

Paquete de R: isocir

Inferencia ISOtónica con datos CIRculares resolviendo problemas de la Biología Molecular.

Sandra Barragán Andrés(1), Cristina Rueda(1), Miguel A. Fernández (1) and Shyamal D. Peddada (2)

(1): Departamento de Estadística e Investigación Operativa
Universidad de Valladolid

(2): National Institute of Environmental Health Sciences
(USA)

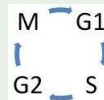
Contenidos

- 1 Motivación del Problema
 - Biología Molecular y Ciclo Celular
- 2 Metodología con Ejemplo
 - Planteamiento general del problema
 - Restricciones de ORDEN
 - Problema de ESTIMACIÓN
 - Problema de CONTRASTE
- 3 Paquete de R: isocir
 - Paquete de R: isocir
 - Funciones del paquete isocir
- 4 Conclusiones
 - Otras aplicaciones
 - Conclusión

Análisis de Datos Circulares bajo Restricciones

La motivación del problema

El análisis de expresiones de genes a lo largo del ciclo celular



Inferencia Con Restricciones

ORDEN: Información Adicional.

Métodos Circulares

DATOS: Puntos en el Círculo.

Análisis de Datos Circulares bajo Restricciones

↓
Paquete de  R: **isocir**

Biología Molecular: La expresión de un gen

Ingredientes:

Biología Molecular: La expresión de un gen

Ingredientes:

Resultados:

Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

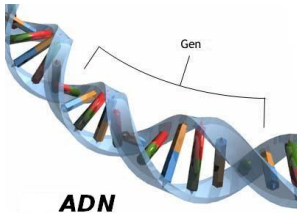


Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

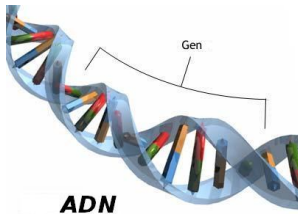


Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

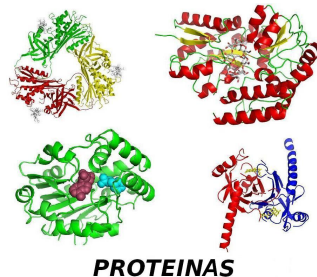
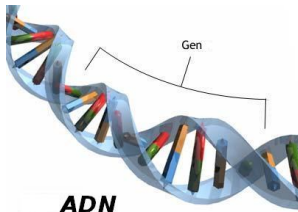


Biología Molecular: La expresión de un gen

Ingredientes:



Resultados:

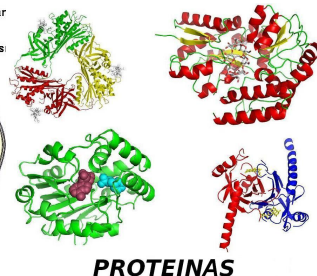
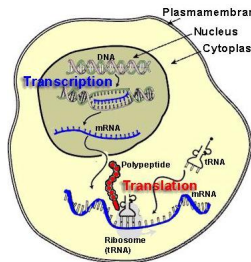
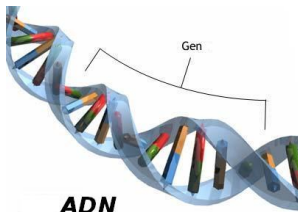


Biología Molecular: La expresión de un gen

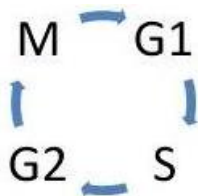
Ingredientes:



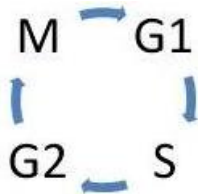
Resultados:



Ciclo Celular



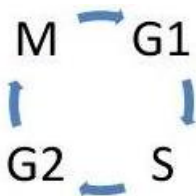
Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

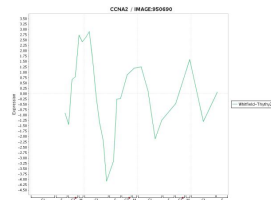
El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.

Ciclo Celular

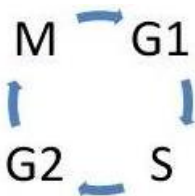


Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.

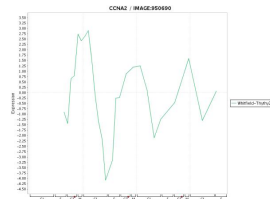


Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

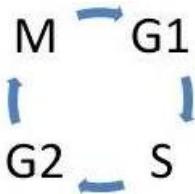
El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



PROBLEMAS BIOLÓGICOS

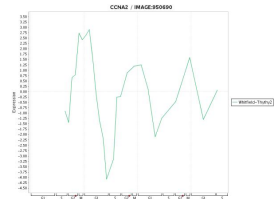
PROBLEMAS ESTADÍSTICOS

Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.

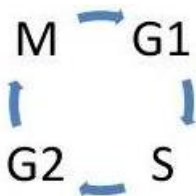


PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**

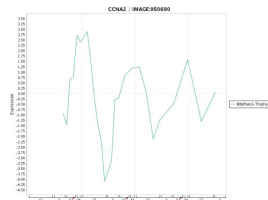
PROBLEMAS ESTADÍSTICOS

Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



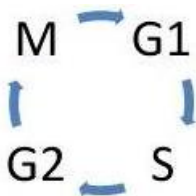
PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**

PROBLEMAS ESTADÍSTICOS

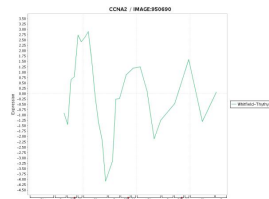


Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



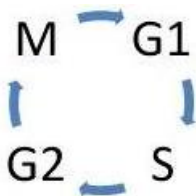
PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**

PROBLEMAS ESTADÍSTICOS

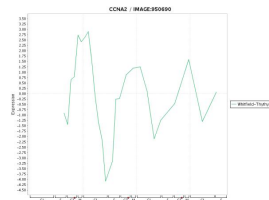
Estimación

Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



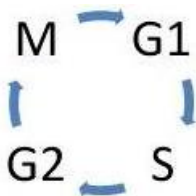
PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**
- Para un conjunto dado de genes, comprobar si **el orden relativo de máximas expresiones se mantiene entre diferentes especies.**

PROBLEMAS ESTADÍSTICOS

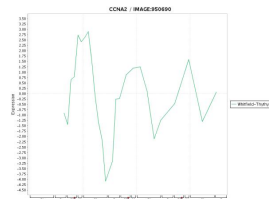
Estimación

Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



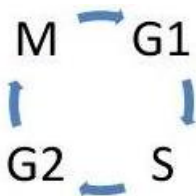
PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**
- **Para un conjunto dado de genes, comprobar si el orden relativo de máximas expresiones se mantiene entre diferentes especies.**

PROBLEMAS ESTADÍSTICOS

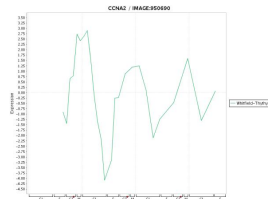
Estimación

Ciclo Celular



Un gen es llamado **gen del ciclo celular** si se expresa cíclicamente a lo largo del ciclo celular.

El momento en el que se expresa con mayor intensidad es llamado **la máxima expresión**.



PROBLEMAS BIOLÓGICOS

- **Determinar la fase del ciclo celular en la que ocurre la máxima expresión de un gen.**
- **Para un conjunto dado de genes, comprobar si el orden relativo de máximas expresiones se mantiene entre diferentes especies.**

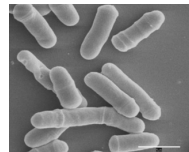
PROBLEMAS ESTADÍSTICOS

Estimación

Contraste

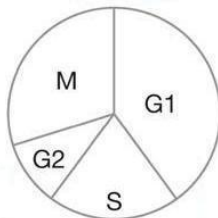
Ejemplo: Planteamiento

Dos especies de levaduras:

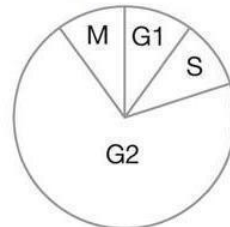
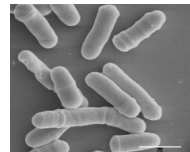


Ejemplo: Planteamiento

Dos especies de levaduras:



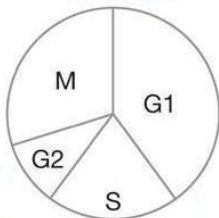
Budding



Fission

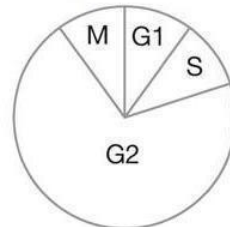
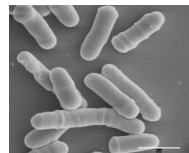
Ejemplo: Planteamiento

Dos especies de levaduras:



Budding

ORDEN



Fission

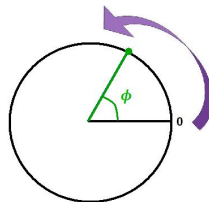
DATOS

Metodología: Notación

Parámetros circulares¹:

$$\phi_1, \dots, \phi_q$$

(ϕ_i : punto que representa en el círculo unidad el momento del ciclo celular en el que ocurre la máxima expresión del gen i).



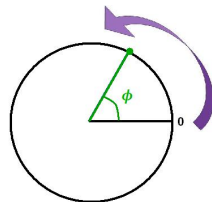
¹MARDIA, K., and JUPP, P. (2000), *Directional Statistics*, Chichester: Wiley.

Metodología: Notación

Parámetros circulares¹:

$$\phi_1, \dots, \phi_q$$

(ϕ_i : punto que representa en el círculo unidad el momento del ciclo celular en el que ocurre la máxima expresión del gen i).



Sean $\theta_1, \dots, \theta_q$ las direcciones medias muestrales y r_1, \dots, r_q las longitudes medias resultantes.

$$\theta_i \rightsquigarrow VM(\phi_i, \kappa_i)$$

- ϕ_i : dirección media
- κ_i : parámetro de concentración.

¹MARDIA, K., and JUPP, P. (2000), *Directional Statistics*, Chichester: Wiley.

Metodología: Restricciones de Orden en el Espacio Circular

Orden Simple

$$C_{OS} = \{\phi \in [0, 2\pi]^q \mid 0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq 2\pi\}$$

Metodología: Restricciones de Orden en el Espacio Circular

Orden Simple

$$C_{OS} = \{\phi \in [0, 2\pi]^q \mid 0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq 2\pi\}$$

Problema con el orden simple en el espacio circular:

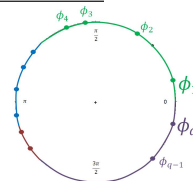
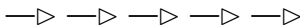
Metodología: Restricciones de Orden en el Espacio Circular

Orden Simple

$$C_{OS} = \{\phi \in [0, 2\pi]^q \mid 0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq 2\pi\}$$

Problema con el orden simple en el espacio circular:

- Sin conexión entre ϕ_1 y ϕ_q



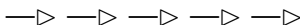
Metodología: Restricciones de Orden en el Espacio Circular

Orden Simple

$$C_{OS} = \{\phi \in [0, 2\pi]^q \mid 0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq 2\pi\}$$

Problema con el orden simple en el espacio circular:

- Sin conexión entre ϕ_1 y ϕ_q



—▷ Solución: **Orden Isotrópico.**

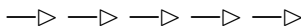
Metodología: Restricciones de Orden en el Espacio Circular

Orden Simple

$$C_{OS} = \{\phi \in [0, 2\pi]^q \mid 0 \leq \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq 2\pi\}$$

Problema con el orden simple en el espacio circular:

- Sin conexión entre ϕ_1 y ϕ_q



→ Solución: **Orden Isotrópico.**

Orden Isotrópico

$$C_{OI} = \{\phi \in [0, 2\pi]^q \mid \phi_1 \leq \phi_2 \leq \dots \leq \phi_q \leq \phi_1\} = \bigcup_{1 \leq l \leq q} C_{SO}^l$$

donde $C_{SO}^l = \{0 \leq \phi_l \leq \phi_{l+1} \leq \dots \leq \phi_q \leq \phi_1 \leq \dots \leq \phi_{l-1} \leq 2\pi\}$

- **No depende del punto inicial del círculo**
- **Es de rotación invariante**



Ejemplo: Aplicación a la Biología Molecular

Orden Isotrópico Parcial (levadura budding)

$$C_{OIP} = \left\{ \phi \in [0, 2\pi]^{16} : \left\{ \begin{array}{c} \phi_{11} \\ \phi_{12} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{21} \\ \phi_{22} \\ \phi_{23} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{array} \right\} \leq \right.$$

$$\left. \leq \left\{ \begin{array}{c} \phi_{41} \\ \phi_{42} \end{array} \right\} \leq \left\{ \phi_{51} \right\} \leq \left\{ \begin{array}{c} \phi_{61} \\ \phi_{62} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{71} \\ \phi_{72} \\ \phi_{73} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{11} \\ \phi_{12} \end{array} \right\} \right\}$$

²PENG X et al.(2005). Identification of Cell Cycle-Regulated Genes in Fission Yeast. *The American Society for Cell Biology*, 16, 1026-1042

Ejemplo: Aplicación a la Biología Molecular

Orden Isotrópico Parcial (levadura budding)

$$C_{OIP} = \left\{ \phi \in [0, 2\pi]^{16} : \left\{ \begin{array}{c} \phi_{11} \\ \phi_{12} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{21} \\ \phi_{22} \\ \phi_{23} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{array} \right\} \leq \right. \\ \left. \leq \left\{ \begin{array}{c} \phi_{41} \\ \phi_{42} \end{array} \right\} \leq \left\{ \phi_{51} \right\} \leq \left\{ \begin{array}{c} \phi_{61} \\ \phi_{62} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{71} \\ \phi_{72} \\ \phi_{73} \end{array} \right\} \leq \left\{ \begin{array}{c} \phi_{11} \\ \phi_{12} \end{array} \right\} \right\}$$

Datos (levadura fission)

Experimentos en el ciclo celular de levaduras **fission**² donde se mide la expresión máxima de los genes.

Matriz de datos: cirgenes (10 experimentos con 16 genes).

²PENG X et al.(2005). Identification of Cell Cycle-Regulated Genes in Fission Yeast. *The American Society for Cell Biology*, 16, 1026-1042

Metodología: Estimador Circular de Regresión Isotónica

El CIRE (del inglés: Circular Isotonic Regression Estimator) es el Estimador Máximo Verosimil Restringido de la dirección media (ϕ) bajo la suposición de un conjunto de orden dado ($\phi \in C$).

CIRE

$$\tilde{\theta} = \arg \min_{\alpha \in C} SCE(\alpha, \theta)$$

donde SCE es la Suma de Errores Circulares definida por

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

³RUEDA, C., FERNANDEZ, M. and 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*, Vol104,n485; pp 338-347.

Metodología: Estimador Circular de Regresión Isotónica

El CIRE (del inglés: Circular Isotonic Regression Estimator) es el Estimador Máximo Verosimil Restringido de la dirección media (ϕ) bajo la suposición de un conjunto de orden dado ($\phi \in C$).

CIRE

$$\tilde{\theta} = \arg \min_{\alpha \in C} SCE(\alpha, \theta)$$

donde SCE es la Suma de Errores Circulares definida por

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

Rueda et al. (2009)³ desarrolló un algoritmo que está implementado en la función **CIREi** del paquete **isocir**.

³RUEDA, C., FERNANDEZ, M. and 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*, Vol104,n485; pp 338-347.

Ejemplo: Aplicación a la Biología Molecular

● ESTIMACIÓN

```
experiment1 <- t(cirgenes[1,])
levels0IP <- c(rep(1,2),rep(2,3),rep(3,3),rep(4,2),
               rep(5,1),rep(6,2),rep(7,3))
```

```
[1] 1 1 2 2 2 3 3 3 4 4 5 6 6 7 7 7
```

```
CIREi(experiment1, levels = levels0IP)
```

```
$cirmmeans
```

```
$SCE [1] 1.059346
```

```
$CIRE
```

$$\left\{ \begin{array}{l} \tilde{\theta}_{11} = 6,255 \\ \tilde{\theta}_{12} = 6,255 \end{array} \right\} \leq \left\{ \begin{array}{l} \tilde{\theta}_{21} = 6,261 \\ \tilde{\theta}_{22} = 6,255 \\ \tilde{\theta}_{23} = 0,0542 \end{array} \right\} \leq \left\{ \begin{array}{l} \tilde{\theta}_{31} = 0,0542 \\ \tilde{\theta}_{32} = 0,0542 \\ \tilde{\theta}_{33} = 1,045 \end{array} \right\} \leq$$

$$\leq \left\{ \begin{array}{l} \tilde{\theta}_{41} = 1,045 \\ \tilde{\theta}_{42} = 1,085 \end{array} \right\} \leq \left\{ \tilde{\theta}_{51} = 1,085 \right\} \leq \left\{ \begin{array}{l} \tilde{\theta}_{61} = 1,288 \\ \tilde{\theta}_{62} = 4,774 \end{array} \right\} \leq \left\{ \begin{array}{l} \tilde{\theta}_{71} = 5,596 \\ \tilde{\theta}_{72} = 4,774 \\ \tilde{\theta}_{73} = 5,209 \end{array} \right\}$$

Metodología: Contraste de Hipótesis

Fernández et al. (2011)⁴ presenta un **test condicional** para desarrollar algunos **test restringidos con hipótesis de orden**.

H_0 : $\phi_i, i = 1, \dots, q$, sigue un orden isotrópico.

H_1 : H_0 no se cumple.

⁴FERNANDEZ, M., RUEDA, C. and PEDDADA, S. (2011), A core set of signature cell cycle genes with relative order of time to peak expression conserved across species, *Accepted for publication in Nucleic Acids Research*.

Metodología: Contraste de Hipótesis

Fernández et al. (2011)⁴ presenta un **test condicional** para desarrollar algunos **test restringidos con hipótesis de orden**.

H_0 : $\phi_i, i = 1, \dots, q$, sigue un orden isotrópico.

H_1 : H_0 no se cumple.

Estadístico Test:

kappa conocido

$$T = 2\kappa SCE(\theta, \tilde{\theta})$$

kappa desconocido

$$T = \frac{2\hat{\kappa} SCE(\theta, \tilde{\theta})}{q}$$

CT: H_0 se rechaza siempre que $T \geq c(m)$ (m : n° conj. de nivel de $\tilde{\theta}$).

$$p - \text{valor} = \begin{cases} pr(\chi_{q-m}^2 \geq t^*)[1 - pr_{\phi^0}(C)] & (\kappa \text{ conocido}) \\ pr(F_{q-m, q-1} \geq t^*)[1 - pr_{\phi^0}(C)] & (\kappa \text{ desconocido}) \end{cases}$$

$pr_{\phi^0}(C)$: probabilidad de C en H_0 bajo la igualdad de los parámetros.

Implementado en la función **CTi** del paquete **isocir**.

⁴FERNANDEZ, M., RUEDA, C. and PEDDADA, S. (2011), A core set of signature cell cycle genes with relative order of time to peak expression conserved across species, *Accepted for publication in Nucleic Acids Research*.

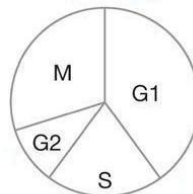
Ejemplo: Aplicación a la Biología Molecular

● CONTRASTE

$H_0 : \phi \in C_{OIP}.$

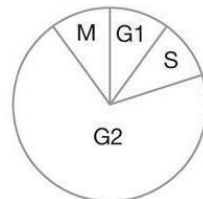
$H_1 : H_0$ no se cumple.

orden: C_{OIP}



Budding

orden: ?



Fission

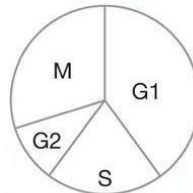
Ejemplo: Aplicación a la Biología Molecular

● CONTRASTE

$H_0 : \phi \in C_{OIP}$.

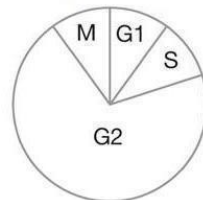
$H_1 : H_0$ no se cumple.

orden: C_{OIP}



Budding

orden: ?



Fission

```
CTi(experiment1, levels = levelsOIP, kappa = 3.958 )
```

```
$pvalue
```

```
[1] 0.2998076
```


Paquete de R: isocir




Paquete de ⁵ : **isocir**

Análisis de Datos Circulares bajo Restricciones
Inferencia **isotónica** con datos **circulares**.



 package **isocir**

- Dependencias: `circular` y `combinat`. Estos paquetes deberán estar instalados en nuestro ordenador antes de cargar `isocir`.

⁵R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 

Paquete de R: isocir en el CRAN (Versión: 1.0)



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[rr](#)

[irtProb](#)

[irtoyo](#)

[isa2](#)

[isdals](#)

[ismev](#)

[isocir](#)

[isopam](#)

[isopat](#)

[isotone](#)

[isva](#)

[iteRates](#)

[iterLap](#)

[iterators](#)

[itertools](#)

[its](#)

Various Coefficients of Interrater Reliability and Agreement

Utilities and Probability Distributions Related to Multidimensional Person Item Response Models (IRT)

Simple interface to the estimation and plotting of IRT models

The Iterative Signature Algorithm

Provides datasets for Introduction to Statistical Data Analysis for the Life Sciences

An Introduction to Statistical Modeling of Extreme Values

Isotonic Inference for Circular data

Isopam (Clustering)

Calculation of isotopic pattern for a given molecular formula

Active set and generalized PAVA for isotone optimization

Independent Surrogate Variable Analysis

Parametric rate comparison

Approximate probability densities by iterated Laplace Approximations

Iterator construct for R

Iterator Tools

Irregular Time Series

Paquete de R: isocir (Versión: 1.0)

Resumen de los componentes de *isocir*

| Funciones | Argumentos |
|--------------|---|
| cirmean | (data) |
| cirSCE | (point1, point2, mrl) |
| mrl | (data) |
| cirkappa | (data) |
| CIREi | (data, levels, isotropic, graphic, stack) |
| CTi | (data, levels, kappa) |

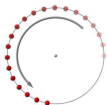
| Datos | Descripción |
|------------|---|
| datareplic | Datos circulares aleatorios con réplicas |
| cirdata | Datos circulares aleatorios |
| cirgenes | Datos de genes de experimentos en el ciclo celular. |

Otros campos de aplicación de **isocir**

- **Ornitología**
- **Endocrinología**
- **Ciencias de la Tierra**
- **Meteorología**
- **Físicas**
- **Psicología**
- **Medicina, Criminología, Análisis de imágenes y mucho más.**



Conclusión



Datos Circulares

+






Información Adicional (Restricciones de Orden)

=

isocir

| Estimación | Contraste |
|------------|-----------|
| CIREi | CTi |

Referencias Básicas

-  MARDIA, K., and JUPP, P. (2000), *Directional Statistics*, Chichester: Wiley.
-  ROBERTSON, T., WRIGHT, F.T. and DYKSTRA, R.L. (1988), *Order Restricted Statistical Inference*, Wiley, New York.
-  RUEDA, C., FERNANDEZ, M. and SHYAMAL, P. (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*, Vol104,n485; pp 338-347.
-  FERNANDEZ, M., RUEDA, C. and SHYAMAL, P. (2011), A core set of signature cell cycle genes with relative order of time to peak expression conserved across species, *Accepted for publication in Nucleic Acids Research*.
-  BARRAGAN, SANDRA (2011) isocir: Isotonic Inference for Circular data. R package version 1.0.
<http://CRAN.R-project.org/package=isocir>.

GRACIAS

por su atención



www.eio.uva.es/~sandra
sandraba@eio.uva.es