Astron. Astrophys. 150, 273–275 (1985)

ASTRONOMY AND ASTROPHYSICS

Data analysis in gamma-ray astronomy: multivariate likelihood method for correlation studies

A. W. Strong

Max-Planck Institut für Extraterrestrische Physik, D-8046 Garching-bei-München, Federal Republic of Germany

Received January 10, accepted March 22, 1985

Summary. A framework for a general likelihood analysis of correlation problems typified by that of gamma rays with other components of the galaxy is presented. The multivariate 'information matrix' contains the essential results for any particular combination of datasets, and its use solves the general problem of how to deal with the interdependence of the errors on the model parameters. By diagonalizing the information matrix and performing various operations any particular set of parameters can be isolated analytically, and explicit formulae for doing this are derived.

The presentation has the advantage of being *concise* and *practical*; it is therefore proposed as a standard for communication of results in this field.

Key words: gamma rays – data analysis

1. Introduction

The study of correlations between gamma ray emission and other constituents of the galaxy is an essential part of the interpretation of such data. Unfortunately the subject has suffered from a "credibility gap" arising from differing methodologies and failure to use adequate statistical methods for what is in essence a problem in multivariate statistics. In their analysis of the COS-B data, the Caravane Collaboration has consistently used the likelihood function in such studies (Lebrun et al., 1982, 1983, Strong et al., 1982, Bloemen et al., 1984a, b) as well as in point source analyses (Pollock et al., 1981, 1985) and this has contributed to the aim of a well-defined and rigorous analysis. The essential point is that once the input data and parameters (the "model") have been defined the likelihood function contains all the essential information for this particular choice of data and model, and use can be made of theorems on the distribution of the likelihood ratio to draw conclusions, for example, on the errors on the parameters. Since the calculation of the likelihood function is not in doubt the discussion of the results is greatly clarified. Clearly it is always necessary to vary the choice of input data and make tests based on physical knowledge in order to judge conclusions based on such techniques, but a correct treatment of the statistical part of the problem is still a prerequisite.

Although the likelihood function is easy to present (e.g. as contours) when the number of variables is very small (1–3), it is impossible when number increases: current studies using 21-cm and CO data with velocity information as well as other components have typically 8 parameters, each of which is necessary to the model and which can be determined from the correlation analysis with varying degrees of accuracy and independence from the other parameters. A general formalism is required to summarize *precisely* the result of such an analysis: an elegant solution is *via* the diagonalized information matrix.

Send offprint requests to: A. W. Strong

2. The information matrix

Suppose the log-likelihood ratio (denoted here by L) is a function n parameters θ_i $(i=1-n)$. Then L can be expanded about its maximum value L_0 in terms of the offsets $\vartheta_i=\Delta\theta_i$ of the parameters from their maximum-likelihood values:

$$L=L_0+\frac{1}{2}\sum\sum\frac{\partial^2 L}{\partial\vartheta_i\,\partial\vartheta_j}\,\vartheta_i\vartheta_j\,. \tag{1}$$

The matrix $H_{ij}=\frac{1}{2}(\partial^2 L/\partial\vartheta_i\,\partial\vartheta_j)$ is called the *information matrix*. The parameter ϑ_i is independently determined only if $H_{ij}=0$ for all $j\neq i$. Using H we can specify the interdependence of all the parameters in the n-dimensional space spanned by ϑ. H_{ij} can be diagonalized by the transformation

$$H^*=Q^{-1}HQ \tag{2}$$

where Q is the matrix of eigenvectors of H. Because Q is symmetric, Q is orthogonal $(Q^{-1}=Q^T)$. In terms of the basis set ϑ the eigenvectors are given by $\underline{e}=Q\underline{\vartheta}$. In terms of the transformed variables the likelihood expansion becomes

$$L=L_0+\sum\lambda_i x_i^2 \tag{3}$$

where λ_i are the eigenvalues of H and where

$$x_i=\sum Q_{ji}\vartheta_j \tag{4}$$

are the components of the offsets of the original parameters in the transformed variables. The eigenvectors are the axes of the likelihood ellipsoid and the lengths of the axes are proportional to the square roots of the eigenvalues.

The advantage of this formalism is that L can be analytically maximised over any subset of the parameters, keeping others fixed as required.

3. Use of the diagonal information matrix

3.1. 1-parameter errors

An important case in practice is the determination of the error on a single parameter when all the others are free. This requires maximising L over all parameters but one, i.e.

maximise $\Delta L = \sum \lambda_i x_i^2$ (5)

subject to $\sum_j Q_{ij} x_j = \vartheta_i$ (6)

Introducing the Lagrangian multiplier Λ we therefore maximise

$$\sum \lambda_j x_j^2 + \Lambda \sum_j Q_{ij} x_j \quad (7)$$

over all x_j giving

$$x_j = -\Lambda \frac{Q_{ij}}{2\lambda_j} \quad (8)$$

Using (6) we get

$$\Lambda = -\frac{\vartheta_i}{\sum_j \frac{Q_{ij}^2}{2\lambda_j}}$$

Substituting in (5) the error estimate corresponding to a given ΔL is

$$\vartheta_i = \left[\Delta L \sum_j \frac{Q_{ij}^2}{\lambda_j} \right]^{1/2} \quad (9)$$

This result avoids the necessity to maximise L over $n-1$ parameters numerically for many values of ϑ_i as is normally done. If one or more of the parameters θ_k is assumed to be known the errors on the other parameters are given by equation (9) omitting terms in λ_k. The errors are then smaller, reflecting the reduction in uncertainty on the ϑ_k.

3.2. Errors on linear combinations of parameters

It is often desirable to obtain limits on linear combinations of parameters; for example we may wish to know the error on the difference of two parameters. The analysis of section (3.1.) can be extended straightforwardly to this case as follows. The required linear combination can be written:

$$\sum a_i \vartheta_i = c \quad (10)$$

where we require the error on the value of c. Using (6) and introducing the Lagrangian multiplier as before we maximise

$$\sum \lambda_j x_j^2 + \Lambda \sum \sum Q_{ij} a_i x_j$$

over x_j giving

$$x_j = -\Lambda \sum_i \frac{Q_{ij} a_i}{2\lambda_j}.$$

Proceeding as before one obtains after some algebra

$$\Delta c = \left[\Delta L \sum \sum \sum \frac{Q_{ij} Q_{kj} a_i a_k}{\lambda_j} \right]^{1/2} \quad (11)$$

which reduces to (9) when $a_i = \delta_{ij}$.

3.3. Projection operators

More generally one can "project out" each parameter ϑ_k in turn, each time reducing the dimensionality of the parameter space by 1. The projection corresponds to maximising over ϑ_k and it can be shown that the projected information matrix H'_{ij} becomes

$$H'_{ij} = H_{ij} - \frac{H_{ik} H_{jk}}{H_{kk}} \quad (i, j \neq k)\,. \quad (12)$$

The magnitude of the terms in the information matrix decreases, reflecting the implicit uncertainty in ϑ_k, as required. By repeating this process any subspace of parameters can be immediately isolated. As a check, note that application of (12) $n-1$ times leads to a scalar which agrees with (9).

3.4. Quality indicator

The n-dimensional hypervolume of the likelihood ellipse, measured by the determinant of H, gives a measure of the "quality" of the analysis since the smaller the volume the more tightly constrained are the parameters. This is useful for comparing different datasets or different treatments of the data. It can also be used to compare the quality of results reported by different authors in the same field.

4. Statistical interpretation

It is not the intention in this paper to review the interpretation of the likelihood function; the reader is referred to Eadie et al. (1982) and Cash (1979) for details. The essential result is that the asymptotic distribution of twice the likelihood ratio ($2\Delta L$ in the present notation) is χ_p^2 where p is the number of parameters held constant in the maximisation. It should be noted that the likelihood ratio is not the only possible way to use the information matrix; alternatives are the use of the statistics $\underline{\vartheta}^T I \underline{\vartheta}$ and $(\partial L/\partial \underline{\vartheta})^T I^{-1} (\partial L/\partial \underline{\vartheta})$ where I is the information matrix at the true values of the parameters, which can be approximated by H as previously described. From the fact that $\underline{\vartheta}$ and $\partial L/\partial \underline{\vartheta}$ have covariance matrices I^{-1} and I respectively, it follows that both these statistics are distributed as χ_p^2. The general advantage of the likelihood ratio is that the "true" information matrix is not required since the theorems refer to the observable H.

5. Example of application of method

The method will be illustrated by its application to the correlation between gamma rays and gas tracers at intermediate galactic latitudes. The data used are described in Strong (1985) to which the reader is referred for details. Three energy ranges were treated in that paper but here I shall consider only the 70–150 MeV range for illustration. There are four parameters in the model: 1. q_1: the emissivity of the atomic gas in units of $10^{-26}\,\mathrm{sr}^{-1}\,\mathrm{s}^{-1}$, 2. q_2: the emissivity of the molecular gas in the same units, 3. I_B^0: the "on-axis" background intensity in units of $10^{-5}\,\mathrm{cm}^{-2}\,\mathrm{sr}^{-1}\,\mathrm{s}^{-1}$, 4. f_{ICS}: a dimensionless factor for the inverse Compton component. A discussion of the maximum-likelihood values themselves is given in Strong (1985); here I discuss only the error estimates. An analysis of the most cos-B data using the present technique can be found in Strong et al. (1985).

Table 1(a) gives the information matrix H, the matrix of eigenvectors Q and the eigenvalues λ_i. The single-parameter errors ($\Delta L = 1$) on the parameters computed as described in Sect. 3.1 are given. Since in this case $2\Delta L$ has a χ_1^2 distribution the "1σ" errors correspond to $\Delta L = 0.5$, and are therefore obtained

Table 1

(a) Information matrix H

155.3409	27.6098	81.5078	32.3011
27.6098	27.1571	18.1488	15.0374
81.5078	18.1488	50.1310	19.7353
32.3011	15.0374	19.7353	13.8205
Matrix of eigenvectors Q (norm = 10,000)			
−1	−4493	−2943	−8434
−3795	2262	−8776	−1856
−2265	−8558	−242	−4642
8970	−1203	−3774	−1959
Eigenvalues			
2.4662	5.3220	24.8670	213.7943
1-Parameter errors			
q_1	q_2	I_B^0	f_{ICS}
0.21	0.31	0.40	0.58

(b) Errors on linear combinations

$\frac{(q_1+q_2)}{2}$	0.202
q_1-q_2	0.35

(c) The 6 2-Dimensional projections (norm = 100) with eigenvalues

	q_1		q_2		I_B^0	
q_2	17	99				
	99	−17				
	9.92	23.2				
I_B^0	−43	90	0	100		
	90	43	100	0		
	5.18	90.3	6.27	10.1		
f_{ICS}	−5	100	−40	91	−30	95
	100	5	91	40	95	30
	2.98	22.7	2.55	23.9	2.82	7.16

from the values given by applying a factor 0.71 (see Eq (9)). To illustrate the use of errors on linear combinations of parameters described in Sect. (3.2), Table 1(b) gives the result for the combinations $(q_1+q_2)/2$ and (q_1-q_2), i.e. the average emissivity and difference in emissivity of the atomic and molecular components. The projection technique of Sect. 3.3 is illustrated by the projection onto the 6 possible 2-D subspaces of the parameters; Table 1(c) shows the corresponding eigenvector matrices and eigenvalues.

The interdependence of the pairs of parameters can be immediately seen: for example q_1 and q_2 are fairly independent while q_2 and f_{ICS} are quite strongly correlated. The accuracy of the method depends on the validity of the approximation of Eq (1). This can be tested by comparing the exact evaluation of the likelihood with that given by (1) for some typical value of ϑ. Taking each parameter to have its maximum likelihood value plus $\frac{1}{4}$ of its ($\Delta L=1$) error gave $\Delta L=4.05$ (exact evaluation) and 4.17 (approximation (1)). The difference is too small to have any appreciable effect on the error estimates and we can conclude that the method is completely satisfactory. (This test should be made for each application since it may fail if the errors are comparable in magnitude to the parameters themselves.)

6. Conclusions

The use of the information matrix is a simple and effective way to summarise the results of a correlation study in cases where the number of variables is too large to allow presentation of the complete likelihood function. All the standard operations involving the use of the likelihood ratio to derive errors and their interdependence can be carried out using this matrix, and explicit formulae for these have been derived. Because of this possibility it is a concise and efficient method of recording the outcome of such a study; unlike other methods *it allows the reader to check the conclusions and perform further analysis if required.* For these reasons it is proposed that it be adopted as a standard way of communicating results of correlation analyses in gamma-ray astronomy and related fields.

References

Bloemen, J. B. G. M., Bennett, K., Bignami, G. F., Blitz, L., Caraveo, P. A., Gottwald, M., Hermsen, W., Lebrun, F., Mayer-Hasselwander, H. A., Strong, A. W., 1984a, *Astron. Astrophys.* **135**, 12

Bloemen, J. B. G. M., Caraveo, P. A., Hermsen, W., Lebrun, F., Maddalena, R. J., Strong, A. W., Thaddeus P.: 1984b, *Astron. Astrophys.* **139**, 37

Cash, W.: 1979, *Astrophys. J.* **228**, 939

Eadie, W. T., Drijard, D., James, F. E., Roos, M., Sadoulet, B.: 1982, Statistical Methods in Experimental Physics, North-Holland Amsterdam, p. 230

Lebrun, F., Bignami, G. F., Buccheri, G., Caraveo, P. A., Hermsen, W., Kanbach, Mayer-Hasselwander, H. A., Paul, J. A., Strong, A. W., Wills, R. D.: 1982, *Astron. Astrophys.* **107**, 390

Lebrun, F., Bennett, K., Bignami, G. F., Bloemen, J. B. G. M., Buccheri, R., Caraveo, P. A., Gottwald, M., Hermsen, W., Kanbach, G., Mayer-Hasselwander, H. A., Montmerle, T., Paul, J. A., Sacco, B., Strong, A. W., Wills, R. D., Dame, T. M., Cohen, R. S., Thaddeus, P.: 1983, *Astron. Astrophys.* **274**, 231

Pollock, A. M. T., Bignami, G. F., Hermsen, W., Kanbach, G., Lichti, G. G., Masnou, J. L., Swanenburg, B. N., Wills, R. D.: 1981, *Astron. Astrophys.* **94**, 116

Pollock, A. M. T., Bennett, K., Bignami, G. F., Bloemen, J. B. G. M., Buccheri, R. Caraveo P. A., Hermsen, W., Kanbach, G., Lebrun, F., Mayer-Hasselwander, H. A., Strong, A. W.: 1985, *Astron. Astrophys.* **146**, 352

Strong, A. W., Bignami, G. F., Bloemen, J. B. G. M., Buccheri, R., Caraveo, P. A., Hermsen, W., Mayer-Hasselwander, H. A., Paul, J. A., Wills, R. D.: 1982, *Astron. Astrophys.* **115**, 404

Strong, A. W.: 1985 *Astron. Astrophys* **145**, 81

Strong, A. W., Bloemen, J. B. G. M., Hermsen, W., Mayer-Hasselwander, H. A.: 1985, *Proc. XIX Int. Cosmic Ray Conf.* OG 3.1-3