

Estadística circular aplicada a la Ecología

Irene Mendoza¹

(1) Departamento de Ecología Integrativa. Estación Biológica de Doñana (CSIC).

Autor para correspondencia: Irene Mendoza [irene.mendoza@ebd.csic.es]

Palabras clave

fenología; actividad diaria; test de Rayleigh; test de Mardia-Watson-Wheeler

Keywords

phenology; diel activity; Rayleigh test; ~~Mardia-Watson-Wheeler test~~ ~~test de Mardia-Watson-Wheeler~~

La estadística circular es una poderosa herramienta que permite analizar de forma apropiada variables que tienen una naturaleza cíclica y para los que la estadística lineal no ~~es correcta~~ ~~resulta apropiada~~. Algunos ejemplos clásicos en Ecología incluyen el estudio de ~~diferentes momentos de la~~ ~~fenología de~~ la floración o fructificación de una especie a lo largo del año, los patrones de actividad diarios de un mamífero, o las direcciones de vuelo de las mariposas. En estos casos, es importante destacar que la designación de un cero es completamente arbitraria y, por consiguiente, también la de valores mayores o menores. Por ejemplo, la diferencia entre el mes de enero y diciembre es de solo un mes en una escala circular, mientras que en una escala lineal, dicha diferencia sería de 12 meses.

Transformación de nuestra variable de naturaleza circular en ángulos

La primera medida que debemos hacer con nuestros datos de naturaleza circular es convertirlos en ángulos. Esto es fácilmente realizable con una simple ecuación (Zar, 1999):

$$a = \frac{(360^\circ)(X)}{k}$$

En la que X representa la variable temporal que queremos convertir en una ~~dirección angular~~ (en grados) y k es la cantidad total de unidades de tiempo de nuestra circunferencia ($k = 365$ para días del año, o $k = 12$ para meses, por ejemplo). De esta manera, el 14 de febrero, que ~~es el día 45º del año~~, ~~corresponde a~~

```
a <- 360*45/365
```

44.38 grados. En el caso de medidas mensuales, los 360° se dividen en 12 sectores de 30° . Por convención, se considera el punto medio del sector como el correspondiente a cada mes, es decir 15° para enero, 45° para febrero, ... y 345° para diciembre.

Comentario [mcr1]: A esto luego lo llamas "valor angular". Tal vez estaría bien decir aquí que son sinónimos o usar solo uno de los dos terminus en el manuscrito.

Comentario [mcr2]: O bien "es el día 45 del año" o "es el 45º día del año"

En todo caso, es curioso que la $^\circ$ puede confundirse con "grados". Tienes incluso otro "45º" escrito cuatro líneas más abajo.

Ejemplos usados y esquema de la nota

En esta nota, vamos a usar dos ejemplos muy sencillos, uno referido a fenología de la fructificación en Nouragues (Guyana francesa) extraído de Mendoza *et al.* (2018) y el otro a los patrones de actividad del agutí y del pecarí de la Isla de Barro Colorado (Rowcliffe *et al.*, 2014), extraídos del paquete `activity` (Rowcliffe, 2019). Estos ejemplos permitirán mostrar cómo hacer representaciones gráficas circulares, calcular métricas circulares básicas y hacer inferencia sobre hipótesis de una y dos muestras. Utilizaremos R (R Core Team, 2020) para todos nuestros análisis, en concreto **los tres de los muchos** paquetes de R que permiten analizar los datos circulares: `Directional` (Tsagris *et al.*, 2020), `CircStats` (Ulric Lund y Claudio Agostinelli, 2018), y `circular` (Agostinelli y Lund, 2017), así como el paquete `ggplot2` (Wickham, 2016) para las representaciones gráficas. Debido a las limitaciones de espacio de esta nota, se recomienda revisar otras referencias para tener más detalle de los métodos, fórmulas exactas y aplicaciones a la Ecología de la estadística circular (Batschelet, 1981; Mardia y Jupp, 2000; Morellato *et al.*, 2010; Staggemeier *et al.*, 2020).

```
library(activity)
library(circular)
library(CircStats)
library(Directional)
library(ggplot2)
```

```
aguti <- subset(activity::BCItime, species == "agouti")
pecari <- subset(activity::BCItime, species == "peccary")
#se seleccionan Los datos del agutí y pecarí de la base de datos
BCItime del paquete activity. El tiempo de muestreo está expresado
como una proporción relativa a las 24 horas de un día.
```

```
frutofrq <- data.frame(meses = c("ene", "feb", "mar", "abr", "may",
"jun", "jul", "ago", "sep", "oct", "nov", "dic"), meses.ang = seq(15,
350, 30), numsp = c(21, 23, 32, 31, 29, 20, 19, 18, 15, 17, 21, 16))
#estos datos representan el número de especies en fruto por mes en
Nouragues (extraído de Mendoza et al. 2018)
```

Representaciones gráficas de datos circulares

El paquete `circular` (Agostinelli y Lund, 2017) permite representar fácilmente los datos circulares que no están en frecuencias convirtiéndolos previamente en un objeto circular usando la función `circularplot` (Fig. 1).

```
aguti.circ <- circular::circular(aguti$time*2*pi, units = "radians",
template = "clock24")
pecari.circ <- circular::circular(pecari$time*2*pi, units = "radians",
template = "clock24")
#convertimos en radianes Los valores temporales

plot (aguti.circ, main = "aguti") #Fig. 1
plot(pecari.circ, main = "pecari")
```

Comentario [mcr3]: Hay más de tres paquetes. Por ejemplo:
Bpnreg
Circglmmbayes
CircMLE
CircSpaceTime
cplots
isocir
NPCirc
Psych
RhythmicAlly
VecStatGraphs2D
Wrapped

Comentario [mcr4]: Creo que es más habitual poner los comentarios antes del código que describen.

Comentario [mcr5]: Decir que "meses.ang es" la dirección angular correspondiente a cada mes.

Comentario [mcr6]: ¿Cómo tienen que estar entonces? Esta frase es poco clara. Mejor decir cómo tiene que estar que cómo no tienen que estar. O mejor, en la subsección anterior, al describir los dos ejemplos, diciendo que uno es de datos brutos y el otro de datos resumidos en frecuencias.

Comentario [mcr7]: Tal como estaba escrito parece que es la función `plot` la que convierte los datos en objeto circular

Sin embargo, esta representación no es especialmente informativa. Es mucho más útil poder representar un histograma circular, en el que la longitud de cada barra representa la frecuencia de cada **medida para cada ángulo (o unidad temporal) dirección (valor) angular**. Para ello, podemos usar la función `windrose` del mismo paquete, pero la opción `con` mayor variedad de parámetros gráficos **la proporciona la función `coord_polar` sin duda es a través de `ggplot2` usando la función `coord_polar`**. Sin embargo, es preciso resaltar que **estos los diagramas de rosa inducen una interpretación tienden a sesgar al lector de los datos**, sobre todo si el número de barras es bajo o el círculo interior pequeño. Se recomienda entonces **representar usar** la raíz cuadrada de las frecuencias **como radios** (Zar, 1999).

```
fig2 <- ggplot(frutofrq, aes(x = meses.ang, y = sqrt(numsp))) +  
geom_bar(stat='identity') + coord_polar()
```

Principales métricas circulares: media angular, desviación estándar angular y vector r

La estadística lineal no sirve para **describir adecuadamente calcular estadísticos de** datos circulares. Para eso, se han desarrollado métricas circulares como la media, varianza y desviación estándar **circulares**. **Una descripción detallada de las mismas se presenta en Para conocer los detalles de las mismas, se recomienda** Zar (1999). Estas métricas son fácilmente calculables usando la función `circ.summary` del paquete `Directional`.

```
aguti.sum <- Directional::circ.summary(aguti$time*2*pi, plot = F)  
  
aguti.sum$mesos #media angular en radianes  
## [1] 2.985616  
  
aguti.sum$circstd #desviación estándar circular en radianes  
## [1] 0.0185485
```

En el caso de datos **resumidos en forma de incluyendo** frecuencias (como ocurre para los datos de fruto de Nouragues), se deben convertir las frecuencias en un vector único que repite cada valor angular el número de veces de su frecuencia. Una vez conseguido este vector, se podrá aplicar la función `circ.summary` como en el caso anterior.

```
fruto <- c(rep(15,21), rep(45,23), rep(75,32), rep(105, 31), rep(135,  
29), rep(165,20), rep(195,19), rep(225,18), rep(255, 15), rep(285,  
17), rep(315, 21), rep(345, 16))  
  
fruto.sum <- Directional::circ.summary(fruto*pi/180, plot = F, rads =  
T) #convertimos en radianes los valores temporales
```

```
fruto.sum
```

Comentario [mcr8]: Antes los has llamado "histogramas circulares". EN todo caso el nombre correcto en español sería "diagrama de rosa de los vientos"

Comentario [mcr9]: Tal vez podrías mencionar aquí el "rango circular" que mencionas luego en el test de Mardia-Watson-Wheeler

Comentario [mcr10]: Tal vez debes explicar que hay que transformar los ángulos en radianes y que aquí lo haces multiplicando por 2π y en el otro ejemplo por $\pi/180$

Comentario [mcr11]: Si no invocas el objeto creado, el lector no puede ver por ejemplo el MLR que comentas debajo

De estas métricas, *MRL* representa **la longitud media resultante el vector de la longitud de la media o vector r** . Es un valor adimensional, **que varía entre 0 y 1 y mide la dispersión de los valores angulares (0 indica dispersión uniforme en todas las direcciones y 1, todas las observaciones apuntando en la misma dirección)**. Este **valor que** se utiliza en fenología como una estima de la concentración de la actividad o la estacionalidad de la fenofase, siendo mayor cuanto más cercano a 1. Así por ejemplo, el valor de $r_{MLR} = 0.99983$ para la actividad de los agutíes nos indica que está muy concentrada en el tiempo. En el caso de la fructificación en Nouragues, al estar los datos agrupados por meses, debemos multiplicar por un factor corrección $c < -(30 * \pi / 360) / \sin(30/2)$ al valor de r_{MLR} que se deriva de la función `circ.summary` $r_{MLR} = 0.06$. El valor **de r** cercano a 0 para estos datos nos indica que la estacionalidad de la fructificación es muy baja.

Comentario [mcr12]: Esto no se entiende, entre otras cosas porque no has mencionado nada de "vectores" hasta ahora (bueno, has mencionado los vectores de datos de R, lo que hace la frase más confusa).

Podrías explicar que en la literature ecológica se le denomina "r" a este pa'ámetro (si es que eso es así).

Tests de uniformidad de una muestra: test de Rayleigh como medida de estacionalidad

La forma de verificar estadísticamente si una distribución circular es significativamente estacional es usando el *test de Rayleigh*. Este test tiene como hipótesis nula que la distribución circular es uniforme, es decir, que no **tiene hay** una dirección **angular preferida media angular**. Para ello, **es un requisito indispensable que la distribución sea unimodal, aunque si los datos fueran bimodales y estuvieran repartidos a lo largo del círculo, podríamos trabajar con las dos medias circunferencias por separado.**

Comentario [mcr13]: EN vez de decir que es un requisite tal vez podrías decir que el test de Rayleigh asume que la distribución sea unimodal, y que en caso contrario hay otros tests como Kuiper's V, Watsons U2 y Rao's spacing tests. Creo que todos los puedes calcular con el paquete circular

Comentario [mcr14]: ¿"...y las modas estuvieran en mitades opuestas del círculo"???

```
CircStats::r.test(fruto, degree = T)

## $r.bar
## [1] 0.1558043
##
## $p.value
## [1] 0.001729289

CircStats::r.test(aguti$time*2*pi, degree = F)

## $r.bar
## [1] 0.5184795
##
## $p.value
## [1] 0
```

Gracias a este test podemos ver que, a pesar de que el valor de r_{MLR} para Nouragues sea bajo, la fructificación es significativamente estacional, con un pico en marzo-mayo.

Tests entre dos distribuciones circulares: test de Mardia-Watson-Wheeler aplicado a los patrones actividad de los animales

Muchos trabajos (ver Monterroso et al. (2014) como ejemplo) analizan los patrones de actividad diaria de diferentes especies animales usando el test no paramétrico de Mardia-Watson-Wheeler (Batschelet, 1981). Este test permite detectar si dos muestras circulares difieren significativamente entre sí de una forma no paramétrica. Para ello, las dos muestras se mezclan, se distribuyen uniformemente y se calculan sus rangos circulares. Si las dos muestras son similares, la distribución de sus rangos también debería serlo, aceptándose la hipótesis nula.

La forma más sencilla de calcular este test es usando la función `watson.wheeler.test` del paquete `circular`.

```
aguti.circ <- circular::circular(aguti$time*2*pi, units = "radians",
template = "clock24")
pecari.circ <- circular::circular(pecari$time*2*pi, units = "radians",
template = "clock24")

circular::watson.wheeler.test(list(aguti.circ, pecari.circ))

##
## Watson-Wheeler test for homogeneity of angles
##
## data: 1 and 2
## W = 2145, df = 2, p-value < 2.2e-16
```

Usamos el ejemplo de los datos de actividad del agutí y el pecarí. El test nos demuestra que, como se podía apreciar por los gráficos de la Fig. 1, el test nos demuestra que el patrón de actividad de las dos especies es significativamente diferente.

Consideraciones finales

En esta nota apenas hemos pincelado-mostrado algunas aplicaciones básicas de la estadística circular a la Ecología, especialmente para la fenología o el estudio de patrones de actividad diaria de animales, pero obviamente hay muchas más posibilidades. No hemos comentado otros tests importantes ni tampoco las correlaciones circular-linear o circular-circular por falta de espacio, por ejemplo. Gracias a la funcionalidad de los paquetes de *R* especializados en estadística circular, estas métricas y análisis están fácilmente disponibles, con lo que el usuario interesado podrá seguir avanzando en la aplicación de una poderosa herramienta estadística.

El código necesario para reproducir este documento se puede consultar en [GitHub](#).

Comentario [mcr15]: Quizás convendría definir brevemente que quiere decir que las muestras mezcladas se "distribuyan uniformemente". Y haber definido anteriormente en qué consiste el "rango circular".

Comentario [mcr16]: Tal vez sería el sitio para recomendar o comentar el libro de Pewsey et al. 2013 *Circular Statistics in R*. También se podría citar bibliografía en español. Yo no soy experto en el tema pero he visto este par de libros: Lahoz-Beltrá, R., Ortega-Escobar, J. & Fernandez-Montraveta, C. (1994). *Métodos estadísticos en biología del comportamiento*. Madrid. Editorial Complutense. Mena, M. 2003. *Aplicaciones de estadística circular a problemas de ciencias naturales*. Akadia.

Agradecimientos

Esta nota es un producto del proyecto TEMPNET, financiado con una beca Marie-Sklodowska Curie (798269 - TEMPNET - H2020-MSCA-IF-2017/H2020-MSCA-IF-2017).

REFERENCIAS

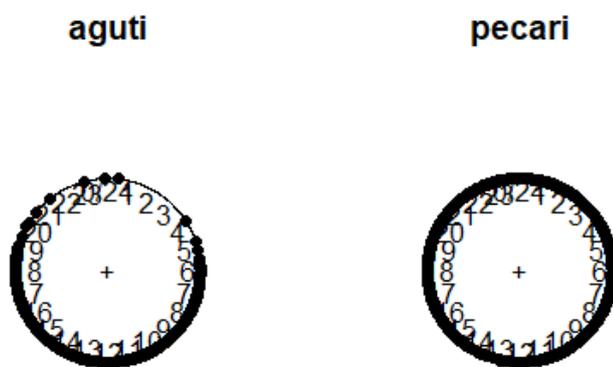
- Agostinelli, C., Lund, U. 2017. *R package circular: Circular Statistics (version 0.4-93)*. CA: Department of Environmental Sciences, Informatics; Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA.
- Batschelet, E. 1981. *Circular Statistics in Biology*. 1.^a ed. Academic Press, London.
- Mardia, K.V., Jupp, P.E. 2000. *Directional Statistics*. 1.^a ed. John Wiley & Sons, Chichester.
- Mendoza, I., Condit, R.S., Wright, S.J., Caubère, A., Châtelet, P., Hardy, I., Forget, P.-M. 2018. Inter-annual variability of fruit timing and quantity at Nouragues (French Guiana): insights from hierarchical Bayesian analyses. *Biotropica* 50: 431-441.
- Monterroso, P., Alves, P.C., Ferreras, P. 2014. Plasticity in circadian activity patterns of mesocarnivores in Southwestern Europe: implications for species coexistence. *Behavioral Ecology and Sociobiology* 68: 1403-1417.
- Morellato, L.P.C., Alberti, L., Hudson, I.L. 2010. Applications of Circular Statistics in Plant Phenology: a Case Studies Approach. En Hudson, I. L., Keatley, M. R. (eds.), *Phenological Research: Methods for Environmental and Climate Change Analysis*, pp. 339-359. Springer Netherlands, Dordrecht.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rowcliffe, J.M., Kays, R., Kranstauber, B., Carbone, C., Jansen, P.A. 2014. Quantifying levels of animal activity using camera trap data. *Methods in Ecology and Evolution* 5: 1170-1179.
- Rowcliffe, M. 2019. *activity: Animal Activity Statistics*.
- Staggemeier, V.G., Camargo, M.G.G., Diniz-Filho, J.A.F., Freckleton, R., Jardim, L., Morellato, L.P.C. 2020. The circular nature of recurrent life cycle events: a test comparing tropical and temperate phenology. *Journal of Ecology* 108:.
- Tsagris, M., Athineou, G., Sajib, A., Amson, E., Waldstein, M.J. 2020. *Directional: Directional Statistics*.
- Ulric Lund, S.-p. original by, Claudio Agostinelli, R. port by. 2018. *CircStats: Circular Statistics, from "Topics in Circular Statistics" (2001)*.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Zar, J.H. 1999. *Biostatistical Analysis*. 4.^a ed. Prentice Hall, New Jersey.

PIES DE FIGURA

Figura 1. Diagrama circular representando la actividad diaria del agutí y el pecarí en la Isla de Barro Colorado. (Datos extraídos de Rowcliffe *et al.* 2014).

Figura 2. Histograma circular representando el número de especies con fruto en cada mes del año en la reserva de Nouragues (Guyana Francesa). Datos extraídos de Mendoza *et al.* 2018. El código para representar esta figura se puede ver en el texto.

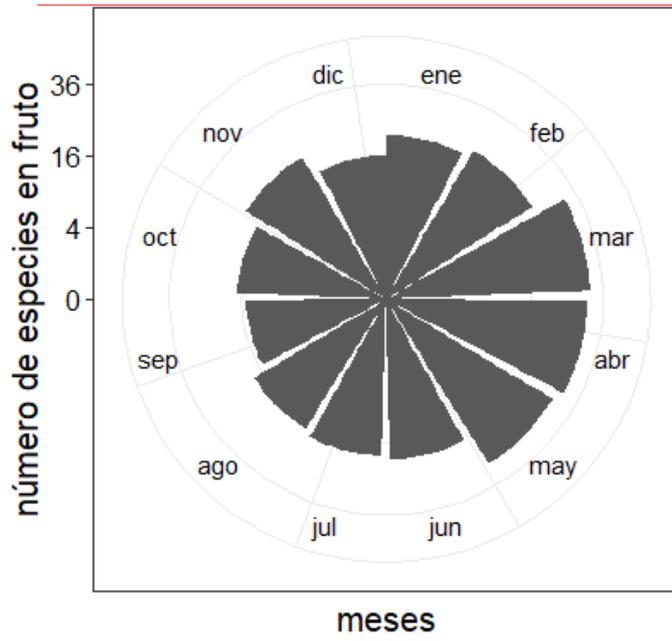
FIGURA 1



Comentario [mcr17]: Deberías conseguir que los numeros no se superpongan (probablemente jugando con `par(mar)` o con `cex`. Además podrías jugar con la transparencia, por ejemplo: `plot(pecari.circ, main = "pecari", col = rgb(0,0,0, alpha = 0.1))`

Figura 1.

FIGURA 2



Comentario [mcr18]: OJO: es el "número de especies en fruto" o la "raíz cuadrada del número de especies en fruto"?

Figura 2.