Conditional Shannon Entropy Method to Detect Periods on Variable Stars

Gabriel Lauffer Ramos[1], Shashi Kanbur[2], Fabricio Ferrari[1]

1- Grupo de Astrofísica Teórica e Computacional - Universidade Federal do Rio Grande
2- State University of New York at Oswego

Introduction

In astronomy, especially in the field of variable stars, usually is necessary to analyze data with unknown period. There are many methods developed to deal with evenly-spaced data but unfortunately the observations are limited to the telescope availability and weather. For that reason, the spacing between consecutive observations may range from hours to months [2] , resulting in a time series unevenly-spaced. The pulsation period of a variable star can tell us many other properties as luminosity, mass, distance, mean density, etc.

The methods used to search for periodicity in astronomical time series may be divided into two groups: Fourier techniques and phase-diagram analysis. The first group includes the classical Fourier transform [4] , and its variations introduced to deal with unevenly sampled data [2] . The second one is based on the analysis of the dispersion of the light curve on the phase space [5, 1] . With a great variety of methods the question arises as to which one is the best.

Each method has certain advantages and disadvantages. For example, while Fourier analysis is faster and extremely efficient, it is much less sensitive to non-sinusoidal functions than the phase-diagram analysis. Also, phase-diagram methods are less affected by randomly occurring gaps in the data, provided that the coverage of the light curve is reasonably uniform in phase space. Even with a great variety of methods, none of them is superior to each other [3]. Here we present a phase-diagram analysis method called Conditional Shannon Entropy. The idea is to develop a fast and reliable method to work with variable stars because the pulsation period tell us many other properties as luminosity, mass, distance, mean density, etc.

Conditional Shannon Entropy

A light curve folded at most trial periods will produce a random arrangement of points.

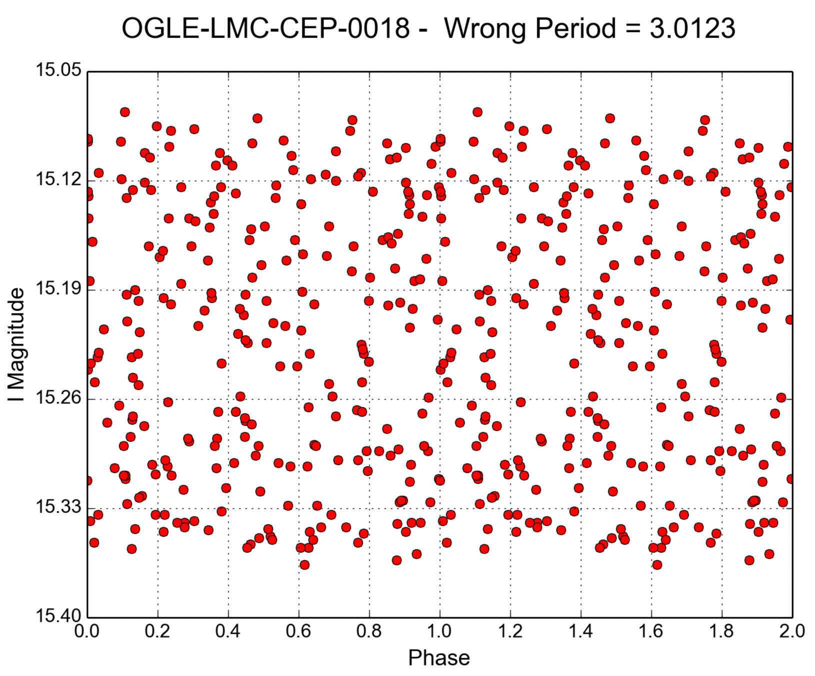

Figure 1 : Data phased with a wrong period

but when folded at the correct period, the light curve will be the most ordered arrangement of data points in the region and so contain the most information about the signal. As entropy measures the lack of information about a system, the correct period minimizes this quantity .

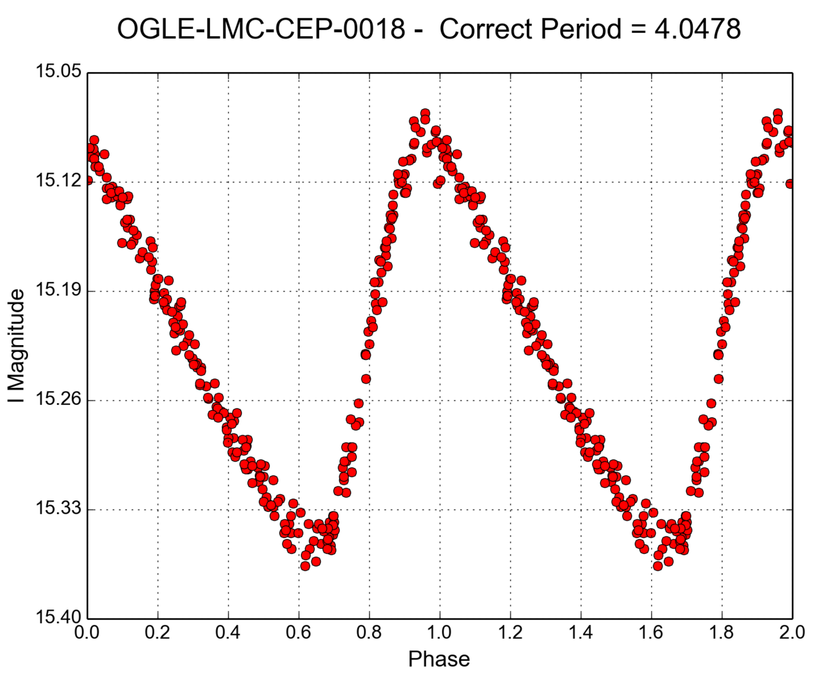

Figure 2 : Data phased with the correct period

The Conditional Shannon Entropy is calculated as follow,

$$H = \sum_{i,j} p(m_i, \phi_j) \ln \left(\frac{p(\phi_j)}{p(m_i, \phi_j)} \right)$$

Where:

- $p(m_i, \phi_j)$ is the probability of m_i given ϕ_j.
- $p(\phi_j)$ is the probability over the jth phase bin.

OGLE-III Data-Set

The OGLE Catalog (*Optical Gravitational Lensing Experiment*) is a online catalog of variable stars. It searches for variable stars in the Magellanic Clouds and the Galactic Bulge.

Methodology

It was created an algorithm in Python3 to calculate the conditional Shannon entropy of two data-sets extracted from the OGLE-III Catalog.

The algorithm works in the following way:

- For each star
  - a period array is created
    - the phase-space is calculated for each period
    - the entropy of the dispersion is calculated and stored
  - The minimum entropy of all periods is found
  - the period associated with the minimum entropy is chosen as the correct one

Results

Star type	Amount	Correct Periods	Percentage
Cepheid	3056	3048	99.74 %
RRLyrae	22651	22075	97.46 %
Total	**25707**	**25123**	**97.73 %**

Table 1 : Amount of analyzed data and correct periods

Conclusion

- The method returns better results for Longer period (Cepheids)
- For small period stars has a slightly smaller rate but a considerable one
- Overall, the Conditional Shannon Entropy has a success rate 97.7% with we conclude as a reliable method to work with variable stars.

Next steps

- Analyze the behavior of the method for simulated data with small periods and different levels of noise
- Analyze T2 stars (Cepheid with periods $\geq$ 100 days)
- Search for multiples periods on the data-set
- Speed up the algorithm

References

[1] P. M. Cincotta, M. Mendez, and J. A. Nunez.
Astronomical Time Series Analysis. I. A Search for Periodicity Using Information Entropy.
The Astrophysical Journal, 449:231, August 1995.

[2] S. Ferraz-Mello.
Estimation of Periods from Unequally Spaced Observations.
The Astronomical Journal, 86:619, April 1981.

[3] M. J. Graham, A. J. Drake, S. G. Djorgovski, A. A. Mahabal, C. Donalek, V. Duan, and A. Maker.
A comparison of period finding algorithms.
Monthly Notices of the RAS, 434:3423–3444, October 2013.

[4] N. R. Lomb.
Least-squares frequency analysis of unequally spaced data.
Astrophysics and Space Science, 39:447–462, February 1976.

[5] A. Schwarzenberg-Czerny.
On the advantage of using analysis of variance for period search.
Monthly Notices of the RAS, 241:153–165, November 1989.

Acknowledgments

I would like to thank Dr. Ashish Mahabal and Dr. Matthew Graham for the support at Caltech, the State University of New York at Oswego for the grant and the International Institute of Education, the Brazilian government and CAPES Foundation for the grant number 88888.060166/2013-00.

gabriellramos@gmail.com, fabricio@ferrari.pro.br