Testing for gene expression rythmicity

Eduardo García-Portugués • Department of Statistics, UC3M
Yolanda Larriba • Department of Statistics, Universidad de Valladolid

Circadian genes in humans coordinate a wide range of physiological processes and mediate the organism's interaction with the external light–dark cycle. Their expression typically exhibits near-24 h oscillations whose phases (timing of peak expression) align with underlying biological functions. Detecting rhythmicity in gene-expression data is essential to understand temporal regulation and to enable synchronization-aware analyses in humans, from body-temperature cycles and endocrine rhythms to applications in personalized medicine.

In practice, statistical testing of rhythmicity is challenging due to limited or uneven coverage across the 24 h cycle, measurement noise, irregular sampling times, and outliers. These issues are particularly acute in transcriptomic data from both controlled experiments and post-mortem cohorts. Existing approaches cast the problem as testing the constancy ("no effect") of an unknown regression function $m : [0, 24) \to \mathbb{R}$ that maps hourly times to expression level, i.e., testing $H_0 : m(x) = c$ versus $H_1 : m(x) \neq c$. Parametric, likelihood-based approaches include the cosinor model (Cornelissen, 2014) and the more flexible Frequency-Modulated Möbius (Rueda et al., 2019). Nonparametric procedures, which relax shape assumptions, include the rank-based JTK-Cycle test (Hughes et al., 2010), the order-restricted test by Larriba et al. (2016), the arc-scanning test by Deschepper et al. (2008), kernel-based lack-of-fit tests on the circle (García-Portugués et al., 2016), and possible adaptations of tests based on the integrated regression function (Escanciano, 2006; Stute, 1997). A systematic, simulation-based evaluation, especially of nonparametric methods, under realistic sampling designs and contamination scenarios remains lacking.

This thesis aims to review, implement, develop, and compare several tests for gene-expression rhythmicity through realistic simulations and applications. The methods to be investigated include the aforementioned parametric and nonparametric approaches, with particular emphasis on implementing the Deschepper et al. (2008) method and, as a methodological contribution, designing and implementing tests based on the integrated regression function adapted to the periodic time domain. All methods will be implemented in R and benchmarked in terms of Type-I error, power, and robustness to irregular sampling distribution times and outliers using biologically-motivated simulation scenarios that include diverse waveforms, noise structures, and small-sample regimes. Finally, the most reliable procedures will be applied to real human post-mortem gene-expression datasets to identify and characterize rhythmic genes across tissues, reporting cross-method concordance, phase distributions, and functional enrichment.

References

Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, *11*(16), 1–24. https://doi.org/10.1186/1742-4682-11-16

Deschepper, E., Thas, O., & Ottoy, J.-P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, *64*(3), 912–920. https://doi.org/10.1111/j.1541-0420.2007.00950.x

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, *22*(6), 1030–1051. https://doi.org/10.1017/S0266466606060506

García-Portugués, E., Van Keilegom, I., Crujeiras, R. M., & González-Manteiga, W. (2016). Testing parametric models in linear-directional regression. *Scandinavian Journal of Statistics*, *43*(4), 1178–1191. https://doi.org/10.1111/sjos.12236

Hughes, M. E., Hogenesch, J. B., & Kornacker, K. (2010). JTK_CYCLE: An efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, *25*(5), 372–380.

Larriba, Y., Rueda, C., Fernández, M. A., & Peddada, S. D. (2016). Order restricted inference for oscillatory systems for detecting rhythmic signals. *Nucleic Acids Research*, *44*(22), e163–e163.

Rueda, C., Larriba, Y., & Peddada, S. D. (2019). Frequency modulated Möbius model accurately predicts rhythmic signals in biological and physical sciences. *Scientific Reports*, *9*(1), 18701. https://doi.org/10.1038/s41598-019-54569-1

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, *25*(2), 613–641. https://doi.org/10.1214/aos/1031833666