

## Fitting with few counts/bin

### Theory

#### ***No background***

Cash (1979) showed that the  $\chi^2$  minimization criterion is a very bad one if any of the observed data bins had few counts. A better criterion is to use a likelihood function:

$$C = 2 \sum_{i=1}^N y(x_i) - y_i \ln y(x_i) + \ln y_i !$$

where  $y_i$  are the observed data and  $y(x_i)$  the values of the function. Minimizing C for some model gives the best-fit parameters. Furthermore, this statistic can be used in the same, familiar way as the  $\chi^2$  statistic to find confidence intervals. One finds the parameter values that give

$$C = C_{min} + N$$

where N is the same number that gives the required confidence for the number of interesting parameters as for the  $\chi^2$  case.

Castor (priv. comm.) has pointed out that a better function to use is :

$$C = 2 \sum_{i=1}^N y(x_i) - y_i + y_i (\ln y_i - \ln y(x_i))$$

This differs from the first function by a quantity that depends only upon the data. In the limit of a large number of counts this second function does provide a goodness-of-fit criterion similar to that of  $\chi^2$  and it is now used in XSPEC. It is important to note that the C-statistic assumes that the error on the counts is pure Poisson, and thus it cannot deal with data that already has been background subtracted, or has systematic errors.

#### ***With background***

Arnaud (2003, ApJ submitted) has extended the method of Cash to include the case when a background spectrum is also in use. Note that this requires the source and background spectra to both be available, it does not work on a background-subtracted spectrum.

Suppose we have an observation which produces  $S_i$  events in the  $i = \{1, N\}$  spectral 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  $t_b$ . If the source count rate in bin  $i$  is  $y_i$  then the new fit statistic is

$$W = 2 \sum_{i=1}^N \{ t_s y_i + (t_s + t_b) f_i - S_i \log(t_s y_i + t_s f_i) - B_i \log(t_b f_i) - S_i (1 - \log S_i) - B_i (1 - \log B_i) \}$$

where

$$f_i = \frac{S_i + B_i - (t_s + t_b) y_i \mp d_i}{2(t_s + t_b)}$$

and

$$d_i = \sqrt{[(t_s + t_b) y_i - S_i - B_i]^2 + 4(t_s + t_b) B_i y_i}.$$

The sign of  $d_i$  in  $f_i$  is chosen so that  $f_i > 0$ .

In the limit of large numbers of counts/bin a second-order Taylor expansion shows that  $W$  tends to

$$\sum_{i=1}^N \left( \frac{[S_i - t_s y_i - t_s f_i]^2}{t_s y_i + t_s f_i} + \frac{[B_i - t_b f_i]^2}{t_b f_i} \right)$$

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

## Practice

### ***No background***

XSPEC's default minimization algorithm is a variant of Marquardt's method (the Levenberg-Marquadt algorithm) described in §11.5 of *Data Reduction and Error Analysis for the Physical Sciences* by Bevington (2002). (The reader is advised that this description is designed to be read in conjunction with Bevington.) This method is appropriate for fitting models that are twice differentiable. We illustrate the implementation of this  $W$  statistic using this method. The algorithm works by finding a matrix  $\alpha_{jk}$  and a vector  $\beta_k$  such that the equation :

$$\beta_k = \sum_j \delta a_j \alpha_{jk}$$

gives sensible values of the change in parameters,  $\delta a_j$ , for the fitting function. Bevington §11.4 gives the derivation of  $\alpha$  and  $\beta$  and shows that  $\beta$  is parallel to the gradient of  $\chi^2$ .

Now the C-statistic has a gradient with respect to the parameters of the fitting function of :

$$-\frac{1}{2} \frac{\partial C}{\partial a_k} = \sum_i \left( \frac{y_i}{y} - 1 \right) \frac{\partial y}{\partial a_k} = \beta_k$$

So, following Bevington, expand  $y(x_i)$  about  $y_0$  :

$$y(x_i) = y_0(x_i) + \sum_j \frac{\partial y_0(x_i)}{\partial a_j} \delta a_j.$$

substitute into C and minimize with respect to the changes in the parameters :

$$\frac{\partial C}{\partial \delta a_k} = 2 \sum_i \left[ \frac{\partial y_0(x_i)}{\partial a_k} - y_i \left( y_0(x_i) + \sum_j \frac{\partial y_0(x_i)}{\partial a_j} \delta a_j \right)^{-1} \frac{\partial y_0(x_i)}{\partial a_k} \right]$$

setting this to zero we obtain, to first order in the parameter changes,

$$\sum_i \left( \frac{y_i}{y_0(x_i)} - 1 \right) \frac{\partial y_0(x_i)}{\partial a_k} = \sum_j \left( \sum_i \frac{y_i}{y_0^2(x_i)} \frac{\partial y_0(x_i)}{\partial a_j} \frac{\partial y_0(x_i)}{\partial a_k} \delta a_j \right)$$

or :

$$\beta_k = \sum_j \delta a_j \alpha_{jk}$$

where:

$$\alpha_{jk} = \sum_i \frac{y_i}{y_0^2(x_i)} \frac{\partial y_0(x_i)}{\partial a_j} \frac{\partial y_0(x_i)}{\partial a_k}$$

These  $\alpha$  and  $\beta$  then are substituted for those used in the  $\chi^2$  case and the algorithm works as required. Note that  $\alpha_{jk}$  is  $-\partial^2 C / \partial a_k \partial a_j$  to first order in partial derivatives in  $y$ , evaluated at  $y_0$ .

There is one further difference in XSPEC between the  $\chi^2$  and likelihood methods, which is caused by the fact that XSPEC uses an analytic formula for setting the model normalization. In the  $\chi^2$  case, this means multiplying the current model by:

$$\sum_i \frac{y_i y(x_i)}{\sigma_i^2} \bigg/ \sum_i \frac{y(x_i)^2}{\sigma_i^2}$$

where  $\sigma_i$  is the error on  $y_i$ . In the likelihood case the corresponding factor is:

$$\sum_i y_i \bigg/ \sum_i y(x_i)$$

### ***With background***

An analogous argument to the above can be followed through for the  $W$  statistic. We need the partial derivatives of  $W$  which are evaluated as follows.

$$\frac{\partial W}{\partial a_j} = 2 \sum_i \left\{ t_s + g_i(t_s + t_b) - \frac{S_i(1 + g_i)}{y_i + f_i} - \frac{B_i g_i}{f_i} \right\} \frac{\partial y_i}{\partial a_j}$$

$$\frac{\partial^2 W}{\partial a_j \partial a_k} = 2 \sum_i \left\{ (t_s + t_b) h_i - \frac{S_i h_i}{y_i + f_i} - \frac{S_i(1 + g_i)^2}{(y_i + f_i)^2} - \frac{B_i h_i}{f_i} - \frac{B_i g_i^2}{f_i^2} \right\} \frac{\partial y_i}{\partial a_j} \frac{\partial y_i}{\partial a_k}$$

where

$$g_i \frac{\partial y_i}{\partial a_j} = \frac{\partial f_i}{\partial a_j} = \frac{1}{2d_i} ((t_s + t_b)y_i - S_i + B_i - d_i) \frac{\partial y_i}{\partial a_j}$$

and

$$h_i \frac{\partial y_i}{\partial a_j} = \frac{\partial g_i}{\partial a_j} = \frac{2(t_s + t_b)S_i B_i}{d_i^3} \frac{\partial y_i}{\partial a_j}$$

Note that in this case there is no analytic formula that can be used to set the model normalization.

### **References**

- Arnaud, K. A. 2003, *Astrophysical Journal*, submitted
- Bevington, P. 2002 *Data Reduction and Error Analysis for the Physical Sciences* Third Edition (McGraw Hill:Columbus)

Cash, W. 1979 Parameter estimation in astronomy through application of the likelihood ratio  
Astrophysical Journal 228, 939