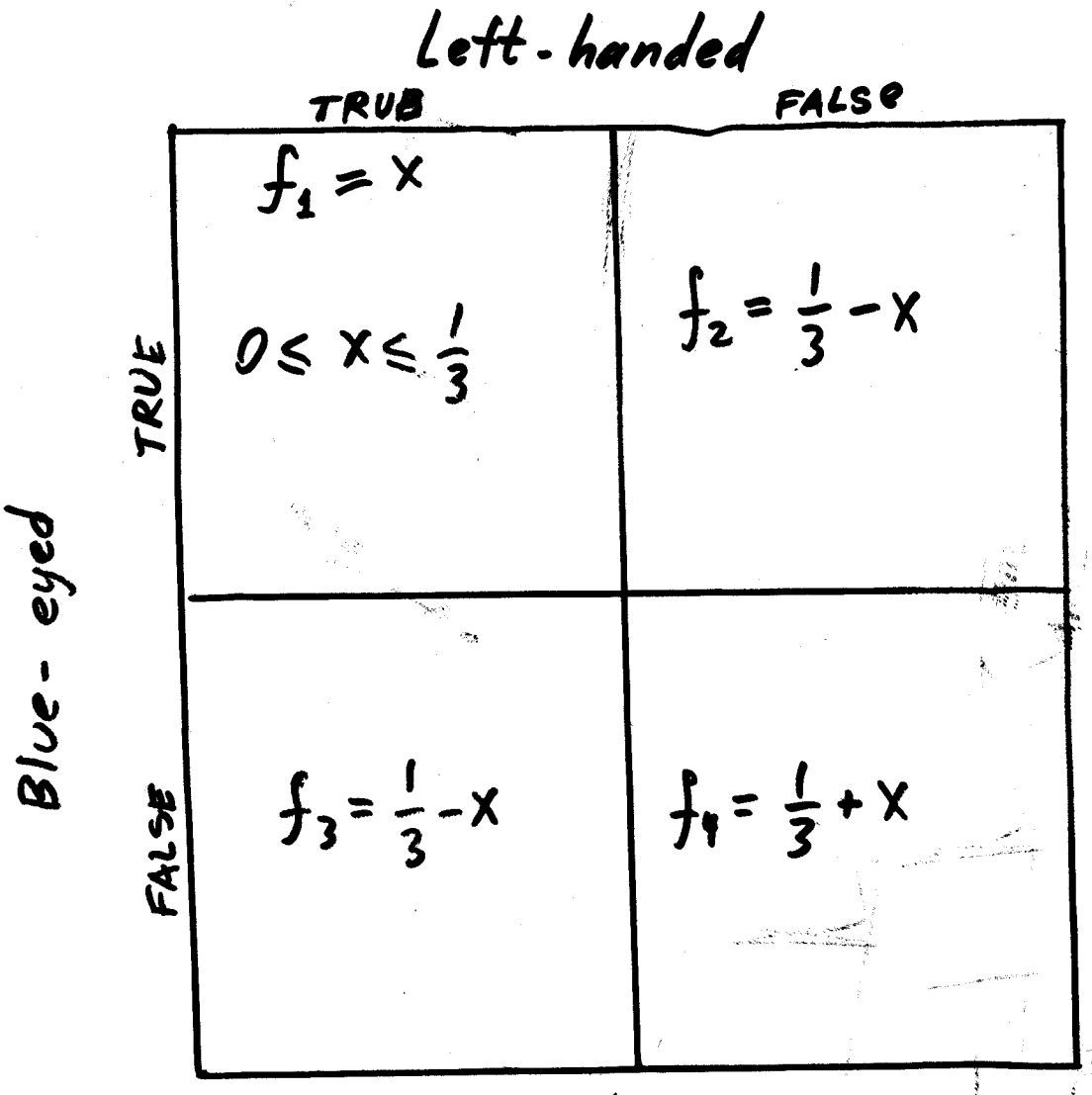
We define *entropy* S as

$$S = - \sum_{i=1}^{n} f_i \log\left(f_i/m_i\right) + const$$

where $f_i \propto N_i$ (in general). As we see,

$$Pr(Scene|I) = Pr(f_1, f_2, \ldots, f_n|\mathbf{m}, I) \propto exp(\alpha S).$$

The Kangaroo Problem.



$$S = -\sum_{i=1}^{4} f_i \log f_i$$

$$S \Rightarrow \max \quad \text{when} \quad x = 1/9$$

(no correlation)

Maximum-Entropy Formalism

Suppose we have some constrains on f_i.

Let's maximize

$$S(f) = -\sum_{i=1}^{n} f_i \log f_i$$

under constraints:

$$\sum_{i=1}^{n} A_{j,i} f_i = a_j, \quad j = 1, 2, \ldots, m; \quad \sum_{i=1}^{n} f_i = 1$$

$A_{i,j}$ is a "point spread function".

We then must maximize the Lagrangian

$$Q(\lambda_1, \ldots, \lambda_m) = -\sum_{i=1}^{n} f_i \log f_i - \lambda_1 \left[\sum_{i=1}^{n} A_{1,i} f_i - a_1\right] - \cdots -$$
$$- \lambda_m \left[\sum_{i=1}^{n} A_{m,i} f_i - a_m\right] - \lambda_0 \left[\sum_{i=1}^{n} f_i - 1\right]$$

The *Partition Function*:

$$Z(\lambda_1, \ldots, \lambda_m) = \sum_{i=1}^{n} \exp(-\lambda_1 A_{1,i} - \cdots - \lambda_m A_{m,i})$$

The Maximum-Entropy Distribution:

$$f_i = Z^{-1}(\lambda_1, \ldots, \lambda_m) \exp(-\lambda_1 A_{1,i} - \cdots - \lambda_m A_{m,i})$$

Also: *(if f_i is probability distribution)*

$$\langle A_j \rangle = \sum_{i=1}^{n} A_{j,i} f_i = a_j = -\frac{\partial}{\partial \lambda_j} \log Z$$

$$\langle A_j A_k \rangle - \langle A_j \rangle \langle A_k \rangle = \sum_{i=1}^{n} A_{j,i} A_{k,i} f_i - \sum_{i=1}^{n} A_{j,i} f_i \sum_{i=1}^{n} A_{k,i} f_i = \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} \log Z$$

Measures of Entropy

1. Shannon's Measure:

$$S(f) = -\sum_{i=1}^{n} f_i \log f_i,$$

or

$$S(f) = -\int f(x) \log f(x)\, dx.$$

2. Generalized Bayes-Shannon Measure:

$$S(f,m) = -\int \{f(x) - m(x) - f(x) \log[f(x)/m(x)]\}\, dx.$$

3. Norm with minus sign:

$$S(f) = -\sum_{i=1}^{n} f_i^2$$

4. Burg's Measure:

$$S(f) = \sum_{i=1}^{n} \log f_i$$

5.

$$S(f) = \sum_{i=1}^{n} \sqrt{f_i}$$

An Example from Radioastronomy

$$Pr(Scene|I) \propto exp(\alpha S).$$

$$Pr(Data|Scene, I) \propto \exp(-\chi^2/2)$$

$$\chi^2 = \sum_{j=1}^{m} \frac{|d_j - a_j|^2}{\sigma_j^2}$$

$$\Pr(Scene|Data, I) \propto \exp(\alpha S - \chi^2/2)$$

Lagrangian

$$Q(\lambda) = -\sum_{i=1}^{n} f_i \log f_i - \frac{\lambda}{2} \sum_{j=1}^{m} \frac{|d_j - a_j|^2}{\sigma_j^2}$$

$$\begin{cases} \frac{\partial Q}{\partial f_i} = 0 \\ \chi^2 = m \end{cases}$$

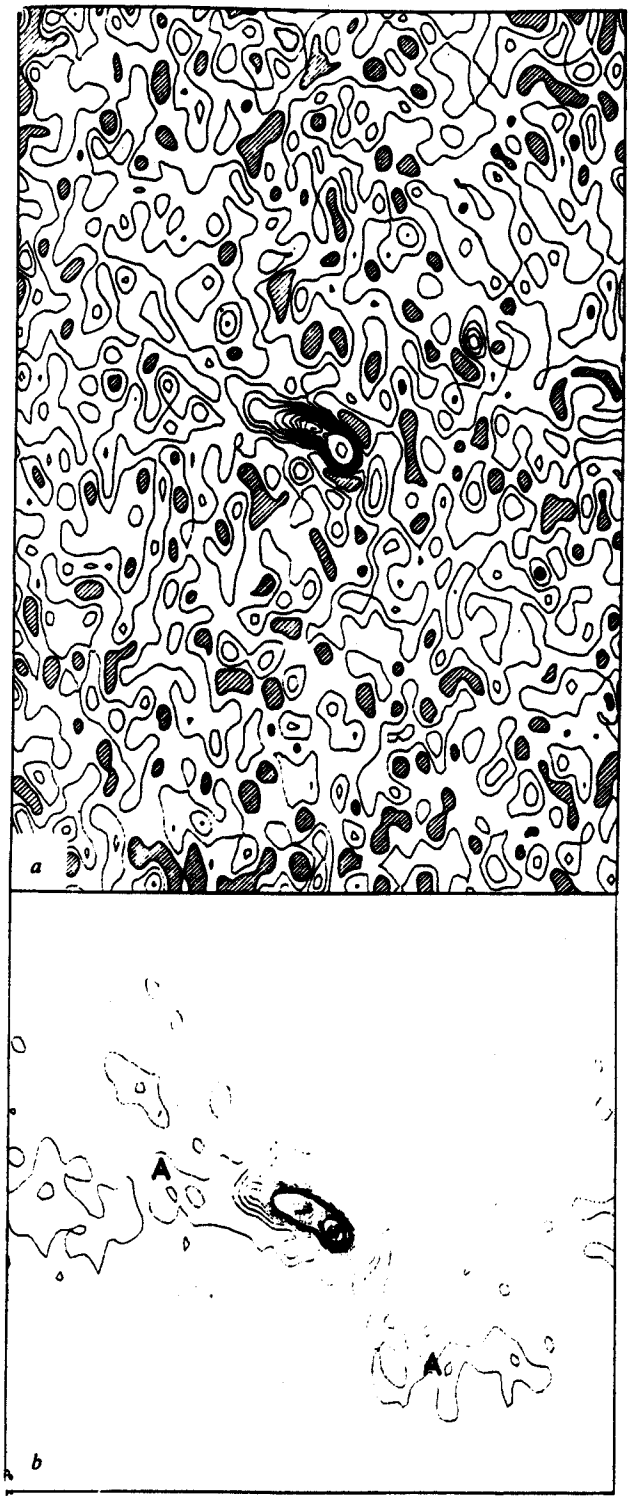


Fig. 2 *a*, Conventional map of 3C66. The contours are confused by fluctuations in the background, which are the result of receiver noise. Negative regions are shown hatched. *b*, Maximum-entropy map of 3C66. Regions of low surface brightness are now visible (A).

constant e^{-1} is related to the problem of the scaling of the data. This problem can be resolved by a more careful investigation of the foundations of the method.

Given E_k, σ_k and λ, this equation determines the intensities m_j on the map. It has been solved numerically by an iterative procedure, starting with the uniform map. Each new iterate is obtained by evaluating the right hand side of equation (1) with M_k equal to the Fourier transform of the previous iterate. We must average successive iterates to obtain convergence. The solution for a map of 128 × 128 grid points takes 3 min on the IBM 370/165 computer of the University Computer Laboratory, Cambridge. We can show that a solution of equation (1) must exist for any $\lambda > 0$, that it is unique and that it makes the quantity $Q(\lambda)$ a global maximum.

The correct value of λ makes χ^2 equal to the number of observations. During our iterative solution, λ is automatically increased until this is achieved by making it dependent on the current value of χ^2. In practice the solution is almost unchanged for a wide range of λ.

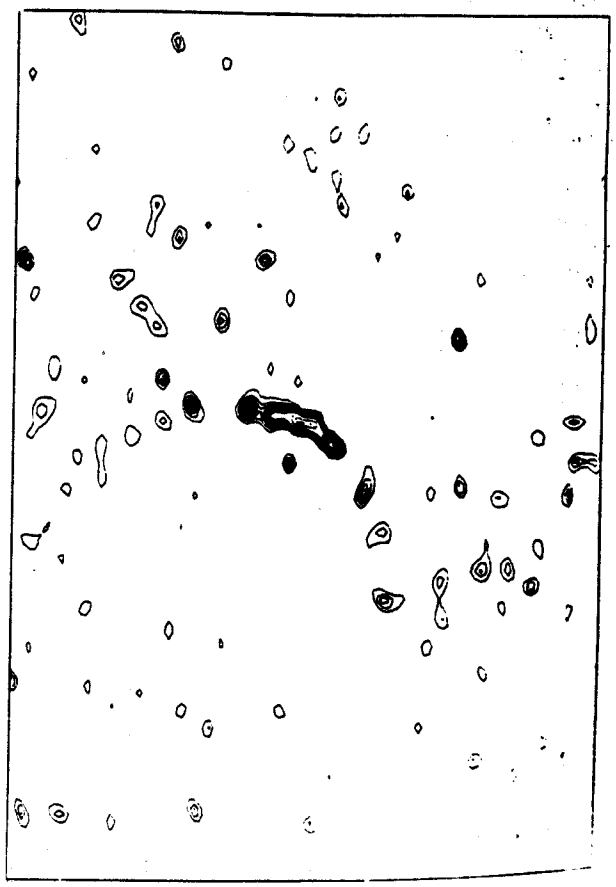
Applications using radioastronomical data

We have made extensive tests with simulated data. In every case the maximum-entropy reconstruction has been very similar to the original and we have found no cases where a misleading conclusion could have been drawn. We therefore present here some results using real observations made by the Cambridge radio telescopes. These data are incomplete because of the finite size of the telescope, because the data may be spoiled by electrical interference, or because the expense of further data collection outweighs the improvement to the final map of the sky. Errors in the data vary from sample to sample because of differing integration times and atmospheric conditions.

Figure 1*a* is a conventional map of the radio galaxy 3C31, observed at 1,407 MHz by the One-Mile Telescope[8]. The dynamic range is about 50:1. The shaded areas indicate negative regions, which arise principally as sidelobes of the main peak. Note the change in contour interval by a factor of 10 after the first 10 contours.

Figure 1*b* shows the maximum-entropy map produced with the same data. All negative sidelobes have been removed and positive ones greatly reduced, because these sidelobes arise from missing Fourier components and can be removed without affecting the fit to the data. The central peak is sharper and the noise background

Fig. 3 An example of the mis-use of the maximum-entropy method. The spurious features, which are not present in Fig. 2*b*, result from an over-optimistic estimate of the noise level in the data.



Literature used:

[1] Kapur, J.N. *Maximum Entropy Models in Science and Engineering*. John Wiley & Sons, 1989.

[2] Duck,B. and V.A. Macaulay, eds. *Maximum Entropy in Action*. Clarendon Press, Oxford, 1991.

[3] Gull, S.F. and G. J. Daniell (1978). *Image reconstruction from incomplete and noisy data.* Nature **272**, 686-690.

[4] Jaynes, E.T. (1985). *Where do we go from here?* In *Maximum Entropy and Bayesian Method in Inverse Problems* ed C.R. Smith and W.T. Grandy Jr. 21-58.

[5] Sivia, D.S. (1990). *Bayesian Inductive Inference, Maximum Entropy & Neutron Scattering*. Los Alamos Science, Summer.