




---

## Sherpa Statistics

---

The following statistics are available in *Sherpa*:

- BAYES
  - CASH
  - $\chi^2$  statistics:
    - ◆ CHI CVAR | CHI PARENT
    - ◆ CHI DVAR
    - ◆ CHI GEHRELS
    - ◆ CHI MVAR
    - ◆ CHI PARENT | CHI CVAR
    - ◆ CHI PRIMINI
  - CSTAT
  - USERSTAT
- 

### BAYES

A Bayesian maximum likelihood function.

In conventional fits involving background data, it is the best-fit background model amplitude in bin  $i$ ,  $B_i$ , that is used when one attempts to ascertain the source model amplitude  $S_i$ . While this is the most probable value of  $B_i$ , it is not the only possible value, and  $S_i$  may change if all possible values of  $B_i$ , weighted by their likelihood, are taken into account. This can be done using Bayes' Theorem. Below, we assume familiarity with the basics of Bayesian statistical methodology; a reader who is not familiar with these basics should consult, e.g., the review text by Loredo (1992, *Statistical Challenges in Modern Astronomy*, ed. Feigelson & Babu, 275).

For simplicity, assume that there are two fitting regions (in time, space, or both), denoted "off-source" and "on-source," with  $N_{i,B}$  and  $N_{i,S}$  counts respectively. Counts are sampled from the Poisson distribution, and the best way to assess the quality of model fits is the use the Poisson likelihood function. Within the on-source region, the Poisson likelihood of the sum  $S_i + B_i$  is

$$p(N_{i,S} | S_i, B_i) = \frac{(S_i + B_i)^{N_{i,S}}}{N_{i,S}!} \exp(-S_i - B_i).$$

We can relate this likelihood to the Bayesian posterior density for  $S_i$  and  $B_i$  using Bayes' Theorem:

$$p(S_i, B_i | N_{i,S}) = p(S_i | B_i) p(B_i) p(N_{i,S} | S_i, B_i) / p(D).$$

The factor  $p(S_i | B_i)$  is the Bayesian prior probability for the source model amplitude, which is assumed to be constant, and  $p(D)$  is an ignorable normalization constant. The prior probability  $p(B_i)$  is treated differently; we can specify it using the posterior probability for  $B_i$  off-source:

$$p(B_i) = \left( A \frac{(AB_i)^{N_{i,B}}}{N_{i,B}!} \right) \exp(-AB_i),$$

where  $A$  is an "area" factor that rescales the number of predicted background counts  $B_i$  to the off-source region.

*IMPORTANT: this formula is derived assuming that the background is constant as a function of spatial area, time, etc. If the background is not constant, the Bayes function should not be used.*

To take into account all possible values of  $B_i$ , we integrate, or marginalize, the posterior density  $p(S_i, B_i | N_{i,S})$  over all allowed values of  $B_i$ :

$$p(S_i | N_{i,S}) = \int_0^{\infty} p(S_i, B_i | N_{i,S}) dB_i.$$

For the constant background case, this integral may be done analytically. We do not show the final result here; see Loredo. The function  $-\log p(S_i | N_{i,S})$  is minimized to find the best-fit value of  $S_i$ . The magnitude of this function depends upon the number of bins included in the fit and the values of the data themselves. Hence one cannot analytically assign a goodness-of-fit measure to a given value of this function. Such a measure can, in principle, be computed by performing Monte Carlo simulations. One would repeatedly sample new datasets from the best-fit model, and fit them, and note where the observed function minimum lies within the derived distribution of minima. (The ability to perform Monte Carlo simulations is a feature that will be included in a future version of *Sherpa*.)

## Note on Background Subtraction

*The background should not be subtracted from the data when this function is used.* The background only needs to be specified, as in this example:

```
sherpa> DATA 1 source.data
sherpa> BACK 1 background.data
sherpa> SOURCE 1 = [sourcemodel]
sherpa> STATISTIC BAYES
sherpa> FIT
...
```

To compare the best-fit source model with the data in a plotting environment, one would have to subtract the background from the raw counts data, as in this example:

```
...
sherpa> FIT
...
sherpa> SUBTRACT
sherpa> PLOT FIT
```

Otherwise, the plotted best-fit model and data will differ by approximately the background amplitude.

If the number of counts in each bin is high ( $> 5$ ), the likelihood is approximately proportional to  $\exp(-\chi^2/2)$ , Hence UNDERFLOW errors can occur if the initial parameter estimate is too far from the location of a local likelihood maximum, or if parameter step sizes are set too large. (This is because during intermediate steps of the computation of  $-\log p(S_i | N_{i,S})$ , absolute likelihoods must be computed, unlike for the case of the CASH statistic, where the log-likelihood terms never need to be exponentiated.) To avoid these errors, it may be necessary to use a  $\chi^2$  statistic to find the approximate solution before using the Bayesian maximum likelihood function.

## Examples

1. Specify the fitting statistic and then confirm it has been set. The optimization method is then changed from "Levenberg–Marquardt" (the default), since that algorithm does not work with this statistic.

```
sherpa> STATISTIC BAYES
sherpa> SHOW STATISTIC
Statistic:          Bayes
sherpa> METHOD POWELL
```

---

## CASH

A maximum likelihood function.

Counts are sampled from the Poisson distribution, and so the best way to assess the quality of model fits is to use the product of individual Poisson probabilities computed in each bin  $i$ , or the likelihood  $\mathcal{L}$ :

$$\mathcal{L} = \prod_i \frac{M_i^{D_i}}{D_i!} \exp(-M_i)$$

where  $M_i = S_i + B_i$  is the sum of source and background model amplitudes, and  $D_i$  is the number of observed counts, in bin  $i$ .

The CASH statistic (Cash 1979, ApJ 228, 939) is derived by (1) taking the logarithm of the likelihood function, (2) changing its sign, (3) dropping the factorial term (which remains constant during fits to the same dataset), and (4) multiplying by two:

$$C = 2 \sum_i [M_i - D_i \log M_i]$$

The factor of two exists so that the change in CASH statistic from one model fit to the next, " $C$ ", is distributed approximately as " $\chi^2$ " when the number of counts in each bin is high ( $> 5$ ). One can then in principle use " $C$ " instead of " $\chi^2$ " in certain model comparison tests. However, unlike  $\chi^2$ , the CASH statistic may be used regardless of the number of counts in each bin.

The magnitude of the CASH statistic depends upon the number of bins included in the fit and the values of the data themselves. Hence one cannot analytically assign a goodness-of-fit measure to a given value of the CASH statistic. Such a measure can, in principle, be computed by performing Monte Carlo simulations. One would repeatedly sample new datasets from the best-fit model, and fit them, and note where the observed CASH statistic lies within the derived distribution of CASH statistics. (The ability to perform Monte Carlo simulations is a feature that will be included in a future version of *Sherpa*.)

## Note on Background Subtraction

*The background should not be subtracted from the data when this statistic is used. It should be modeled simultaneously with the source, as in this example:*

```
sherpa> DATA source.data
sherpa> BACK background.data
sherpa> SOURCE = [source model]
```

```
sherpa> BG = [background model]
sherpa> STATISTIC CASH
sherpa> FIT
```

## Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CASH
sherpa> SHOW STATISTIC
Statistic:          Cash
```

## $\chi^2$ statistics:

Chi-square statistic.

The chi-square statistic is

$$\chi^2 = \sum_i \frac{[N_{i,S} - B_i(x_i, p_B) - S_i(x_i, p_S)]^2}{\sigma_i^2}$$

where  $N_{i,S}$  is the total number of observed counts in bin  $i$  of the on-source region;  $B_i(x_i, p_B)$  is the number of predicted background model counts in bin  $i$  of the on-source region (zero for background-subtracted data), rescaled from bin  $i$  of the off-source region, and computed as a function of the model argument  $x_i$  (e.g., energy or time) and set of background model parameter values  $p_B$ ;  $S_i(x_i, p_S)$  is the number of predicted source model counts in bin  $i$ , as a function of the model argument  $x_i$  and set of source model parameter values  $p_S$ ; and  $\tilde{A}_i$  is the error in bin  $i$ .

The options for assigning  $\tilde{A}_i$  are described in the documentation for [CHI\\_DVAR](#), [CHI\\_GEHRELS](#), [CHI\\_MVAR](#), and [CHI\\_PRIMINI](#). In each of these files,  $N_{i,B}$  is the total number of observed counts in bin  $i$  of the off-source region;  $A_B$  is the off-source "area," which could be the size of the region from which the background is extracted, or the length of a background time segment, or a product of the two, etc.; and  $A_S$  is the on-source "area."

In the analysis of PHA data,  $A_B$  is the product of the `BACKSCAL` and `EXPTIME` FITS header keyword values, provided in the file containing the background data.  $A_S$  is computed similarly, from keyword values in the source data file.

Note that in the current version of *Sherpa*, it is assumed that there is a one-to-one mapping between a given background region bin and a given source region bin. For instance, in the analysis of PHA data, it is assumed that the input background counts spectrum is binned in exactly the same way as the input source counts spectrum, and any filter applied to the source spectrum automatically applied to the background spectrum. This means that currently, the user cannot, e.g., specify arbitrary background and source regions in two dimensions and get correct results. This will be changed in a future version of *Sherpa*.

(However, this limitation only applies when analyzing background data that have been entered with the `BACK` command. One can always enter the background as a separate dataset and jointly fit the source and background regions.)

## CHI CVAR | CHI PARENT

Chi-square statistic with constant variance computed from the counts data.

In some applications, analysts have seen fit to assume that the variance is constant for each bin. For this choice of statistic, the variance is assumed to be the mean number of counts, or

$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^N N_{j,S} + \left( \frac{A_S}{A_B} \right)^2 N_{j,B},$$

where  $N$  is the number of on-source (and off-source) bins included in the fit. The background term appears only if a background region is specified *and* background subtraction is done.

See the section on [C<sup>2</sup> statistics](#) for more information, including definitions of the additional quantities shown in the equation.

### Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CHI CVAR
sherpa> SHOW STATISTIC
Statistic:           Chi-Squared Constant Variance
```

## CHI DVAR

Chi-square statistic with variance computed from the data.

If the number of counts in each bin is large ( $> 5$ ), then the shape of the Poisson distribution from which the counts are sampled tends asymptotically towards that of a Gaussian distribution, with variance

$$\sigma_i^2 = N_{i,S} + \left( \frac{A_S}{A_B} \right)^2 N_{i,B}.$$

The background term appears only if a background region is specified *and* background subtraction is done. See the section on [C<sup>2</sup> statistics](#) for more information, including definitions of the additional quantities shown in the equation.

### Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CHI DVAR
sherpa> SHOW STATISTIC
Statistic:           Chi-Squared Data Variance
```

## CHI GEHRELS

Chi-square statistic with the Gehrels variance function.

This is the *Sherpa* default statistic.

If the number of counts in each bin is small ( $< 5$ ), then we cannot assume that the Poisson distribution from which the counts are sampled has a nearly Gaussian shape. The standard deviation (i.e., the square-root of the variance) for this low-count case has been derived by Gehrels (1986):

$$\sigma_{i,S} = 1 + \sqrt{N_{i,S} + 0.75}.$$

Higher-order terms have been dropped from the expression; it is accurate to approximately one percent. If one does not perform background subtraction, then  $\tilde{A}_i = \tilde{A}_{i,S}$ ; otherwise, one may use standard error propagation to estimate that

$$\sigma_i^2 = \sigma_{i,S}^2 + \left(\frac{A_S}{A_B}\right)^2 \sigma_{i,B}^2.$$

The background term appears only if a background region is specified *and* background subtraction is done. See the section on [C<sup>2</sup> statistics](#) for more information, including definitions of the additional quantities shown in the equation.

### Note on Background Subtraction

*We have not determined the accuracy of the latter expression, thus the user should proceed with caution when subtracting background from the raw data when using this statistic.* An approach preferable to background subtraction is to model the background and data simultaneously.

### Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CHI GEHRELS
sherpa> SHOW STATISTIC
Statistic:                Chi-Squared Gehrels
```

## CHI MVAR

Chi-square statistic with variance computed from model amplitudes.

This statistic is equivalent to [CHI\\_DVAR](#), except that the variance is estimated using the background and source model amplitudes rather than the observed counts data:

$$\sigma_i^2 = S_i + \left(\frac{A_S}{A_B}\right)^2 B_{i,\text{off}},$$

where  $B_{i,\text{off}}$  is the background model amplitude in bin  $i$  of the off-source region. See the section on [C<sup>2</sup> statistics](#) for more information, including definitions of the additional quantities shown in the equation.

## Note on Background Subtraction

The background should not be subtracted from the data when this statistic is used. [CHI\\_MVAR](#) underestimates the variance when fitting background-subtracted data.

## Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CHI MVAR
sherpa> SHOW STATISTIC
Statistic:          Chi-Squared Model Variance
```

## CHI PARENT | CHI CVAR

The statistic [CHI\\_CVAR](#) is equivalent; see that section for further information.

## CHI PRIMINI

Chi-square statistic with Primini variance function.

The  $\chi^2$  statistic is a biased estimator of model parameters (unlike the likelihood functions). In an attempt to remove this bias, Kearns, Primini, & Alexander (1995, ADASS IV, 331) use a scheme dubbed "Iterative Weighting" (IW; see Wheaton *et al.* 1995, ApJ 438, 322), in which

$$\sigma_i^2 = S(i, x_i, pS^{j-1}) + \left( \frac{A_N}{A_B} \right)^2 B_{\text{off}}(i, x_i, pB^{j-1}),$$

where  $j$  is the number of iterations that have been carried out in the fitting process,  $B_{\text{off}}$  is the background model amplitude in bin  $i$  of the off-source region, and  $pS^{j-1}$  and  $pB^{j-1}$  are the set of source and background model parameter values derived during the iteration previous to the current one.

In addition to reducing parameter estimate bias, it can be used even when the number of counts in each bin is small ( $< 5$ ), although the user should proceed with caution.

## Note on Background Subtraction

The background should not be subtracted from the data when this statistic is used. [CHI\\_PRIMINI](#) underestimates the variance when fitting background-subtracted data.

See the section on  [\$\chi^2\$  statistics](#) for more information, including definitions of the additional quantities shown in the equation.

## Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CHI PRIMINI
```

```
WARNING: with CHI PRIMINI, displayed error bars will only be
        correct after a fit is performed with the current filter.
sherpa> SHOW STATISTIC
Statistic:          Chi-Squared Primini
```

## CSTAT

A maximum likelihood function.

The CSTAT statistic is equivalent to the *XSPEC* implementation of the Cash statistic.

Counts are sampled from the Poisson distribution, and so the best way to assess the quality of model fits is to use the product of individual Poisson probabilities computed in each bin  $i$ , or the likelihood  $\mathcal{L}$ :

$$\mathcal{L} = \prod_i \frac{M_i^{D_i}}{D_i!} \exp(-M_i)$$

where  $M_i = S_i + B_i$  is the sum of source and background model amplitudes, and  $D_i$  is the number of observed counts, in bin  $i$ .

The CSTAT statistic (Cash 1979, ApJ 228, 939) is derived by (1) taking the logarithm of the likelihood function, (2) changing its sign, (3) dropping the factorial term (which remains constant during fits to the same dataset), (4) adding an extra data-dependent term, and (4) multiplying by two:

$$C = 2 \sum_i [M_i - D_i + D_i(\log D_i - \log M_i)]$$

The factor of two exists so that the change in CSTAT statistic from one model fit to the next,  $C$ , is distributed approximately as  $C^2$  when the number of counts in each bin is high ( $> 5$ ). One can then in principle use  $C$  instead of  $C^2$  in certain model comparison tests. However, unlike  $C^2$ , the CSTAT statistic may be used regardless of the number of counts in each bin.

The advantage of CSTAT over *Sherpa*'s implementation of CASH is that one can assign an approximate goodness-of-fit measure to a given value of the CSTAT statistic, i.e., the observed statistic, divided by the number of degrees of freedom, should be of order 1 for good fits.

### Note on Background Subtraction

*The background should not be subtracted from the data when this statistic is used. It should be modeled simultaneously with the source, as in this example:*

```
sherpa> DATA source.data
sherpa> BACK background.data
sherpa> SOURCE = [source model]
sherpa> BG = [background model]
sherpa> STATISTIC CSTAT
sherpa> FIT
```



## Examples

1. Specify the fitting statistic and then confirm it has been set.

```
sherpa> STATISTIC CSTAT
sherpa> SHOW STATISTIC
Statistic:          Cstat
```

---

## USERSTAT

User implemented statistic.

It is possible for the user to create and implement his or her own model, own optimization method, and own statistic function within *Sherpa*. The [User Models, Statistics, and Methods Within \*Sherpa\*](#) chapter of the [Sherpa Reference Manual](#) has more information on this topic.

The tarfile `sherpa_user.tar.gz` contains the files needed to define the `userstat`, e.g. Makefiles and Implementation files, plus example files, and it is available from the *Sherpa* threads page: [Data for \*Sherpa\* Threads](#).



The Chandra X-Ray Center (CXC) is operated for NASA by the Smithsonian Astrophysical Observatory.  
60 Garden Street, Cambridge, MA 02138 USA.  
Smithsonian Institution, Copyright © 1998–2006. All rights reserved.

URL:  
<http://cxc.harvard.edu/sherpa3.4/statistics/index.html>  
Last modified: 11 January 2007

