Astronomy and Computing 23 (2018) 83–91

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Estimation of error on the cross-correlation, phase and time lag between evenly sampled light curves

R. Misra [a], A. Bora [b,*], G. Dewangan [a]

[a] IUCAA, Post Bag 4, Ganeshkhind, Pune 411007, India
[b] Department of Applied Sciences (Physical Science division),GUIST, Gauhati University, Guwahati 781014, India

ARTICLE INFO

Article history:
Received 11 April 2017
Accepted 12 March 2018
Available online 16 March 2018

Keywords:
Accretion
X-ray binaries
AGN
Analytical method

ABSTRACT

Temporal analysis of radiation from Astrophysical sources like Active Galactic Nuclei, X-ray Binaries and Gamma-ray bursts provides information on the geometry and sizes of the emitting regions. Establishing that two light-curves in different energy bands are correlated, and measuring the phase and time-lag between them is an important and frequently used temporal diagnostic. Generally the estimates are done by dividing the light-curves into large number of adjacent intervals to find the variance or by using numerically expensive simulations.

In this work we have presented alternative expressions for estimate of the errors on the cross-correlation, phase and time-lag between two shorter light-curves when they cannot be divided into segments. Thus the estimates presented here allow for analysis of light-curves with relatively small number of points, as well as to obtain information on the longest time-scales available. The expressions have been tested using 200 light curves simulated from both white and $1/f$ stochastic processes with measurement errors. We also present an application to the *XMM-Newton* light-curves of the Active Galactic Nucleus, Akn 564. The example shows that the estimates presented here allow for analysis of light-curves with relatively small ($\sim$ 1000) number of points.

1. Introduction

Establishing that two light-curves, measured in different energy bands, are correlated with each other is an important temporal diagnostic for various kinds of Astrophysical sources, especially for Active Galactic Nuclei (AGN) and X-ray binaries. The detection and measurement of the level of correlation constrain the radiative processes active in the source and can be used to validate (or rule out) models based on spectral analysis. Phase and time-lags detected for correlated light-curves can provide further insight into the geometry and size of the emitting region. Often in these applications, the light curves available for analysis are of short duration and have measurement errors. The true temporal behaviour of a source can only be established if there are robust estimates of the errors on the cross-correlation, phase and time-lags.

It is important to emphasise that a cross-correlation analysis between two finite length light-curves will not provide an accurate measure of the correlation between them, even in the absence of measurement errors. Intrinsic stochastic fluctuations in the light curves will induce an error on the cross-correlation measured. An estimate of the significance and error of the cross-correlation detected, should take into account both, measurement errors as well as statistical fluctuations.

A standard method to estimate the error on the cross-correlation involves dividing the light curves into several equal segments and finding the cross-correlation for each. Then the net cross-correlation is given by the average of the different segments, and the variance is quoted as an error. For example, this technique is implemented by the "crosscor" tool in Xrones of the high energy astrophysics software *HEASOFT*.[1] The method is reliable only if the light curves can be divided into a large number of segments ($\gg$10) and each segment is sufficiently long and not dominated by measurement errors. The temporal behaviour of many astrophysical systems depends on the time-scales of the analysis and hence by using this method, one loses information on the behaviour of the system on time-scales comparable to the length of the original data. In AGN, the time scale involved is long comparable to the length of observation in many cases, hence it is not practical to divide the light curve in segments. Moreover, there does not seem to be any established way by which this method can be extended to get an estimate of the time-lag between the light curves and its error.

* Corresponding author.
E-mail addresses: rmisra@iucaa.in (R. Misra), abora.80@gmail.com (A. Bora), gulabd@iucaa.in (G. Dewangan).

1 http://heasarc.gsfc.nasa.gov/docs/software/lheasoft/.

https://doi.org/10.1016/j.ascom.2018.03.002
2213-1337/

These deficiencies can be overcome by using a Monte Carlo technique where one simulates a large number of pairs of light curves having the same *assumed* temporal properties and with the same measurement errors as the original pair. The results of the original pair can be compared with the simulated ones to ascertain the confidence level of the cross-correlation and time-lag. The simulated light curves should take into account the stochastic fluctuations of the light curves and not just the measurements errors. Indeed, when the light curve is sampled unevenly and with measurement errors changing in time, the Monte Carlo technique may be the only way to obtain reliable estimates (for e.g. Peterson et al., 1998). Monte Carlo technique is numerically expensive and hence are not practical for analysis of a large sets of data. More importantly, the results depend on the subjectivity of the assumed temporal properties of the system. For example, to ascertain the errors on an observed cross-correlation and time-lag value, the simulations are generally done with the assumption that these are the true intrinsic values. Similar assumptions have to be made on the shape of the power spectra of the light curves.

As pointed out and discussed extensively by Welsh (1999), an analytical estimate of the variance on cross-correlation is not straight forward. In the literature, there is an analytical estimate for the cross-correlation known as Bartlett's equation (Bartlett, 1955) which is not often used in Astronomical contexts. This method is available in the "crosscorrelation" function in the IMSL numerical libraries.[2] The error is accurate only when the complete knowledge of the cross-correlation and auto-correlation functions are available. Its effectiveness for short duration light curves is uncertain. Moreover, this error estimate does not naturally translate into error estimates for the phase and time lag between the light curves.

Complete information regarding the temporal relation between two light curves can be obtained by computing the coherence and time-lag as a function of Fourier frequency. A detailed description of the technique as well as physical interpretation is given by Nowak et al. (1999). The two light curves are divided into many segments and for each segment a Fourier transform is undertaken and coherence and phase lag as a function of frequency is estimated. For the different segments, the coherence and phase lags are averaged and their errors can be estimated analytically. Such detailed information can only be obtained for long light curves which can be split into several segments. In the absence of such rich data, statistically significant results can be obtained by averaging over Fourier frequencies. Indeed, from this view point, the cross-correlation is, in some sense, the average of the coherence over all frequencies. However, computing the error on the cross-correlation using the error estimates for the coherence is not straight forward. First, the averaging has to be appropriately weighted by the power in each frequency bin. Second, the error estimate for the coherence is reliable only if the error itself is small, which is the case when many segments are averaged and not necessarily true for the coherence at a single frequency bin obtained from a single segment.

In this work, we present an expression for the cross-correlation between two evenly sampled light curves. The error estimate is based on the Fourier transforms of the light curves and then averaging over different frequency modes. As the method does not require the light curve to divide into multiple segments, it will be useful in the analysis of short duration light curves when it cannot be divided into statistically significant segments. The expression has been presented in Section 2 and the same has been verified by simulations with and without measurement errors. Section 3 highlights the difficulties in estimating a time-lag and its error using the standard method of finding the peak of the cross-correlation function. The cross-Correlation phasor is introduced in Section 4 which leads to an estimate of the phase lag between the light curves. In the same section, a technique is introduced by which one can measure the time lag and its error. In this method the time-lag measured can be even smaller than the sampling time bin of the light curves. The complete fully self contained algorithm is presented in Section 5 for easy reference. As an example, in Section 6, the technique is applied to the *XMM-Newton* light curves of the highly variable and well studied AGN, Akn 564. In Section 7, the summary and discussion includes a list of important assumptions on which the technique is based and provides examples when the assumptions may not be valid.

2. Analytical error estimate of cross-correlation

2.1. Light curves without measurement errors

We first consider an idealised case of two light curves, X and Y, without measurement errors. The two light curves are of length N, which have recorded the count rates in $j = 0, 1, 2, \ldots, N-1$ discrete equally spaced time intervals, Δt. The mean is subtracted from each of them. Further it is assumed that they are partially linearly dependent on each other by A, such that we have

$$X = x_j \tag{1}$$

$$Y = z_j + Ax_j \tag{2}$$

where x_j and z_j are time-series produced by two independent stochastic processes. Each time series can be conveniently represented, in frequency domain k by its discrete Fourier transform, $\tilde{X}_k$, defined as

$$\tilde{X}_k = \sum_{j=0}^{N-1} X_j \exp(2\pi ijk/N) \tag{3}$$

and a power spectrum is estimated as $P_{Xk} \equiv (2/N)|\tilde{X}_k|^2$. Here the normalisation constant of the power spectrum is used as suggested by Leahy et al. (1983). For a stationary system, the ensemble average (i.e. average of an infinite number of realisations) of the power, $\langle P_{Xk}\rangle$, is a characteristic of the stochastic process. A power derived from a single time series, P_{Xk} is only an estimator of its value. In particular the real and imaginary parts of $\tilde{X}_k$ varying independently can be derived from two independent Gaussian distributions (Timmer and Koenig, 1995). The standard deviation of P_{Xk} from $\langle P_{Xk}\rangle$ is roughly equal to $\langle P_{Xk}\rangle$ i.e. the power estimate from a single light curve has nearly 100% sampling variation. The variance $\sigma_X^2 \equiv \sum P_{Xk}$ is again an estimate of the ensemble averaged variance $\langle\sigma_X^2\rangle = \sum\langle P_{Xk}\rangle$ (Van der Klis, 1989) where $k = -N/2, \ldots, N/2-1$ and $k \neq 0$.

One can define the cross-correlation estimate of the two time series as

$$C_{XY} = \frac{c_{XY}}{\sqrt{\sigma_X^2\sigma_Y^2}} \tag{4}$$

where

$$c_{XY} == \frac{1}{N}\sum_{j=0}^{N-1} X_j Y_j = \frac{1}{N^2}\sum_{k=-N/2}^{N/2-1} \tilde{X}_k\tilde{Y}_k^*. \tag{5}$$

Here $\tilde{X}_k$ and $\tilde{Y}_k$ are Discrete Fourier transforms of X_j and Y_j, respectively, and

$$\sigma_X^2 = \frac{1}{N}\sum_{j=0}^{N-1} X_j^2 = \frac{1}{N^2}\sum_{k=-N/2}^{N/2-1} |\tilde{X}_k|^2$$

$$\sigma_Y^2 = \frac{1}{N}\sum_{j=0}^{N-1} Y_j^2 = \frac{1}{N^2}\sum_{k=-N/2}^{N/2-1} |\tilde{Y}_k|^2. \tag{6}$$

[2] http://www.vni.com.