1 Testing for Source Variability Using the Kolmogorov-Smirnov (K-S) Test

The K-S Test is used to test whether two distributions are consistent, by comparing their cumulative distribution functions (CDFs). Under the null hypothesis that the two CDFs are consistent, the maximum difference between them is governed by a probability distribution which is, in general, independent of the forms of the CDFs in question, and is easily evaluated for large sample sizes[1].

In the context of L3 Source Variability, we may use the K-S Test to test the hypothesis that a source is non-variable within an OBI. The appropriate test to use is the One-Sample (because one of the CDFs is a model), Two-Sided (because the statistic we use is $max\left|CDF_1 - CDF_2\right|$) test.

The model CDF corresponds to the hypothesis that the source has a constant rate $\overline{R}$. Assuming constant exposure and no data gaps, it is given by

$$CDF_M(t) = \int_{t_1}^{t} \overline{R}\,dt \Big/ \int_{t_1}^{t_2} \overline{R}\,dt = (t - t_1)/(t_2 - t_1) \qquad (1)$$

where t_1 and t_2 correspond to the begin and end times of the good time interval (GTI).

Because there appear to be some limitations in using the K-S Test for binned data ([2]), we recommend using the set of unbinned, time-ordered, event arrival times $\{t_i^{event}, i = 1 \ldots N\}$ rather than a binned light curve to generate the sample CDF. This is then given by

$$CDF_S(t) = \{\text{number of events with } t_i^{event} \le t\}/N \qquad (2)$$

The situation is shown in Illustration 1.

1.1 Sensitivity to Different Types of Variability

Simulations are required to establish the sensitivity of the K-S test to different types of source variability. Each type of variability may be expressed by a model rate $R_V(t)$, with a model CDF given by

$$CDF_V(t) = \int_{t_1}^{t} R_V(t)\,dt \Big/ \int_{t_1}^{t_2} R_V(t)\,dt \quad . \qquad (3)$$

Sets of n simulated arrival times are generated by drawing n random numbers from a uniform distribution and determining t_i such that $R.N_i = CDF_V(t_i)$. The fraction of these sets which do not support the constant source hypothesis at a given confidence level represents the test's sensitivity to that type of variability at that confidence level. Typically, these sensitivities will depend on the number of events in the sample and the confidence level, and may also depend on other parameters of the $R_V(t)$.

1.1.1 Step-Function Variability

In the case the variability is modelled as

$$R_V(t) = \begin{matrix} C_1 & t \le t_{step} \\ (1+\alpha)C_1 & t > t_{step} \end{matrix} \quad (4)$$

The CDF for this kind of variability is shown in Illustration 2.
As might be expected, the sensitivity to this kind of variability depends not only on the confidence level, but also on the amplitude of the step. The following table illustrates the sensitivity at the 90% confidence level for representative values of n and α .

	α		
n	0.1	0.5	1
30	11	25	50 (25)
100	12	55	93 (66)
300	18	94	100 (99)
1000	38	100	100 (100)

Table 1 Sensitivity of K-S Test to Step Variability

Each entry in Table 1 represents the percentage of total simulations for that value of n and α which are identified as variable at the 90% confidence level. Ten thousand simulations were run for each entry. Arrival times were defined in the range $t_0 \le t \le t_1$ with $t_{step} = 0.5$. One may define the arrival times in this fashion, since the K-S test is not affected by scaling of the x-axis. However, it should be noted that sensitivity is not independent of x. The numbers in parentheses for $\alpha = 1$, for example, represent the sensitivity if $t_{step} = 0.25$, and demonstrate a significant decrease in sensitivity. This is consistent with the fact that the K-S test appears most sensitive at the median of the probabilty distribution (midpoint of the CDF). Variants of the K-S statistic, such as the Kuiper statistic, do not suffer from this problem, and should be considered as a replacement for L3.

1.1.2 Burst Variability

This is TBD.

1.2 Corrections for Variable Exposure

If the events in the source event list are selected from an aperture that is dithering over a region of variable exposure, then the model CDF for a constant source in Equation 1 should be replaced by

$$CDF_M(t) = \int_{t_1}^{t} \bar{R}E(t)\,dt \Big/ \int_{t_1}^{t_2} \bar{R}E(t)\,dt \quad (5)$$

where $\bar{R}$ may be considered the photon flux in $photons - cm^{-2} - s^{-1}$ and $E(t)$ is the time-varying exposure, in $cm^2 - s$ in the region. The latter may include effects of dithering over regions of the exposure map with different values or dithering off-chip, as well as data gaps due to GTIs.
The effectiveness of this approach needs to be verified through simulations.

1.3 Corrections for Variable Background

TBD.

Bibliography

1: Press, W.H. et al., 1992, Numerical Recipes in C, 2nd Edition, Cambridge University Press.
2: Conover, W.J., , 1972, J. Amer. Statist. Assoc., 591.

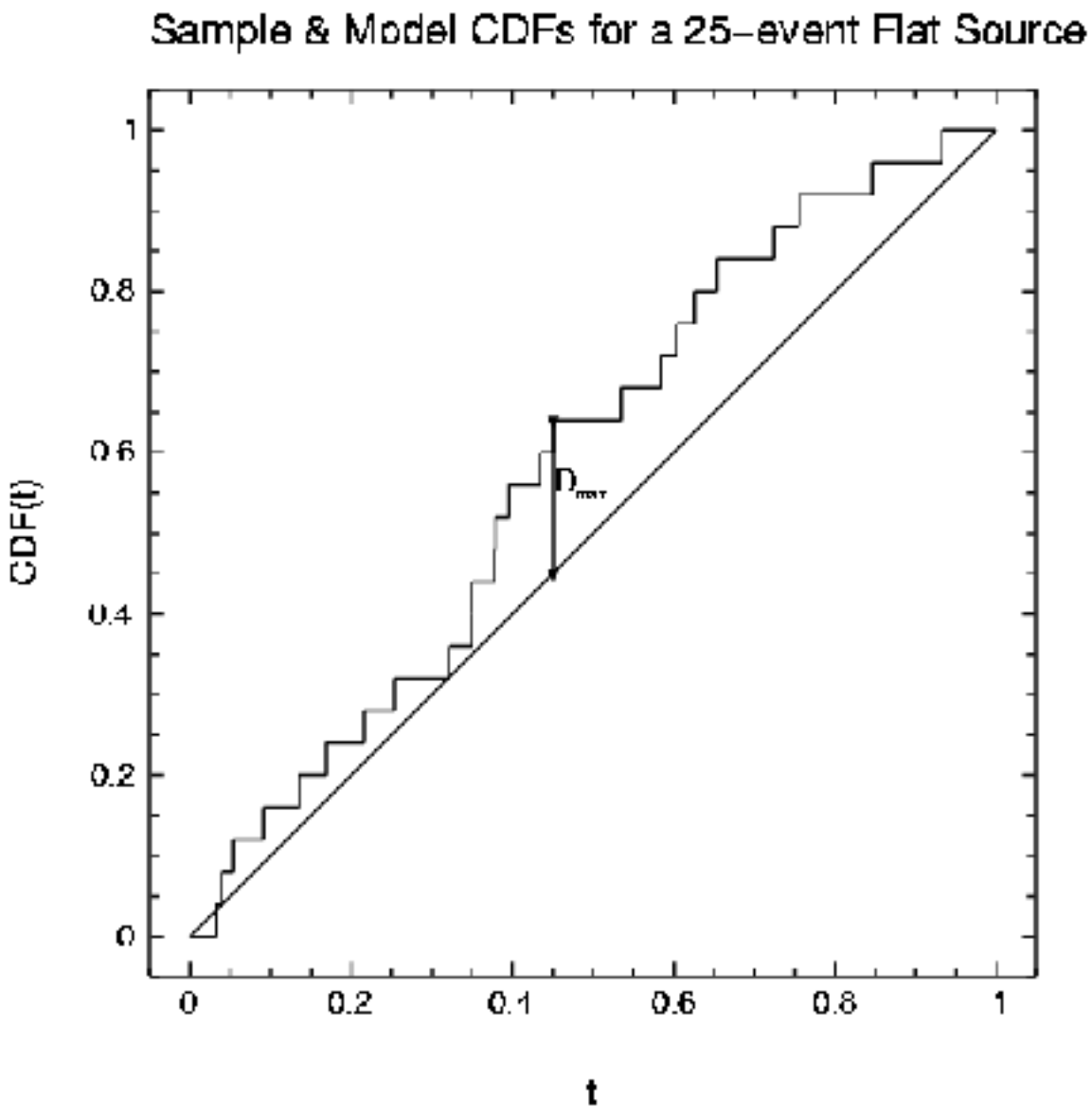

Illustration 1 Sample and model CDFs for a constant source. The sample CDF is generated from 25 event arrival times randomly sampled from a uniform distribution in the range $0 \le t \le 1$ ***. The quantity*** D_{max} ***is the statistic used in the K-S test.***

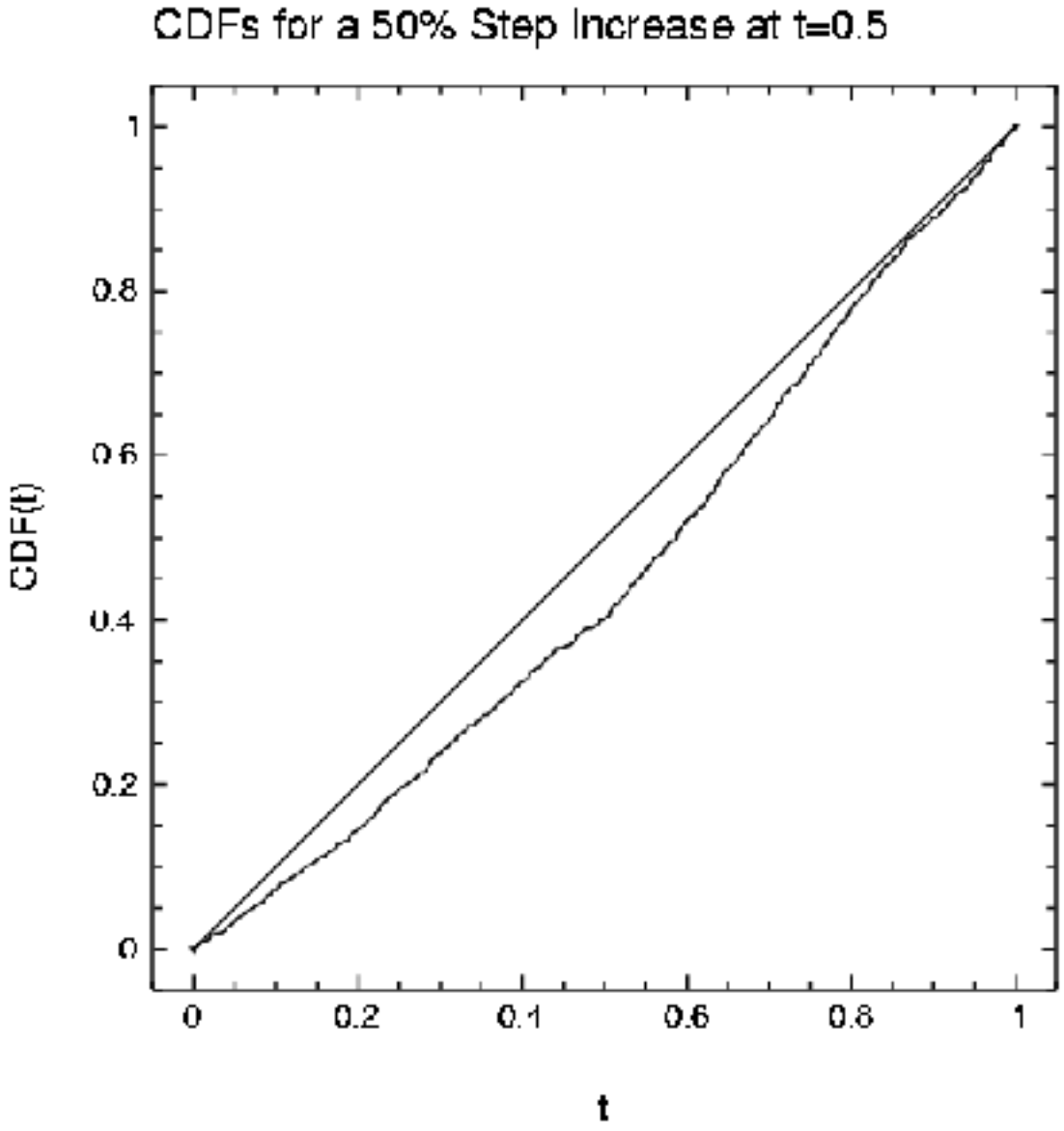

Illustration 2 Sample CDF for step function variability. The CDF is generated from 1000 event arrival times randomly sampled from a step function CDF in the interval from $0 \le t \le 1$, ***with a 50% step increase at*** $t = 0.5$. ***The model CDF for a constant source is also shown.***