isocir: An R Package for Constrained Inference using Isotonic Regression for Circular Data, with an Application to Cell Biology.

Sandra Barragán
Universidad de Valladolid

Miguel A. Fernández
Universidad de Valladolid

Cristina Rueda
Universidad de Valladolid

Shyamal Das Peddada
National Institute of
Environmental Health Sciences

Abstract

In many applications one may be interested in drawing inferences regarding the order of a collection of points on a unit circle. Due to the underlying geometry of the circle standard constrained inference procedures developed for Euclidean space data are not applicable. Recently, statistical inference for parameters under such order constraints on a unit circle was discussed in Rueda *et al.* (2009); Fernández *et al.* (2012). In this paper we introduce an R package called **isocir** which provides a set of functions that can be used for analyzing angular data subject to order constraints on a unit circle. Since this work is motivated by applications in cell biology, we illustrate the proposed package using a relevant cell cycle data.

Keywords: Circular Data, Circular Order, CIRE, Conditional Test, R package **isocir**, R.

---

1. Introduction

Circular data arise in a wide range of contexts, such as in geography, cell biology, circadian biology, endocrinology, ornithology, etc (cf Zar 1999 or Mardia *et al.* 2008). This work is motivated by applications in cell cycle biology where one may want to draw inferences regarding angular parameters subject to order constraints on a unit circle. A cell cycle among eukaryotes follows a well-coordinated process where cells go through cycle four phases of distinct biological functions, namely, G1, S, G2 and M phase (see Figure 1).

Genes participating in the cell division cycle are often called cell cycle genes. A cell cycle gene has a periodic expression with its peak expression occurring just before its biological function (Jensen *et al.* 2006). Since the periodic expression of a cell cycle gene can be mapped onto a unit circle, the angle corresponding to its peak expression is known as the phase angle of the gene (see Liu *et al.* (2004) and Figure 2).
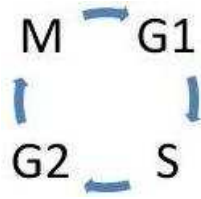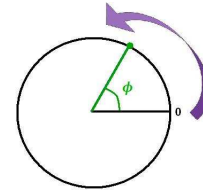
Figure 1: Phases of a cell cycle.

Figure 2: Phase angle (ϕ).

Since cell cycle is fundamental to the growth and development of an organism, cell biologists have been interested in understanding various aspects of cell cycle that are evolutionarily conserved. For instance, they would like to identify genes whose relative order of peak expressions is evolutionarily conserved. In order to solve such problems Rueda *et al.* (2009) introduced an order restriction on the unit circle, called circular order and extended the notion of isotonic regression estimator to circular parameter space by defining the circular isotonic regression estimator (CIRE). Using CIRE, Fernández *et al.* (2012) developed a formal statistical theory and methodology for testing whether the circular order of peak expression of a subset of cell cycle genes is conserved across multiple species.

These statistical methods may have numerous other applications apart from cell cycle. With the increase in human survival rates, there is considerable interest in understanding neurological diseases related to aging such as the Alzheimer's disease (AD) and the Parkinson's disease (PD). An important aspect of such neurological diseases is the disruption of circadian clock and genes participating in it. Researchers are interested in testing differences in the phases of expression of circadian genes in normal and AD patients (Cermakian *et al.* 2011). Methodology discussed in this paper can be used for analyzing such data. Other areas of applications include; the study of migratory patterns and directions of birds (Cochran *et al.* 2004), the changes in the wind directions (Bowers *et al.* 2000), directional fluctuations in the atmosphere (van Doorn *et al.* 2000), psychology (studies of mental maps or monitoring data (Kibiak and Jonas 2007), the orientation of ridges in fingerprints or magnetic maps (Boles and Lohmann 2003).

Motivated by the wide range of applications and the non-existence of a user friendly software, in Section 3 we introduce the **isocir** package programmed in R environment, R Development Core Team (2012), which can be downloaded from `http://cran.r-project.org/web/packages/isocir/`. The package provides functions which can be used for drawing inferences regarding the order of a collection of points on a unit circle. In Section 2 we describe the statistical problem and the methodology of Rueda *et al.* (2009) and Fernández *et al.* (2012). The **isocir** package is illustrated in Section 4 using the motivating cell cycle gene expression data. Some concluding remarks are provided in Section 5.

2. Angular parameters under order constraints

2.1. Circular order restriction

Let $\overline{\theta} = (\overline{\theta}_1, \overline{\theta}_2, \ldots, \overline{\theta}_q)$ where each $\overline{\theta}_i$, $i = 1, 2, \ldots, q$ is the sample circular mean of a random sample of size n_i from a population with unknown circular mean ϕ_i. All angles are defined in a counter clockwise direction relative to a given pole. The mean resultant lengths for each population are denoted as $r_1, \ldots, r_q$ (see Mardia and Jupp (2000) for the definition of circular mean and mean resultant length of a set of angles). Then the problem of interest is to draw statistical inferences on ϕ_i, $i = 1, 2, \ldots, q$, subject to the constraint that the angles $\phi_1, \phi_2, \ldots, \phi_q$ are in a counter clockwise order on a unit circle. Thus ϕ_1 is "followed" by ϕ_2 which is "followed" by $\phi_3, \ldots, \phi_q$ is "followed" by ϕ_1. More precisely, we shall denote this simple circular order among angular parameters as follows:

$$C_{sco} = \{\phi = (\phi_1, \phi_2, \ldots, \phi_q) \in [0, 2\pi]^q : \phi_1 \preceq \phi_2 \preceq \cdots \preceq \phi_q \preceq \phi_1\}. \tag{1}$$

It is important to note that the order among the angular parameters is invariant under changes in location of the pole (the initial point of the circle). Unlike the Euclidean space, points on a unit circle wrap around. That is, starting at the pole, by traveling 2π radians around the circumference of the circle one would return to pole. For this reason the circular order among points on a unit circle is preserved even if the location of the pole is shifted. This is why Rueda *et al.* (2009) and Fernández *et al.* (2012) refer to the circular order C_{sco} as isotropic order. As in this paper we will also consider more general circular order restrictions, from now on we will refer to C_{sco} as simple circular order.

As a consequence of the geometry, a circle can never be linearized and hence methods developed for Euclidean space data are not applicable to circular data. The problem is even more challenging when the angular parameters are constrained by an order around the circle, such as C_{sco}. General methodology for circular data, when there are no constraints on the angular parameters, can be found in the book Mardia and Jupp (2000), among others. Constrained inference for circular data is rather recent (Rueda *et al.* (2009) and Fernández *et al.* (2012)). As noted in Rueda *et al.* (2009), standard Euclidean space methods such as the pool adjacent violators algorithm (PAVA) used for computing isotonic regression (see Robertson *et al.* (1988) for details) cannot be applied to circular data. For example, when a cell biologist is investigating a large number of cell cycle genes, it may be difficult to ascertain the circular order among all cell cycle genes under consideration. However, based on the underlying biology, the investigator may a priori know the circular order among groups of genes but not the order among genes within each group. In such situations a partial circular order, C_{pco} as defined below, can be used.

$$C_{pco} = \left\{ \phi \in [0, 2\pi]^q : \left\{ \begin{array}{c} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{l_1} \end{array} \right\} \preceq \left\{ \begin{array}{c} \phi_{l_1+1} \\ \phi_{l_1+2} \\ \vdots \\ \phi_{l_1+l_2} \end{array} \right\} \preceq \cdots \preceq \left\{ \begin{array}{c} \phi_{l_1+\ldots+l_{L-1}+1} \\ \phi_{l_1+\ldots+l_{L-1}+2} \\ \vdots \\ \phi_{l_1+\ldots+l_L} \end{array} \right\} \preceq \left\{ \begin{array}{c} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{l_1} \end{array} \right\} \right\}. \tag{2}$$

In this case we have L sets of parameters with l_j angular parameters in set j and $q = \sum_{j=1}^{L} l_j$. Order among the parameters within a set is not known but every parameter in a given set is "followed" by every parameter in the next set.

2.2. Estimation and testing under circular order restrictions

Analogous to PAVA for Euclidean data, Rueda *et al.* (2009) derived a circular isotonic regression estimator (CIRE) for estimating angular parameters $(\phi_1, \phi_2, \ldots, \phi_q)$ subject to a circular order. The CIRE of $\phi = (\phi_1, \phi_2, \ldots, \phi_q)$, under the constraint $(\phi_1, \phi_2, \ldots, \phi_q)' \in C_{sco}$, is given by:

$$\widetilde{\phi} = \arg \min_{\phi \in C_{sco}} SCE(\phi, \overline{\theta}), \tag{3}$$

where $SCE(\phi, \overline{\theta})$, defined below, is the sum of circular errors (SCE), a circle analog to the Sum of Squared Errors (SSE) used for Euclidean data;

$$SCE(\phi, \overline{\theta}) = \sum_{i=1}^{q} r_i (1 - cos(\phi_i - \overline{\theta}_i)), \tag{4}$$

where r_i are the mean resultant lengths. The CIRE is implemented in the function `CIRE` of the package **isocir**.

Just as the normal distribution is commonly used for the Euclidean space data, the von-Mises distribution is widely used for describing angular data on a unit circle. Accordingly, for $i = 1, 2, \ldots, q$, throughout this paper we assume that $\overline{\theta}_i$ are independently distributed according to von-Mises distribution, denoted as $M(\phi_i, \kappa)$ where ϕ_i is the mean direction and κ is the concentration parameter (see Mardia and Jupp (2000)). Under such an assumption, Fernández *et al.* (2012) developed a conditional test for testing the following hypotheses:

$$H_0 : \quad \phi \in C_{sco}$$
$$H_1 : \quad H_0 \text{ is not true.}$$

The conditional test statistic is given by $T^* = \frac{2\hat{\kappa} SCE(\widetilde{\phi}, \overline{\theta})}{q}$, where $\hat{\kappa}$ is the estimator of κ, $\widetilde{\phi}$ is the CIRE computed under H_0. The estimate $\widetilde{\phi}$ determines a partition of $\wp = \{1, \ldots, I\}$ into sets of coordinates on which $\widetilde{\phi}$ is constant. These sets are called level sets. The rejection region for the conditional α-level test is given by:

$$\text{Reject } H_0 \text{ if } T^* \geq c(m),$$

where m is the number of level sets for $\widetilde{\phi}$ and, for large values of κ, the approximate critical value $c(m)$ is chosen so that:

$$pr(F_{q-m,q-1} \geq c(m)) = \frac{\alpha}{1 - 1/(q-1)!}, \tag{5}$$

where $F_{q-m,q-1}$ represents the central F random variable with $(q-m, q-1)$ degrees of freedom. The above test statistic is proportional to a chi-square test when κ is known. For details one may refer to Fernández *et al.* (2012). The above methodology can be modified to test

$$H_0 : \quad \phi \in C_{pco}$$
$$H_1 : \quad H_0 \text{ is not true}$$

by replacing $1/(q-1)!$ by $(l_1! l_2! \cdots l_L!)/(q-1)!$.

The simulation study performed in Fernández *et al.* (2012) suggests that the power of this test is quite reasonable. Notice that, for a given data set, the p-value obtained by using the above methodology may serve as a useful goodness of fit criterion when comparing two or more plausible circular orders among a set of angular parameters. Larger values may suggest that the estimations are closer to the presumed circular order. Thus the statistical methodology developed in Fernández *et al.* (2012) can be used not only for testing relative order among the parameters. It can be also useful for selecting a "best fitting" circular order among several circular order candidates.

These tests are implemented in the function `cond.test` in the R package **isocir** introduced in the next Section.

3. Package isocir

We start this section by giving some background on R packages for isotonic regression and analysis of circular data. We then describe the structure of our package **isocir** and illustrate it by some examples.

3.1. Related packages

As isotonic regression is a well-known and widely used technique there are many packages in R for performing isotonic regression, such as:

- **isotone** (de Leeuw *et al.* 2011): Active set and generalized PAVA for isotone optimization.

- **Iso** (Turner 2009): Functions to perform isotonic regression.

- **ordMonReg** (Balabdaoui *et al.* 2009): Compute least squares estimates of one bounded or two ordered isotonic regression curves.

Similarly, there are several packages in R for analyzing circular data, such as:

- **CircStats** (Lund and Agostinelli 2009): The implementations of the Circular Statistics from "Topics in circular Statistics" Jammalamadaka and SenGupta (2001). It is an R port from the S-plus library with the same name.

- **circular** (Agostinelli and Lund 2011): This package expands in several ways the **CircStats** package.

Since none of the existing packages for circular data are applicable for analyzing circular data under constraints, in this article we introduce the software package "**iso**tonic inference for **cir**cular data", with the acronym **isocir**, for analyzing circular data under constraints. Our package depends on **circular** (see (Agostinelli and Lund 2011)) and **combinat** (see (Chasalow 2010)). These packages should be installed in the computer before loading **isocir**.

3.2. Package structure

For the convenience of the reader, we summarize all the functions, arguments and descriptions of our package **isocir** in Table 1.

Functions	Arguments	Description
sce	(arg1, arg2, meanrl)	SCE
mrl	(data)	Mean resultant length
CIRE	(data, groups, circular)	Calculates the CIRE
cond.test	(data, groups, kappa)	Conditional test
isocir	(cirmeans,SCE,CIRE,pvalue,kappa)	S3 object of class isocir
is.isocir	(x)	Checks class isocir
print.isocir	(x,decCIRE,decpvalue,deckappa,...)	Prints an object of class isocir
plot.isocir	(x, option,...)	Plots an object of class isocir

Table 1: Summary of the components of the package **isocir**.

In the following we describe each function of the software in detail.

- **Functions sce() and mrl()**

The auxiliary function sce computes the sum of circular errors between a given q-dimensional vector (denoted by arg1) and one or more q-dimensional vectors (denoted by arg2). The function mrl computes the mean resultant length for the input data.

- **Function CIRE()**

Using the methodology developed in Rueda *et al.* (2009), this R function computes the CIRE for a given circular order (1) or a partial order (2). The arguments of this function are summarized in Table 2. The input variable data is a matrix where each column contains

Arguments	*Values*
data	vector or matrix with the data
groups	the groups of the order
circular	=TRUE(by default) / =FALSE

Table 2: Arguments of the CIRE function.

the vector of unconstrained angular means corresponding to each replication. If there is only one replication then data is a vector. The position i in the vector groups contains the group number to which the parameter ϕ_i belongs to. The logical argument circular sets whether the order is wrapped around the circle, i.e., circular order (circular = TRUE) or not, i.e., simple order (circular = FALSE). For example, the simple order cone in the circle $C_{so} = \{\phi = (\phi_1, \phi_2, \ldots, \phi_q) \in [0, 2\pi]^q : 0 \le \phi_1 \le \phi_2 \le \cdots \le \phi_q \le 2\pi\}$ would be a non circular order. The output of this function is an object of class isocir (explained later) containing the circular isotonic regression estimator ($\widetilde{\phi}$), the unrestricted circular means ($\overline{\theta}$) and the corresponding sum of circular error ($SCE(\widetilde{\phi}, \overline{\theta})$).

- **Function cond.test()**

This function performs the conditional test and computes the corresponding p-value for the following hypotheses:

H_0 :   The angles $\phi_1, \ldots, \phi_q$ follow a (simple or partial) circular order.

H_1 : H_0 is not true.

The arguments of this function appear in Table 3 and are explained below. Arguments `data`

Arguments	κ known	κ unknown
`data`	numeric vector	matrix (as many columns as replications)
`groups`	numeric vector with the groups of the order to be tested	
`kappa`	positive numeric value	(NULL)
`biasCorrect`	(NULL)	=TRUE(by default) / =FALSE

Table 3: Arguments of the `cond.test` function.

and `groups` are same as those in function `CIRE` although in this function `groups` is the order to be tested instead of the known order. The argument `kappa` is needed only when `data` is a vector. If there are no replications in `data`, the value of κ must be set by the user. Even when there are replicated data, if the user knows the value of κ, it may be introduced and it will be taken into consideration to perfom the conditional test. When κ is unknown and there are replicated data, the parameter is internally estimated by maximum likelihood and $\hat{\kappa}$ is shown in the output. The `biasCorrect` is related to the estimation of κ. If `biasCorrect=TRUE` the bias correction appearing in Mardia and Jupp (2000) p. 87 is performed in the estimation of κ. The output of this function is an object of class `isocir` (explained below) with all the results from the conditional test: The CIRE ($\widetilde{\phi}$), the unrestricted circular means ($\overline{\theta}$), the SCE ($SCE(\widetilde{\phi}, \overline{\theta})$), the kappa value (estimated or introduced) and the p-value of the conditional test.

- **Class `isocir`**

Finally we describe the `isocir` class. The `isocir` function creates the S3 objects of class `isocir` which is a list with the following elements:

`$cirmeans` is a list with the unrestricted circular means. Notice that when the argument `data` is a vector, these values match exactly with the input. However, if there are replicated data, the argument `data` is a matrix and `$cirmeans` contains the corresponding unrestricted circular means ($\overline{\theta}_1, \ldots, \overline{\theta}_q$).

`$SCE` is the value of the Sum of Circular Errors ($SCE(\widetilde{\phi}, \overline{\theta})$).

`$CIRE` is a list with the Circular Isotonic Regression Estimator ($\widetilde{\phi}$) obtained under the order defined by the `groups` argument.

`$kappa` the value of kappa (either set by the user or estimated).

`$pvalue` the p-value of the conditional test obtained from the function `cond.test`.

These objects of class `isocir` are the output of the functions `CIRE` and `cond.test`. The last two elements of the list (`$kappa` and `$pvalue`) are `NULL` if the object comes from the function `CIRE`. Otherwise, if the object comes from the function `cond.test` not only there are the results of the conditional test (`$kappa` and `$pvalue`) but also an attribute called `estkappa`

will inform (or rather remind) the user if the value in `$kappa` has been internally estimated or introduced as a known input.

Some S3 methods have also been defined for the class `isocir`:

- `isocir(cirmeans = NULL, SCE = NULL, CIRE = NULL, pvalue = NULL, kappa = NULL)`: This function creates an object of class `isocir`.

- `is.isocir(x)`: This function checks whether the object `x` is of class `isocir`.

- `print.isocir(x, decCIRE, decpvalue, deckappa, ...)`: This S3 method is used to print an object `x` of class `isocir`. The number of decimal places can be chosen.

- `plot.isocir(x, option = c("CIRE", "cirmeans"), ...)`: This S3 method is used to plot an object `x` of class `isocir`. The argument `option` gives the user the option to plot the points of the Circular Isotonic Regression Estimator (by default) or the unrestricted circular means.

3.3. Examples

In this section we provide examples to illustrate **isocir**.

Example 1

Suppose the observed angular means of eight populations are given by:

$\overline{\theta}_1 = 0.025$; $\overline{\theta}_2 = 1.475$; $\overline{\theta}_3 = 3.274$;
$\overline{\theta}_4 = 5.518$; $\overline{\theta}_5 = 2.859$;
$\overline{\theta}_6 = 5.387$;
$\overline{\theta}_7 = 4.179$; $\overline{\theta}_8 = 1.962$.

We illustrate **isocir** for estimating the 8 population angular parameters under the following partial circular order constraint:

$$\left\{ \begin{array}{c} \phi_1 \\ \phi_2 \\ \phi_3 \end{array} \right\} \preceq \left\{ \begin{array}{c} \phi_4 \\ \phi_5 \end{array} \right\} \preceq \{\phi_6\} \preceq \left\{ \begin{array}{c} \phi_7 \\ \phi_8 \end{array} \right\} \preceq \left\{ \begin{array}{c} \phi_1 \\ \phi_2 \\ \phi_3 \end{array} \right\}$$

These data are a set of random circular data called `cirdata` in our package and they can be used by calling as below:

```
R> data("cirdata")
R> cirdata
```

```
[1] 0.025 1.475 3.274 5.518 2.859 5.387 4.179 1.962
```

Since in this example, there are no replications, we provide data in a vector format. The groups of the order are defined as follows:

```
R> orderGroups <- c(1, 1, 1, 2, 2, 3, 4, 4)
```

Thus we obtain CIRE using the function `CIRE` as follows:

```
R> example1CIRE <- CIRE(cirdata, groups = orderGroups,
+                        circular = TRUE)
```

The output is saved in `example1CIRE` and the printed output is as follows:

```
R> example1CIRE

 Circular Isotonic Regression Estimator (CIRE):
    0.993 1.475 3.066
    5.056 3.066
    5.056
    5.056 0.993
 Sum of Circular Errors: SCE =  1.428
Invisible: Unrestricted circular means;
           these can be obtained via $cirmeans
```

Thus the constrained estimates satisfy the required order as follows:

$$\left\{ \begin{array}{l} \widetilde{\phi}_1 = 0.993 \\ \widetilde{\phi}_2 = 1.475 \\ \widetilde{\phi}_3 = 3.066 \end{array} \right\} \preceq \left\{ \begin{array}{l} \widetilde{\phi}_4 = 5.056 \\ \widetilde{\phi}_5 = 3.066 \end{array} \right\} \preceq \{ \widetilde{\phi}_6 = 5.056 \} \preceq \left\{ \begin{array}{l} \widetilde{\phi}_7 = 5.056 \\ \widetilde{\phi}_8 = 0.993 \end{array} \right\}$$

where $\widetilde{\phi}$ is the Circular Isotonic Regression Estimator of ϕ. Results may be displayed graphically by setting `plot(example1CIRE)`. When done so, a plot with the points of the CIRE is produced. To see the plot for the unrestricted estimates the argument `option = "cirmeans"` can be used (i.e., `plot(example1CIRE, option = "cirmeans")`).

Example 2 (κ unknown (replications needed))

Using the data in our package called `datareplic` we demonstrate the use of the function `cond.test` when κ is unknown. As remarked earlier, when κ is unknown we need replicate data to estimate κ. The file `datareplic` is a matrix where each column contains the values of a replication and each row the angles observed at each population mean. In this example we have 8 populations and hypotheses regarding the corresponding 8 parameters are as follows:

$$H_0 : \quad \phi_1 \preceq \phi_2 \preceq \phi_3 \preceq \phi_4 \preceq \phi_5 \preceq \phi_6 \preceq \phi_7 \preceq \phi_8 \preceq \phi_1$$
$$H_1 : \quad H_0 \text{ is not true.}$$

We take the data from the package and set the groups of the order in the argument `groups`.

```
R> data("datareplic")
R> orderGroups2 <- c(1:8)
```

Since replicate data are available, we do not include the argument `kappa` as we want the function to estimate it. Moreover, we correct the bias in the estimation of κ, so we set `biasCorrect = TRUE`. Thus we have the following code:

```
R> example2test <- cond.test(datareplic, groups = orderGroups2,
+                            biasCorrect = TRUE)
R> example2test


 Circular Isotonic Regression Estimator (CIRE):
    1.223
    1.223
    1.223
    3.130
    4.194
    4.194
    5.541
    1.223
 Sum of Circular Errors: SCE =  1.532
Invisible: Unrestricted circular means;
           these can be obtained via $cirmeans
 pvalue =  0.0034
 kappa =  3.72
Kappa has been estimated
```

The result is the p-value defined in (5). Since the `p-value = 0.0034` we reject the null hypothesis and conclude that the parameters do not satisfy the specified circular order.

If the user is interested in printing the unrestricted circular means $\overline{\theta}$ then the following command is used: `example2test$cirmeans`. The result is a list that is saved in the same format as CIRE. Since each group in the circular order has a single element, it is convenient to use the vector format. Hence we have:

```
R> round(unlist(example2test$cirmeans), digits = 3)


[1] 0.753 1.764 6.173 3.131 4.469 3.920 5.542 2.367
```

4. Application to analysis of the cell cycle gene expression data

As noted in the Introduction there has been considerable discussion in the literature on the conservation of various aspects of cell cycle genes (Fernández *et al.* 2012), particularly between two yeast species, namely, *S. Cerevisiae* (budding yeast) and *S. Pombe* (fission yeast). Using the 10 published budding yeast data sets (Rustici *et al.* 2004; Oliva *et al.* 2005; Peng *et al.* 2005), we illustrate the **isocir** package to test the null hypothesis that 16 fission yeast genes, namely, *ssb1, cdc22, msh6, psm3, rad21, cig2, mik1, h3.3, hhf1, hht3, hta2, htb1, fkh2, chs2, sid2* and *slp1* satisfy the same circular

order as their budding yeast orthologs (RFA1, RNR1, MSH6, SMC3, MCD1, CLN2, SWE1, HHT2, HHF1, HHT1, HTA2, HTB2, FKH1, CHS2, DBF2 and CDC20) whose circular order is obtained from Cyclebase (`http://www.cyclebase.org`) and published literature. Thus we test the following hypothesis:

$$H_0 : \phi_{ssb1} \preceq \phi_{cdc22} \preceq \phi_{msh6} \preceq \phi_{psm3} \preceq \phi_{rad21} \preceq \phi_{cig2} \preceq \phi_{mik1} \preceq \phi_{h3.3} \preceq$$
$$\preceq \phi_{hhf1} \preceq \phi_{hht3} \preceq \phi_{hta2} \preceq \phi_{htb1} \preceq \phi_{fkh2} \preceq \phi_{chs2} \preceq \phi_{sid2} \preceq \phi_{slp1} \preceq \phi_{ssb1} \qquad (6)$$

$H_1 : H_0$ is not true.

For each of the 10 experimental data sets, the unconstrained estimates of the phase angles of the above 16 fission yeast genes appearing in Table 5 were obtained using the Random Periods Model (Liu *et al.* 2004). The R code for that software can be obtained from `http://www.niehs.nih.gov/research/atniehs/labs/bb/staff/peddada/index.cfm`.

Notice that there are no replicated data here, since the experiments were not performed under the same experimental conditions. It appears that the 10 experiments were not synchronized (i.e., cells were probably not arrested at the same point in the cell cycle). For this reason, from Table 5 it appears that there is a large variability in the estimates of phase angles of each of the 16 genes. Even though there may be large variability in the estimated values, our interest is in the relative order of phase angles among the 16 genes which does not rely on the location of the pole and hence does not rely on the synchronization. As there are no replicated data, we have a single observation for each of the 16 fission yeast genes in each experiment and, therefore, the values in Table 5 play the role of the unrestricted circular mean in each experiment. Consequently we suppose that,

$$\overline{\theta}_{ij} \sim^{ind} M(\phi_{ij}, \kappa_j), i = 1, 2, \ldots, 16, j = 1, 2, \ldots, 10,$$

where $\overline{\theta}_{ij}$ is the unrestricted circular mean of the gene i in the experiment j.

Since the 10 experiments may not considered as replications of each other, we performed a separate test for each experiment. Moreover, as explained in Fernández *et al.* (2012) we assume that the concentration parameter κ_j depends on the experiment but not on the gene. The reason for this is that out of the two sources of uncertainty, one specific to the gene and another one due to the experiment (and therefore common to all genes within the experiment), the former source maybe considered negligible relative to the latter as the number of time points used in each time course experiment is fairly large for any specific gene.

The κ_j values considered for this example are obtained using Fernández *et al.* (2012). The procedure used for the computation of these values comes from an analysis of variance type methodology. Under the assumptions made before, the model for the circular means is:

$$\phi_{ij} = \mu + \alpha_i + \beta_j,$$

where α_i is the gene effect and β_j is the experiment effect. The proposed model is analogous to the standard two-way analysis of variance model and is fully detailed in the supplementary material of Fernández *et al.* (2012).

For each of the 10 experiments, we test the hypothesis (6) using the function `cond.test` that is in our software. The following code gives the p-values for each experiment. Results are summarized in Table 5.

```
R> data("cirgenes")
R> kappas <- c(2.64773, 3.24742, 2.15936, 4.15314, 4.54357,
+              29.07610, 6.51408, 14.19445, 5.66920, 11.12889)
R> allresults <- list()
R> resultIsoCIRE <- matrix(ncol = ncol(cirgenes),
+                          nrow = nrow(cirgenes))

R> SCEs <- vector(mode = "numeric", length = nrow(cirgenes))
R> pvalues <- vector(mode = "numeric", length = nrow(cirgenes))

R> for (i in 1 : nrow(cirgenes)) {
+    k <- kappas[i]
+    genes <- as.numeric(cirgenes[i, !is.na(cirgenes[i, ])])
+    allresults[[i]] <- cond.test(genes, kappa = k)
+    resultIsoCIRE[i, !is.na(cirgenes[i, ])] <-
+        unlist(allresults[[i]]$CIRE)
+    SCEs[i] <- allresults[[i]]$SCE
+    pvalues[i] <- allresults[[i]]$pvalue
+ }
```

From the p-values in Table 5, we see that the null hypothesis cannot be rejected in any of the 10 experiments even at a level of significance as high as 0.20. Therefore, it seems plausible that the peak expressions of these 16 genes in *S. Pombe* (fission yeast) follow the same order as their *S. Cerevisiae* (budding yeast) orthologs.

5. Conclusions

In this paper the R package **isocir** has been presented. This package provides useful tools for drawing inferences from circular data under order restrictions. There are two main functions (`CIRE` and `cond.test`). The first one computes the CIRE, the circular version of the widely known isotonic regression in R^q. The second one is designed for testing circular hypotheses using a conditional test. We have also created the class `isocir` in order to properly save all the results. Although we illustrated the proposed methodology using an example from cell biology, the proposed software can be applied to a wide range of contexts. For example, biologists working on circadian clocks may be interested in the testing for the conservation of circular order among circadian genes between two tissues (e.g., Liu *et al.* 2006).

We would like to emphasize that the field of constrained inference on a unit circle is in its infancy and is wide open for new developments both in methods as well as applications. As observed in the introduction, such constrained inference problems arise naturally in many applications. Therefore we expect the software described in this paper to be widely used by researchers working in such areas.

Acknowledgments

References

Agostinelli C, Lund U (2011). *circular: Circular Statistics.* CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA. R package version 0.4-3, URL https://r-forge.r-project.org/projects/circular/.

Balabdaoui F, Rufibach K, Santambrogio F (2009). *OrdMonReg: Compute Least Squares Estimates of One Bounded or Two Ordered Isotonic Regression Curves.* R package version 1.0.2, URL http://CRAN.R-project.org/package=OrdMonReg.

Boles L, Lohmann K (2003). "True Navigation and Magnetic Maps in Spiny Lobsters." *Nature*, **421**, 60–63.

Bowers J, Morton I, Mould G (2000). "Directional Statistics of the Wind and Waves." *Applied Ocean Research*, **22**, 13–30.

Cermakian N, Lamont E, Bourdeau P, Boivin D (2011). "Circadian Clock Gene Expression in Brain Regions of Alzheimer's Disease Patients and Control Subjects." *Journal of Biological Rhythms*, **26**, 160–170.

Chasalow S (2010). *combinat: Combinatorics Utilities.* R package version 0.0-8, URL http://CRAN.R-project.org/package=combinat.

Cochran W, Mouritsen H, Wikelski M (2004). "Migrating Songbirds Recalibrate Their Magnetic Compass Daily from Twilight Cues." *Science*, **304**, 405–408.

de Leeuw J, Hornik K, Mair P (2011). *isotone: Active Set and Generalized PAVA for Isotone Optimization.* R package version 1.0-1, URL http://CRAN.R-project.org/package=isotone.

Fernández M, Rueda C, Peddada S (2012). "Identification of a Core Set of Signature Cell Cycle Genes Whose Relative Order of Time to Peak Expression is Conserved Across Species." *Nucleic Acids Research*, **40**(7), 2823–2832. URL http://nar.oxfordjournals.org/content/40/7/2823.

Jammalamadaka S, SenGupta A (2001). *Topics in Circular Statistics*. World Scientific.

Jensen JL Jensen T, Lichtenberg U, Brunak S, Bork P (2006). "Co-Evolution of Transcriptional and Post-Translational Cell-Cycle Regulation." *Nature*, **443**, 594–597.

Kibiak T, Jonas C (2007). "Applying Circular Statistics to the Analysis of Monitoring Data." *European Journal of Psychological Assessment*, **23**, 227–237.

Liu D, Peddada S, Li L, Weinberg C (2006). "Phase Analysis of Circadian-Related Genes in Two Tissues." *BMC Bioinformatics*, **7**, 87.

Liu D, Umbach D, Peddada S, Li L, Crockett P, Weinberg C (2004). "A Random Periods Model for Expression of Cell-Cycle Genes." *PNAS*, **101**(19), 7240–7245.

Lund U, Agostinelli C (2009). **CircStats**: *Circular Statistics, from "Topics in Circular Statistics" (2001)*. R package version 0.2-4, URL http://CRAN.R-project.org/package=CircStats.

Mardia K, Hughes G, Taylor C, Singh H (2008). "A Multivariate von Mises Distribution with Applications to Bioinformatics." *Canadian Journal of Statistics*, **36**, 99–109.

Mardia K, Jupp P (2000). *Directional Statistics*. John Wiley & Sons.

Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J (2005). "The Cell-Cycle-Regulated Genes of Schizosaccharomyces Pombe." *Plos. Biology*, **3**, 1239–1260.

Peng X, Karuturi R, Miller L, Lin K, Jia Y, Kondu P, Wang L, Wong L, Liu E, Balasubramanian M, Liu J (2005). "Identification of Cell Cycle-Regulated Genes in Fission Yeast." *The American Society for Cell Biology*, **16**, 1026–1042.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Robertson T, Wright F, Dykstra R (1988). *Order Restricted Statitical Inference*. John Wiley & Sons.

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." *Journal of the American Statistical Association*, **104**(485), 338–347.

Rustici G, Mata J, Kivinen K, Lio P, Penkett C, Burns G, Hayles J, Brazma A, Nurse P, Bahler J (2004). "Periodic Gene Expression Program of the Fission Yeast Cell Cycle." *Nature Genetics*, **36**, 809–817.

Turner R (2009). **Iso**: *Functions to Perform Isotonic Regression*. R package version 0.0-8, URL http://CRAN.R-project.org/package=Iso.

van Doorn E, Dhruva B, Sreenivasan K, Cassella V (2000). "Statistics of Wind Direction and Its Increments." *Physics of Fluids*, **12**, 1529–1534.

Zar J (1999). *Biostatistical Analysis*. Prentice Hall.

Unrestricted circular means

Experiments \ Genes	ssb1 $\bar\theta_1$	cdc22 $\bar\theta_2$	msh6 $\bar\theta_3$	psm3 $\bar\theta_4$	rad21 $\bar\theta_5$	cig2 $\bar\theta_6$	mik1 $\bar\theta_7$	h3.3 $\bar\theta_8$	hhf1 $\bar\theta_9$	hht3 $\bar\theta_{10}$	hta2 $\bar\theta_{11}$	htb1 $\bar\theta_{12}$	fkh2 $\bar\theta_{13}$	chs2 $\bar\theta_{14}$	sid2 $\bar\theta_{15}$	slp1 $\bar\theta_{16}$
1.Oliva cdc	0.202	0.218	6.262	5.765	0.893	5.612	6.257	1.178	0.912	1.200	0.971	1.289	5.298	5.597	4.252	5.209
2.Oliva elut1	2.940	3.262	2.811	2.848	1.603	2.382	1.709	4.689	4.355	4.717	4.419	4.397	1.600	1.819	1.751	2.519
3.Oliva elut2	0.440	0.447	5.258	6.206	4.381	5.458	6.045	1.541	0.727	6.115	0.352	0.687	3.935	3.970	5.836	5.896
4.Peng cdc	3.328	3.565	3.387	2.806	3.194	3.260	3.026	4.779	4.694	4.755	4.816	4.675	2.685	2.770	2.885	2.422
5.Peng elut	3.333	3.912	3.894	3.444	3.648	3.970	4.296	5.189	5.060	5.144	5.216	5.243	3.338	3.607	3.082	3.185
6.Rust cdc1	1.965	2.151	2.034	2.029	1.742	2.073	1.730	3.130	2.993	3.085	3.063	2.873	1.281	1.179	1.906	1.237
7.Rust cdc2	1.809	2.208	1.415	1.351	1.964	1.940	1.978	3.745	3.584	3.670	3.479	3.590	1.383	1.456	1.064	1.396
8.Rust elut1	—	1.457	1.288	—	1.530	1.374	1.379	2.420	2.279	2.409	2.311	2.245	1.010	1.146	—	1.091
9.Rust elut2	2.214	1.786	1.730	1.987	1.878	1.882	3.071	2.704	2.788	2.814	2.909	2.740	1.352	1.441	1.275	1.420
10.Rust elut3	2.340	2.702	2.704	2.526	2.979	2.319	2.284	3.773	3.567	3.636	3.465	3.431	1.981	1.716	2.523	2.118

Table 4: Initial S. Pombe phase angle data for each experiment.

Experiments \ Genes	CIRE under circular order																SCE	p-value
	$\tilde{\phi}_1$	$\tilde{\phi}_2$	$\tilde{\phi}_3$	$\tilde{\phi}_4$	$\tilde{\phi}_5$	$\tilde{\phi}_6$	$\tilde{\phi}_7$	$\tilde{\phi}_8$	$\tilde{\phi}_9$	$\tilde{\phi}_{10}$	$\tilde{\phi}_{11}$	$\tilde{\phi}_{12}$	$\tilde{\phi}_{13}$	$\tilde{\phi}_{14}$	$\tilde{\phi}_{15}$	$\tilde{\phi}_{16}$		
1.Oliva cdc	6.257	6.257	6.257	6.257	0.054	0.054	0.054	1.045	1.045	1.085	1.085	1.289	5.069	5.069	5.069	5.209	1.270	0.6658
2.Oliva elut1	2.526	2.526	2.526	2.526	2.526	2.526	2.526	4.515	4.515	4.515	4.515	4.515	1.600	1.785	1.785	2.519	1.218	0.7214
3.Oliva elut2	5.849	5.849	5.849	5.849	5.849	5.849	6.045	0.598	0.598	0.598	0.598	0.687	3.935	3.970	5.836	5.849	2.660	0.2437
4.Peng cdc	3.225	3.225	3.225	3.225	3.225	3.225	3.225	4.736	4.736	4.749	4.749	4.749	2.685	2.693	2.693	2.693	0.248	0.9983
5.Peng elut	3.333	3.725	3.725	3.725	3.725	3.970	4.296	5.124	5.124	5.144	5.216	5.243	3.302	3.302	3.302	3.302	0.156	0.9850
6.Rust cdc1	1.961	1.961	1.961	1.961	1.961	1.961	1.961	3.029	3.029	3.029	3.029	3.029	1.230	1.230	1.571	1.571	0.213	0.4142
7.Rust cdc2	1.693	1.693	1.693	1.693	1.952	1.952	1.978	3.614	3.614	3.614	3.614	3.614	1.301	1.301	1.301	1.396	0.296	0.9536
8.Rust elut1	—	1.373	1.373	—	1.427	1.427	1.427	2.333	2.333	2.333	2.333	2.333	1.010	1.118	—	1.118	0.028	0.9992
9.Rust elut2	1.909	1.909	1.909	1.916	1.916	1.916	2.837	2.837	2.837	2.837	2.837	2.837	1.352	1.358	1.358	1.420	0.125	0.9992
10.Rust elut3	2.340	2.585	2.585	2.585	2.585	2.585	2.585	3.574	3.574	3.574	3.574	3.574	1.849	1.849	2.321	2.321	0.269	0.8748

Table 5: CIRE, SCE and p-values for each experiment.

Affiliation:

Sandra Barragán, Miguel A. Fernández, Cristina Rueda
Departamento de Estadística e Investigación Operativa
Instituto de Matemáticas (IMUVA)
Universidad de Valladolid
Valladolid, Spain
E-mail: sandraba@eio.uva.es,
miguelaf@eio.uva.es,
crueda@eio.uva.es
URL: http://www.eio.uva.es/~sandra/
http://www.eio.uva.es/~miguel/
http://www.eio.uva.es/~cristina/

Shyamal D. Peddada
Biostatistics Branch
National Institute of Environmental Health Sciences
Research Triangle Park
NC 27709, USA
E-mail: peddada@niehs.nih.gov,
URL: http://www.niehs.nih.gov/research/atniehs/labs/bb/staff/peddada/index.cfm