Applying the Discrete Correlation Function

Lowell M. Boone and Thomas G. Spears
Department of Physics, The College of Wooster

While fitting observational data to time-dependent emission models is rapidly becoming the preferred mode of temporal analysis, it is often desirable (and even necessary) to invoke a model-independent technique for a given set of observations. The Discrete Correlation Function (DCF), as described by Edelson & Krolik (1988), appears frequently in the literature as the analysis technique of choice for ascertaining correlations in the high-energy emissions of astrophysical objects. The DCF is particularly appealing because it does not rely on interpolation, and is thus better suited for the sparsely sampled data that are common in higher-energy observations. In an effort to better understand its results, we have undertaken studies of the DCF as applied to artificial signals with known correlations.

Simulation Basics.

Simple simulated light curves were constructed as a combination of noise components (independent random flux values drawn from a Gaussian distribution), and well-defined trends. "Pure" light curves were then "corrupted" to simulate less idealized observations. First, the time stamp of each simulated flux measurement was randomized to plus or minus one half the idealized sampling interval. Then roughly 50% of the data points were removed to reflect the sparse sampling characteristic of the data under consideration.

The advantage in using this simplistic approach to light curve modeling is that the autocorrelation function for random noise should be a simple delta function. Thus, the signature of a meaningful noise correlation in the DCF is relatively easy to identify. However, we are mindful that other constructions, such as first-order autoregressive models with measurement noise, or even simple random walks appear to more accurately describe light curves from astrophysical sources (Koen 2003, Spears 2004).

The Moving Average.

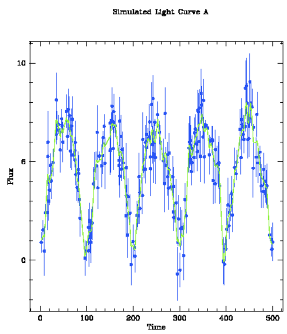
To correct for the correlation function's bias to low-frequency trends (Vio & Wamsteker 2001), the data are "whitened" prior to calculating the DCF. Such pre-whitening is discussed in Diggle (1990), and is accomplished by subtracting a moving flux average from each data point. For unevenly spaced data, we adopt the following expression for the moving average at any given point:

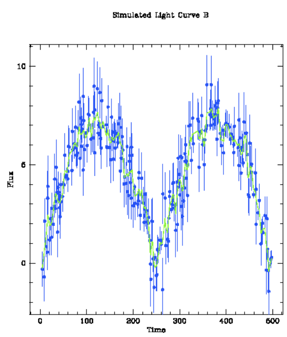
$$\bar{\phi} = \frac{\sum_i \phi_i \exp\left(-\frac{\Delta t_i^2}{2\sigma^2}\right)}{\sum_i \exp\left(-\frac{\Delta t_i^2}{2\sigma^2}\right)}$$

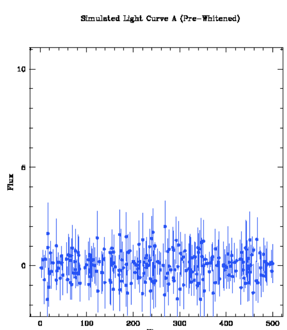
where Δt_i is the time difference between the data point in question and the ith observation. The variance (σ^2) in the Gaussian term may be used to control the sensitivity of the average to the behavior of the data in the vicinity of the point in question. In this investigation, σ^2 was arbitrarily assigned a value of four.

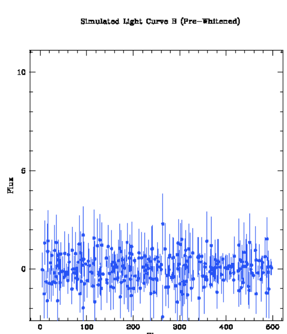
References.

C. Koen 2003, MNRAS **344**,798.
R. A. Edelson & J. H. Krolik 1988, ApJ **333**,646.
R. Vio & W. Wamsteker 2001, PASP, **113**,86.
P. J. Diggle 1990, *Time Series: A Biostatistical Intro*, Clarendon Press.
T. G. Spears 2004, Senior Thesis, The College of Wooster.









Pre-Whitening Data to Enhance the DCF Performance

The discrete correlation function (DCF) of a light curve, as described by Edelson & Krolik (1988), is given by the expression below. Note that the terms representing the measurement error (e_a and e_b) have been suppressed due to problems arising from their interpretation and subsequent inclusion in the DCF calculation. This is a known issue with the DCF, and most authors simply use the expression we present here.

$$DCF(\tau) = \left\langle \frac{(a_i - \bar{a})(b_j - \bar{b})}{\sigma_a \sigma_b} \right\rangle_{(t_i - t_j)\sim\tau}$$

The average is taken for all pairs of data points (i and j) for which: $\tau - \Delta\tau/2 \le t_i - t_j < \tau + \Delta\tau/2$

While the DCF works very well for well-populated, but unevenly sampled data sets, its performance (and the performance of its competitors) is much poorer when the data become sparsely sampled. Unfortunately, this is precisely the situation that often arises when one wishes to look for correlations that involve very high energy (VHE) gamma-ray data. Due to observational restrictions (relatively low duty cycles of ground-based imaging detectors), and comparatively low signal to noise ratios, gamma-ray light curves are rarely sampled at a rate or quality that permits raw analysis with the DCF.

In an effort to improve the performance of the DCF on sparsely sampled data, we have experimented with the notion of "whitening" the data prior to imposing the DCF analysis. This amounts to subtracting long-term trends from the data, leaving only the "white noise" component of the light curve (see "The Moving Average").

The top two frames (left) show simulated light curves whose noise components (see "Simulation Basics") are correlated, but shifted by two time units. The underlying trends are sinusoidal bumps, intended to represent uncorrelated long-term trends. The trends for each curve were calculated (see "The Moving Average"), and then subtracted from the data. The calculated trends appear in green, and the trend-subtracted ("pre-whitened") data appear in the lower left-hand panels.

The DCF was constructed for both the raw light curves and the pre-whitened data. These functions appear in the top right and bottom right frames respectively. Note that the strong trends seem to wash out the highly correlated noise component, resulting in a DCF that is fairly consistent with zero correlation. But the pre-whitened data show a strong and rather unmistakable correlation at an offset of two time units.