Construction of Circular-Circular Graphical Network via Simulated and Real Data

Elif Doğan Dar and Vilda Purutçuoğlu

Department of Statistics, Middle East Technical University, Ankara, Turkey

Circular/angular data contains directional information which can represent angles or a recurrent phenomenon that repeats in fixed time intervals. Traditional statistical theory does not take into account the circular nature of such data. Therefore, all statistical theory has to be built from the ground up according to the topology of the circular data which is inherently different from the linear data. This project aims to explore the interactions among circular random variables. For this purpose, circular correlation measures or model-based approaches can be used, i.e., many circular regression models have been suggested in the literature where relations between a dependent variable and independent variables are investigated [1,2]. In this work, we build a circular network with circular variables as nodes and interactions among them as edges. We regress each node on others by using regression model suggested by Kim and SenGupta (2017) [2], and declare an edge between dependent and independent variables if corresponding regression coefficient is statistically significant. Coefficients can be estimated by using Least Circular Mean Square Estimation (LCMSE), which corresponds to minimizing sample mean circular distance instead of usual mean square of Euclidean distance. For inference, we use the fact that these estimators are asymptotically normal, however, there are no closed form formula for standard errors. Therefore, bootstrap methods can be used to estimate standard errors. In case that sample is not large enough, these bootstrap samples can also be used for inference as a non-parametric method. We apply a step-wise variable selection method to eliminate edges which are not statistically significant.

We first demonstrated the proposed approach on simulated data. The data were simulated from multivariate normal distribution with a given variance-covariance matrix. Since zeros in the inverse of covariance matrix (precision matrix) implies conditional independence, we give a precision matrix with zeros for non-edges between variables. Then, we wrap generated data around circle and produce a sample coming from wrapped normal distribution. We estimate the network structure via the proposed circular model using this data and compare it with the ground-truth network. We also examined the performance of the proposed approach on real circadian gene data. Circadian genes are expressed in a recurrent way, hence circular in nature. We use microarray data of circular genes taken at 2-3 hours time intervals for at least 1 day, transform this data to angular data via sinusoidal waves [3] and slice it at any given time point. Then, we estimate the network structure of these genes by applying the proposed method and compare with known interaction structure in the literature. We also examine the time effect by comparing networks produced at different sliced time points.

Keywords: Circular data, least circular mean square estimation, graphical network

[1] J. Jha, A. Biswas, "Multiple circular–circular regression," Statistical Modelling, 2017, 17 (3), 142–171.

[2] S. Kim, S. SenGupta, "Multivariate-multiple circular regression," Journal of Statistical Computation and Simulation, 2017, 87 (7), 1277-1291.

[3] D.S. Hickey, J.L. Kirkland, S.B. Lucas, M. Lye "Analysis of circadian rhythms by fitting a least squares sine curve," Computers in Biology and Medicine, 1984, 14 , 217-223.