1976A&A....52..307C

Astron. & Astrophys. 52, 307–308 (1976)

ASTRONOMY
AND
ASTROPHYSICS

Research Note

Generation of Confidence Intervals for Model Parameters in X-ray Astronomy

W. Cash

Space Science Laboratory and Department of Physics, University of California, Berkeley, California 94720

Received May 10, revised July 24, 1976

Summary. We prove that a procedure which utilizes the χ^2 statistic to generate confidence intervals for parameter values of models being fit to observed data is applicable to models which are nonlinear as well as those which are linear and hence is a suitable tool for X-ray astronomy. The procedure is applicable to the construction of a confidence interval on subsets of the model parameters as well as the full set.

Key words: parameter estimation — error analysis — X-ray astronomy

Recently there has been considerable discussion in the astronomical literature regarding the problem of generating confidence intervals for the values of parameters of models being fitted to observed X-ray spectral data (Margon et al., 1975; Kellogg et al., 1975; Lampton et al., 1976; and Avni, 1976). A statistic these authors have found appropriate for many applications in X-ray astronomy is

$$S=\sum_{i=1}^{N}\frac{(D_i-F_i)^2}{F_i}$$

where D_i is the number of observed counts in bin i and F_i is the number of counts predicted by the model for bin i. It is well known that the parameter values which yield the minimum S for a given data set are the values most likely to be the correct values, but procedures for finding the extent of the acceptable range of values are less well understood.

One method which uses the S statistic to generate a joint confidence region for all parameters of a model is: first, find the minimum value of S when the model is fitted to the data and the p parameters used in calculating F_i are varied; then find the locus of points in parameter space for which S is greater than the minimum by an amount $\chi_p^2(\alpha)$. This increase, $\chi_p^2(\alpha)$, is the value at which the variate of the χ^2 function with p degrees of freedom has significance α [i.e. probability α of being greater than $\chi_p^2(\alpha)$]. It is well known that for linear models the region enclosed by such a locus will contain with probability $1-\alpha$ the set of true parameters (cf. Eadie et al., 1971, p. 207). Recently, Lampton et al. (1976) have proved that this method is also applicable to nonlinear models.

Avni (1976) employed an extension of this procedure which provides for generation of intervals for some subset of the parameters when the model is linear. He suggested that the procedure might also hold for nonlinear models and performed numerical simulations to demonstrate the procedure's validity for a specific nonlinear case, absorption in extra-galactic X-ray source spectra. In the present paper we show that an application of a theorem by Wilks (1938) furnishes a general proof of the applicability of Avni's assertion to models, linear and nonlinear.

Let us investigate a model which has p parameters $\theta_1, \ldots, \theta_p$. Of these let us say that $\theta_{q+1}, \ldots, \theta_p$ are uninteresting, but necessary, for the purposes of the experiment (e.g. background intensity). Let $\theta_1^T, \ldots, \theta_p$ exist and be the true values of the parameters values which are known only to nature. If we constrain $\theta_1, \ldots, \theta_q$ to be $\theta_1^T, \ldots, \theta_q^T$ and vary only the parameters $\theta_{q+1}, \ldots, \theta_p$, when we fit the model to the data we will find some value of S which is a minimum in the $p-q$ dimensional subspace. Call this value $(S_{\min})_{p-q}^T$. If we vary all the parameters $\theta_1, \ldots, \theta_p$ we will find some set of p parameters which will generate a minimum value of S when fit to the data. Call this value $(S_{\min})_p$. Now, the theorem of Wilks establishes with remarkable generality that the value of the quantity $(S_{\min})_{p-q}^T-(S_{\min})_p$ will be distributed as χ^2 with q degrees of freedom, as the number of observed events becomes large.

This directly implies that $(S_{\min})_{p-q}^T-\chi_q^2(\alpha)$ will exceed $(S_{\min})_p$ with probability α. Or, the set of points in the q dimensional parameter space which has $(S_{\min})_{p-q}-(S_{\min})_p$ greater than $\chi_q^2(\alpha)$ will contain $\theta_{q+1}^T, \ldots, \theta_p^T$ with probability α.

Thus we have the theorem:

In the limit where the number of observed events becomes large, the true value of a q parameter subset of the p parameters of the model will, with probability $1-\alpha$, be enclosed by the locus of all points in the q dimensional parameter space where $(S_{\min})_{p-q}=(S_{\min})_p + \chi_q^2(\alpha)$.

As model linearity has at no point been required, this proof is valid for nonlinear as well as linear models. In the case where all parameters are interesting ($p=q$) this theorem reduces to the formulation of Lampton et al. for the generation of joint confidence intervals. When $p \neq q$ the theorem proves the formulation hypothesized by Avni: If at each point in the space of q interesting parameters S is minimized with respect to the $p-q$ uninteresting parameters, the formalism of Lampton et al. for the joint estimation of q parameters may be applied.

The suitability of the S statistic and the applicability of Wilks' theorem place some restrictions on the general application of the procedure. Lampton et al. have presented a detailed discussion of suitability of the S statistic for the joint estimation of parameters, and the restrictions on the above procedure are identical. Still, two points of caution need emphasis here. First, the finite sample sizes used in this technique will generate a statistic S which will be distributed as $[\chi^2+0(1/\sqrt{n})]$ where $0(1/\sqrt{n})$ is a term of size $\sim 1/\sqrt{n}$ and n is the number of events (photons) observed in the data. Thus care should be exercised in the interpretation of S when n is low (<20) or when S changes very slowly with the parameters. Second, the model must be checked to verify that everywhere the procedure is to be applied, the values of the best fit parameters in question converge to their true values as the number of events observed becomes large, i.e. all the parameters of the model must be finite and well defined everywhere in the region of interest.

Acknowledgements. I wish to thank R. Malina and M. Lampton for the important conversations that led to the eventual formulation of this work. This project was supported by NASA Grant NGR 05-003-450.

References

Avni, Y.: 1976, *Astrophys. J.* (in press)
Eadie, W. T., Drijard, D., James, F. E., Roos, M., Sadoulet, B.: 1971, North-Holland Publishing Co., Amsterdam
Kellogg, E. M., Baldwin, J. R., Koch, D.: 1975, *Astrophys. J.* **199**, 299
Lampton, M., Margon, B., Bowyer, S.: 1976, *Astrophys. J.* **208** (in press)
Margon, B., Lampton, M., Bowyer, S., Cruddace, R.: 1975, *Astrophys. J.* **197**, 25
Wilks, S. S.: 1938, *Ann. Math. Stat.* **9**, 60