bkgoptrate

June 2, 2019

Abstract

This task calculates the optimum ratio (or count), based on the given (background) time-series FITS and prints it to STDOUT, as well as adds FITS header keywords if requested.

1   Instruments/Modes

bkgoptrate is not XMM-specific: it can be applied to any FITS file.

2   Use

pipeline processing	yes
interactive analysis	yes

3   Description

In general any data include the time-variable background. If the background during a certain and limited period in the observation is too high compared to the averaged ratio of the intensity of the source to background, then those period with high background had better be excluded from the data. As the background level sometimes flares up during X-ray observations, we call it background (BKG) flare-filtering.

Now it is not a trivial question what is the optimum level of the BKG-flare-filtering.

3.1   Theory

A possible and good algorithm to achieve the objective, which this task follows, is to search for the point, at which the maximum signal-to-noise (S/N) ratio is achieved for the given background time-series after the bins above a threshold are excluded.

The S/N ratio is described as

$$S/N = \frac{SourceCount}{\sqrt{BackgroundCount}},$$

(1)

providing the standard Poisson statistics are applicable.

Let us define Z^* as the reference threshold count per time bin, and Exp^* as the exposure time for the bin. When we use i subscripts to represent each time bin, the above formula is rewritten as

$$\text{S/N} \;=\; \frac{\Sigma_{i(\text{Cnt}>Z_i)}\text{Exp}_i \times \text{SourceRate}}{\sqrt{\Sigma_{i(\text{Cnt}>Z_i)}\text{Cnt}_i}} \tag{2}$$

$$Z_i \;\equiv\; Z^* \times \frac{\text{Exp}_i}{\text{Exp}^*}, \tag{3}$$

$$\tag{4}$$

where SourceRate, Exp_i, Cnt_i and Z_i are the count rate of the source, exposure, count of the given time-series and threshold count in each bin i, respectively.

What we here search for is the optimum Z^*, which is the background level, with which S/N becomes the maximum. In principle it requires a minimization (or maximization) method for search. We assume SourceRate is constant, then we can ignore SourceRate in comparing S/N ratios for different thresholds (Z^*):

$$\text{S/N} \;\propto\; \frac{\Sigma_{i(\text{Cnt}>Z_i)}\text{Exp}_i}{\sqrt{\Sigma_{i(\text{Cnt}>Z_i)}\text{Cnt}_i}} \tag{5}$$

$$\text{S/N} \;\propto\; \frac{\Sigma_{i(\text{Cnt}>Z_i)}\text{Exp}_i}{\sqrt{\Sigma_{i(\text{Cnt}>Z_i)}\text{Rate}_i \times \text{Exp}_i}}, \tag{6}$$

where the second form represents the one with the rate instead of count of the input time-series, and Z_i is defined in Eq. 3.

If all the time bins have the same exposure, then this formula is further simplified:

$$\text{S/N} \;\propto\; \frac{\Sigma_{i(\text{Cnt}>Z^*)}i}{\sqrt{\Sigma_{i(\text{Cnt}>Z^*)}\text{Cnt}_i}} \tag{7}$$

$$\text{S/N} \;\propto\; \frac{\Sigma_{i(\text{Rate}>Z^*)}i}{\sqrt{\Sigma_{i(\text{Rate}>Z^*)}\text{Rate}_i}}. \tag{8}$$

3.2 Practical behaviour and command-line parameters

This task reads the background time-series FITS-Table (`tssettabname`) for the specified X and Y columns (`xcol` and `ycol`) for TIME and RATE (or COUNT), respectively, searches for the optimum background count (or rate) per bin, then if found, prints the value (refered to `BGOPTCUT`, hereafter) to standard output (`STDOUT`).

If `doinsertkwds`='yes', it also adds the two header attributes in the input time-series dataset: `BGOPTCUT` for the value `BGOPTCUT` and `FRACTLFT` for the fraction of the number of the resultant time bins to the original one[1].

If `dooutputsntab`='yes', it also outputs the FITS file containing the table of the diagnostic information of the calculation (see Table 1), where the FITS and table names are specified in `snsettabname`. The further information is written in the hearder of the table extention (the first extention) of the FITS file (see Table 1).

In practice, all the bins for Time are assumed to be the same in size, that is, duration.

---

[1]The 'original' number of time bins here means that after the specified pre-filters, such as `lowercutoffcount` and/or `endtime`, are applied.

Table 1: The column names of the diagnostic output FITS file

Keyword	Description
Table columns	
BKGRATECUT	The trial threshold for the background in unit of counts/unitTime.
BKGRAWCUT	The same as BKGRATECUT but in the 'raw' unit; e.g., if the original is in unit of counts, 'raw' unit is counts.
BKGRATE	The maximum rate in the data in unit of counts/unitTime that satisfies the threshold of BKGRATECUT. Therefore this value should be (very) slightly lower than the corresponding BKGRATECUT.
SN_RATIO	Signal-to-noise ratio; n.b., this used to be S/N up to Ver.1.1.
N_BIN	The number of bins that satisfy the threshold.
Header keywords in the Table extention	
BGOPTCUT	As explained (Optimum threshold of the background rate).
FRACTLFT	As explained in the text; NBINOPT/NBINAFTF in the following notations.
NBINOPT	The number of bins if BGOPTCUT is applied.
NBINAFTF	The number of bins after the initial filtering is applied.
NBINORIG	The number of the original bins.

This task uses a very simple minimization method; it finds the bin(s) that has the highest rate then calculate the S/N with those bins excluded, and compares with the previous S/N, namely the S/N with those bins included. This procedure continues until the newly calculated S/N is lower than the the previous value, then it prints out the previous highest rate (NOT the count), or strictly speaking the value that is slightly higher than the previous highest rate (so that the users do not have to worry about whether to use "larger/smaller than" or "equal to or larger/smaller than" in the later processing). The assumption here is the S/N as a function of Z^* is monotonic.

The unit of the time-series can be significant. Users can specify it with tsstyle=(auto—rate—count). If it is "auto" (Default), this task determines tsstyle (1) according to ycol if specified, (2) or if not, this task searches for the name of the column of RATE, COUNT and COUNTS in this order and determines ycol first, then tsstyle accordingly.

Although the size of each time bin is assumed to be the same, it can be defined in the different column in the table (in the same FITS extension as the X(time) and Y(count) axes), typically FRACEXP, which can be explicitly specified with fracexpcol.

Users can also control how this task treats the FRACEXP column with fracexpstyle=(auto—calc—threshold—none). If fracexpstyle=none this task simply ignores FRACEXP column and assumes all the bins have the same duration[2]. If fracexpstyle=threshold, then this task ignores the bins, of which FRACEXP is smaller than the user-specified parameter, fracexplower. If fracexpstyle=calc, then this task takes FRACEXP into account in the search. If fracexpstyle=auto (Default), then this task searches for the column fracexpcol (FRACEXP in default), and sets fracexpstyle to be (1) "calc" if found, else (2) "none".

Users can also specify the start and end times to consider in the search for the optimum background level with the parameters of starttime and endtime.

Users can also specify the bins to be ignored if the count (or rate) of the bin is lower than the value, lowercutoffcount. Note that the bins with negative count (or rate) or with undefined value are always ignored.

Users can also specify the ratio of the minimum time to be left where the further search stops, with mintimeratio (0.05 in default). Namely if the search for the optimum background cuts continues to the extent smaller than the fraction of mintimeratio (5% in default) in the exposure to the original

---

[2]If a bin has an undefined value of count or rate, then the bin is ignored.

exposure after all the other filtering is performed is left to be valid (in other words if over 95% in default of the exposure is regarded as high-background area), then this task stops further search for the optimum background level and instead raises an error (`noThresholdFound`).

Finally, if the large number of bins have the smallest and identical count, then chances are the maximum S/N is achieved when only those bins are used. For example, it is the case for the perfectly uniform time-series (although it must be an extremely rare case, if ever, with the real data). In that case, this task either (1) raises an error (`noThresholdFound`) preceded by a warning (`recommendedLowerMinTimeRatio`), or (2) returns the value slightly higher than the particular value preceded by a warning (`regardLowestBinsAsOptimum`) if all the time-bins are examined (due to a very small `mintimeratio` given) and if those time-bins occupy more than a third of the number of entire time-bins.

4   Parameters

This section documents the parameters recognized by this task (if any).

Parameter	Mand	Type	Default	Constraints
tssettabname	yes	dataset		

Name of the input FITS data set and table, the latter of which can be added, following the ':', such as, 'input1.ds:EVENTS'.

xcol	no	string	TIME	

Column name of X-axis.

ycol	no	string	default	

Column name of Y-axis.

fracexpcol	no	string	FRACEXP	

Column name of FRACEXP.

tsstyle	no	string	auto	auto—rate—count

Style of the input time-series.

withstarttime	no	boolean	no	

Whether start time is specified (Default: No). If yes, the task reads `starttime`.

starttime	yes	real		

Start time to evaluate.

withendtime	no	boolean	no	

Whether end time is specified (Default: No). If yes, the task reads `endtime`.

endtime	yes	real		

End time to evaluate.

fracexpstyle	no	string	auto	auto—calc—threshold—none

Style of handling method of FRACEXP. If `threshold`, the task reads `fracexplower`.

fracexplower	yes	real		$0 \geq f \geq 1$

Lower threshold for FRACEXP to be used in the calculation.

withlowercutoffcount	no	boolean	no	

Whether the lower cutoff count (or rate) is specified. If yes, the task reads `lowercutoffcount`.

lowercutoffcount	yes	real	0.0	≥ 0

The lower threshold count or ratio in search for the optimum value, that is, any bin having the count (or ratio) equal to, or smaller than, this value is not taken into account in the search. This value has to be either zero or positive.

withmintimeratio	no	boolean	no	

Whether the minimum time ratio is specified. If yes, the task reads `mintimeratio`.

mintimeratio	yes	real	0.05	≥ 0

The ratio of the acceptable minimum number of (time) bins (or row numbers in the FITS table) to the initial number of bins after other filtering is applied, during the search for the optimum value. If the search does not find the optimum threshold with larger number of bins, then the task raises an error and exits.

ignorelowcnttail	no	boolean	no	

If yes, the lower-count tail is ignored and the best local maximum is used to derive the output threshold, when smaller the background count (rate) threshold is, better the statistics will be. Otherwise, in those cases, an error `noThresholdFound` is issued, unless the number of bins that contains the smallest count consists more than the third of the entire light-curve.

doinsertkwds	no	boolean	no	

Whether inserting the header attributes in the input time-series dataset. If yes, the (header of the) input time-series dataset will be modified.

dooutputsntab	no	boolean	no	

Whether outputting the table of signal-to-noise ratio etc (see text for detail) to the cutoff rate/count.

snsettabname	no	dataset	sntobkgcut.ds:SN_TO_BKGCUT	

Name of the output table of signal-to-noise ratio to the cutoff rate/count in the form of "FILE.fits:TABLENAME".

5 Errors

This section documents warnings and errors generated by this task (if any). Note that warnings and errors can also be generated in the SAS infrastructure libraries, in which case they would not be documented here. Refer to the index of all errors and warnings available in the HTML version of the SAS documentation.

noThresholdFound *(error)*
> Failed to find the optimum background (in the given constraints).

regardLowestBinsAsOptimum *(error)*
> Regarded the lowest bins as the optimum background level (see text for detail).

allMaskFalse *(error)*
> After filtering with various thresholds (GTI, Start/End-Times, Lower-Cutoff-Count, Min-Time-Ratio), no significant bin is left. As a typical case, all the bins in the input time-series are uniformly zero.

recommendedLowerMinTimeRatio *(warning)*

> *corrective action:* The lowest bins looks like the optimum background level in the given `mintimeratio` (see text for detail).

negativeLowerCutoffY *(warning)*

> *corrective action:* Specified `lowercutoffcount` is negative.

lowestBinsHaveBestSN *(warning)*

> *corrective action:* If `ignorelowcnttail` is true, and if the lowest count bin shows the best signal-to-noise ratio.

noSignificantBin *(warning)*

> *corrective action:* The same as the error `allMaskFalse`.

6 Input Files

The input FITS needs not be XMM time-series but any FITS tables.

1. (Mandatory) `tssettabname`: the input FITS-table.

7 Output Files

1. The header of `tssettabname` is modified if `doinsertkwds`='yes'.
2. The diagnostic information is output to the FITS table `snsettabname` if `dooutputsntab`='yes'.

8 Algorithm

9 Comments

-

References