Maximum entropy data analysis: another derivation of S-χ^2

1988 J. Phys. A: Math. Gen. 21 L63

(http://iopscience.iop.org/0305-4470/21/2/001)

J. Phys. A: Math. Gen. **21** (1988) L63–L65. Printed in the UK

LETTER TO THE EDITOR

Maximum entropy data analysis: another derivation of $S-\chi^2$

R Lieu

Astrophysics Group, Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

Received 9 November 1987

Abstract. An earlier work suggested a statistic $S-\chi^2$ as the key to maximum entropy data analysis in the simplest experimental situations. The present letter reinforces such a claim by offering an alternative proof. Slightly more elaborate inference procedures are also treated, together with a critique of currently available techniques.

A recent paper (Lieu *et al* 1987a) addresses the problem of maximum entropy as a method of inference in the analysis of data from measurements which are inevitably subject to errors. The reasoning yielded a statistic $S-\chi^2$, which is a precise expression for the entropy. The procedure of maximum $S-\chi^2$ generates a unique output distribution, and applications to simulated data having a variety of signal–noise combinations were carried out. Here we seek a more complete mathematical justification of such an undertaking.

The basic problem concerns the best estimate of a set of probabilities

$$\{p_i,\ i=1,2,\ldots,r\} \qquad \sum_{i=1}^{r} p_i = 1$$

subject to some prior knowledge

$$\{\bar{p}_i,\ \pm\sigma_i,\ i=1,2,\ldots,r\} \qquad \sum_{i=1}^{r} \bar{p}_i = 1$$

where $\{\bar{p}_i\}$ are the experimental data and $\{\sigma_i\}$ their associated random errors (not necessarily Gaussian).

Following Boltzmann, we solve the problem by imagining that the game of probabilities is repeated N times, where N is large, and working out the most likely distribution

$$\{n_i = Np_i,\ i=1,2,\ldots,r\} \qquad \sum_{i=1}^{r} n_i = N.$$

According to the central limit theorem, the distribution of an individual variate n_i is given by

$$P(n_i) = (2\pi N)^{-1/2}\sigma_i^{-1}\exp\left(-\frac{(n_i-\bar{n}_i)^2}{2N\sigma_i^2}\right)$$

0305-4470/88/020063+03\$02.50

and this is valid even though p_i may have random errors which are non-Gaussian. Hence, the conditional probability of having a distribution identical to the data, $\{\bar{n}_i, i=1,2,\ldots,r\}$, given a hypothesis population $\{n_i, i=1,2,\ldots,r\}$, is the following:

$$P(\text{data/hypothesis}) = (2\pi N)^{-r/2}(\sigma_1\sigma_2\ldots\sigma_r)^{-1}\exp\left(-\sum_{i=1}^{r}\frac{(n_i-\bar{n}_i)^2}{2N\sigma_i^2}\right).$$

What is needed, however, is P(hypothesis/data). This is related to the former by $P(\text{hypothesis/data}) \propto P(\text{data/hypothesis}) \times P(\text{hypothesis})$ where the constant of proportionality is $1/P(\text{data})$. Now the probability of a hypothesis distribution is obtainable by standard entropy arguments (i.e. enumeration of outcomes) as

$$P(\text{hypothesis}) = \frac{1}{r^N}\frac{N!}{n_1!n_2!\ldots n_r!}.$$

The most likely distribution in the presence of the available information is then determined by maximising $P = P(\text{hypothesis/data})$, or

$$P = \frac{N!}{n_1!n_2!\ldots n_r!}\exp\left(-\sum_{i=1}^{r}\frac{(n_i-\bar{n}_i)^2}{2N\sigma_i^2}\right)$$

(note that some constants of propotionality are omitted here). Using Stirling's formula of $N!$ in $\log P$, it is quite easy to show that

$$S_d = S - \chi^2$$

where

$$S = -\sum_{i=1}^{r} p_i \log p_i$$

and

$$\chi^2 = \sum_{i=1}^{r}\frac{(p_i-\bar{p}_i)^2}{2\sigma_i^2}$$

is the correct statistic to maximise.

Although this result was obtained previously, it is possible to extend the theory a little. Thus instead of requiring the entire set of probabilities $\{p_i, i=1,2,\ldots,r\}$, we may also find the best estimate of $p = p_1 + p_2$ quite simply by not distinguishing between the first and second bins. Then

$$P(\text{data/hypothesis}) = (2\pi N)^{-(r-1)/2}(\sigma_1^2+\sigma_2^2)^{-1/2}(\sigma_3\ldots\sigma_r)^{-1}$$
$$\times\exp\left(-\frac{(n-\bar{n}_1-\bar{n}_2)^2}{2N(\sigma_1^2+\sigma_2^2)}-\sum_{i=3}^{r}\frac{(n_i-\bar{n}_i)^2}{2N\sigma_i^2}\right)$$

and

$$P(\text{hypothesis}) = \frac{1}{(r-1)^N}\frac{N!}{n!n_3!\ldots n_r!}$$

where $n = n_1 + n_2$. The unique solution of $\{p = p_1 + p_2, p_3, \ldots, p_r\}$ is obtained by maximising the statistic

$$-\left(p\log p + \sum_{i=3}^{r} p_i \log p_i + \frac{(p-\bar{p}_1-\bar{p}_2)^2}{2(\sigma_1^2+\sigma_2^2)} + \sum_{i=3}^{r}\frac{(p_i-\bar{p}_i)^2}{2\sigma_i^2}\right).$$

The procedure can be generalised to the sum of any combinations of the p_i.

The author does not agree with the manner in which problems of a similar nature have been tackled in contemporary literature. The Gull–Daniell–Skilling method gives the false impression that the quality of image data can be significantly improved at no expense. Both the rationale and actual performance of this method have been criticised (Lieu *et al* 1987a, b). Another recent work proposes the maximum of $NS-\chi^2$ (Jaynes 1984). This is surprising because, N, the number of (imaginary) trial runs, is large beyond measure. The error lies in the calculation of P(data/hypothesis). This is proportional to $e^{-N\chi^2}$, not $e^{-\chi^2}$, because each fictitious trial is equivalent to one measurement of the probability distribution. After N measurements, the standard deviation must be reduced by $\sqrt{N}$.

References

Jaynes E T 1984 *4th Ann. Workshop on Bayesian/Maximum Entropy Methods in Geophysical Inverse Problems, Calgary*

Lieu R, Hicks R B and Bland C J 1987a *J. Phys. A: Math. Gen.* **20** 2379

—— 1987b *Mon. Not. R. Astron. Soc.* **229** 49P