Analysis of Circular Data from Cell Cycle Transcripts Using Order Restricted Statistical Procedures

Miguel A. Fernández, Sandra Barragán, Cristina Rueda
Dpt. Statistics and Operations Research, Universidad de Valladolid
Valladolid (47005), Spain
E-mail: miguelaf@eio.uva.es, sandraba@eio.uva.es, crueda@eio.uva.es

Shyamal D. Peddada
Biostatistics Branch, National Institute of Environmental Health Sciences
RTP, NC (27709), United States of America
E-mail: peddada@niehs.nih.gov

The problem

Circular data arise in many applications such as, meteorology (wind directions), physics (orbits of planets), neuroscience (orientation selectivity), psychology (studies of mental maps), criminology, political sciences, geography and many others. The books by Fisher (1995) and Mardia and Jupp (2000) contain a good number of statistical techniques for descriptive and inferential analysis of circular data.

Depending upon the scientific question, there are often constraints on the parameters of the circular data that may be important when performing statistical inferences. Often these constraints are in the form of "inequalities", known as order restrictions. Order restrictions in the context of Euclidean space data has over 50 years of literature with four books written on the subject (c.f. Robertson et al. (1988) and Silvapulle and Sen (2005), and references therein). However, until recently (Rueda et al. (2009) and Fernandez et al. (2011)), there did not exist any literature on order restricted inference for circular data.

The work in Rueda et al. (2009) and Fernandez et al. (2011) was motivated by problems encountered in cell biology. Due to the topology of the space, methodology developed for Euclidean space data do not directly carry over to the circular case. Notice that, unlike the Euclidean space, the points wrap around on a unit circle. This means that if we start at a point and travel around the circunference we will return to the starting point. Thus a circle cannot be cut into a straight line without destroying this important feature of a circle.

In fact, one of the first questions is to define appropriate order on a circle. Notice that, in many applications, the starting point on the circumference of a circle (i.e. the pole of a circle) is arbitrarily chosen. This means that in these applications the order should be rotation invariant. With this in mind, in Rueda et al (2009) we considered the isotropic order. The components of a vector of parameters $\phi = (\phi_1, ..., \phi_p)$ *are said to follow an isotropic order (denoted as* $\phi_1 \prec ... \prec \phi_p \prec \phi_1$*) if* ϕ_i *follows* ϕ_{i-1} *and is followed by* ϕ_{i+1} *, where* $\phi_0 \equiv \phi_p$ *and* $\phi_{p+1} \equiv \phi_1$*.*

The pool adjacent violators algorithm, commonly used for constrained inference problems in Euclidean space, is not applicable for estimating parameters under isotropic order constraint on a unit circle. This is because the Cauchy mean value property is not satisfied by angular mean. To deal with this complexity, Rueda et al. (2009) introduced a modification to PAVA for circular data and developed circular isotonic regression estimators (CIRE) which are the restricted maximum likelihood estimators (RMLE) if the angular data are assumed to follow the von-Mises distribution with a common concentration parameter κ*.*

Using CIRE developed in Rueda et al. (2009), in Fernandez et al. (2011) we discussed the problem of testing the null hypothesis that a vector of angular parameters satisfy an isotropic order against the alternative that they do not. Since the likelihood ratio test can be computationally intensive *as p increases, in Fernandez et al. (2011) we developed a conditional test which has a simple chi-square approximation for large values of the common concentration parameter κ. The proposed conditional test is computationally simple and has a good power against interesting alternatives. In that paper we also prove that the test is asymptotically an α level test which resulted to be a quite involved property due to the fact that the least favourable configuration of parameters under the null hypotheses is not attained at the usual equality point.*

In order to make all these procedures available to researchers, we have developed an R library named isocir containing all necessary procedures to make isotropic inferences for circular data. This library is fully described in Barragán et al. (2011).

We illustrate the proposed methodology using cell cycle data obtained from three cell-cycle experiments, namely, Rustici et al. (2004), Oliva et al. (2005) and Peng et al. (2005).

Acknoledgments

MAF's, SB's and CR's research was partially supported by Spanish MCI grant MTM2009-11161. SDP's research [in part] was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Institute of Environmental Health Sciences (NIEHS) (Z01 ES101744-04).

REFERENCES

Barragán, S., Fernández, M. , Rueda, C. & Peddada, S. (2011). isocir: An R package for Isotonic Inference in Circular data. An application to the analysis of cell cycle expression data. *Submitted.*

Fernández, M., Rueda, C. & Peddada, S. (2011). Conservation of Isotropic Order among a Core Set of Cell-cycle Genes between Budding Yeast, Fission Yeast and Humans. *Submitted.*

Fisher, N.I. (1995). Statistical Analysis of Circular Data. Cambridge University Press

Mardia K. & Jupp P (2000). Directional Statistics. Wiley.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S. et al. (2005) The cell cycle-regulated genes of Schizosaccharomyces pombe. *PLoS Biol. 3(7), 1239-1260.*

Peng, X., Karuturi, R.K.M., Miller, L.D., Lin, K. et al. (2005) Identification of Cell cycle-regulated Genes in Fission Yeast. *Molec. Biol. Cell, 16, 1026-1042.*

Robertson T, Wright F, Dykstra R (1988). Order Restricted Statistical Inference. Wiley.

Rueda, C., Fernández, M. & Peddada, S. (2009). Estimation of Parameters subject to Order Restrictions on a Circle with Application to Estimation of Phase Angles of Cell-cycle Genes. *JASA, 104(495), 338-347.*

Rustici, G., Mata, J., Kivinen, K., Lió, P. et al. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nature Genet., 36, 809-817.*

Silvapulle, M.J. & Sen, P.K. (2005). Constrained Statistical Inference. Wiley.