ekstest

November 4, 2014

Abstract

ekstest reads the FITS file containing the EPIC source and background time series' produced by **epiclccorr** and when required its associated Good Timing Interval (GTI) file produced by **tabgtigen**. It eliminates all bins that contain null values as well as those that are not in the GTIs (if the file is provided). It then performs a variety of variability tests, including the Kolmogorov-Smirnov probability of constancy test, chi-squared probability of constancy test, flare tests and on the light curve to determine whether the source is variable and writes the various statistics and probabilities that the source is not variable and the number of good bins used to determine these values into the header of a new version of the file or to the screen. No tests are carried out if there are insufficient good bins.

1 Instruments/Modes

Instrument	Mode
EPIC MOS	IMAGING, TIMING
EPIC PN	IMAGING, TIMING

2 Use

pipeline processing	yes
interactive analysis	yes

3 Description

ekstest reads the FITS file containing the EPIC source and background time series' produced by **epiclccorr** or **elcbuild** and its associated Good Timing Interval (GTI) file produced by **tabgtigen**, if required. It eliminates all bins that contain null or negative values as well as those that are not in the GTIs, if the GTI file is provided. It can perform any or all of the following variability tests on the light curve to determine whether the source is variable, however if insufficient good bins remain, no variability tests are carried out.

- a **Kolmogorov-Smirnov test** carried out either on: the cumulative probability functions of the observed count distribution and the observed background distribution; or the cumulative probability function of the observed net count distribution and the cumulative

time distribution. It then writes the Kolmogorov-Smirnov statistic, the probability that the source is not variable and the number of good bins used to determine these values into the header of the input file or into the header of a new file and/or to the screen. Keywords with 'NULL' values indicate that there were too few bins for the test to be carried out. It must be noted that only the Kolmogorov-Smirnov test with the cumulative probability function of the observed net count distribution and the cumulative time distribution can be carried out at the same time as the other tests.

- a **chi-squared probability of constancy test** on the source+background light curve. It then writes the chi-squared statistic, the probability that the source is not variable and the number of good bins used to determine these values into the header of the input file or into the header of a new file and/or to the screen. Keywords with 'NULL' values indicate that there were too few bins for the test to be carried out.
- a **fractional variability amplitude test** on either the net light curve or the total (source+background) lightcurve (Edelson et. al., 2002, ApJ, 568, 610, Vaughan et. al., 2003, MNRAS, 345, 1271 and see Sec. 9). The user can choose whether to use the net or total lightcurve via the boolean parameter 'netlightcurve', unless the chi-squared test is run at the same time, in which case **ekstest** will automatically use the total lightcurve. A warning is raised however in this case. **ekstest** writes the fractional varability amplitude (F_{var}) and the error on this value, along with the number of good bins used to determine these values into the header of the input file or into the header of a new file and/or to the screen. A value of -1 indicates that the noise of the data is much greater than the scatter of the data. Keywords with 'NULL' values indicate that there were too few bins for the test to be carried out.
- a **flare test** on the net lightcurve. This searches for the highest peak in the lightcurve, which is over a threshold value provided by the user (given in sigma above the median value of the lightcurve), with an integer number of consecutive bins over the same threshold, also given by the user. The average sigma of the flare, above the median value of the lightcurve along with the number of good bins in the lightcurve are written either into the header of the input file or into the header of a new file and/or to the screen. Keywords with 'NULL' values indicate that there were too few bins for the test to be carried out.
- a **variation test** between the beginning and the end of the lightcurve. The user provides the percentage of the lightcurve to be used at each end and the difference in sigma between the mean values of the two ends along with the number of good bins in the lightcurve are written either into the header of the input file or into the header of a new file and/or to the screen. Keywords with 'NULL' values indicate that there were too few bins for the test to be carried out.

The input FITS files must contain (net) count rates, background count rates and the associated errors. The time series must follow a regular binning scheme (i.e. equispaced time bins) [1]. The bin width is given by the keyword `TIMEDEL`.

3.1 Examples

To run the task **ekstest** using the input FITS file 'lightcurve.ds' and its associated GTI file 'gti.ds', with the single distribution Kolmogorov-Smirnov test and and the flare test where the flare must be at least 3σ above the mean and at least 5 bins wide and then write the results both into a new FITS file called 'test.fit' and also write the results to the screen, a command such as,

```
ekstest set=lightcurve.ds newoutset=yes outset=test.fit screen=yes
gtiset=gti.ds kstest=yes twodists=no flaretest=yes flarefactor=3 flarebins=5
```

can be used.

4 Parameters

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

Parameter	Mand	Type	Default	Constraints
set	yes	dataset	No default	

Name of the FITS file containing the time series. Case sensitive.

newoutset	no	boolean	no	

The name of a new output file if required. This can also be the original name if you want to overwrite the input file.

outset	no	dataset	No default	

Name of the output FITS file.

screen	no	boolean	no	

Do you want print the results to the screen?

gtis	no	boolean	no	

Do you want to provide a GTI file?

gtiset	no	dataset	No default	

Name of the FITS file containing the GTIs. Case sensitive.

kstest	no	boolean	no	

Do you want to perform the Kolmogorov-Smirnov test?

twodists	no	boolean	no	

Do you want to use the two distribution test?

chisquaretest	no	boolean	no	

Do you want to carry out a chi-squared (Pearsons) test?

fracvartest	no	boolean	no	

Do you want to carry out a fractional varability amplitude test?

netlightcurve	no	boolean	yes	

Do you want to carry out the fractional varability amplitude test on the net lightcurve?

flaretest	no	boolean	no	

Do you want to carry out a flare test?

flarefactor	no	integer	3	

Factor (in sigma) above the median which indicates a flare

flarebins	no	integer	3	

Number of bins required to be consecutive to define the flare

variationtest	no	boolean	no	

Do you want to carry out a variation test?

percentage	no	integer	20	

Percentage of the lightcurve (beginning and end) that you want to use to determine the variations

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.

Keyword TIMEDEL missing in Input File *(error)*
The TIMEDEL keyword is missing in the timeseries FITS file, which is necessary to determine the binning factor of the data

Null or negative Bin Width in Input File *(error)*
The TIMEDEL keyword is null or negative in the timeseries FITS file. This should be a positive value.

Different number of columns between RATE, ERROR, BACKV and BACKE *(error)*
There should be one RATE column, one ERROR column, one BACKV column and one BACKE column for each lightcurve in the input timeseries FITS file.

Different number of rows in columns between RATE, ERROR, BACKV and BACKE *(error)*
The columns in the timeseries FITS file have differing lengths. They should all be the same length

Number of light-curves contained in a file must not exceed 9 *(error)*
There are too many lightcurves contained in the timeseries FITS file - 9 is the maximum allowed.

No data point received *(error)*
Less than two good values in the timeseries FITS file.

Negative count rates in Input File *(error)*
The total count rates are negative and should not be.

Negative background values in Input File *(error)*
The background count rates are negative and should not be.

Require new output file name *(error)*
newoutset should be set to yes and a new output filename given or screen should be set to yes

Can not carry out two distribution KS test and chi-squared test *(error)*
The program is limited to the carry out only the cumulative probability function of the observed net count distribution and the cumulative time distribution with the other variability tests

Not enough bins to process variability tests *(warning)*
Less than two good values remain after removing those bins outside of the GTIs and those bins with both source and background values that are null or negative. Hence it is impossible

to carry out the variability test.
corrective action: Use a different input file

Keyword CONTENT missing in Input File *(warning)*
No valid CONTENT keyword is contained in the header of the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

CONTENT parameter for input file is ... *(warning)*
The CONTENT keyword is not the expected parameter.
corrective action: Verify that this is a valid fits file

Keyword HDUCLASS missing in Input File *(warning)*
The keyword is missing in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

HDUCLAS is not equal to OGIP *(warning)*
The keyword is not as expected in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

HDUCLAS1 is not equal to LIGHTCURVE *(warning)*
The keyword is not as expected in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

HDUCLAS2 is not equal to TOTAL *(warning)*
The keyword is not as expected in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

Keyword TSTART missing in Input *(warning)*
The keyword is missing in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

Keyword EXPOSURE missing in Input File *(warning)*
The keyword is missing in the input timeseries FITS file which could indicate that there is a problem with the input timeseries FITS file.
corrective action: Verify that this is a valid fits file

Null or negative exposure time in Input File *(warning)*
The EXPOSURE keyword is null or negative in the timeseries FITS file. This should be a positive value.
corrective action: Verify that this is a valid fits file

Keyword TSTART missing in GTI File *(warning)*
The keyword is missing in the input GTI FITS file which could indicate that there is a problem with the input GTI FITS file.
corrective action: Verify that this is a valid fits file

Keyword TSTOP missing in GTI File *(warning)*
The keyword is missing in the input GTI FITS file which could indicate that there is a problem with the input GTI FITS file.
corrective action: Verify that this is a valid fits file

6 Input Files

1. EPIC FITS Time Series [1] (generated by **epiclccorr** or **elcbuild**).
2. EPIC FITS GTI file [1] (generated by **tabgtigen**).

7 Output Files

1. EPIC FITS Time Series [1]

8 Algorithm

```
Subroutine ekstest

  Recover parameters (File name, GTI file)

  Read source time series FITS file :
     Get dataset(s) and table.
     Check important keyword consistency (notably TSTART and TIMEDEL).
     Call a warning or error if necessary.

  Recover all light curves included in table :
     Net source rates and errors and background rates and errors are recorded in arrays
     of dimension Nlightcurve * Nbins.

   For each light curve :

     Delete gaps in data (when the IEEE NaN constant is found) and outside of GTIs.

     Perform variability tests on rebinned counts and background or the net source counts and the cu
        Calculate mean count rate and variance.
        Test the null hypothesis (the source is not variable) with the :
            - Kolmogorov-Smirnov test
            - chi-squared test
            - fractional variability amplitude test
            - flare test
            - variation test

  End for

  Write variability test results into header and/or to screen.

  Release memory.

End subroutine ekstest
```

9 Comments

The fractional variability amplitude (F_{var}) test (Edelson et. al., 2002, ApJ, 568, 610 and Vaughan et. al., 2003, MNRAS, 345, 1271) is now implemented as of 3rd April 2012.

$$F_{var} = \sqrt{\frac{S^2 - \langle\sigma^2_{err}\rangle)}{\langle x\rangle^2}}$$

where,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \langle x\rangle)^2$$

and N is the number of bins, x_i the net rate for the ith data point and $\langle x\rangle$ is the mean of the net rate ($= \frac{1}{N}\sum_{i=1}^{N}(x_i)$) and

$$\langle\sigma^2_{err}\rangle = \frac{1}{N}\sum_{i=1}^{N}(NetRateErr_i)^2$$

The error on F_{var} is then given by :

$$err(F_{var}) = \frac{1}{2F_{var}}\sqrt{\left(\sqrt{\frac{2}{N}}\frac{\langle\sigma^2_{err}\rangle}{\langle x\rangle^2}\right)^2 + \left(\sqrt{\frac{\langle\sigma^2_{err}\rangle}{N}}\frac{2F_{var}}{\langle x\rangle}\right)^2}$$

References

[1] SSC. XMM Survey Science Centre to Science Operations ICD for SSC Products. Technical Report XMM-SOC-ICD-0006-SSC Issue 2.1, SSC, Mar 2000.