Parameter estimation in X-ray astronomy in the Chandra and XMM-Newton era

K.A. Arnaud[1]

Laboratory for High Energy Astrophysics, NASA/GSFC, Greenbelt MD 20771

[1]also Astronomy Department, University of Maryland

ABSTRACT

We build on the work of Wachter, Leach & Kellogg (1979) and of Cash (1979) to develop a statistic that can be used for parameter estimation when modelling *Chandra* and *XMM-Newton* data. This statistic is particularly useful when there are a small number of photons detected in both the source and background observations. We illustrate the use of the statistic in fitting models to X-ray spectra from *Chandra* and propose a new way of plotting such spectra.

Subject headings: methods: statistical — X-rays: general

1. Introduction

Many applications in X-ray astronomy require fitting some parameterized model to a dataset. We take our favored theoretical model, convolve it with the instrumental response to make a simulated dataset, compare the simulated to real dataset using some statistic, then vary the model parameters till the best match of simulated to real dataset is achieved. This is most familiar in the analysis of X-ray spectra but the same technique can be used on images or timing data.

In the early years of X-ray astronomy there was some debate about the appropriate statistic to use to find the "best-fit" values of the parameters and their confidence regions (Cruddace et al. 1972; Margon et al. 1975; Kellogg, Baldwin & Koch 1975). The fit statistic now routinely used is

$$S = \sum_i (S_i - B_i - M_i)^2 / (\sqrt{(\sigma_S)_i^2 + (\sigma_B)_i^2})$$

where S_i are the source observation counts in the $i = \{1, , N\}$ data bins, B_i the background observation counts, and M_i the simulated model counts. $(\sigma_S)_i^2$ and $(\sigma_B)_i^2$ are the variance on the source and background counts, respectively. The variances are customarily estimated by S_i and B_i. S is usually referred to as the "χ^2 statistic" although it has only a tenuous link to the original χ^2 statistic defined by Pearson and Fisher (Babu & Feigelson 1996).

The S statistic is familiar and is easy to calculate but is not without its difficulties. Wheaton et al. (1995) show that in the linear problem (ie where the model depends linearly on the parameters to be estimated), using the number of observed counts as the variance introduces a systematic bias. These authors present a way to estimate the variance that does not introduce a bias and show that in this case the least-square estimator is the best possible. The non-linear problem is more complicated and is the focus of this letter.

It is well-known that the S statistic fails when there are few counts in each data bin. A typical response has been to rebin the data so that each data bin contains a large enough number of counts. This loses information and may introduce a bias depending on the exact rebinning method. An alternative to rebinning is to modify S so that it performs better in the low count regime. Working by analogy from the linear problem, the failure of the S statistic is likely due to estimating the variance by the counts. A downward fluctuation in the counts will be weighted more heavily than a comparable upward fluctuation, biassing the model downwards. Churazov et al. (1996) proposed to escape from this problem by estimating the variance for a given data bin by the average counts from surrounding bins. Their Monte Carlo simulations demonstrated that this worked well for the cases they considered.

While it is possible to work with a modified S, doing so is justified by Monte Carlo simulations, which must be performed for each problem under consideration. A better solution is to follow standard statistical methods and construct a maximum-likelihood estimator based on the Poisson distribution of the detected counts. This was done by Cash (1979) in a much-cited paper. While Cash's statistic (C) has been widely used, a common argument against it is that the data cannot be background-subtracted, requiring models to be generated to include the effects of background. This objection was overcome by Wachter, Leach & Kellogg (1979). Their paper is much less well-known than that of Cash, probably because Wachter et al. did not frame their results in a form familiar to X-ray astronomers. The purpose of this brief letter is firstly to bring Wachter, Leach & Kellogg (1979) to greater attention and secondly to reformulate its results in more familiar guise. By combining the the insights of Cash and of Wachter et al. we show how to perform parameter estimation in the low count limit when background is significant. Our results are simple to apply to existing software and have been implemented in XSPEC.

The method presented here follows the Frequentist statistical paradigm. van Dyk et al. (2000) have recently discussed a Bayesian analysis approach. We will provide our own Bayesian solution in a subsequent paper.

2. Maximum-likelihood estimator

Suppose we have an observation which produces S_i events in the $i = \{1, , N\}$ data bins (e.g. spectral channels, image elements, or time bins) in an exposure time of t_s. This observation includes events from the source of interest along with background events. Further suppose that we perform a background observation which generates B_i events in an exposure time t_b. Then, if the observed counts are Poisson variables we can write the joint probability in bin i as :

$$P_i = (t_s m_i + t_s b_i)^{S_i} (t_b b_i)^{B_i} e^{-t_s m_i - t_s b_i - t_b b_i} / S_i! / B_i!$$

where m_i and b_i are the true mean rates for the source and background.

Multiplying these probabilities for all N bins we get the likelihood, L. For convenience we take the logarithm of L and then multiply by -2 to get a statistic that we wish to minimize :

$$-2\log L = 2\sum\{t_s m_i + (t_s + t_b)b_i - S_i \log(t_s m_i + t_s b_i) - B_i \log(t_b b_i) + \log S_i! + \log B_i!\}$$

Since we don't care about the background we model it by N parameters, β_i, such that $b_i = \beta_i$. Now we solve for the β_i by minimizing $-\log L$ through $\partial(-\log L)/\partial\beta_i = 0$ to give a quadratic in b_i whose solution is :

$$\hat{b}_i = \frac{S_i + B_i - (t_s + t_b)m_i + a_i}{2(t_s + t_b)}$$

where :

$$a_i = \sqrt{[(t_s + t_b)m_i - S_i - B_i]^2 + 4(t_s + t_b)B_i m_i}$$

and the positive root is chosen to ensure that b_i is positive. Now let us write $f_i(m_i) = \hat{b}_i$ and substitute back into the equation for the log likelihood to give the new statistic to be minimized :

$$-2\log L = 2\sum\{t_s m_i + (t_s + t_b)f_i - S_i \log(t_s m_i + t_s f_i) - B_i \log(t_b f_i) + \log S_i! + \log B_i!\}$$

J. Castor (1995, private communication) has pointed out that Cash's C statistic can be converted to a likelihood ratio by dividing the likelihood of the current model by the likelihood of a perfect model (where the model exactly equals the observation). This ratio is asymptotically distributed as χ^2. The same method can be used with our current work, leading to the new statistic :

$$W = 2\sum\{t_s m_i + (t_s + t_b)f_i - S_i \log(t_s m_i + t_s f_i) - B_i \log(t_b f_i) - S_i(1 - \log S_i) - B_i(1 - \log B_i)\}$$

A second-order Taylor expansion can be used to show that in the limit of large numbers of counts/bin W is equal to :

$$\sum \frac{[S_i - t_s m_i - t_s f_i]^2}{t_s m_i + t_s f_i} + \frac{[B_i - t_b f_i]^2}{t_b f_i}$$

which is distributed as χ^2 with $(N-M)$ degrees of freedom, where the model m_i has M parameters (including the normalization).

It is straightforward to modify the Levenberg-Marquardt algorithm (Press et al. 1992) to minimize W and hence derive the "best-fit" parameters[1]. To calculate the confidence regions we use the insight of Cash (1979) and note that the difference between the minimum value of W when all parameters are free to vary and the minimum value of W when a subset of size P of the parameters are fixed is distributed as χ^2 with P degrees of freedom[2]. We can, therefore, derive confidence regions based on W by the standard method used with the S or C statistics (Avni 1976; Press et al. 1992).

3. Spectral fitting example

As an example we consider the analysis of a *Chandra* ACIS-S X-ray spectrum. We also take the opportunity to present a new way of plotting data with small numbers of photons covering a wide energy range. Our example source is taken from a 100 ksec observation of the Hawaii SSA13 field (Mushotzky et al. 2000). For the source chosen we collect about 50 photons in the $0.5-7.0$ keV energy range within the selected region of radius $3''$. Approximately half of these events are source and the rest background. We have a background spectrum extracted from a large source-free region of the detector.

We model the source as a power-law modified by photo-electric absorption due to matter at the redshift (2.625) of the source. Using the W statistic we find the best-fit values

---

[1]Note that while the S and C statistics have analytic solutions for the model normalization this is not available for W so the convergence to the best-fit can be slow if the initial parameters are chosen badly.

[2]This is true to $O(n^{-1/2})$ where n is the total number of photons

for the parameters then calculate the 90% confidence interval on the intrinsic column to be $1.1 \times 10^{22} - 1.5 \times 10^{23}$ cm^{-2}. In comparison, if we use the S statistic and rebin the data to have at least 10 counts in each bin we lose enough information that we can only constrain the intrinsic column to be less than 3.6×10^{23} cm^{-2}.

Figure 1 shows the spectrum and the best-fit model in an unfamiliar guise. With as few as 50 photons over the energy range $0.5 - 7$ keV none of the standard ways of plotting the spectrum yield anything which can be easily interpreted. If we plot all the channels then the entire plot is covered by error bars but if we rebin to produce points with reasonable size error bars then we lose too much information. Our solution is to produce an integral plot almost exactly as though we were performing a Kolmogorov-Smirnov test. The only modification is that we increment the plot at each energy there is a source photon but also decrement it for a photon in the background spectrum. The result is a background-subtracted integral counts plot. We normalize the graph to unity to make it easier to show multiple spectra on the same plot.

4. Conclusions

We have argued, based on the work of Wachter, Leach & Kellogg (1979), that the maximum likelihood estimator for a source in the presence of background provides a straightforward way to perform parameter estimation when fitting models to *Chandra* and *XMM-Newton* data. Indeed, since the statistic W asymptotes to something distributed as χ^2 there is no reason not to use it for all data whose only source of variance is counting statistics.

In this letter we have considered only the problem of finding best estimates of the model parameters and assigning confidence regions. This assumes that the model used is correct. Babu & Feigelson (1996) provide a discussion of methods to determine whether the model and data are consistent. We lean towards the use of Monte Carlo methods : given the model simulate the data many times and find the fraction of simulations which give a W value worse than that obtained with the real data.

We thank Eric Feigelson for spurring us on to take statistics more seriously, John Castor for communicating his modification to Cash's C statistic, and the referee, William Wheaton, for clarifying for us the reason why the S statistic fails.

REFERENCES

Avni, Y. 1976, ApJ, 210, 642

Babu, G.J. & Feigelson, E.D. 1996 Astrostatistics (London:Chapman & Hall)

Cash, W. 1979, ApJ, 228, 939

Churazov, E., Gilfanov, M., Forman, W., Jones, C. 1996, ApJ, 471, 673

Cruddace, R., Bowyer, S., Lampton, M., Mack, J. & Margon, B. 1972, ApJ, 174, 529

Kellogg, E., Baldwin, J.R. & Koch, D. 1975, ApJ, 199, 299

Margon, B., Lampton, M., Bowyer, S. & Cruddace, R. 1975, ApJ, 197, 25

Mushotzky, R.F., Cowie, L.L., Barger, A.J. & Arnaud, K.A. 2000, Nature, 404, 459

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992, Numerical Recipes: The Art of Scientific Computing (Cambridge: Cambridge University Press)

van Dyk, D.A., Connors, A., Kashyap, V.L. & Siemiginowska, A. 2000, ApJ, in press (astro-ph/0008170)

Wachter, K., Leach, R. & Kellogg, E. 1979, ApJ, 230, 274

Wheaton, W.A., Dunklee, A.L., Jacobson, A.S., Ling, J.C., Mahoney, W.A. & Radocinski, R.G. 1995, ApJ, 438, 322

This manuscript was prepared with the AAS LaTeX macros v5.0.

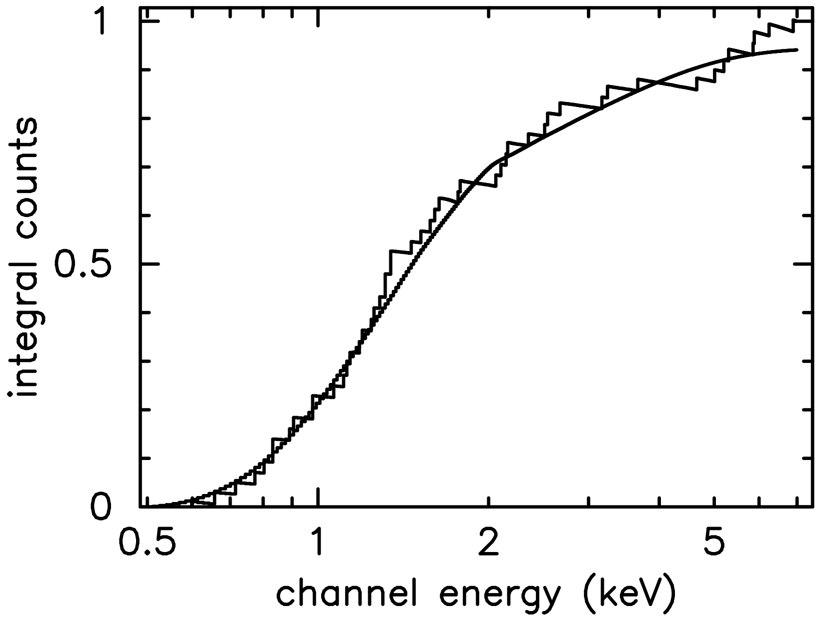

Fig. 1.— An example integrated counts plot (jagged curve) and best-fit model. This source has intrinsic photo-electric absorption.