Tests for Periodicity in Short Time Series and Their Approximate Distributions

Yuanhao Lai[a,1], Ian McLeod[b,2]

[a] *Western University, London, Ontario, Canada*
[b] *Western University, London, Ontario, Canada*

Abstract

Fisher's *g* test and several other tests for periodicity are compared for short time series. An accurate approximation to the distribution function of the test statistics is derived using response surface regression. To fit this regression, a large simulation was carried out on a large cluster computing facility, SHARCNET, using the R programming environment. Methods of ensuring an accurate fit of the response surface regression are discussed. The final cumulative distribution functions are parameterized by a relatively small number of parameters enabling efficient computation. A complete R package available on CRAN for periodicity testing is described.

Keywords: Accurate approximate cumulative distribution functions, Cluster computing with R, Fisher's *g* test, Microarray time series experiments, Periodicity tests, Response surface regression

[1]Email: ylai72@uwo.ca
[2]Email: aim@stats.uwo.ca

1. Introduction

The problem of detecting periodicity in short time series arises in microarray time series experiments and there are potentially many other possible areas where it is of interest to detect periodicity in short signals. In the case of microarray experiments, DNA expression is obtained at regular time intervals to obtain n consecutive time series observations for each of the many thousands of genes on the microarray. It is of interest to find which genes may exhibit cyclic or periodic behavior. In many cases $10 \leq n \leq 100$ and so the problem is one of testing for periodicity in short time series. This is illustrated in a microarray time series experiment with a type of yeast, C. crescentus, (Wichert et al., 2004). The left panel in Fig. 1 shows a four parameter harmonic regression model fitted to $n = 11$ consecutive observations for the gene identified as ORF06806. The fitted model four parameter harmonic regression, shown in Fig. 1 strongly suggests periodicity. The residual-fit plot is shown in the right panel in Fig. 1. Overall these plots suggest the model is reasonable but the p-value from Fisher's test for periodicity (Fisher, 1929) is only 11.9%. Four improved tests for periodicity in short time series are discussed in the next section and a computationally accurate and efficient method is developed using response surface regression (RSR) to compute their p-values.

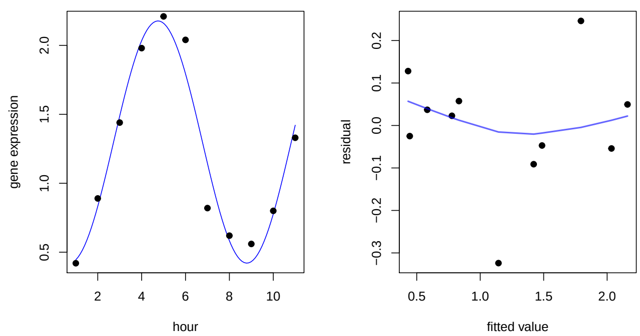

Figure 1: Fitted harmonic regression (left), $\hat{z}_t = \hat{\mu} + \hat{A}\cos(2\pi\hat{\lambda}t) + \hat{B}\sin(2\pi\hat{\lambda}t)$, where $t = 1, \ldots, 11$, $\hat{\mu} = 1.2991$, $\hat{A} = -0.7528$, $\hat{B} = -0.4523$, $\hat{\lambda} = 0.1235$,and residual-fit plot (right). The estimated frequency $\hat{\lambda} = 0.1235$ implies a period of about 8 hours.

While longer signals are of great interest in many applications and extensive research has been carried out on this problem (Quinn and Hannan, 2001), Fig. 1 demonstrates that the problem of periodicity detection in short series is of practical interest.

$\boxed{\text{to be continued ...}}$

References

Fisher, R.A., 1929. Tests of significance in harmonic analysis. Proceedings of the Royal Society of London. Series A , 54–59.

Quinn, B.G., Hannan, E.J., 2001. The Estimation and Tracking of Frequency. Cambridge University Press.

Wichert, S., Fokianos, K., Strimmer, K., 2004. Identifying periodically expressed transcripts in microarray time series data. Bioinformatics 20, 5–20. doi:10.1093/bioinformatics/btg364.