



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**EL CONCEPTO DE SIMETRÍA EN
ESTADÍSTICA CIRCULAR**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

SARA IVETH TORRES ZAMBRANO



**DIRECTOR DE TESIS:
MAT. MARGARITA ELVIRA CHÁVEZ CANO**

Cd. Universitaria, D. F. 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno
Torres
Zambrano
Sara Iveth
55 26905112
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
303241221
2. Datos del tutor
Mat.
Chávez
Cano
Margarita Elvira
3. Datos del sinodal 1
Dra.
Alonso
Reyes
María del Pilar
4. Datos del sinodal 2
M. en C.
Flores
Díaz
José Antonio
5. Datos del sinodal 3
Act.
Sánchez
Villarreal
Francisco
6. Datos del sinodal 4
M. en C.
Santana
Cobian
David Josafat
7. Datos del trabajo escrito
El concepto de simetría en estadística circular
81 p.
2016

Con todo mi cariño, respeto y gratitud dedico esta tesis a mis padres; los cuales siempre me han brindado todo su amor, apoyo y comprensión en cada paso de mi vida.

Quiero darle primero gracias a Dios, que es el que me ha levantado de todas las caídas que tuve en el proceso; pero sobre todo es el que ha hecho posible que llegue a esta meta, poniendo en mi camino a todas esas personas que directa o indirectamente me han brindado su ayuda todos estos años.

A la Mat. Margarita Chávez de una manera muy especial, ya que con ella aprendí el gusto por la estadística; y es ella quien me ha llevado de la mano en la realización de este proyecto tan importante en mi vida.

A esta gran institución que es la UNAM, que me ha cobijado y me ha dado, a través de muy buenos profesores, una de las más grandes herramientas para el futuro "el conocimiento".

Finalmente quiero dar gracias a dos personas muy importantes en mi vida; a mi prima hermana, que como buena hermana mayor ha sabido apoyarme y aconsejarme; y a mi gran amor Beto, que ha complementado mi vida y me ha hecho crecer como persona.

Índice

<i>Introducción</i>	<i>iii</i>
Métodos descriptivos	1
1.1 Representación gráfica de los datos	2
1.1.1 La gráfica de los datos primarios	2
1.1.2 Histogramas	3
1.1.3 Diagramas de tallo y hojas	6
1.2 Medidas descriptivas	7
1.2.1 Medidas de tendencia central	8
1.2.2 Momentos trigonométricos de la muestra	10
1.2.3 Medidas de dispersión	11
1.2.4 Medidas de asimetría y curtosis	13
El concepto de simetría	15
2.1 Introducción	15
2.1.1 Distintos tipos de simetría	16
2.1.2 Conceptos relacionados	18
2.2 Descripción matemática de la simetría	19
2.2.1 Simetría en la estadística	20
Distribuciones circulares	24
3.1 Introducción	24
3.1.1 Distribuciones envueltas	25
3.2 Momentos trigonométricos	25
3.3 Distribuciones simétricas	29

3.3.1	Distribución uniforme	30
3.3.2	Distribución von Mises	32
3.3.3	La distribución cardioide	35
3.3.4	Distribución normal envuelta	36
3.3.5	Distribución Cauchy envuelta	38
3.4	<i>Distribuciones asimétricas.....</i>	39
3.5	<i>Simetría en la estadística circular</i>	41
Pruebas de simetría para datos circulares		44
4.1	<i>Introducción.....</i>	44
4.2	<i>Análisis exploratorio.....</i>	45
4.3	<i>Pruebas lineales adaptadas para datos circulares.....</i>	46
4.3.1	La prueba de los rangos con signo de Wilcoxon	47
4.3.2	La prueba del signo.....	49
4.4	<i>Prueba ómnibus para simetría circular reflexiva</i>	50
4.5	<i>Dos ejemplos prácticos</i>	53
4.5.1	Truenos.....	53
4.5.2	Tortugas	56
Conclusiones		61
Apéndices		
A1	63
B1	67
B2	68
B3	69
C1	73
Bibliografía		81

Introducción

Los datos circulares, como comúnmente se les conoce, surgen en diferentes áreas como lo son la geología, la paleontología, la geografía, la meteorología, la biología, la medicina, etc. Y se presentan principalmente de dos maneras, como direcciones; en el estudio de temas relacionados con migración, rutas de escape, exploración, entre otros; o en general como cualquier evento que se produce repetidamente dentro de un período de tiempo.

Es posible ver a cada observación circular θ , como un punto sobre la circunferencia de un círculo de radio uno; de tal forma que si un vector unitario es trazado en el plano apuntando a esta observación, θ representará el ángulo formado (en dirección contraria a las manecillas del reloj) por el vector con el eje x . También es posible representar una observación circular mediante coordenadas cartesianas, o bien como coordenadas polares. Dependiendo del estudio que se esté realizando, la medición de los datos circulares puede ser en grados ($0^\circ < \theta \leq 360^\circ$) o en radianes ($0 < \theta \leq 2\pi$).

En la fase exploratoria del análisis estadístico de cualquier conjunto de datos, se consideran algunas interrogantes preliminares con relación a su comportamiento, como por ejemplo si los datos parecen ser uniformes, unimodales o multimodales, y si hay características atípicas, como lo es la presencia de valores extremos. La simetría es una propiedad que si bien no puede describir por completo el comportamiento de los datos, si puede aportar herramientas suficiente para el análisis y más aun para la realización de un modelo matemático que ajuste a los datos que se están analizando.

Retomando que para diversas áreas de la investigación científica es necesaria la aplicación de un análisis estadístico a un conjunto de datos, donde alguna de sus variables posee características circulares; y que la simetría es uno de los principales temas de apoyo en el análisis exploratorio, surge el tema de tesis "El concepto de simetría en estadística circular". Éste está basado en el trabajo hecho por Arthur Pewsey en el artículo que lleva por título "Testing circular symmetry".

Esta tesis se desarrolla en cuatro capítulos, el primero llamado "Métodos descriptivos" aborda el tema de la presentación, descripción, análisis e interpretación de una colección de datos circulares, con el fin de obtener conclusiones sobre los datos, únicamente si se conocen los elementos de la muestra. Así pues, brinda herramientas para tener una idea inicial del comportamiento de los datos y con esto poder proponer un modelo para su distribución, sobre todo para saber si la muestra parece proceder de una distribución simétrica.

"El concepto de simetría" es el segundo capítulo, el cual contiene información de cómo a lo largo de los años la simetría se ha convertido en un concepto integral, por lo que hablar de ella conlleva a varios temas, varios puntos de vista y sobre todo diversas áreas de estudio; ya que surge en el hombre más como una forma de ver la vida, que como un concepto más de apreciación de las cosas. Se puntualiza la importancia que tiene en las matemáticas, y para fines de esta tesis; se define en la estadística circular junto con las consecuencias que tiene aceptar o rechazar el supuesto de simetría sobre la distribución de la variable aleatoria circular.

Una vez que el concepto de simetría en estadística circular es un poco más claro, se considera el tema de las "Distribuciones Circulares" en el tercer capítulo; esto con el fin de entender el comportamiento de las distribuciones simétricas en estadística circular. Los principales temas a tratar dentro de este capítulo son: primeramente cómo es una función de distribución en estadística circular; segundo, la importancia que tienen los momentos trigonométricos en el comportamiento de una distribución circular; y finalmente se da a conocer un conjunto de distribuciones simétricas, al igual que algunas familias de distribuciones asimétricas y su relación entre sí.

Finalmente con el fin de darle un uso práctico, el último capítulo titulado "Pruebas de simetría para datos circulares" conjunta toda la información; además, se dan dos muestras aleatorias de datos circulares donde se tiene el supuesto de simetría en la distribución de sus datos; y finalmente se prueba si la hipótesis es cierta, o bien debe ser rechazada. Es por esto que dentro de este capítulo se detallan distintas pruebas de simetría circular, tomando en cuenta los requerimientos que necesita cada una de ellas, y considerando también la conjetura a la que se llegará con el resultado obtenido.

Capítulo 1

Métodos descriptivos

La investigación, como instrumento fundamental de las ciencias naturales, se basa en los principios de observación y razonamiento; es así como el análisis de datos juega un papel muy importante dentro de ella, ya que de este modo es posible obtener información de interés con una determinada confiabilidad. Este análisis de datos requiere de la estadística como una de sus principales herramientas; por lo que los investigadores necesitarán del manejo eficiente de los conceptos, técnicas y procedimientos estadísticos para que aunados con el conocimiento especializado de su área puedan dar respuesta a sus interrogantes.

Las ciencias naturales, utilizan a la estadística circular como apoyo para la resolución de problemas específicos; y es aquí donde se aborda el tema de los métodos descriptivos como primer acercamiento al análisis de un conjunto de datos circulares. Ya que éstos forman parte de la estadística descriptiva, para definirlos, primero se tiene que saber lo que es esta estadística para datos circulares.

La estadística descriptiva para datos circulares al igual que para datos lineales; se refiere a la presentación, descripción, análisis e interpretación de una colección de datos. Dentro de la estadística circular, la estadística descriptiva utiliza los métodos descriptivos con el fin de obtener de un conjunto de datos conclusiones sobre sí mismos sin sobrepasar el conocimiento proporcionado por éstos; es decir, podrán utilizarse para resumir o describir cualquier conjunto de datos circulares siempre y cuando se conozcan los elementos de la muestra.

1.1 Representación gráfica de los datos

La representación gráfica de los datos es muy importante ya que permite tener una idea inicial de las características de la muestra. Por ejemplo, si se quiere saber si una muestra parece provenir de una distribución uniforme, unimodal, o multimodal; bastará con ver la distribución de los datos a lo largo del círculo unitario, de tal modo que se deberá observar si existe alguna acumulación en direcciones específicas, o bien si están dispersos uniformemente alrededor del círculo.

De la misma forma se pueden proponer diversos modelos para la distribución de los datos, así pues si se cuenta con una muestra que parece proceder de una distribución simétrica unimodal; será posible suponer que proviene de una distribución von Mises. Pero si no se observa el comportamiento de los datos antes de emitir una afirmación acerca del modelo y se acepta que la muestra proviene de una distribución von Mises; y al visualizar los datos se observa que la muestra procede de una distribución multimodal, entonces se habrá cometido un gran error.

1.1.1 La gráfica de los datos primarios

Este tipo de representación gráfica consiste en la elaboración de un diagrama de dispersión de un conjunto de observaciones circulares; es decir, cada observación es representada como un punto sobre el círculo unitario, como se observa en la figura 1.1. Aunque es la representación más simple de los datos circulares, la elaboración de la gráfica de los datos primarios es importante sobre todo para evitar cometer errores en los supuestos del modelo.

Tiene algunas ventajas en comparación con otras gráficas de los datos; ya que el histograma por ejemplo puede ocultar características pequeñas pero importantes en la distribución de los datos, como la existencia de un pequeño grupo modal o un punto extremo cuya existencia puede ser de suma importancia en el análisis de los datos. Sin embargo en otros casos, como cuando se trabaja con una gran cantidad de datos en forma agrupada, será más apropiado dibujar un histograma.

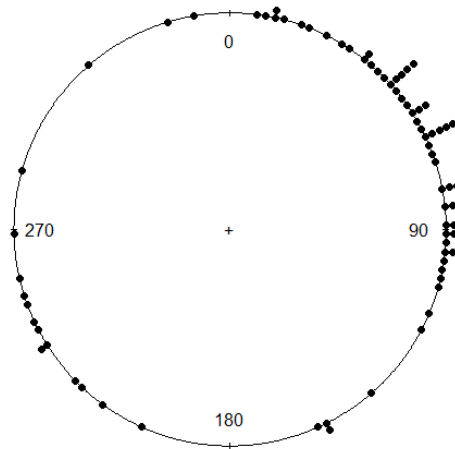


Figura 1.1 Diagrama de dispersión que representa la orientación que tomaron 76 tortugas después de haber puesto sus huevos¹.

1.1.2 Histogramas

En estadística circular existen dos formas básicas de representar los histogramas; lineales y angulares, y a su vez los histogramas angulares se dividen en: histograma angular simple y diagrama de rosa. Para cualquier tipo de histograma de datos circulares es necesario primeramente saber la elaboración de un histograma de datos lineales, ya que básicamente siguen el mismo proceso. Este proceso consiste: primero, en determinar el rango de los datos; segundo, en obtener el número de clases (existen distintos métodos, uno de ellos es sacando la raíz cuadrada del número de datos) y con esto establecer el ancho de clase (dividiendo el rango entre el número de clases); y finalmente graficar, tomando en cuenta que la altura de cada barra del histograma está dada por la frecuencia de la clase, mientras que los límites horizontales son los límites de cada clase.

Histogramas lineales: Un histograma lineal se construye como si los datos fueran lineales y no circulares (véase la figura 1.2), con la diferencia de que el punto de partida es seleccionado arbitrariamente. Si es más conveniente trabajar con un histograma lineal y se cuenta con un histograma angular, será posible hacer una

¹ Fisher N. I. (1993). Statistical Analysis Of Circular Data. Cambridge University Press.

transformación; cortando el histograma circular en un punto conveniente del círculo y desenrollando el histograma circular en un histograma lineal con un intervalo de 360° de ancho. Es importante considerar que si los datos son unimodales el corte debe ser casi opuesto a la moda, de tal forma que el centro del histograma lineal este cerca de la moda.

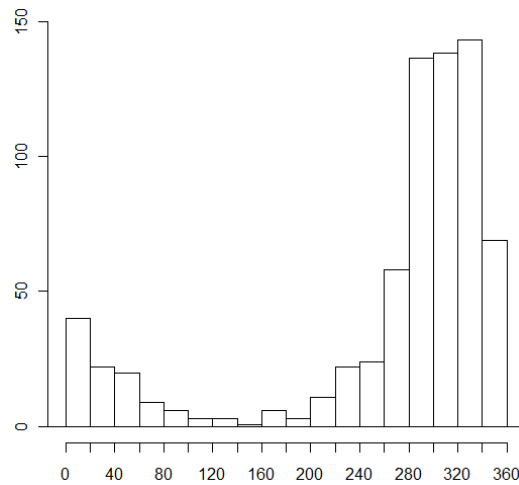


Figura 1.2 Histograma lineal de la dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos².

Histograma angular simple: La construcción de este tipo de histograma es análoga al histograma lineal, más específicamente, se obtiene envolviendo un histograma lineal alrededor de un círculo; de manera que cada barra del histograma quede centrada en el punto medio del intervalo de la clase correspondiente. El histograma angular simple puede representarse ya sea por medio de barras rectangulares con la finalidad de representar adecuadamente las frecuencias, o por medio de líneas cuando las observaciones representadas se refieren al análisis de sólo un espécimen.

² Mardia K. V (1975). Statistics Of Directional Data. London: Academic Press.

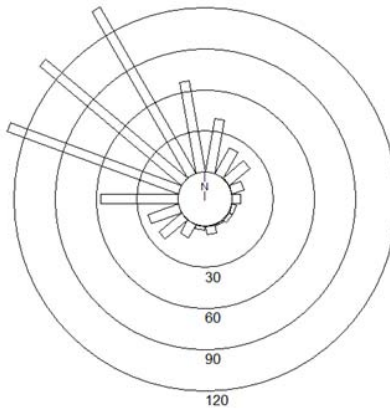


Figura 1.3 Histograma angular de la dirección en que desaparecen en el horizonte 714 Ánades Reales Británicos.³

Diagrama de rosa: El diagrama de rosa es la forma más común de los histogramas angulares. Su característica principal es que sustituye por sectores las barras del histograma angular simple. El área de cada sector será proporcional a la frecuencia de la clase correspondiente, puesto que el radio de cada sector es proporcional a la raíz cuadrada de la frecuencia relativa de las observaciones.

Existen varios puntos a considerar en relación a las ventajas y desventajas que ofrece un histograma. La virtud principal de representar los datos mediante algún tipo de histograma es su sencillez de construcción, lo cual es notorio al utilizar otro tipo de representación gráfica como la densidad no paramétrica estimada. Tanto el histograma como la densidad no paramétrica estimada persiguen la misma finalidad, la cual es representar la distribución de los datos mediante un adecuado alisamiento; en el caso del histograma, el alisamiento se verá reflejado en una cuidadosa elección del ancho de clase.

Por otra parte, los histogramas implican una elección arbitraria del rango de datos, lo cual puede modificar la distribución de los datos. Este defecto puede subsanarse mediante la repetición de un ciclo completo del histograma. Un aspecto quizá más importante en la interpretación de un histograma lineal es el punto de

³ ídem.

partida; por ejemplo, si el punto de partida es cercano a la moda daría la impresión errónea de que los datos son bimodales, como se muestra en la figura 1.2. Y finalmente, la elección del ancho de clase influirá de manera importante en la representación gráfica de cualquier tipo de histograma, tal como se aprecia en la figura 1.4; ya que una desafortunada elección puede resultar en una grave distorsión de la información en la muestra: sobre el número, tamaño y ubicación de los grupos modales; y por ende se puede tener un desempeño pobre del histograma. La elección del punto de partida y el ancho de clase dependerá principalmente de la distribución y del tamaño del conjunto de datos que se tengan.

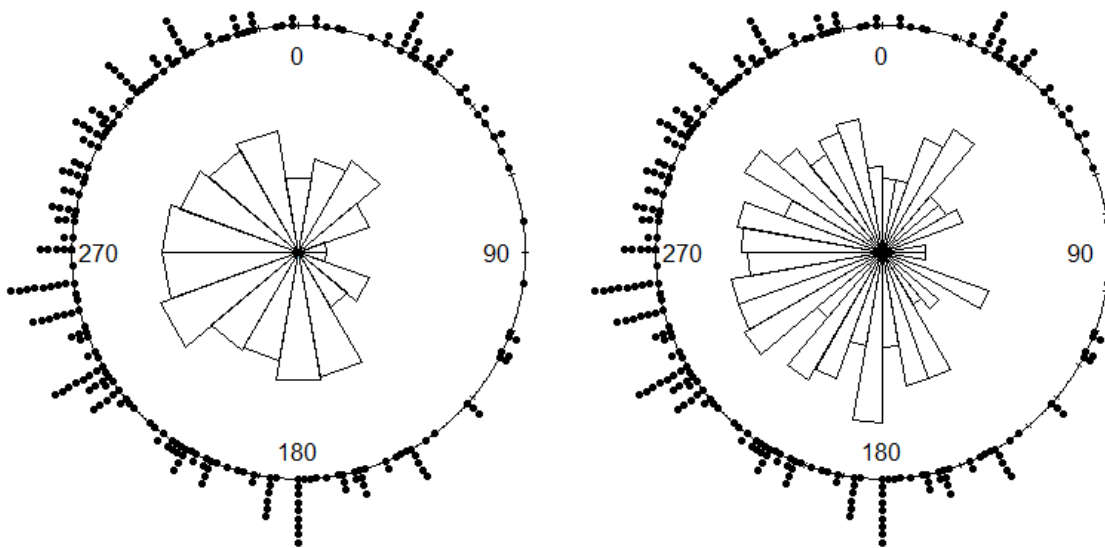


Figura 1.4 Diagramas de Rosa que muestran el tiempo de llegada (en horas) a la unidad de terapia intensiva de 254 pacientes⁴.

1.1.3 Diagramas de tallo y hojas

Se puede ver a este tipo de representación gráfica como una especie de histograma, donde el valor original de los datos es conservado formando una especie de barra. Esencialmente cada valor de datos es dividido en una hoja (normalmente el último dígito de la derecha) y un tallo (los otros dígitos restantes). Los valores del tallo se

⁴ Fisher N. I. (1993) *Ibidem*.

grupos modales, entre otras. Dichas medidas persiguen reducir en pocas cifras significativas ciertas características del conjunto de datos, de tal forma que se obtenga una comparación más precisa de la que se puede conseguir con las representaciones gráficas.

1.2.1 Medidas de tendencia central

Las medidas de tendencia central son indicadores estadísticos que suelen situarse hacia el centro de la distribución de los datos, de tal forma que permiten analizar los datos en torno a un valor central. Entre las medidas de tendencia central se tienen a las direcciones: media, mediana y moda.

En estadística circular el proceso para el cálculo del primer momento trigonométrico de la muestra, mejor conocido como la dirección media muestral m , está basado en una metodología totalmente diferente que en estadística de datos lineales, ya que no es posible calcular una media aritmética en los datos direccionales. En consecuencia, se define un proceso adecuado de tal forma que la dirección media apunte al centro de masa de la distribución de los datos. Este procedimiento consiste en representar cada observación como un vector sobre el círculo unitario y obtener el vector resultante de la suma vectorial de todas las observaciones; para poder así calcular las coordenadas polares asociadas a la dirección media.

Sea $\theta_1, \dots, \theta_n$ una muestra de datos circulares con vectores unitarios asociados e_1, \dots, e_n , tal que $\|e_i\| = 1$ ($\forall i = 1, \dots, n$); entonces, la longitud R del vector resultante para la muestra estará dada por:

$$R = \left\| \sum_{i=1}^n e_i \right\|; \quad (R \geq 0)$$

Ahora bien, la cantidad R se encuentra en el intervalo $(0, n)$; mientras que la longitud media resultante \bar{R} asociada al primer momento trigonométrico de la muestra y definida como:

$$\bar{R} = R/n$$

se encuentra en el intervalo $(0,1)$.

Tomando en cuenta que los componentes rectangulares para cada vector unitario e_i están representados como:

$$x_i = \cos \theta_i, \quad y_i = \text{sen } \theta_i$$

Los componentes rectangulares para el primer momento trigonométrico, en adelante dirección media, serán:

$$a_1 = (1/n) \sum_{i=1}^n \cos \theta_i, \quad b_1 = (1/n) \sum_{i=1}^n \text{sen } \theta_i$$

De esta forma el ángulo $\bar{\theta}$ asociado al vector resultante, llamado el ángulo medio de la muestra, está dado por las ecuaciones:

$$\cos \bar{\theta} = a_1/\bar{R}, \quad \text{sen } \bar{\theta} = b_1/\bar{R}$$

Al resolver estas dos ecuaciones se tiene:

$$\bar{\theta} = \begin{cases} \tan^{-1}(b_1/a_1) & \text{si } a_1 \geq 0 \\ \pi + \tan^{-1}(b_1/a_1) & \text{si } a_1 < 0 \end{cases} \quad \text{donde } \tan^{-1}(\theta) \in [-\pi/2, \pi/2]$$

Es así como las coordenadas polares asociadas a la dirección media m están dadas por:

$$m = (\bar{R}, \bar{\theta})$$

Una manera sencilla de obtener la mediana de la muestra consiste en elegir un eje que divida a la muestra en dos grupos, de tal manera que la mitad de los datos se encuentren en un lado del eje y la otra mitad en el otro lado. A este eje se le denomina dirección mediana; y al ángulo que forma, ángulo mediano. Cuando el tamaño de la muestra es impar, la dirección mediana pasará a través de un punto de la muestra; mientras que cuando es par, la dirección mediana pasará en medio de dos puntos de la muestra. Este tipo de procedimiento es fácil de realizar mientras los datos se

concentran en un arco del círculo sustancialmente menor que el círculo completo; esto cuando sea posible identificar los dos extremos de los datos.

Otra forma de calcular el ángulo mediano $\tilde{\theta}$ y que también se lleva a cabo cuando los dos extremos de los datos pueden ser identificados, consiste en aplicar un algoritmo como si los datos fueran lineales. Aunque si la masa de los datos coincide con 0° ; se deberá hacer una rotación inicial de los datos, posteriormente se aplicará el algoritmo y finalmente se deberá rotar de vuelta la respuesta. De manera general $\tilde{\theta}$ se podrá encontrar⁶ al minimizar la función:

$$d(\theta) = \pi - \frac{1}{n} \sum_{i=1}^n |\pi - |\theta_i - \theta||$$

Existen diversos puntos que se deben de considerar en el cálculo de la mediana. Primero, que para datos multimodales la dirección mediana no puede ser definida de forma única; sin embargo es única para datos considerablemente unimodales. Y segundo, para una distribución simétrica la dirección mediana estará localizada en el eje de simetría.

La dirección modal de la muestra, análogamente a la moda obtenida de datos lineales, se define como cualquier dirección que corresponda a una concentración máxima local de los datos. Un procedimiento para el cálculo del ángulo $\tilde{\theta}$ asociado a la dirección modal para una muestra de observaciones discretas consiste en obtener el punto medio del sector con mayor frecuencia en un diagrama de rosa; este procedimiento pierde consistencia debido a que el grupo modal puede variar considerablemente dependiendo de la ubicación y el ancho de clase.

1.2.2 Momentos trigonométricos de la muestra

Sea $\theta_1, \dots, \theta_n$ una muestra de datos circulares, se define de manera general para $p = 1, 2, \dots$, el p -ésimo momento trigonométrico de la muestra m_p , como las coordenadas polares

$$m_p = (R_p, \bar{\theta}_p)$$

⁶ Ídem.

donde R_p es la longitud resultante del p -ésimo momento trigonométrico de la muestra tal que:

$$R_p^2 = a_p^2 + b_p^2$$

donde

$$a_p = (1/n) \sum_{i=1}^n \cos p\theta_i, \quad b_p = (1/n) \sum_{i=1}^n \sen p\theta_i$$

de tal forma que su ángulo polar asociado se calcula como:

$$\bar{\theta}_p = \begin{cases} \tan^{-1}(b_p/a_p) & \text{si } a_p \geq 0 \\ \pi + \tan^{-1}(b_p/a_p) & \text{si } a_p < 0 \end{cases} \quad \text{donde } \tan^{-1}(\theta) \in [-\pi/2, \pi/2]$$

Al igual que en estadística lineal, en estadística circular también están definidos los p -ésimos momentos trigonométricos centrales de la muestra m'_p , con componentes rectangulares:

$$\bar{a}_p = (1/n) \sum_{i=1}^n \cos p(\theta_i - \bar{\theta}), \quad \bar{b}_p = (1/n) \sum_{i=1}^n \sen p(\theta_i - \bar{\theta})$$

y longitud resultante R'_p , tal que:

$$R'_p{}^2 = \bar{a}_p{}^2 + \bar{b}_p{}^2$$

1.2.3 Medidas de dispersión

Un rasgo principal de los datos es su dispersión, la cual se refiere a la concentración y a la variabilidad de los valores de la distribución. Las medidas de dispersión brindan información sobre en qué medida los valores de la muestra se concentran alrededor de la media; y hasta qué punto las medidas de tendencia central son representativas como síntesis de la información. Entre las medidas de dispersión se distinguen a la varianza circular, a la desviación estándar circular y al rango circular de la muestra.

En los casos cuando los datos de la muestra están comprendidos en un solo grupo, o cuando la muestra es unimodal, la longitud media resultante \bar{R} servirá como una medida de concentración. En el caso extremo cuando todos los puntos de la muestra caen en el mismo punto del círculo unitario, se tendrá una $\bar{R} = 1$, éste es el caso de máxima concentración; conforme la concentración de los datos en torno a cierta dirección disminuya, el valor de \bar{R} también lo hará. En el extremo inferior cuando $\bar{R} = 0$, no existe una concentración en torno a una sola dirección; aunque si los datos no están comprendidos en un solo grupo, una $\bar{R} = 0$ no implicará que haya dispersión uniforme alrededor del círculo.

Considerando a la longitud media resultante \bar{R} como una medida de concentración, e indirectamente como una medida de dispersión, de tal forma que \bar{R} disminuye de 1 a 0 cuando la dispersión aumenta; se define a la varianza circular de la muestra como:

$$V = 1 - \bar{R}$$

Cuanto menor sea el valor de la varianza circular, la distribución estará mayormente concentrada en un arco determinado. A diferencia de una varianza para datos lineales, se tiene que $0 \leq V \leq 1$. Y finalmente si $V = 1$, es decir si $\bar{R} = 0$, no implicará necesariamente una distribución dispersa al máximo.

La desviación estándar circular de la muestra equivalente a la desviación estándar para datos lineales, se define como:

$$v = [2(1 - \bar{R})]^{1/2}$$

A medida que la longitud media resultante \bar{R} disminuya, la desviación estándar circular v aumentará. Cuando $\bar{R} \approx 1$ una buena aproximación estará dada por:

$$v \approx (-2 \ln \bar{R})^{1/2}$$

Del primero y segundo momento trigonométrico se obtiene otra medida de dispersión llamada la dispersión circular de la muestra $\hat{\delta}$, definida como:

$$\hat{\delta} = (1 - R_2)/(2\bar{R}^2)$$

Esta nueva medida desempeña un papel importante en el cálculo del intervalo de confianza para una dirección media.

Finalmente el rango circular de la muestra, otra medida de dispersión, se define como la longitud del arco más pequeño que contiene a todas las observaciones de la muestra. El proceso para el cálculo de esta medida consiste en obtener las estadísticas de orden para la muestra $\theta_1, \dots, \theta_n$, como si los datos fueran lineales, de tal forma que $\theta_{(1)} \leq \dots \leq \theta_{(n)}$. Posteriormente se definen las medidas de los arcos adyacentes como:

$$T_i = \theta_{(i+1)} - \theta_{(i)}; \quad i = 1, \dots, n - 1$$

$$T_n = 2\pi - \theta_{(n)} + \theta_{(1)}$$

Por lo tanto el rango circular ω está dado por:

$$\omega = 2\pi - \max(T_1, \dots, T_n)$$

1.2.4 Medidas de asimetría y curtosis

Las medidas de asimetría y curtosis para datos circulares describen la distribución de la muestra, de tal manera que permiten identificar la forma en que las observaciones se concentran alrededor de la moda. Por lo que este tipo de medidas son útiles solo para distribuciones unimodales. Utilizando el primer y segundo momento trigonométrico de la muestra, se definen las medidas de asimetría y curtosis respectivamente como:

$$\hat{g}_1 = [R_2 \text{sen}(\bar{\theta}_2 - 2\bar{\theta})]/(1 - \bar{R})^{3/2}$$

$$\hat{g}_2 = [R_2 \cos(\bar{\theta}_2 - 2\bar{\theta}) - \bar{R}^4]/(1 - \bar{R})^2$$

tal que $\bar{\theta}_2$ es el ángulo asociado al segundo momento trigonométrico.

Así pues para los datos pertenecientes a una muestra de una distribución simétrica unimodal, \hat{g}_1 será aproximadamente cero. El término que determina básicamente la asimetría de una muestra circular es:

$$R_2 \text{sen}(\bar{\theta}_2 - 2\bar{\theta})$$

Ya que para muestras simétricas esta expresión desaparece.

En el caso de la medida de curtosis \hat{g}_2 , ésta será casi cero para una distribución unimodal con un pico normal. Si $\hat{g}_2 > 0$ la distribución puede ser descrita como leptocúrtica; es decir, que los datos se concentran en un arco pequeño (la distribución tiene un pico grande en la moda). Mientras que si $\hat{g}_2 < 0$ la distribución puede ser descrita como platicúrtica; esto es que los datos se concentran de manera relativamente uniforme en todo el arco que contiene a las observaciones de la muestra (la distribución es considerablemente plana en la moda).

Capítulo 2

El concepto de simetría

2.1 Introducción

La palabra simetría es un término que a lo largo de los años ha sufrido diversas modificaciones, debido a que dicha palabra toma un significado específico de acuerdo a las vivencias que cada persona tenga; es así que lo que para una persona es significado de proporción, para otra es significado de belleza, o en muchos casos perfección. A pesar de la subjetividad relacionada con este término, es importante destacar que desde hace miles de años la palabra simetría ya estaba asociada como una característica que reflejaba cierta semejanza respecto a la medida de las cosas. De hecho, la etimología de esta palabra de origen griego surge a partir de la combinación de palabras “syn”, que significa en conjunto o simultáneamente; y “metrios” que significa en la medida adecuada o en la proporción correcta.

La simetría es un concepto integral que al igual que el concepto de orden, armonía, ritmo, analogía, entre otros, aparece muy a menudo tanto en la vida cotidiana como en diversas áreas de estudio; debido a la necesidad que tiene el ser humano de organizar y definir todo lo que existe a su alrededor. Tales conceptos son utilizados como un principio de organización que sirve de ayuda para dar un amplio estudio interdisciplinario del mundo. Además, a través de la simetría se puede tener una idea de cómo las artes y las ciencias, las diferentes disciplinas y diferentes culturas humanas; se han enriquecido conceptualmente entre sí.

De manera general, el significado de la palabra simetría puede verse ya sea como un fenómeno, como un concepto, o como una operación; así pues, cuando se habla de aquello que se considera como simétrico con base en las experiencias y conocimientos adquiridos, se dirá que se habla de simetría como un fenómeno. Si en cambio, se hace referencia a todo aquello relacionado con la simetría vista como un fenómeno, entonces se estará abordando el término simetría como un concepto; o si a su vez se

está considerando todo aquello que hace posible que el fenómeno de la simetría se desarrolle; se estará hablando de la simetría como una operación.

2.1.1 Distintos tipos de simetría

Aunque existen diversos casos en los que la simetría se puede hacer presente, es la reflexión el ejemplo más conocido dentro de la geometría. Este tipo de simetría consiste en reflejar una figura plana con relación a un eje, conocido comúnmente como eje de simetría; dicho proceso da origen a una figura reflejada en la que tanto los puntos de la nueva figura como los puntos de la figura original estarán a la misma distancia del eje, aunque en direcciones opuestas. En otras palabras, la figura reflejada mantendrá tanto la forma de la figura original como su tamaño, sus ángulos internos y su color; el único cambio que hará la reflexión será modificar la orientación de la figura, de tal modo que la izquierda se intercambia por la derecha y viceversa. Las características de la figura original que permanecen sin cambios con respecto a su reflejo son las cualidades que identifican este tipo de simetría. Una propiedad de este tipo de simetría es que; si una figura que fue reflejada mediante un determinado eje es nuevamente reflejada con respecto al mismo eje de simetría, entonces se regresará nuevamente a la figura original. Si se desea reflejar una figura en una dimensión adicional será necesario reflejar la figura pero ahora mediante un plano, el cual actuará como un espejo.

Otro tipo muy usual de operación simétrica es la rotación, la cual consiste en girar una figura en torno a un eje; al igual que la reflexión, la rotación conservará las características internas de la figura y adicionalmente también se conserva la distancia que existe de los puntos al eje. Si una figura es rotada por un ángulo α de tal forma que al rotarla n -veces dicha figura vuelve a su forma original; entonces a este tipo de operación se le denominará n -veces simetría; en general $\alpha = 360^\circ/n$, donde n es un número natural.

La traslación ocupa el tercer lugar entre las operaciones simétricas más usuales; esta operación consiste en desplazar una figura geométrica a lo largo de una línea recta, manteniendo siempre un período constante y una misma dirección lo que da origen a una serie repetida de la figura original. En otras palabras, el proceso

mantiene todas las características de la figura, salvo su ubicación; es esta uniformidad lo que hace de esta operación una operación simétrica.

Finalmente dentro de las operaciones simétricas más comunes está la reflexión de deslizamiento, este tipo de simetría puede verse como una operación combinada; ya que el objeto a ser reflejado es primero deslizado sobre una línea horizontal para posteriormente ser reflejado a partir de un eje con la misma dirección del deslizamiento. Las características que preservará la figura están formadas por las características que corresponderían a la simetría de traslación y de reflexión por separado, de hecho en la cristalografía donde es muy usual este tipo de simetría este tipo de operación se considera como una operación de traslación y reflexión por separado.

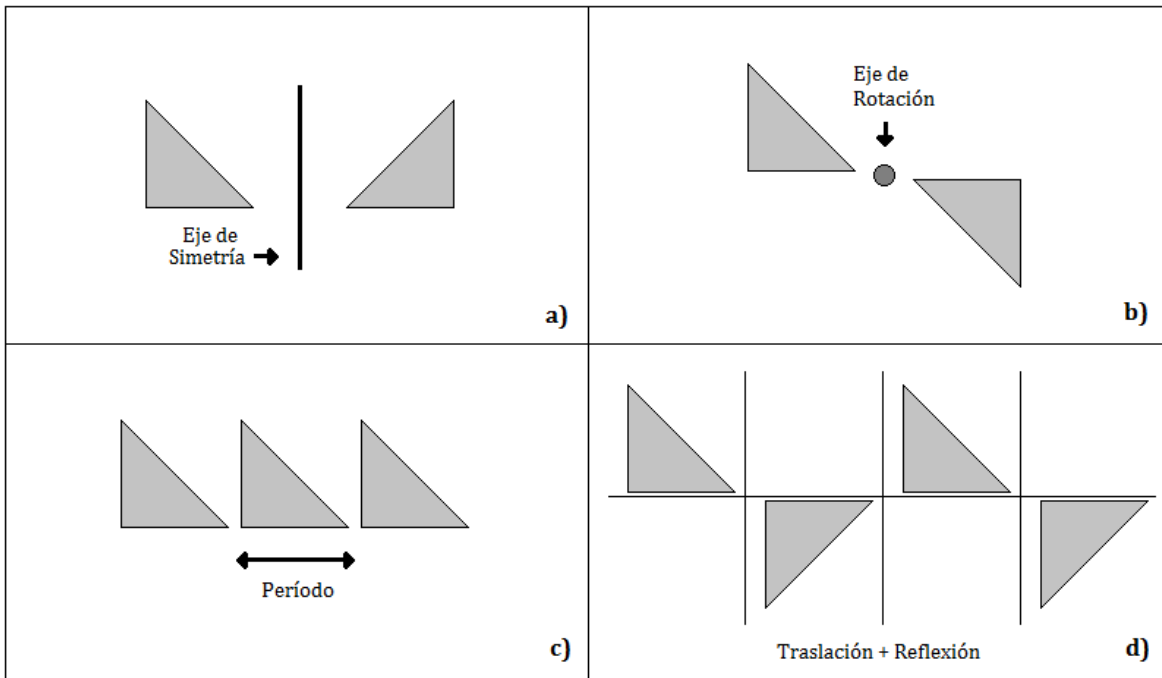


Figura 2.1. a) Reflexión de un triángulo. b) Rotación de un triángulo 180°. c) Traslación de un triángulo. d) Reflexión de deslizamiento de un triángulo.

Una característica de los distintos tipos de simetría geométrica es que utilizan como herramienta principal una determinada operación matemática conocida como

transformación. En el proceso de estas transformaciones ciertas propiedades de la figura se conservan invariantes. En la vida existe una gran diversidad de simetrías e invariancias vistas en forma de transformaciones, los criterios que las agrupan van desde los que sólo son aplicables en ciertos casos hasta los que son válidos en un sentido más amplio. Estos últimos reciben el nombre de leyes, en la cristalografía por ejemplo aparecen diversas leyes de simetría.

La simetría no sólo se presenta en figuras u operaciones geométricas, y tampoco específicamente mediante características geométricas; en general, se puede hablar de simetría si algún objeto no necesariamente geométrico permanece invariante en alguna de sus características ante algún tipo de operación o transformación. Esto hace posible poder entender la invariancia bajo la reflexión de carga, en la física; o bajo el intercambio de colores, en el arte. De manera general siempre que cualquier característica permanezca constante se podrá hablar de simetría. Por ejemplo, una magnitud física, el ritmo de un poema o el tema de una pieza de música podrán ser considerados como objeto de estudio de la simetría.

A pesar de que el fenómeno de la simetría aparece de manera muy frecuente dentro del área de las matemáticas, ésta es sólo una de las diversas áreas en las que el fenómeno de la simetría se hace presente. En las ciencias, algunos ejemplos de la presencia del fenómeno de la simetría son: en la física, las leyes de conservación, las simetrías y asimetrías de la estructura de la materia; en la astronomía, la investigación de los eventos periódicos y la elaboración de los calendarios; en la biología, la estructura del ADN y las moléculas de ARN, las simetrías morfológicas y funcionales de los seres vivos; en la psicología, las pruebas de personalidad y la asimetría del cerebro. Además, su representación en la arquitectura de la obra musical, la rima, la armonía de las proporciones y el color, entre otros muchos aspectos demuestran también su aparición en el arte.

2.1.2 Conceptos relacionados

Para un mejor entendimiento del concepto de simetría es necesario tener conocimiento de los conceptos que se generan a partir de éste. Un primer concepto es la asimetría, la cual denota la ausencia de simetría; se dice que un objeto presenta asimetría cuando ninguna de sus características muestra la presencia de simetría. En

cambio, cuando en un objeto se hace notar la presencia de simetría pero ésta no es perfecta o tiene algún defecto; se dice que existe disimetría. La antisimetría es otro concepto estrechamente relacionado; se dice que un objeto presenta antisimetría cuando ante un tipo de operación o transformación, alguna característica de dicho objeto es preservada de tal modo que se transforma en su forma opuesta.

En el lejano oriente, por ejemplo, algunos símbolos muestran simetría en algunas de sus características geométricas; mientras que en otras características no geométricas se presentan pares opuestos; tal es el caso del símbolo del yin-yang, donde se observa que la figura presenta una simetría rotacional de tal forma que al rotar el símbolo 180° alrededor de su centro la forma de la figura se conserva pero su color cambia de blanco a negro y de negro a blanco, es decir, que en cuanto a sus colores se dice que la figura es antisimétrica. De igual forma obras filosóficas de origen chino utilizaron la antisimetría para describir muchas dicotomías que en su interpretación pasaron a formar parte de su visión del mundo.

Es tan grande el número de simetrías que se encuentran en la naturaleza que a veces se podría ver todo con la perspectiva de si un objeto es o no simétrico. Debido a esto es importante resaltar que tanto la simetría como la asimetría en la vida cotidiana no existen de manera perfecta, cualquier objeto o fenómeno no puede estar libre de simetría, y en general se puede decir que nada es totalmente asimétrico sino que en ciertas características es simétrico y en otras asimétrico.

2.2 Descripción matemática de la simetría

Tanto las simetrías geométricas como las invariancias bajo distintos cambios, no necesariamente geométricos, son fenómenos que pueden ser descritos mediante el uso de una teoría matemática unificada. La complejidad de esta teoría matemática radica en que mientras más unificada sea la descripción que se busca para caracterizar a la simetría, más alto será el nivel de abstracción necesario debido a que el objetivo principal es modelar matemáticamente la mayor parte de simetrías teóricamente posibles.

El concepto de grupo ha demostrado ser la herramienta más útil para la descripción matemática de las simetrías, debido a que cualquier cosa puede ser considerada como un elemento de un grupo, ya sea un número, un evento, un objeto, etc. Esto asegura que la simetría puede ser interpretada mediante cualquier conjunto de fenómenos; además la operación definida por los elementos del grupo puede ser cualquier tipo de operación que satisfaga los requerimientos de los cuatro axiomas de un conjunto, lo que hace posible no restringirse a sólo operaciones aritméticas básicas.

Con el cuadrado se puede ejemplificar un conjunto de ocho operaciones simétricas las cuales dan como resultado un cuadrado equivalente de tal manera que no se podría distinguir las cifras antes y después de cada operación. Estas operaciones son: i) las reflexiones obtenidas mediante las líneas trazadas entre los puntos medios de los lados opuestos como ejes de simetría, ii) las reflexiones obtenidas tomando las líneas trazadas entre los vértices opuestos; iii) las rotaciones generadas al tomar como referencia el centro del cuadrado alrededor de 90° , 180° , 270° y 360° ; iv) la operación inversa de cada una de las operaciones anteriores generada como la misma operación pero en reversa; y finalmente v) la operación de dejar el cuadrado intacto (neutro del elemento). Es así que con todas las operaciones anteriores es posible representar un conjunto, ya que se satisfacen los cuatro axiomas que debe cumplir un conjunto.

Otro ejemplo es el que el matemático suizo Andreas Speiser señala en su libro *Die Theorie der Gruppen von Endlicher Ordnung*; al brindar una detallada explicación de la teoría de conjuntos de las simetrías ornamentales; él sostiene que la presentación de los cinco poliedros regulares fue el mayor logro de la geometría deductiva en el sistema creado por los griegos y canonizada en los *Elementos* de Euclides.

2.2.1 Simetría en la estadística

Hablar de simetría en el área de la estadística conlleva al análisis de distintas hipótesis; cada una de las cuales se apoya en un modelo con parámetros específicos. Este análisis plantea cuestiones importantes; ya que en una función de distribución la simetría resulta ser una de las características de real interés estadístico.

Es fácil definir gráficamente el concepto de simetría para una distribución; de esta manera, se puede decir que una distribución es simétrica si la mitad izquierda de la gráfica de la función de densidad de probabilidad es la imagen reflejada de la mitad derecha de la gráfica. Existen diversas distribuciones que pueden ejemplificar este concepto; en el caso discreto, la distribución binomial es simétrica siempre que la probabilidad de éxito sea igual a $1/2$, mientras que la distribución uniforme discreta es siempre simétrica, como se observa en la figura 2.2 donde la línea punteada representa la línea alrededor de la cual las distribuciones son simétricas. La distribución normal estándar ilustra el concepto gráfico en el caso continuo, como se ve en la figura 2.3.

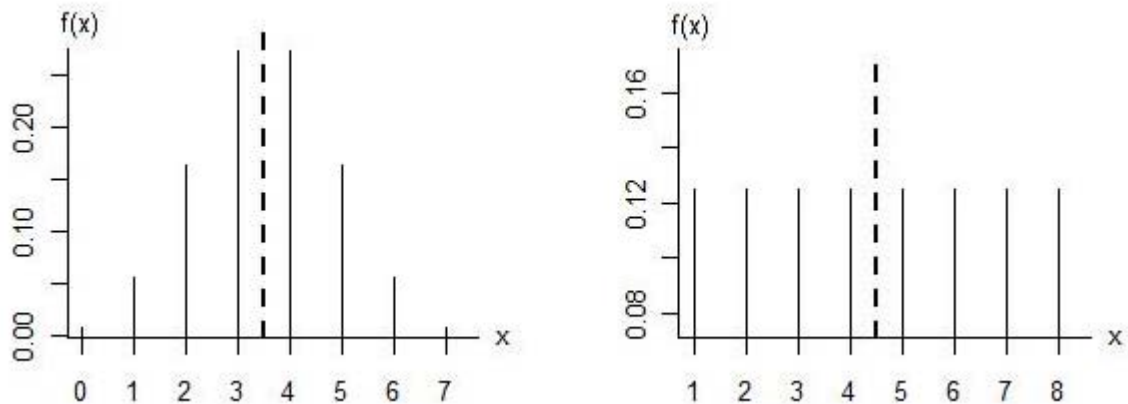


Figura 2.2. Las distribuciones: binomial (a la izquierda) y uniforme discreta (a la derecha).

Debido a que el concepto gráfico de simetría suele ser algo abstracto, es necesario contar con una definición más precisa de simetría de manera que se puedan abarcar los distintos escenarios que presenta la simetría. Para esto será necesario definir una variable aleatoria X , con función de distribución $F(x)$ y función de densidad de probabilidad $f(x)$.

Definición 1: Los métodos estadísticos clásicos suponen que $f(x)$ es una función simétrica sobre algún valor desconocido μ , si $f(\mu - x) = f(\mu + x)$ para toda x .

Definición 2: La distribución de la variable aleatoria X es simétrica respecto de la recta $x = c$, para alguna constante c , si la probabilidad de $\{X \leq c - x\}$ es igual a la probabilidad de $\{X \geq c + x\}$ para cada valor posible de x .

De acuerdo a la definición 2, se puede observar en la figura 2.2 la recta $x = c$ para la cual cada una de las distribuciones es simétrica, para el caso de la distribución binomial $c = 3.5$ y la definición es fácilmente comprobable para toda x en los números reales; en el caso de la distribución uniforme discreta $c = 4.5$.

Definición 3: Se dice que la función de distribución $F(x)$ es simétrica con respecto al origen si

$$F(x) + F(-x) = 1 \quad \text{Para toda } x$$

Dentro del análisis de una variable aleatoria en estudio, el supuesto de simetría para dicha variable será un punto clave; ya que si bien algunas veces no se puede saber la distribución exacta de una variable aleatoria a menudo es posible suponer que la distribución es simétrica. Esta hipótesis no es tan fuerte como el supuesto de normalidad; ya que mientras que todas las distribuciones normales son simétricas, no todas las distribuciones simétricas son normales.

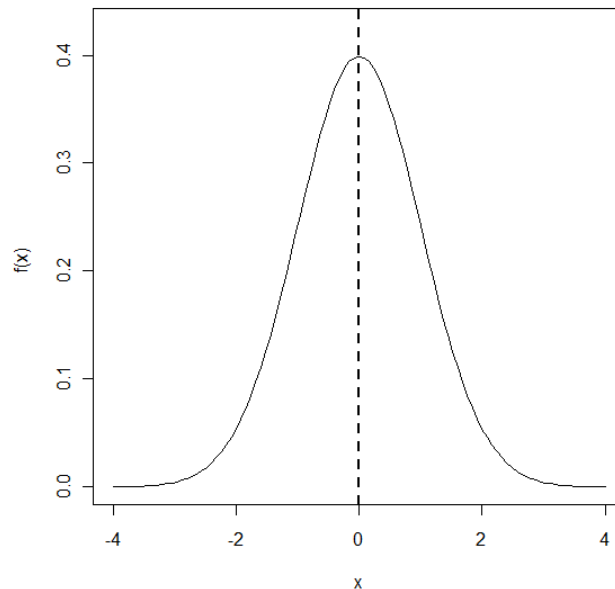


Figura 2.3. La distribución normal estándar.

En la figura 2.3 se muestra la distribución normal estándar la cual es un caso particular de la distribución normal; mediante esta función de densidad se puede observar la simetría que presentan las distribuciones normales.

Es importante probar la hipótesis de simetría ya que aceptar o rechazar dicho supuesto trae consigo diversos efectos sobre la función de distribución de la variable aleatoria en estudio. En resumen, algunas de las consideraciones generales para el análisis de simetría en una variable aleatoria son las siguientes:

1. Puede haber consecuencias importantes en la interpretación de los resultados, ya que si la función de distribución de una variable aleatoria no es simétrica; entonces ya no hay sólo una medida natural de localización y por lo tanto se tendrá que decidir cuál será la medida de localización que se va a considerar o analizar para dicha variable en estudio.
2. Debido a que las técnicas robustas siguen suponiendo simetría, si la función de distribución de los datos que se están analizando no es simétrica y para el análisis de dichos datos es necesario aplicar dichas técnicas; entonces una manera de tratar los datos es eligiendo una transformación de una familia paramétrica flexible que haga simétricos a los datos, realizar una prueba de simetría, y luego aplicar los procedimientos deseados a los datos que ya se han establecido simétricos.
3. Si el supuesto de simetría en el modelo es aceptado cualquier inferencia asociada a la mediana será válida también para la media y viceversa; debido a que si una función de densidad es simétrica se tiene como consecuencia que la media (si existe) coincide con la mediana; esto debido a que ambas se encuentran exactamente en medio de la distribución, es decir, en la línea de simetría.
4. Si el supuesto de simetría es significativo la distancia de la mediana es una medida significativa; y por tanto la distancia entre dos observaciones es una medida significativa. Como resultado, la escala de medición es más que sólo ordinal, es de intervalo.

Capítulo 3

Distribuciones circulares

3.1 Introducción

Una característica principal de las distribuciones circulares es que en teoría son continuas; además en contraste con las distribuciones de la estadística lineal, las distribuciones circulares son siempre finitas y su rango se encuentra entre 0 y 2π . La función de densidad de probabilidad $f(\theta)$, asociada a la distribución de la variable aleatoria circular θ , es una función continua que cumple con las siguientes propiedades:

1. Es siempre positiva:

$$f(\theta) \geq 0; \quad 0 \leq \theta < 2\pi$$

2. La integral de $f(\theta)$ sobre todo su rango es igual a 1:

$$\int_0^{2\pi} f(\theta) d\theta = 1$$

3. Es una función periódica con respecto a 2π :

$$f(\theta) = f(\theta + k \cdot 2\pi); \quad \forall k \in \mathbb{Z}$$

La relación que existe entre la función de densidad $f(\theta)$ y la función de distribución $F(\theta)$, de la variable aleatoria circular θ , se obtiene mediante la siguiente igualdad:

$$F(\theta) = \int_0^{\theta} f(\phi) d\phi$$

En particular, $F(\theta)$ cumple que:

$$F(0) = 0 \quad y \quad F(2\pi) = 1$$

3.1.1 Distribuciones envueltas

Una distribución circular puede ser considerada como una distribución bivariada sujeta a la restricción de que la probabilidad total (masa total) estará dispersa sobre el círculo unitario. Con base en esta restricción y mediante diversos procesos, algunas distribuciones circulares pueden ser formadas a partir de distribuciones lineales ya conocidas.

Uno de estos procesos consiste en envolver una distribución lineal alrededor de la circunferencia del círculo unitario; reduciendo las posibles observaciones angulares al intervalo $[0, 2\pi]$, y agregando las probabilidades que caen dentro de este mismo sector del círculo. Es decir, si x es una variable aleatoria con función de distribución $F(x)$; la variable aleatoria x_w de la distribución envuelta está dada por:

$$x_w = x(\text{mod } 2\pi)$$

Asimismo la función de distribución de x_w se podrá ver como:

$$F_{x_w}(\theta) = \sum_{k=-\infty}^{\infty} \{F(\theta + 2\pi k) - F(2\pi k)\}; \quad 0 \leq \theta < 2\pi$$

Similarmente si x tiene una función de densidad de probabilidad $f(x)$; la función de densidad de probabilidad de θ será:

$$f_{x_w}(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k)$$

3.2 Momentos trigonométricos

En el estudio de las distribuciones circulares el cálculo de los momentos es equivalentemente útil como en las distribuciones lineales, las cuales son

caracterizadas a través de sus momentos. Los momentos trigonométricos, llamados así ya que se obtienen mediante funciones trigonométricas, se definen considerando la periodicidad que identifica a una distribución circular.

Sea $f(\theta)$ la función de densidad de probabilidad de una distribución circular, se tiene entonces que para cada vector unitario con componentes:

$$x = \cos \theta, \quad y = \operatorname{sen} \theta$$

existe un valor asignado de $f(\theta)$. En consecuencia el primer momento trigonométrico de la distribución circular esta dado por:

$$\alpha_1 = \int_0^{2\pi} x f(\theta) d\theta = \int_0^{2\pi} \cos \theta f(\theta) d\theta$$

$$\beta_1 = \int_0^{2\pi} y f(\theta) d\theta = \int_0^{2\pi} \operatorname{sen} \theta f(\theta) d\theta$$

En este caso α_1 y β_1 son los componentes de la dirección media, la cual apunta al centro de masa. Es así como el primer momento trigonométrico de la distribución circular es llamado también la dirección media de la distribución, y se denota como:

$$\varphi_1 = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}$$

A diferencia de los momentos de una distribución lineal, el primer momento trigonométrico es una magnitud vectorial. Mediante coordenadas polares se puede ver a φ_1 como:

$$\varphi_1 = (\rho_1, \mu_1)$$

donde ρ_1 es la longitud de la dirección media y μ_1 su ángulo polar, llamado ángulo medio.

En general para $p = 1, 2, \dots$, se define el p -ésimo momento trigonométrico de la variable aleatoria circular θ como:

$$\varphi_p = \begin{bmatrix} \alpha_p \\ \beta_p \end{bmatrix}$$

donde

$$\alpha_p = \int_0^{2\pi} \cos(p\theta) f(\theta) d\theta, \quad \beta_p = \int_0^{2\pi} \operatorname{sen}(p\theta) f(\theta) d\theta$$

Las coordenadas polares asociadas a cada momento trigonométrico de θ se denotan como:

$$\varphi_p = (\rho_p, \mu_p)$$

Dado que ρ_p es la longitud de φ_p , es decir,

$$\rho_p = \|\varphi_p\| = (\alpha_p^2 + \beta_p^2)^{1/2}$$

si $\rho_p = 0$, no existirá un ángulo polar μ_p asociado a φ_p ; en caso contrario cuando $\rho_p > 0$, dicho ángulo podrá ser determinado como:

$$\mu_p = \begin{cases} \tan^{-1}(\beta_p/\alpha_p) & \text{si } \alpha_p \geq 0 \\ \pi + \tan^{-1}(\beta_p/\alpha_p) & \text{si } \alpha_p < 0 \end{cases}$$

Para el primer momento trigonométrico, esto es para $p = 1$, se denotará de manera simple a μ_1 como μ y a ρ_1 como ρ .

Al igual que las distribuciones lineales cuentan con momentos centrales, las distribuciones circulares contarán con momentos trigonométricos centrales. De modo que el p -ésimo momento trigonométrico central de θ para $p = 1, 2, \dots$ se define como:

$$\varphi'_p = \rho_p e^{i(\mu_p - p\mu)}$$

El p -ésimo momento trigonométrico central también podrá verse como el p -ésimo momento trigonométrico de $\theta - \mu$, de modo que:

$$\varphi'_p = \begin{bmatrix} \alpha'_p \\ \beta'_p \end{bmatrix}$$

donde

$$\alpha'_p = \int_0^{2\pi} \cos p(\theta - \mu) f(\theta) d\theta, \quad \beta'_p = \int_0^{2\pi} \operatorname{sen} p(\theta - \mu) f(\theta) d\theta$$

Utilizando los primeros momentos trigonométrico se obtienen distintas medidas que son de interés en el análisis de las distribuciones circulares. Tal es el caso del primer momento, donde al considerar que ρ toma valores entre 0 y 1 ($0 \leq \rho \leq 1$) se podrá deducir que si ρ se aproxima a 1; la masa total estará concentrada alrededor de la dirección media. Sin embargo, si la masa total no está concentrada alrededor de alguna dirección específica pero está extendida uniformemente o con más de una moda, ρ será pequeña. Así pues, se puede considerar a ρ como una medida de concentración alrededor de la dirección media.

Debido a que en ocasiones es más conveniente trabajar con una medida de dispersión que con una medida de concentración; se plantea el concepto de la desviación estándar circular como:

$$\sigma = [2(1 - \rho)]^{1/2}$$

Aun así es visible la relación que existe entre la medida de concentración ρ y la medida de dispersión σ ; de tal modo que mientras una se incrementa, la otra disminuye, y viceversa.

Otra forma de medir la desviación estándar circular es a través de la expresión:

$$s = (-2 \ln \rho)^{1/2}$$

Cuando $\rho \approx 1$, σ y s tienen aproximadamente el mismo valor ($\sigma \approx s$). Mientras que para valores pequeños de ρ , σ y s diferirán considerablemente una de la otra; de modo que cuando $\rho \rightarrow 0$, $s \rightarrow \infty$ y σ toma valores finitos.

Una medida de dispersión asociada a la distribución es la dispersión circular, definida como:

$$\delta = (1 - \rho_2)/(2\rho^2)$$

La cual desempeña un papel importante dentro de la inferencia estadística para la dirección media.

Tomando en consideración que el segundo momento trigonométrico de una distribución es obtenido duplicando los ángulos en módulo 2π para cada valor posible de θ ; de tal modo que para esta nueva distribución la dirección media es precisamente el segundo momento trigonométrico de la distribución original, se plantea entonces que el grado de asimetría dependerá del término:

$$A = \rho_2 \text{sen}(\mu_2 - 2\mu)$$

Considerando que para una distribución simétrica duplicar los ángulos no deshace la simetría; se observa que tanto el doble del primer momento trigonométrico como el segundo momento trigonométrico apuntarán hacia el mismo centro de masa, dicho de otra forma $\mu_2 = 2\mu$. Por lo que el término anterior desaparece si la distribución es simétrica.

La medida de curtosis para una distribución depende de la expresión:

$$K = \rho_2 \cos(\mu_2 - 2\mu)$$

Mediante esta expresión se podrá saber si la distribución en estudio es relativamente plana en la moda o por el contrario tiene un pico pronunciado.

3.3 Distribuciones simétricas

Este conjunto de distribuciones está conformado principalmente por la distribución uniforme, la von Mises, la cardioide, la normal envuelta y la Cauchy envuelta; como su nombre lo indica se caracterizan por ser simétricas; y a excepción de la distribución uniforme todas las demás distribuciones son similares en apariencia, debido a que son

unimodales. Sin embargo los puntos de inflexión para cada una de estas distribuciones diferirán considerablemente, como se observa en la figura 3.1.

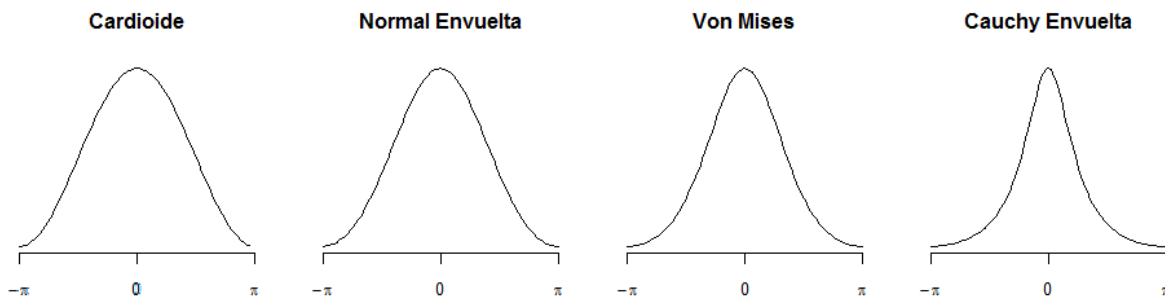


Figura 3.1 Comparación de las distribuciones circulares simétricas y unimodales. Todas estas distribuciones están centradas alrededor de $\mu = 0$, con $\rho = 0.45$.

Una característica importante de señalar es que cada distribución ya sea la distribución von Mises, la cardioide, la normal envuelta o la Cauchy envuelta pueden converger a la distribución uniforme con una elección adecuada de sus parámetros. Y a su vez se podrá tener una aproximación de la distribución von Mises mediante la distribución Cauchy envuelta o mediante la distribución normal envuelta; dependiendo de los parámetros de cada distribución involucrada.

3.3.1 Distribución uniforme

Si todas las direcciones posibles entre 0 y 2π son elegidas con la misma probabilidad, o dicho de otra manera, si la probabilidad es uniforme sobre cualquier sector del círculo; entonces la distribución es llamada uniforme. En consecuencia la función de densidad de probabilidad para esta distribución será constante sobre toda la circunferencia del círculo unitario, como se ve en la figura 3.2. Por lo tanto la función de densidad de una distribución uniforme para una variable aleatoria circular θ se define como:

$$f(\theta) = \frac{1}{2\pi}; \quad \text{para } 0 \leq \theta < 2\pi$$

Su función de distribución está dada por:

$$F(\theta) = \frac{\theta}{2\pi}; \quad \text{para } 0 \leq \theta < 2\pi$$

Los momentos para la distribución uniforme son:

$$\alpha_p = \beta_p = 0; \quad \text{para } p \geq 1$$

En tal caso se tiene que la dispersión circular es:

$$\delta = (1 - \rho_2)/(2\rho^2) = \infty$$

Utilizando su propiedad de simetría, el centro de masa de la distribución coincide exactamente con el origen 0; por lo que $\rho = 0$ y por ende la dirección media no existe. Por otro lado se dice que no existe un punto de concentración alrededor de alguna dirección específica. Todos estos elementos que describen el comportamiento de esta función de probabilidad hacen que la distribución uniforme sea propuesta como hipótesis nula para un conjunto de datos circulares. Algunas de las hipótesis alternativas contra las que se compara esta hipótesis nula son la existencia de la dirección media, la presencia de una o varias modas, entre otras.

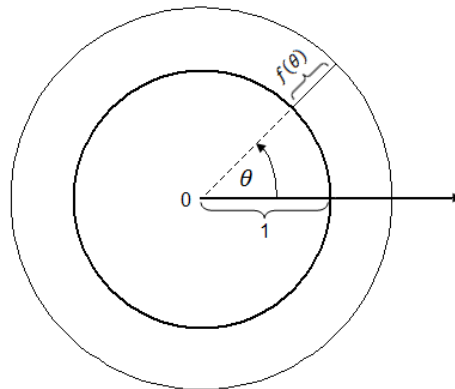


Figura 3.2 La función de densidad de probabilidad para la distribución uniforme.

3.3.2 Distribución von Mises

Este tipo de distribución es el modelo más utilizado para las muestras unimodales de datos circulares, de tal modo que se dice que una variable aleatoria θ sigue una distribución von Mises si su función de densidad de probabilidad está dada por:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)]; \quad \text{para } 0 \leq \theta < 2\pi$$

con los parámetros $\kappa \geq 0$ y $0 \leq \mu < 2\pi$. Donde $I_0(\kappa)$ es la función modificada de Bessel de orden cero especificada como:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos\theta) d\theta = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{\kappa}{2}\right)^{2r}$$

El valor de θ para el cual la función $f(\theta)$ alcanza su máximo y mínimo dependerá directamente del máximo y mínimo de la función coseno. Es por esto que cuando $\theta = \mu$, $f(\theta)$ alcanza su máximo y por consiguiente μ es la moda, lo cual se puede observar en la figura 3.3. La antimoda de la distribución se localiza en $\theta = \mu \pm \pi$, ya que para ese valor de θ la función coseno alcanza su mínimo. Esta distribución cuenta con la propiedad de simetría a causa de la simetría del coseno. Si además se toma en cuenta que la distribución es simétrica con respecto a su moda y que asimismo es unimodal; se dice que μ es también la dirección media.

Para $\kappa = 0$ la distribución von Mises se convierte en una distribución uniforme; mientras que cuando $\kappa \rightarrow \infty$ se tendrá una tendencia a la distribución puntual concentrada en la dirección μ . Conforme κ aumenta, la concentración de la distribución alrededor de la dirección media también aumentará; esto se induce al observar que la razón de $f(\mu)$ (la densidad de probabilidad evaluada en la moda) con respecto a $f(\mu \pm \pi)$ (la densidad de probabilidad evaluada en la antimoda) está dada por $e^{2\kappa}$; por lo que κ es llamado el parámetro de concentración.

En resumen el centro de masa está dado por la dirección media φ_1 , con un ángulo polar asociado μ y una longitud resultante ρ , definida como:

$$\rho = |\varphi_1| = A_1(\kappa)$$

Su función de distribución está dada por:

$$F(\theta) = \frac{1}{2\pi I_0(\kappa)} \int_0^\theta \exp[\kappa \cos(\phi - \mu)] d\phi$$

Sus momentos son:

$$\alpha_p = A_p(\kappa) \cos(p\mu) \quad y \quad \beta_p = A_p(\kappa) \sin(p\mu); \quad \text{para } p \geq 1$$

Y su dispersión circular cuando $\mu = 0$ está dada como:

$$\delta = [\kappa A_1(\kappa)]^{-1}$$

donde la función $A_p(\kappa)$ y la función modificada de Bessel $I_p(\kappa)$ se definen como:

$$A_p(\kappa) = \frac{I_p(\kappa)}{I_0(\kappa)}; \quad \kappa \geq 0, \quad p = 1, 2, \dots$$

$$I_p(\kappa) = \sum_{r=0}^{\infty} \frac{(\kappa/2)^{2r+p}}{[(r+p)!r!]^{-1}}; \quad p = 1, 2, \dots$$

Esta distribución fue introducida por von Mises para estudiar las desviaciones que se daban al medir los pesos atómicos. Actualmente la distribución von Mises es frecuentemente usada en el análisis de datos circulares, y en muchos aspectos es análoga en el círculo a la distribución normal en la recta. Por esta razón Gumbel, Greenwood y Durand también llamaron a esta distribución la distribución normal circular; sin embargo no es muy usual este nombre ya que causa confusión con la distribución normal bivariada con desviación estándar $\sigma_1 = \sigma_2$ y coeficiente de correlación $\rho_{1,2} = 0$; dado que las líneas de contorno de la función de densidad de probabilidad son círculos.

Existe una relación entre la distribución von Mises y la distribución normal bivariada, la cual fue descubierta por Downs en 1976. Para esto se debe considerar un caso especial de la distribución normal bivariada con centro en $(m, 0)$, desviación estándar $\sigma_1 = \sigma_2 = 1$ y coeficiente de correlación $\rho_{12} = 0$. De esta forma las líneas de contorno interceptan al círculo unitario, dando como resultado una distribución circular, en este caso la distribución von Mises. En otras palabras, dado el caso especial de la distribución normal bivariada planteada anteriormente; su distribución condicional bajo la restricción $x^2 + y^2 = 1$ es una distribución von Mises con parámetro de concentración $\kappa = m$.

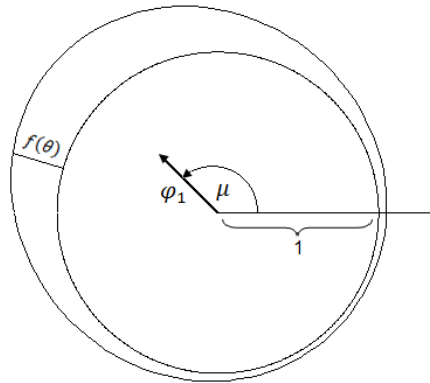


Figura 3.3 Distribución von Mises con centro de masa φ_1 y su ángulo polar asociado $\mu = 3\pi/4$ y $\kappa = 1.2$.

Las distribuciones von Mises, normal envuelta, Cauchy envuelta y la cardioide son marcadamente parecidas, especialmente las dos primeras; por lo que si se desea diferenciar entre la distribución von Mises y la normal envuelta, se necesitará un tamaño de muestra realmente grande. En la práctica se utiliza aquella distribución que sea más conveniente de acuerdo al modelo y al tipo de estudio que se requiera realizar sobre las observaciones. Por ejemplo, la distribución von Mises es empleada con mayor frecuencia debido a que la inferencia estadística es más fácil.

3.3.3 La distribución cardioide

La función de densidad de probabilidad para una variable aleatoria Θ que sigue una distribución cardioide con longitud resultante para la dirección media ρ y ángulo medio μ está dada por:

$$f(\theta) = \frac{1}{2\pi} + \frac{\rho}{\pi} \cos(\theta - \mu); \quad 0 \leq \rho \leq \frac{1}{2}$$

La representación polar de esta densidad es una curva cardioide, de ahí el nombre de esta distribución; aunque también es llamada distribución coseno puesto que es derivada de la función coseno sobre el intervalo $[-\pi/2, \pi/2]$. Alcanza su máximo cuando $\cos(\theta) = 1$, es decir cuando $\theta = \mu$, como en la figura 3.4. Considerando que la distribución es simétrica unimodal; se dice que la dirección media, la dirección mediana y la dirección modal son iguales.

Su función de distribución está dada por:

$$F(\theta) = \frac{\theta}{2\pi} + \frac{\rho}{\pi} \text{sen}(\theta - \mu); \quad \text{para } 0 \leq \theta < 2\pi$$

Los momentos para esta distribución se puntualizan como:

$$\alpha_p = 0; \quad \text{para } p \geq 2$$

$$\beta_p = 0; \quad \text{para } p \geq 1$$

Por lo tanto la dispersión circular δ se reduce a:

$$\delta = 1/(2\rho^2)$$

Cuando $\rho = 0$ la distribución cardioide se reduce a la distribución uniforme; y por tanto una ρ pequeña representará una desviación ligera de la distribución uniforme.

Esta distribución fue presentada por Jeffreys en 1948. Una de sus aplicaciones es la siguiente: considere un disco circular, una bandeja por ejemplo, en la cual se dejan caer algunas canicas mientras es agitada sobre su propio eje. Si la bandeja se encuentra en una posición horizontal, la probabilidad de que una canica se salga de la bandeja se distribuye uniformemente con respecto a cualquier azimut θ . Conforme la bandeja sea inclinada en la dirección $\theta = \mu$, la probabilidad de que alguna canica salga de la bandeja se asemejará a la forma de la distribución cardioide.

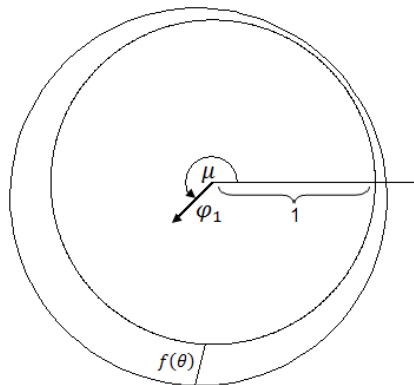


Figura 3.4 La distribución cardioide para $\mu = 5\pi/4$ y $\rho = 0.4$. La dirección media para esta distribución está representada por φ_1 y su ángulo polar asociado por μ .

3.3.4 Distribución normal envuelta

Al envolver la distribución normal ordinaria alrededor del círculo unitario, como se plantea en las distribuciones envueltas, se tiene como resultado a la llamada distribución normal envuelta. Su función de densidad de probabilidad con el ángulo medio μ está dada por:

$$f(\theta) = (2\pi)^{-1/2}\sigma^{-1} \sum_{k=-\infty}^{\infty} \exp [-(\theta + 2\pi k - \mu)^2/2\sigma^2]; \quad 0 \leq \theta < 2\pi, \quad 0 \leq \rho \leq 1$$

donde σ^2 es la varianza de la distribución normal ordinaria utilizada para la distribución normal envuelta.

La longitud de su dirección media es:

$$\rho = \exp(-\sigma^2/2)$$

Los momentos de esta distribución son:

$$\alpha_p = \rho^{p^2} \cos(p\mu) \quad \text{y} \quad \beta_p = \rho^{p^2} \sen(p\mu); \quad \text{para } p \geq 1$$

De ahí que la dispersión circular cuando $\mu = 0$ es:

$$\delta = (1 - \rho^4)/(2\rho^2)$$

Con relación a las características generales de la forma de la distribución normal envuelta, se dice que es simétrica unimodal respecto de μ y que cuenta con dos puntos de inflexión como se observa en la figura 3.5. Además cuando $\sigma \rightarrow \infty$ la distribución normal envuelta tiende a la distribución uniforme, mientras que cuando $\sigma \rightarrow 0$ la distribución tiende a la distribución puntual concentrada en la dirección μ .

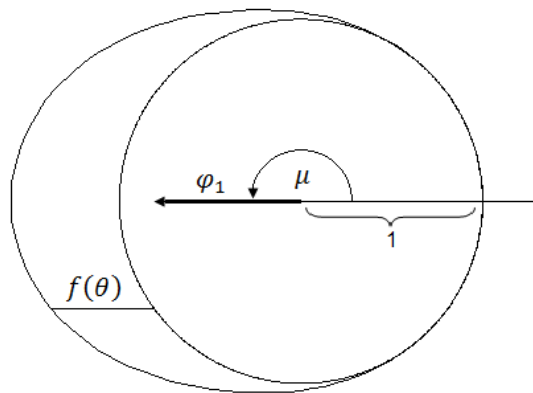


Figura 3.5 La función de densidad de probabilidad para la distribución normal envuelta con $\mu = \pi$ y $\sigma^2 = 0.21$. Su dirección media está representada por φ_1 y su ángulo polar asociado por μ .

Esta distribución ha sido estudiada por Zernkie, Wintner y Lévy, entre otros. Se tiene conocimiento que ya en el año 1917 Schmidt utilizaba la distribución normal envuelta como un modelo para datos geológicos. Hass-Lorentz relacionó a esta

distribución con el movimiento browniano en el círculo, de tal modo que si en una caminata aleatoria a tiempo continuo se cumple lo siguiente: i) una partícula comienza en $\theta = 0$ al tiempo $t = 0$, ii) la partícula se mueve a distancias infinitesimales en períodos de tiempo infinitesimal, y iii) la media es cero y la varianza es $\sigma^2 = ct$ al tiempo t ; entonces la función de densidad de probabilidad de la posición de la partícula al tiempo t está dada por la densidad de una normal envuelta con $\mu = 0$ y $\sigma^2 = ct$.

3.3.5 Distribución Cauchy envuelta

Siguiendo el proceso de obtener una distribución envuelta a partir de una distribución lineal, en este caso la distribución Cauchy; se tiene nuevamente una distribución circular simétrica unimodal. Esta distribución es llamada distribución Cauchy envuelta y su función de densidad de probabilidad es:

$$f(\theta) = \frac{1}{2\pi} \cdot \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}; \quad 0 \leq \theta < 2\pi, \quad 0 \leq \rho \leq 1$$

donde μ es el ángulo medio y ρ es la longitud de la dirección media.

Su función de distribución está dada por:

$$F(\theta) = \frac{1}{2\pi} \cos^{-1} \left(\frac{(1 + \rho^2) \cos(\theta - \mu) - 2\rho}{1 + \rho^2 - 2\rho \cos(\theta - \mu)} \right); \quad 0 \leq \theta < 2\pi$$

Para la distribución Cauchy envuelta los momentos están dados como:

$$\alpha_p = \rho^p \quad y \quad \beta_p = 0; \quad \text{para } p \geq 1$$

De ahí que la dispersión circular es:

$$\delta = (1 - \rho^2)/(2\rho^2)$$

Cuando $\rho \rightarrow 0$ la distribución Cauchy envuelta converge a la distribución uniforme; ahora bien cuando $\rho \rightarrow 1$, la distribución estará concentrada en la dirección μ .

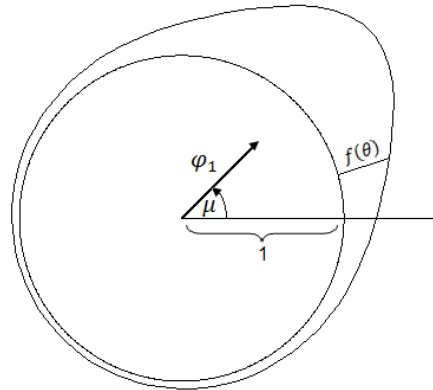


Figura 3.6 La distribución Cauchy envuelta con $\mu = \pi/4$ y $\rho = 0.6$. La dirección media para esta distribución está representada por φ_1 y su ángulo polar asociado por μ .

Esta distribución fue introducida por Lévy en 1939. Una aplicación indirecta de la distribución Cauchy envuelta es en la simulación de datos de una distribución von Mises. También ha sido propuesta como alternativa a la distribución von Mises para modelar datos simétricos unimodales; además la estimación por máxima verosimilitud de μ representa una alternativa robusta a la dirección media muestral.

3.4 Distribuciones asimétricas

Aunque en general la mayor parte de aplicaciones de la estadística circular se encuentran en las distribuciones simétricas, en ocasiones será necesario el uso de distribuciones asimétricas. Una de las familias de distribuciones asimétricas está dada por:

$$f(\theta) = \frac{1}{2\pi} + \frac{k}{2\pi} \text{sen}(\theta + v \text{sen } \theta) \quad (1)$$

donde $0 < |k| \leq 1$ y $0 < |v| \leq 1$ son los parámetros de la función. Además de ser asimétrica esta familia de distribuciones es unimodal, debido a la condición de v . Un ejemplo de esta familia de distribuciones asimétricas está dado por $k = 0.9$ y $v = -0.5$, como se ve en la figura 3.7.

Si se aplica la transformación $\theta = \psi + \pi/2$ y se usa que $\text{sen}(\psi + \pi/2) = \text{cos}(\psi)$, se obtiene que la función de densidad de probabilidad cambia a:

$$f(\theta) = \frac{1}{2\pi} + \frac{k}{2\pi} \text{cos}(\psi + v \text{cos} \psi) \quad (2)$$

De esta forma, esta familia de distribuciones asimétricas puede estar dada por dos senos como en (1) o por dos cosenos como en (2). Es necesario que el manejo de los parámetros para esta familia de distribuciones asimétricas sea cuidadoso; ya que una mala elección de los parámetros puede desembocar en la pérdida de la característica esencial de estas distribuciones, la asimetría. Lo anterior a causa de que si $k = 0$ la densidad para esta familia de distribuciones se convierte en una distribución uniforme; o si $v = 0$ la función de densidad de probabilidad se reduce a la distribución cardioide.

Una segunda familia de distribuciones asimétricas está dada por la función de densidad de probabilidad:

$$f(\theta) = c \cdot \exp[k \text{cos}(\theta - v \text{cos} \theta)]$$

con parámetros k , v y un coeficiente c .

En esta familia de distribuciones asimétricas el parámetro que se restringe es v , ya que para $v = 0$ la asimetría desaparece y se tiene como resultado una distribución von Mises con parámetro de concentración k . Para $v \neq 0$ el coeficiente c es determinado con la segunda propiedad de la función de densidad de probabilidad para una variable aleatoria circular; esto es, se debe satisfacer que la integral de $f(\theta)$ sobre todo su dominio sea igual a 1.

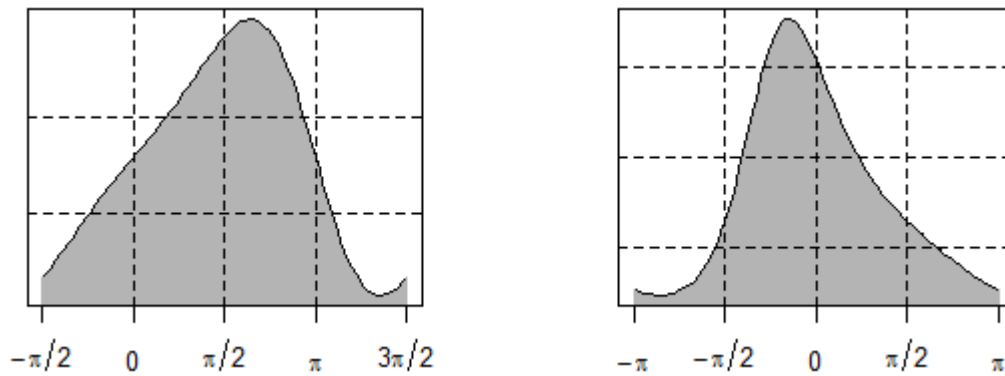


Figura 3.7 Ejemplos de distribuciones asimétricas.

A la izquierda de la figura 3.7, un ejemplo de la primera familia de distribuciones asimétricas con $k = 0.9$ y $v = -0.5$; y a la derecha un ejemplo de la segunda familia con $c = 0.13$, $k = 1$ y $v = 0.52$.

3.5 Simetría en la estadística circular

Al igual que en la simetría para datos lineales, la simetría para datos circulares puede definirse en primera instancia de forma gráfica; de este modo se dice que una distribución circular es simétrica si respecto a una dirección específica la parte de la gráfica de la función de densidad en dirección positiva es el reflejo de la gráfica en dirección negativa.

Como se observa en la figura 3.8 gracias a las líneas punteadas en cada imagen; la distribución von Mises ilustra bien la definición de simetría, ya que lo es independientemente del valor de sus parámetros, en este caso respecto a $\theta = 3\pi/4$ y $\theta = 7\pi/4$. Otra distribución simétrica es la uniforme, la cual presenta dicha propiedad para cualquier dirección dentro del círculo unitario.

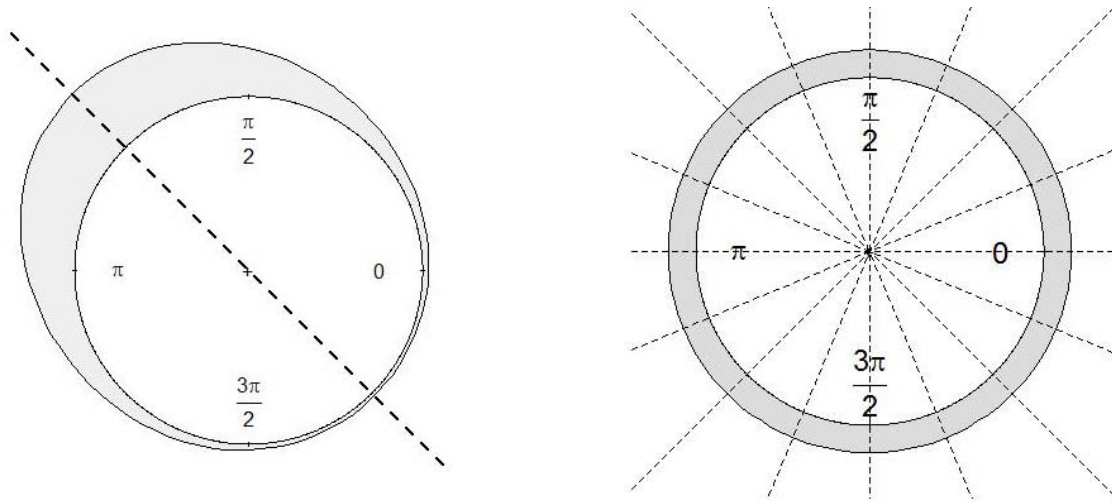


Figura 3.8. La distribuciones von Mises (a la izquierda) y uniforme (a la derecha).

Para establecer en estadística circular un concepto más conciso de simetría; es necesario definir una variable aleatoria circular θ , con función de densidad de probabilidad $f(\theta)$ y función de distribución $F(\theta)$; además de suponer que el rango de θ es $-\pi < \theta \leq \pi$. Así pues, se establece que la distribución de θ es simétrica alrededor de $\theta = 0$, si la distribución es invariante bajo la transformación

$$\theta' = -\theta(\text{mod } 2\pi)$$

es decir bajo la reflexión en la dirección cero. De manera general, si la función de densidad de θ es simétrica alrededor de ϕ se tiene que:

$$f(\theta) = f(2\phi - \theta)$$

Además, si la función de densidad es simétrica alrededor de $\theta = \phi$; entonces es también simétrica alrededor de $\theta = \phi + \pi$.

El supuesto de simetría en el análisis de una variable aleatoria circular es algo esencial; ya que si bien cuando se prueba dicho supuesto no se tienen las herramientas suficientes para hacer una conjetura sobre determinada distribución

sobre la variable aleatoria; el aceptar o rechazar la hipótesis de simetría trae distintos efectos que pueden ayudar a comprender mejor el comportamiento de la variable.

Si bien se encuentran diversas similitudes entre la estadística lineal y la estadística circular con relación al efecto que tiene la simetría sobre la distribución de la variable aleatoria en estudio; es importante notar que existen casos donde no se pueden aplicar las propiedades que tiene la distribución simétrica de una variable aleatoria lineal, tal es el caso del cambio de escala de ordinal a intervalo.

Tomando en cuenta que si la variable aleatoria en estudio tiene asociada una distribución unimodal y el supuesto de simetría es aceptado; entonces las direcciones media, mediana y modal son iguales. Esto trae como consecuencias los siguientes puntos:

1. Cualquier característica o propiedad relacionada con la mediana podrá ser utilizada para la media y la moda.
2. Los momentos centrales seno son cero, es decir $\bar{\beta}_p = 0$ para toda p en los números naturales positivos.

Capítulo 4

Pruebas de simetría para datos circulares

4.1 Introducción

Un problema que ha sido considerado en diferentes estudios es probar la hipótesis de que una distribución es simétrica con respecto a un valor determinado; contra la alternativa, que este valor está desplazado a la derecha (o a la izquierda) o bien, contra la hipótesis alternativa de dos colas. Un problema diferente que también ha recibido atención es probar si la distribución es simétrica con respecto a un valor específico; contra la alternativa, que la distribución no es simétrica; o que en cambio, el parámetro de localización es erróneo.

A lo largo de los años se han propuesto diversas pruebas de simetría para datos lineales; en esencia estas pruebas se dividen con respecto a si el centro de simetría es conocido o no. Dentro de este último grupo de procedimientos, específicamente para probar con el enfoque no paramétrico la hipótesis de simetría alrededor de un centro desconocido; destacan las pruebas de Gupta (1967), Gastwirth (1971), Finch (1977), Davis & Quade (1978) Randles, Fligner, Policello & Wolfe (1980), Antille, Kersting & Zucchini (1982), Bhattacharya, Gastwirth y Wright (1982), Boos (1982), Koziol (1983), Heathcote (1987), Schuster y Barker (1987), y Cabilio & Masaro (1996). Romano (1988, 1989) por ejemplo, considera el uso de pruebas de aleatoriedad y bootstrap para varias hipótesis no paramétricas incluyendo la de simetría.

La simetría en la estadística circular es un concepto más complicado, ya que la simetría no está definida unívocamente; debido a su tamaño compacto e isometrías de rotación y reflexión. De esta forma se vuelven tres, en lugar de dos, los posibles escenarios para las pruebas de simetría. La primera de estas pruebas, simetría cíclica o n -veces simetría, no tiene equivalente en estadística lineal. Jupp & Spurr (1983)

consideraron este problema introduciendo procedimientos basados en la prueba de rangos. La segunda prueba fue examinada por Schach (1969), ésta considera la simetría reflexiva alrededor de un eje específico. Schach deriva resultados a nivel local con pruebas de rangos localmente más potentes, contra otras alternativas de rotación.

La tercera prueba, la cual plantea la simetría reflexiva alrededor de una dirección central desconocida; puede ser analizada de manera informal usando un procedimiento gráfico, el cual evalúa si una distribución unimodal es simétrica reflexiva en relación a la mediana. Otra opción sugerida puede ser basada en una prueba estándar, como la de rangos con signo de Wilcoxon donde el eje de simetría es la mediana de la muestra. Sin embargo esto no es del todo apropiado, puesto que deja de ser no-paramétrica al estimar el centro de simetría; de esta forma el nivel de significancia y la potencia de la prueba son más bajos que sus niveles nominales.

4.2 Análisis exploratorio

Es la fase inicial del análisis estadístico de una muestra aleatoria de datos circulares $\theta_1, \dots, \theta_n$, en la cual algunos de los métodos descriptivos son utilizados para identificar las características cualitativas de la muestra; tales como unimodalidad, multimodalidad y simetría. Esta etapa exploratoria se apoya principalmente en métodos gráficos, como la gráfica de los datos primarios, la cual es particularmente útil en la identificación de la forma de la distribución.

En este contexto, cuando se desea postular la hipótesis nula de simetría respecto de su dirección mediana y además es posible suponer que la muestra se ha extraído de una distribución unimodal; se puede probar dicha hipótesis de manera informal usando un procedimiento gráfico simple. Este procedimiento gráfico consiste en calcular el ángulo mediano $\tilde{\theta}$, y a continuación los valores:

$$z_i = \frac{1}{2}(\theta_i - \tilde{\theta}), \quad -\frac{1}{2}\pi \leq z_i \leq \frac{1}{2}\pi; \quad i = 1, \dots, n$$

Posteriormente se deberán ordenar los valores de forma ascendente, de tal forma que:

$$z_{(1)} \leq \dots \leq z_{(n)}$$

Y finalmente se graficará:

$$\left(\text{sen } z_{(i)}, \text{sen} \left(\frac{1}{2}(z_{(i)} + z_{(n+1-i)}) \right) \right); \quad i = 1, \dots, n$$

En esta prueba gráfica se supone que si la distribución subyacente de la muestra $\theta_1, \dots, \theta_n$ es simétrica respecto a su dirección mediana; los puntos deberán graficarse aproximadamente a lo largo de una banda horizontal cercana al eje de la abscisa. Es decir, que si los puntos distan demasiado de este eje, ya sea que el valor de las ordenadas para cada punto sea muy grande o muy pequeño; se rechazará la hipótesis nula de simetría respecto de su dirección mediana.

Si se cuenta con una muestra grande donde los datos se encuentran muy dispersos, es recomendable hacer una modificación; ya que se pueden perder algunos detalles en los extremos de la gráfica. Esta modificación consiste en obtener z_i de la siguiente forma:

$$z_{(i)} = \frac{1}{4}(\theta_i - \tilde{\theta}), \quad -\frac{1}{4}\pi \leq z_i \leq \frac{1}{4}\pi; \quad i = 1, \dots, n$$

La metodología de la gráfica y la decisión de rechazo seguirán siendo las mismas.

4.3 Pruebas lineales adaptadas para datos circulares

Si los datos analizados se distribuyen a lo largo de un arco reducido; entonces, es posible aplicar cualquiera de las pruebas estándar de simetría alrededor de un centro desconocido. Para esto se aplica una rotación adecuada de los datos y un desenvolvimiento de éstos sobre la línea; sin embargo, si los datos se concentran sobre toda la circunferencia del círculo estas pruebas no pueden ser empleadas, ya que las estadísticas y las distribuciones teóricas supuestas no son aplicables. A pesar de esto, es posible hacer diversas modificaciones de tal modo que los procedimientos estándar sean aplicables para el caso circular.

4.3.1 La prueba de los rangos con signo de Wilcoxon

Cuando se desea probar la hipótesis nula de que la dirección cero es el eje de simetría para una muestra aleatoria de datos circulares $\theta_1, \dots, \theta_n$, contra la hipótesis alternativa de un eje desplazado; es posible utilizar la prueba habitual de los rangos con signo de Wilcoxon, aunque se deberán de hacer ciertas modificaciones de tal forma que pueda ser ajustable para el caso circular.

Una vez que se tienen las observaciones ordenadas $\theta_{(1)}, \dots, \theta_{(n)}$; el primer paso para la realización de la prueba es identificar las observaciones que se encuentran en la parte superior de la circunferencia, para después realizar una transformación a las observaciones restantes, de tal modo que:

$$\theta_i^* = \begin{cases} \theta_{(i)} & i = 1, \dots, s \\ 2\pi - \theta_{(i)} & i = s + 1, \dots, n \end{cases}$$

donde s es el número de observaciones que se localizan en la parte superior de la circunferencia.

Posteriormente se procede a hacer el cálculo de los rangos para la serie $\{\theta_{(1)}, \dots, \theta_{(s)}, \theta_{(s+1)}^*, \dots, \theta_{(n)}^*\}$, de manera que:

$$R_i = \begin{cases} 0 & \text{si } \theta_{(i)} \notin \text{semicírculo superior} \\ \text{rango de } \theta_{(i)} \text{ en la serie} & \text{si } \theta_{(i)} \in \text{semicírculo superior} \end{cases}$$

Si dos o más $\theta_{(i)}$'s son iguales; se asigna el promedio de los rangos correspondientes a cada $\theta_{(i)}$. Los datos con valor igual a la dirección cero deberán eliminarse, y el tamaño de muestra n se reduce proporcionalmente en los cálculos. Entonces la estadística de prueba T está dada por:

$$T = \sum_{i=1}^n R_i$$

La prueba plantea que cada θ_i tiene la misma probabilidad de caer en la parte superior o inferior del semicírculo; más específicamente, la probabilidad de que θ_i pertenezca al semicírculo superior es igual a la probabilidad de que no pertenezca.

Además se utiliza el supuesto de que la distribución de la muestra es simétrica con respecto a la dirección cero, ya que sin el supuesto de simetría es posible que el valor de las R_i 's tienda a ser muy grande o muy pequeño. Es por esto que se rechaza la hipótesis nula de simetría si la estadística de prueba T es demasiado grande o demasiado pequeña.

Los valores críticos para esta prueba dependerán del tamaño de muestra n , y están dados de la siguiente forma:

- a) Si $n \leq 20$, entonces para una prueba al nivel de significancia α , se rechaza la hipótesis nula de simetría si:

$$T > w_{1-\frac{\alpha}{2}} \quad o \quad T < w_{\frac{\alpha}{2}}$$

donde w_α es el valor crítico α para el tamaño de muestra n dado en el Apéndice B2.

- b) Si $n > 20$, se obtiene el nivel de significancia mediante la siguiente aproximación:

$$x_\alpha \cong \frac{T - n(n+1)/4}{[n(n+1)(2n+1)/24]^{1/2}}$$

donde x_α es el cuantil α de una distribución normal estándar dado en el Apéndice B1.

Schach (1969) demostró que la eficiencia de la prueba de Wilcoxon con relación a la mejor prueba para una distribución von Mises con parámetros $\kappa \geq 0$ y $0 \leq \mu < 2\pi$ está dada por:

$$eficiencia(Prueba\ Wilcoxon/Mejor\ Prueba) = \frac{3}{\pi^2} \frac{\{I_0(\kappa)e^{-\kappa} - I_0(2\kappa)\}^2}{\kappa I_0(\kappa)^3 I_1(\kappa)}$$

y que el límite de la eficiencia para la prueba de Wilcoxon cuando $\kappa \rightarrow 0$ es $6/\pi^2$ y $3/\pi$ cuando $\kappa \rightarrow \infty$.

4.3.2 La prueba del signo

Al igual que la prueba de Wilcoxon, la prueba del signo es para probar simetría respecto a la dirección cero. El procedimiento de la prueba del signo, ajustado a una muestra aleatoria de datos circulares $\theta_1, \dots, \theta_n$, consiste en establecer como referencia al semicírculo superior, SS ; de tal modo que cada observación pueda ser distinguida en relación a si pertenece o no a la parte superior de la circunferencia.

Posteriormente se especifica la siguiente función indicadora:

$$I_{SS}(\theta_i) = \begin{cases} 0 & \text{si } \theta_i \notin \text{semicírculo superior} \\ 1 & \text{si } \theta_i \in \text{semicírculo superior} \end{cases}$$

Por lo que la estadística de prueba N está dada por:

$$N = \sum_{i=1}^n I_{SS}(\theta_i) = \text{número de observaciones } \theta_i \in \text{semicírculo superior}$$

La prueba define dos eventos: cuando θ_i pertenece al semicírculo superior, y cuando no pertenece. La idea de la prueba consiste en comparar la probabilidad de estos dos eventos, tomando en cuenta la hipótesis de simetría alrededor de la dirección cero; por tanto se espera que el número de θ_i 's que pertenecen al semicírculo superior se distribuya binomial, y que la probabilidad que tiene θ_i de pertenecer o no a dicho semicírculo sea la misma. Es por esto que se cuenta el número de observaciones que caen dentro del semicírculo superior, y si el número es demasiado grande o demasiado pequeño la hipótesis nula de simetría es rechazada.

La decisión de rechazo está dada de acuerdo al tamaño de muestra n de la siguiente manera:

- a) Para $n \leq 20$, se rechaza la hipótesis nula de simetría al nivel de significancia α si:

$$N \geq n - t \quad \text{ó} \quad N \leq t$$

donde t es el cuantil $\alpha/2$ de una distribución binomial $(n, 1/2)$ dado en el Apéndice B3.

- b) Para $n > 20$ se utilizará la decisión de rechazo anterior, pero ahora calculando t a través del cuantil $\alpha/2$ de una distribución normal estándar $x_{\alpha/2}$, dado en el Apéndice B1.

$$t = \frac{1}{2}(n + x_{\alpha/2}\sqrt{n})$$

Al igual que para la prueba de Wilcoxon, Schach (1969) demostró que la eficiencia de la prueba del signo con relación a la mejor prueba para una distribución von Mises con parámetros $\kappa \geq 0$ y $0 \leq \mu < 2\pi$ está dada por:

$$\text{eficiencia}(\text{Prueba del Signo/Mejor Prueba}) = \frac{(e^\kappa - e^{-\kappa})^2}{\pi^2 I_0(\kappa) \kappa I_1(\kappa)}$$

y que el límite de la eficiencia para la prueba del signo cuando $\kappa \rightarrow 0$ es $8/\pi^2$ y $2/\pi$ cuando $\kappa \rightarrow \infty$. Estos resultados indican que para muestras grandes y un parámetro κ pequeño, la prueba del signo es preferible a la prueba de Wilcoxon.

4.4 Prueba ómnibus para simetría circular reflexiva

En esta prueba se plantea un procedimiento para probar simetría reflexiva para datos circulares, sin importar que los datos estén o no distribuidos a lo largo de todo el círculo unitario; y sin que se pierda la confiabilidad que métodos como la prueba gráfica o las adaptaciones de las pruebas estándares llegan a perder por las peculiares características que se presentan en los datos circulares.

Sea $\theta_1, \dots, \theta_n$ una muestra aleatoria de una distribución circular con dirección media μ y una longitud resultante para la media ρ ; el procedimiento de la prueba de simetría reflexiva para datos circulares está basado en \bar{b}_2 , el componente rectangular del segundo momento trigonométrico centrado de la muestra; el cual es una medida para el sesgo circular propuesto por Batschelet en 1965. El siguiente teorema es utilizado para el desarrollo de la prueba, éste proporciona la distribución asintótica para \bar{b}_2 bajo ciertas condiciones generales.

Teorema 1: Sea $\theta_1, \dots, \theta_n$ variables aleatorias independientes e idénticamente distribuidas de una distribución circular con longitud resultante para la media $\rho \in (0, 1)$. Si $\rho \neq 0$, la dirección media, μ , existe y es única. De esta forma

$$E(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n} \left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4} \right) + O(n^{-3/2}) \quad (1)$$

y

$$\text{Var}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - 2\bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho} \left\{ \bar{\alpha}_3 + \frac{\bar{\alpha}_2(1 - \bar{\alpha}_2)}{\rho} \right\} \right] + O(n^{-3/2}) \quad (2)$$

donde $\bar{\alpha}_p = E\{\cos p(\theta - \mu)\}$ y $\bar{\beta}_p = E\{\sin p(\theta - \mu)\}$ son los p-ésimos momentos coseno y seno alrededor de μ , respectivamente. Cuando $n \rightarrow \infty$.

$$\frac{\bar{b}_2 - E(\bar{b}_2)}{\sqrt{\text{Var}(\bar{b}_2)}} \xrightarrow{D} N(0,1)$$

si $\text{Var}(\bar{b}_2) > 0$, con $E(\bar{b}_2)$ y $\text{Var}(\bar{b}_2)$ dadas por (1) y (2). La prueba del teorema se encuentra en el Apéndice A1.

Para una distribución con función de densidad $f(\theta)$, la cual es simétrica reflexiva alrededor de una dirección central ϕ , se tiene que $f(\theta) = f(2\phi - \theta)$ y claramente $f(\theta)$ es también simétrica alrededor de $\phi + \pi$. Si además $\rho \neq 0$, entonces $\mu = \phi$ y $E(\bar{b}_2) = \bar{\beta}_2 = 0$. En este contexto la estimación de la varianza para $\bar{\beta}_2$ es obtenida con (2) usando $\bar{\beta}_2 = 0$ y para ρ y $\bar{\alpha}_p$ usando sus estimadores \bar{R} y $\bar{\alpha}_p$, respectivamente. Por lo tanto, una distribución asintótica para la prueba de \mathcal{H}_0 puede basarse en la estadística estudentizada:

$$\frac{\bar{b}_2}{\sqrt{\widehat{\text{Var}}(\bar{b}_2)}} \quad (3)$$

donde

$$\widehat{\text{Var}}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{\alpha}_4}{2} - 2\bar{\alpha}_2 + \frac{2\bar{\alpha}_2}{\bar{R}} \left\{ \bar{\alpha}_3 + \frac{\bar{\alpha}_2(1 - \bar{\alpha}_2)}{\bar{R}} \right\} \right]$$

Se compara el valor absoluto de (3) con el cuantil de una distribución normal estándar para rechazar la hipótesis de simetría a favor de algún sesgo alternativo.

Para definir las pruebas de aleatoriedad y bootstrap basadas en la estadística (3), será necesario el proceso de simetrizar una muestra. Este proceso produce una muestra simétrica de tamaño $2n$ mediante el aumento de la muestra original con los valores $2\hat{\phi} - \theta_i$, donde $\hat{\phi}$ es una estimación de la dirección central dada por $\bar{\theta}$.

Una vez que se ha producido una muestra simetrizada, las muestras para aleatoriedad y para bootstrap se pueden generar a partir de ésta, y retroalimentarse de sus marcos muestrales. La muestra de aleatoriedad es formada por el remuestreo de un elemento al azar de cada uno de los n pares $(\theta_i, 2\hat{\phi} - \theta_i)$, mientras que la muestra bootstrap es generada con un remuestreo con reemplazo de n elementos al azar de la muestra simetrizada. Para muestras pequeñas, pueden llevarse a cabo pruebas exactas utilizando una enumeración completa de la distribución de remuestreo. Para muestras grandes, la técnica de reducción de varianza del remuestreo balanceado puede ser empleada si el almacenamiento de grandes arreglos no es un problema. La hipótesis \mathcal{H}_0 es rechazada si el valor de (3) para la muestra original es considerado extremo al compararse con su distribución de remuestreo.

Como se ha definido hasta ahora, alguna versión de la prueba se puede utilizar con los datos de la mayoría de las distribuciones subyacentes, pero no con todas. La condición de que $\rho \neq 0$ excluye la distribución uniforme y algunas distribuciones multimodales. Esto último incluye todas las distribuciones multimodales, que son cíclicamente simétricas o tienen más de un eje de simetría reflexiva, o ambas cosas. Sin embargo, si se cree posible que $\rho = 0$ para alguna distribución multimodal subyacente, y el número de modas de la distribución m se puede establecer previamente, entonces una versión de esta prueba puede ser aplicada después de haber empleado el mecanismo de envolver m -veces el círculo en sí mismo. Este proceso consiste en multiplicar cada valor de los datos por m , produciendo así una muestra de una distribución con dirección media única definida. De hecho, este procedimiento puede ser utilizado con los datos de cualquier distribución multimodal, y si $\rho \neq 0$ sería conveniente aplicar la prueba tanto a los datos originales como a los datos envueltos y luego comparar los resultados. El otro extremo, cuando $\rho = 1$, corresponde al caso cuando la distribución está acumulada en un punto.

4.5 Dos ejemplos prácticos

Con la finalidad de poder entender mejor la metodología de cada prueba y las conclusiones que se obtienen de cada una de éstas, se presenta a continuación dos ejemplos prácticos de muestras aleatorias donde se asumirá la hipótesis nula de simetría sobre la distribución de sus datos.

4.5.1 Truenos

Se tiene una muestra de datos agrupados en intervalos de dos horas con 725 observaciones, las cuales muestran el número de veces en que se escuchó un trueno en un día en Kew (Inglaterra) durante los veranos de 1910 hasta 1935.⁷

Hora	Ángulo	Frecuencia	Hora	Ángulo	Frecuencia
00--	0° --	26	12--	180° --	133
02--	30° --	24	14--	210° --	149
04--	60° --	14	16--	240° --	122
06--	90° --	15	18--	270° --	80
08--	120° --	14	20--	300° --	61
10--	150° --	65	22--	330° --	22
			Total		725

Tabla 4.1 Número de veces en que se registró un trueno en un día en Kew (Inglaterra) durante los veranos de 1910 hasta 1935.

A continuación en la figura 4.1 se presentan dos histogramas de la muestra, de los cuales se puede inducir que los datos provienen de una distribución asimétrica. Visualizando el histograma angular simple y tomando en cuenta que el histograma lineal puede presentar erróneamente la presencia de dos modas, se puede asumir como una primera hipótesis, que los datos de la muestra provienen de una distribución unimodal asimétrica.

⁷ Mardia K. V (1975) *Ibíd.*

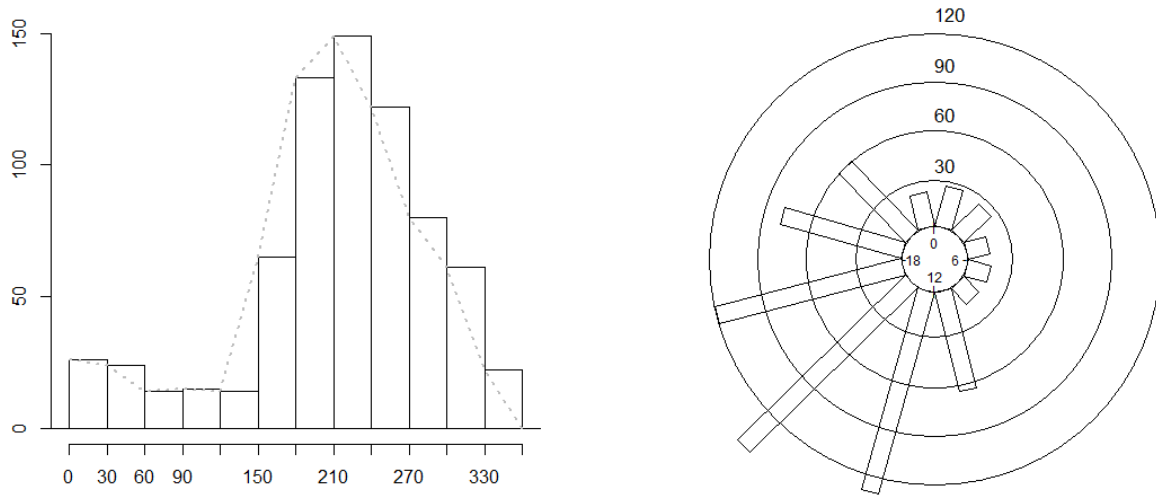


Figura 4.1 Histogramas: lineal (izquierda) y angular simple (derecha)

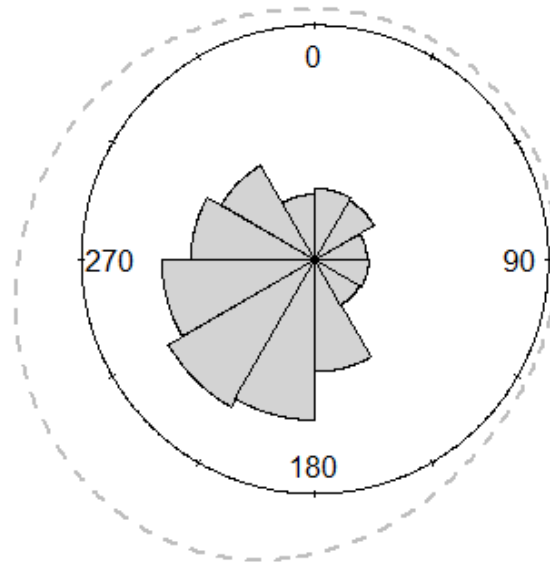


Figura 4.2 Diagrama de rosa con la estimación no paramétrica de su densidad.

Una vez que se ha observado el diagrama de rosa con la estimación no paramétrica de la densidad de la muestra, es posible visualizar con mayor facilidad que la muestra es unimodal y que además presenta un sesgo hacia los 180°, es decir hacia las 12 horas del día.

Prueba gráfica

A pesar de ser un procedimiento simple e informal para los datos de la muestra aleatoria, es posible asumir que la hipótesis nula de simetría respecto de su dirección mediana debe ser rechazada, ya que los puntos no pueden concentrarse en una sola banda horizontal cercana al eje x , como se puede ver en la figura 4.3.

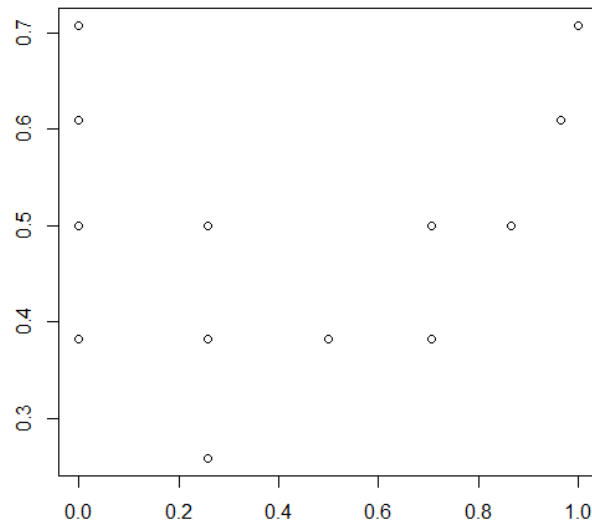


Figura 4.3 Prueba Gráfica.

Pruebas lineales adaptadas para datos circulares

En la aplicación de estas pruebas es necesario poder establecer la dirección cero como el eje de simetría para la muestra aleatoria en estudio; en este caso se toma el ángulo mediano $\tilde{\theta}$ como estimador para el eje de simetría ϕ , y dado que $\tilde{\theta}$ no coincide con la dirección cero se realiza una rotación a los datos de la muestra, de tal modo que el eje de simetría para esta nueva muestra aleatoria ϕ' (estimado como el ángulo mediano de los datos rotados) coincide exactamente con la dirección cero. Una vez que es posible establecer la dirección cero como el eje de simetría para la muestra aleatoria en estudio, se aplican las pruebas a los datos de la muestra y se obtiene lo siguiente:

- En la prueba de Wilcoxon se tiene una significancia de $\alpha = 0.0053$.

- En la prueba del signo para una $\alpha = 0.05$ se obtiene una $N = 414$, lo cual hace que

$$414 > 388.8868 \quad \text{o bien} \quad N \geq n - t$$

Tomando en cuenta los resultados obtenidos en las pruebas anteriores es posible suponer que los datos provienen de una distribución asimétrica ya que el nivel de significancia es muy bajo y los valores de las estadísticas pertenecen a la zona de rechazo.

Prueba ómnibus para simetría circular reflexiva

Se obtuvieron valores muy pequeños del p-value para las pruebas basadas en el componente rectangular \bar{b}_2 ; tanto para la prueba original como para las pruebas de aleatoriedad y bootstrap basadas en una muestra simetrizada de tamaño $2n$ con 3000 simulaciones. Por lo tanto se rechaza la hipótesis nula de simetría reflexiva en la distribución de los datos.

Prueba	P-value	Decisión
Ómnibus	0.00002251	Rechazar
Aleatoriedad	0.00166611	Rechazar
Bootstrap	0.00033322	Rechazar

Tabla 4.3. Decisión de rechazo de las pruebas para simetría circular reflexiva.

Una vez que se han tomado en cuenta los resultados obtenidos de todas las pruebas planteadas en este capítulo, se puede decir con cierta certeza que los datos de la muestra provienen de una distribución circular asimétrica.

4.5.2 Tortugas

Se cuenta con una muestra aleatoria de datos circulares, donde se presenta la orientación que tomaron 76 tortugas después de haber puesto sus huevos⁸. La medición de estas observaciones se hizo en sentido de las manecillas del reloj y se tomó como punto de partida la dirección norte.

⁸ Fisher N. I. (1993) *Ibíd.*

Dirección en grados									
8	9	13	13	14	18	22	27	30	34
38	38	40	44	45	47	48	48	48	48
50	53	56	57	58	58	61	63	64	64
64	65	65	68	70	73	78	78	78	83
83	88	88	88	90	92	92	93	95	96
98	100	103	106	113	118	138	153	153	155
204	215	223	226	237	238	243	244	250	251
257	268	285	319	343	350				

Tabla 4.4 Orientación que tomaron 76 tortugas después de haber puesto sus huevos.

La figura 1.5 del capítulo 1 que muestra el diagrama de tallo y hoja de esta muestra permite tener una idea general del comportamiento de los datos, pero en esencia no permite hacer alguna suposición específica de la distribución de sus datos.

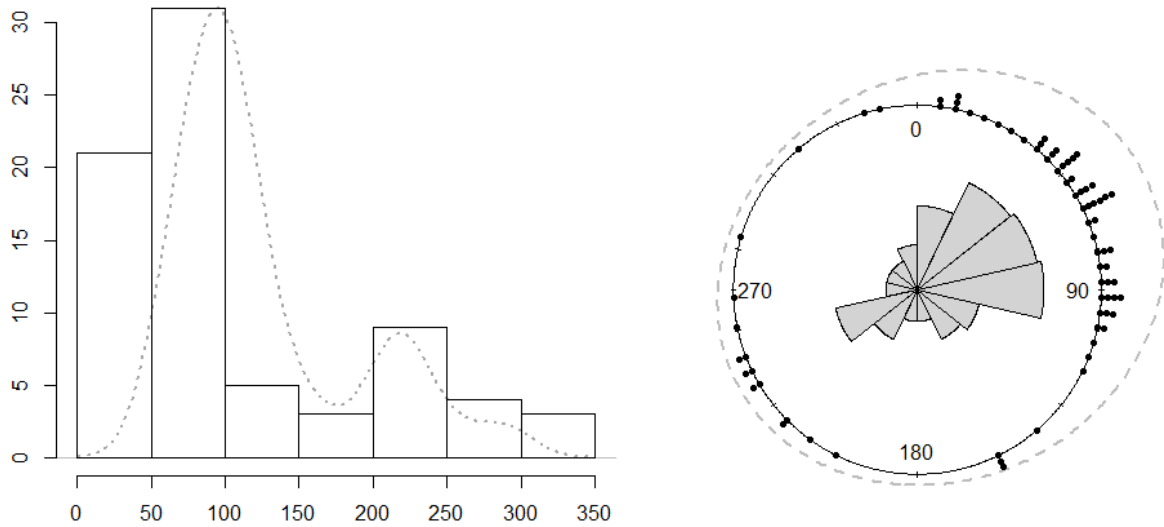


Figura 4.4 Histograma lineal (izquierda) y diagrama de rosa con su densidad estimada (derecha).

Es hasta que se observa el histograma lineal que se puede tener visualmente una idea más concreta de los datos, y así poder suponer que los datos provienen de una distribución bimodal y posiblemente simétrica. En el diagrama de rosa es más fácil visualizar las modas que presenta la distribución y el comportamiento antípoda de éstas; ya con la estimación no paramétrica de densidad es posible ver que la distribución es simétrica y que existe sólo un eje de simetría.

Prueba gráfica

Debido a que la prueba gráfica toma el supuesto de que la muestra se ha extraído de una distribución unimodal, y asumiendo que la distribución de los datos es bimodal; se tendrá que hacer una transformación a los datos utilizando el mecanismo de envolver 2-veces el círculo en sí mismo, es decir, se tendrá que multiplicar cada valor de los datos por 2, produciendo así una muestra con una distribución unimodal. Una vez que la muestra es unimodal, es posible realizar la prueba gráfica y observar que la mayoría de los puntos pueden delimitarse en una banda horizontal paralela al eje x ; a pesar de esto no es posible aceptar el supuesto de simetría en la distribución de los datos de la muestra, ya el comportamiento de los puntos en la gráfica no es del todo uniforme.

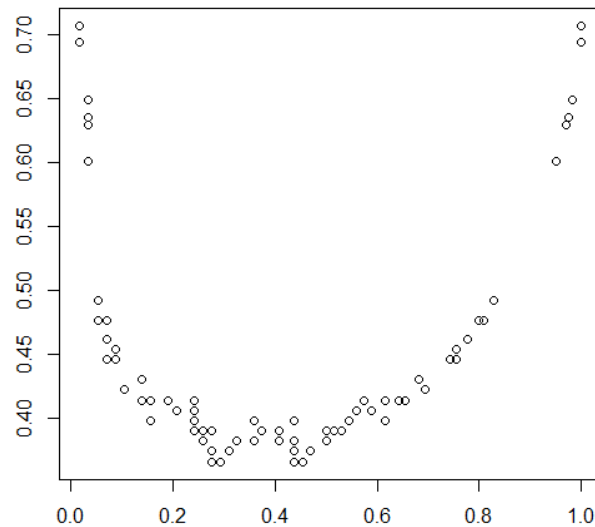


Figura 4.5 Prueba Gráfica.

Pruebas lineales adaptadas para datos circulares

Tomando en cuenta que las pruebas lineales adaptadas para datos circulares no suponen que la muestra en estudio es unimodal, tan solo suponen que la distribución de la muestra es simétrica con respecto a la dirección cero; y dado que la muestra es bimodal antípoda, es decir, que presenta un sólo eje de simetría, es posible realizar esta adaptación de las pruebas lineales a los datos de la muestra. Solamente será necesario hacer una rotación a los datos de la muestra de tal modo que el eje de simetría para la muestra aleatoria (estimado como el ángulo mediano de los datos) coincida exactamente con la dirección cero, ya que al igual que en el ejemplo anterior el ángulo mediano $\tilde{\theta}$ no coincide con la dirección cero. Una vez que la dirección cero es considerada como el eje de simetría se realizan las pruebas sobre los datos de la muestra, y se obtiene que:

- En la prueba de Wilcoxon se tiene una significancia de $\alpha = 0.69156$.
- En la prueba del signo para una $\alpha = 0.05$ se obtiene una $N = 40$, lo cual hace que

$$29.45672 < 40 < 46.54328 \quad \text{ie.} \quad t < N < n - t$$

Es posible no rechazar la hipótesis nula de que la muestra proviene de una distribución simétrica; ya que se obtuvieron valores favorables tanto en la estadística de prueba como en el nivel de significancia.

Prueba ómnibus para simetría circular reflexiva

Al igual que en las adaptaciones de las pruebas lineales, en estas pruebas basadas en el componente rectangular \bar{b}_2 se obtuvieron valores favorables del p-value, como se puede apreciar en la tabla 4.6; es por esto que se aceptará que la muestra aleatoria proviene de una distribución simétrica reflexiva alrededor de la dirección central ϕ .

Prueba	P-value	Decisión
Ómnibus	0.7442	No rechazar
Aleatoriedad	0.7514	No rechazar
Bootstrap	0.7474	No rechazar

Tabla 4.6. Decisión de rechazo de las pruebas para simetría circular reflexiva.

Haciendo una recopilación de los resultados que arrojaron todas las pruebas planteadas, se puede decir que la muestra no presenta evidencia para rechazar la hipótesis nula de simetría en la distribución de los datos; y a su vez todas las pruebas muestran valores que indica que es posible aceptar el supuesto de simetría en la distribución de los datos, por lo que se puede concluir que los datos de la muestra proviene de una distribución simétrica reflexiva alrededor de la dirección central ϕ .

Conclusiones

Cuando se está analizando una muestra, es necesario saber cómo se distribuyen sus datos para posteriormente poder hacer conjeturas sobre éstos, tomando en cuenta dicho comportamiento. El tema "el concepto de simetría en estadística circular" surge por la importancia que tiene el supuesto de simetría sobre la distribución de los datos que pertenecen a una muestra con atributos circulares; ya que aceptar o rechazar el supuesto de simetría en datos circulares trae consigo importantes consecuencias que podrán ser de ayuda o que delimitarán las inferencias que se deseen hacer sobre los datos que se estén analizando.

El propósito de esta tesis fue poder entender de manera detallada las implicaciones que tiene el supuesto de simetría sobre la distribución de los datos de la muestra; una vez teniendo este conocimiento, establecer diversas pruebas para poder aceptar o rechazar esta hipótesis, detallando una a una las características de cada prueba. Y finalmente dar a conocer dos ejemplos prácticos en los cuales se analiza el supuesto de simetría sobre la distribución de sus datos.

Al analizar los dos ejemplos prácticos y comparar los resultados que arroja cada una de las pruebas de simetría, se llega a las siguientes conclusiones:

- I. Es muy útil que antes de hacer cualquier tipo de supuesto sobre los datos, se realicen diversas representaciones gráficas de éstos; ya que para poder aplicar algunas de las pruebas, se deben de cumplir con ciertos supuestos de la distribución de los datos; y al realizar un análisis exploratorio, se puede observar si es necesario realizar una transformación sobre los datos de la muestra, antes de poder aplicar dichas pruebas.
- II. La prueba gráfica es de interpretación ya que la decisión de rechazo sólo establece que los datos deben de caer dentro de una banda horizontal cercana al eje x, pero el qué tan grande podrá ser el ancho de esta banda quedará a juicio de quien está realizando la prueba. Otro problema que presenta esta prueba es cuando la mayoría de los datos cae en una banda cercana al eje x pero se observa que los datos siguen cierta tendencia, es

decir que los datos no están distribuidos uniformemente sobre la banda de control; es entonces cuando dependerá mucho del juicio del que hace la prueba si decide aceptar o rechazar la hipótesis de simetría.

- III. Las pruebas lineales adaptadas para datos circulares resultan ser muy prácticas, ya que si no se cuenta con los paquetes estadísticos adecuados, el procedimiento puede llevarse a cabo en cualquier hoja de cálculo. Tomando en cuenta además la eficiencia de estas pruebas, es posible considerar las pruebas lineales adaptadas para datos circulares como un buen método para analizar el supuesto de simetría.
- IV. La prueba ómnibus es más robusta y el único supuesto que tiene sobre el comportamiento de los datos es que $\rho \neq 0$; lo cual facilita la aplicación de la prueba, ya que puede ser aplicada sin la necesidad de que se cumplan diversos supuestos que requieren las pruebas adaptadas a datos circulares. La realización de la prueba es muy sencilla utilizando la paquetería de R, y gracias a que la decisión de rechazo se basa en el p-value, es más confiable aceptar o rechazar la hipótesis de simetría sobre la distribución de los datos de la muestra.

Apéndice A1

Teorema 1: Sea $\theta_1, \dots, \theta_n$ variables aleatorias independientes e idénticamente distribuidas de una distribución circular con longitud resultante para la media $\rho \in (0, 1)$. Si $\rho \neq 0$, la dirección media, μ , existe y es única. De esta forma

$$E(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n} \left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4} \right) + O(n^{-3/2}) \quad (1)$$

y

$$\text{Var}(\bar{b}_2) = \frac{1}{n} \left[\frac{1 - \bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - 2\bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho} \left\{ \bar{\alpha}_3 + \frac{\bar{\alpha}_2(1 - \bar{\alpha}_2)}{\rho} \right\} \right] + O(n^{-3/2}) \quad (2)$$

donde $\bar{\alpha}_p = E\{\cos p(\theta - \mu)\}$ y $\bar{\beta}_p = E\{\sin p(\theta - \mu)\}$ son los p-ésimos momentos coseno y seno alrededor de μ , respectivamente. Cuando $n \rightarrow \infty$.

$$\frac{\bar{b}_2 - E(\bar{b}_2)}{\sqrt{\text{Var}(\bar{b}_2)}} \xrightarrow{D} N(0,1)$$

donde $\text{Var}(\bar{b}_2) > 0$, con $E(\bar{b}_2)$ y $\text{Var}(\bar{b}_2)$ dadas por (1) y (2).

Demostración

Si $\rho \neq 0$ es posible suponer que después de una rotación se puede obtener $\bar{\theta} = \tan^{-1}(b_1/a_1)$ con probabilidad 1.⁹ De tal modo que

$$\begin{aligned} \bar{b}_2 &= \frac{1}{n} \sum_{i=1}^n \sin 2(\theta_i - \bar{\theta}) = \frac{1}{n} \sum_{i=1}^n (\sin 2\theta_i \cos 2\bar{\theta} - \cos 2\theta_i \sin 2\bar{\theta}) \\ &= \frac{1}{n} \cos 2\bar{\theta} \sum_{i=1}^n \sin 2\theta_i - \frac{1}{n} \sin 2\bar{\theta} \sum_{i=1}^n \cos 2\theta_i = b_2 \cos 2\bar{\theta} - a_2 \sin 2\bar{\theta} \\ &= \frac{(a_1^2 - b_1^2)b_2 - 2a_1b_1a_2}{(a_1^2 + b_1^2)} \end{aligned}$$

⁹ Mardia (1972) Ibídem.

Aplicando el teorema del límite central, la distribución asintótica para los momentos a_1, b_1, a_2 y b_2 es normal con

$$E(a_1) = \alpha_1, E(b_1) = \beta_1, E(a_2) = \alpha_2, E(b_2) = \beta_2,$$

varianzas

$$\begin{aligned} \sigma_{11} = \text{var}(a_1) &= \frac{(1 + \alpha_2 - 2\alpha_1^2)}{2n}, & \sigma_{22} = \text{var}(b_1) &= \frac{(1 - \alpha_2 - 2\beta_1^2)}{2n}, \\ \sigma_{33} = \text{var}(a_2) &= \frac{(1 + \alpha_4 - 2\alpha_2^2)}{2n}, & \sigma_{44} = \text{var}(b_2) &= \frac{(1 - \alpha_4 - 2\beta_2^2)}{2n}, \end{aligned} \quad (4)$$

y covarianzas

$$\begin{aligned} \sigma_{12} = \sigma_{21} = \text{cov}(a_1, b_1) &= \frac{(\beta_2 - 2\alpha_1\beta_1)}{2n}, \\ \sigma_{13} = \sigma_{31} = \text{cov}(a_1, a_2) &= \frac{(\alpha_1 + \alpha_3 - 2\alpha_1\alpha_2)}{2n}, \\ \sigma_{14} = \sigma_{41} = \text{cov}(a_1, b_2) &= \frac{(\beta_1 + \beta_3 - 2\alpha_1\beta_2)}{2n}, \\ \sigma_{23} = \sigma_{32} = \text{cov}(b_1, a_2) &= \frac{(\beta_3 + \beta_1 - 2\beta_1\alpha_2)}{2n}, \\ \sigma_{34} = \sigma_{43} = \text{cov}(a_2, b_2) &= \frac{(\beta_4 - 2\alpha_2\beta_2)}{2n}, \end{aligned} \quad (5)$$

Usando la expansión de Taylor para aproximar la distribución de \bar{b}_2 el segundo momento trigonométrico de la muestra, y utilizando el segundo momento trigonométrico poblacional $\bar{\beta}_2$ tal que

$$\bar{\beta}_2 = \frac{(\alpha_1^2 - \beta_1^2)\beta_2 - 2\alpha_1\beta_1\alpha_2}{(\alpha_1^2 + \beta_1^2)} = \frac{(\alpha_1^2 - \beta_1^2)\beta_2 - 2\alpha_1\beta_1\alpha_2}{\rho^2} \quad (6)$$

A continuación se presentan algunas identidades útiles en las siguientes simplificaciones:

$$\begin{aligned} \bar{\alpha}_2\rho^2 &= (\alpha_1^2 + \beta_1^2)\alpha_2 + 2\alpha_1\beta_1\beta_2, \\ \bar{\alpha}_3\rho^3 &= \alpha_1(\alpha_1^2 - 3\beta_1^2)\alpha_3 + \beta_1(3\alpha_1^2 - \beta_1^2)\beta_3, \end{aligned}$$

(7)

$$\begin{aligned}\bar{\beta}_3 \rho^3 &= \alpha_1(\alpha_1^2 - 3\beta_1^2)\beta_3 - \beta_1(3\alpha_1^2 - \beta_1^2)\alpha_3, \\ \bar{\alpha}_4 \rho^4 &= (\alpha_1^4 - 6\alpha_1^2\beta_1^2 + \beta_1^4)\alpha_4 + 4\alpha_1\beta_1(\alpha_1^2 - \beta_1^2)\beta_4.\end{aligned}$$

Sea $\xi_1 = \alpha_1$, $\xi_2 = \beta_1$, $\xi_3 = \alpha_2$, $\xi_4 = \beta_2$, se define $g = \bar{\beta}_2$; entonces las primeras derivadas de g con respecto a ξ_i pueden ser expresadas como

$$g^{(1)} = \frac{2\beta_1\bar{\alpha}_2}{\rho^2}, g^{(2)} = -\frac{2\alpha_1\bar{\alpha}_2}{\rho^2}, g^{(3)} = -\frac{2\alpha_1\beta_1}{\rho^2}, g^{(4)} = \frac{(\alpha_1^2 - \beta_1^2)}{\rho^2}, \quad (8)$$

y las segundas derivadas parciales son

$$\begin{aligned}g^{(11)} &= \frac{4\beta_1(\beta_1\beta_2 + \alpha_1\alpha_2) - 8\alpha_1\beta_1\bar{\alpha}_2}{\rho^4}, \\ g^{(12)} &= \frac{8\alpha_1\beta_1\beta_2 + 2\alpha_2(\alpha_1^2 - 3\beta_1^2) - 8\beta_1^2\bar{\alpha}_2}{\rho^4}, \\ g^{(13)} &= \frac{2\beta_1(\alpha_1^2 - \beta_1^2)}{\rho^4}, \\ g^{(14)} &= \frac{4\alpha_1\beta_1^2}{\rho^4}, \\ g^{(22)} &= \frac{4\alpha_1(\beta_1\alpha_2 - \alpha_1\beta_2) + 8\alpha_1\beta_1\bar{\alpha}_2}{\rho^4}, \\ g^{(23)} &= -\frac{2\alpha_1(\alpha_1^2 - \beta_1^2)}{\rho^4}, \\ g^{(24)} &= -\frac{4\alpha_1^2\beta_1}{\rho^4},\end{aligned} \quad (9)$$

donde $g^{(i)} = \partial g / \partial \xi_i$ y $g^{(ij)} = \partial^2 g / \partial \xi_i \partial \xi_j$. Las aproximaciones de la media y la varianza de $\bar{\beta}_2$ están dados por

$$E(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 g^{(ij)} \sigma_{ij} + O(n^{-3/2})$$

y

$$\text{Var}(\bar{b}_2) = \sum_{i=1}^4 \sum_{j=1}^4 g^{(i)} g^{(j)} \sigma_{ij} + O(n^{-3/2}).$$

Sustituyendo las derivadas parciales de (8) y (9) y las varianzas y covarianzas de (4) y (5), y haciendo uso de las identidades de (6) y (7), es posible obtener después de algunas simplificaciones algebraicas la media y la varianza de \bar{b}_2 como en (1) y (2).

La distribución asintótica normal de \bar{b}_2 se desprende de la distribución asintótica normal de los momentos a_1, b_1, a_2 y b_2 , considerando que el término en (1) es finito y el supuesto de que el término en (2) es positivo; el teorema queda demostrado.

Apéndice B1

Distribución Normal: x_α es el cuantil α de una distribución normal estándar¹⁰.

x_α	α	x_α	α	x_α	α
-3.7190	0.0001	-0.4677	0.32	0.5244	0.70
-3.2905	0.0005	-0.4399	0.33	0.5534	0.71
-3.0902	0.001	-0.4125	0.34	0.5828	0.72
-2.5758	0.005	-0.3853	0.35	0.6128	0.73
-2.3263	0.01	-0.3585	0.36	0.6433	0.74
-2.1701	0.015	-0.3319	0.37	0.6745	0.75
-2.0537	0.02	-0.3055	0.38	0.7063	0.76
-1.9600	0.025	-0.2793	0.39	0.7388	0.77
-1.8808	0.03	-0.2533	0.40	0.7722	0.78
-1.7507	0.04	-0.2275	0.41	0.8064	0.79
-1.6449	0.05	-0.2019	0.42	0.8416	0.80
-1.5548	0.06	-0.1764	0.43	0.8779	0.81
-1.4758	0.07	-0.1510	0.44	0.9154	0.82
-1.4395	0.075	-0.1257	0.45	0.9542	0.83
-1.4051	0.08	-0.1004	0.46	0.9945	0.84
-1.3408	0.09	-0.0753	0.47	1.0364	0.85
-1.2816	0.10	-0.0502	0.48	1.0803	0.86
-1.2265	0.11	-0.0251	0.49	1.1264	0.87
-1.1750	0.12	0.0000	0.50	1.1750	0.88
-1.1264	0.13	0.0251	0.51	1.2265	0.89
-1.0803	0.14	0.0502	0.52	1.2816	0.90
-1.0364	0.15	0.0753	0.53	1.3408	0.91
-0.9945	0.16	0.1004	0.54	1.4051	0.92
-0.9542	0.17	0.1257	0.55	1.4395	0.925
-0.9154	0.18	0.1510	0.56	1.4758	0.93
-0.8779	0.19	0.1764	0.57	1.5548	0.94
-0.8416	0.20	0.2019	0.58	1.6449	0.95
-0.8064	0.21	0.2275	0.59	1.7507	0.96
-0.7722	0.22	0.2533	0.60	1.8808	0.97
-0.7388	0.23	0.2793	0.61	1.9600	0.975
-0.7063	0.24	0.3055	0.62	2.0537	0.98
-0.6745	0.25	0.3319	0.63	2.1701	0.985
-0.6433	0.26	0.3585	0.64	2.3263	0.99
-0.6128	0.27	0.3853	0.65	2.5758	0.995
-0.5828	0.28	0.4125	0.66	3.0902	0.999
-0.5534	0.29	0.4399	0.67	3.2905	0.9995
-0.5244	0.30	0.4677	0.68	3.7190	0.9999
-0.4959	0.31	0.4959	0.69		

¹⁰ Conover W. J. (1971). Practical Nonparametric Statistics. John Wiley & Sons, Inc.

Apéndice B2

Wilcoxon: w_α es el cuantil α de la estadística de prueba Wilcoxon¹¹.

n	w_{.005}	w_{.01}	w_{.025}	w_{.05}	w_{.10}	w_{.20}	w_{.30}	w_{.40}	w_{.50}	n(n+1)/2
4	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.5	15
6	0	0	1	3	4	6	8	9	10.5	21
7	0	1	2	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.5	45
10	4	6	9	11	15	19	22	25	27.5	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.5	91
14	13	16	22	26	32	39	44	48	52.5	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.5	153
18	28	33	41	48	56	66	73	80	85.5	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	82	91	98	105	210

¹¹ Ídem.

Apéndice B3

Distribución Binomial: t tiene una distribución binomial con parámetros (n, p) ¹².

n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5	n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5
1	0	0.9000	0.8000	0.7000	0.6000	0.5000	8	0	0.4305	0.1678	0.0576	0.0168	0.0039
1	1	1.0000	1.0000	1.0000	1.0000	1.0000	8	1	0.8131	0.5033	0.2553	0.1064	0.0352
2	0	0.8100	0.6400	0.4900	0.3600	0.2500	8	2	0.9619	0.7969	0.5518	0.3154	0.1445
2	1	0.9900	0.9600	0.9100	0.8400	0.7500	8	3	0.9950	0.9437	0.8059	0.5941	0.3633
2	2	1.0000	1.0000	1.0000	1.0000	1.0000	8	4	0.9996	0.9896	0.9420	0.8263	0.6367
3	0	0.7290	0.5120	0.3430	0.2160	0.1250	8	5	1.0000	0.9988	0.9887	0.9502	0.8555
3	1	0.9720	0.8960	0.7840	0.6480	0.5000	8	6	1.0000	0.9999	0.9987	0.9915	0.9648
3	2	0.9990	0.9920	0.9730	0.9360	0.8750	8	7	1.0000	1.0000	0.9999	0.9993	0.9961
3	3	1.0000	1.0000	1.0000	1.0000	1.0000	8	8	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.2401	0.1296	0.0625	9	0	0.3874	0.1342	0.0404	0.0101	0.0020
4	1	0.9477	0.8192	0.6517	0.4752	0.3125	9	1	0.7748	0.4362	0.1960	0.0705	0.0195
4	2	0.9963	0.9728	0.9163	0.8208	0.6875	9	2	0.9470	0.7382	0.4628	0.2318	0.0898
4	3	0.9999	0.9984	0.9919	0.9744	0.9375	9	3	0.9917	0.9144	0.7297	0.4826	0.2539
4	4	1.0000	1.0000	1.0000	1.0000	1.0000	9	4	0.9991	0.9804	0.9012	0.7334	0.5000
5	0	0.5905	0.3277	0.1681	0.0778	0.0313	9	5	0.9999	0.9969	0.9747	0.9006	0.7461
5	1	0.9185	0.7373	0.5282	0.3370	0.1875	9	6	1.0000	0.9997	0.9957	0.9750	0.9102
5	2	0.9914	0.9421	0.8369	0.6826	0.5000	9	7	1.0000	1.0000	0.9996	0.9962	0.9805
5	3	0.9995	0.9933	0.9692	0.9130	0.8125	9	8	1.0000	1.0000	1.0000	0.9997	0.9980
5	4	1.0000	0.9997	0.9976	0.9898	0.9688	9	9	1.0000	1.0000	1.0000	1.0000	1.0000
5	5	1.0000	1.0000	1.0000	1.0000	1.0000	10	0	0.3487	0.1074	0.0282	0.0060	0.0010
6	0	0.5314	0.2621	0.1176	0.0467	0.0156	10	1	0.7361	0.3758	0.1493	0.0464	0.0107
6	1	0.8857	0.6554	0.4202	0.2333	0.1094	10	2	0.9298	0.6778	0.3828	0.1673	0.0547
6	2	0.9842	0.9011	0.7443	0.5443	0.3438	10	3	0.9872	0.8791	0.6496	0.3823	0.1719
6	3	0.9987	0.9830	0.9295	0.8208	0.6563	10	4	0.9984	0.9672	0.8497	0.6331	0.3770
6	4	0.9999	0.9984	0.9891	0.9590	0.8906	10	5	0.9999	0.9936	0.9527	0.8338	0.6230
6	5	1.0000	0.9999	0.9993	0.9959	0.9844	10	6	1.0000	0.9991	0.9894	0.9452	0.8281
6	6	1.0000	1.0000	1.0000	1.0000	1.0000	10	7	1.0000	0.9999	0.9984	0.9877	0.9453
7	0	0.4783	0.2097	0.0824	0.0280	0.0078	10	8	1.0000	1.0000	0.9999	0.9983	0.9893
7	1	0.8503	0.5767	0.3294	0.1586	0.0625	10	9	1.0000	1.0000	1.0000	0.9999	0.9990
7	2	0.9743	0.8520	0.6471	0.4199	0.2266	10	10	1.0000	1.0000	1.0000	1.0000	1.0000
7	3	0.9973	0.9667	0.8740	0.7102	0.5000	11	0	0.3138	0.0859	0.0198	0.0036	0.0005
7	4	0.9998	0.9953	0.9712	0.9037	0.7734	11	1	0.6974	0.3221	0.1130	0.0302	0.0059
7	5	1.0000	0.9996	0.9962	0.9812	0.9375	11	2	0.9104	0.6174	0.3127	0.1189	0.0327
7	6	1.0000	1.0000	0.9998	0.9984	0.9922	11	3	0.9815	0.8389	0.5696	0.2963	0.1133
7	7	1.0000	1.0000	1.0000	1.0000	1.0000	11	4	0.9972	0.9496	0.7897	0.5328	0.2744

n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5	n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5
11	5	0.9997	0.9883	0.9218	0.7535	0.5000	14	1	0.5846	0.1979	0.0475	0.0081	0.0009
11	6	1.0000	0.9980	0.9784	0.9006	0.7256	14	2	0.8416	0.4481	0.1608	0.0398	0.0065
11	7	1.0000	0.9998	0.9957	0.9707	0.8867	14	3	0.9559	0.6982	0.3552	0.1243	0.0287
11	8	1.0000	1.0000	0.9994	0.9941	0.9673	14	4	0.9908	0.8702	0.5842	0.2793	0.0898
11	9	1.0000	1.0000	1.0000	0.9993	0.9941	14	5	0.9985	0.9561	0.7805	0.4859	0.2120
11	10	1.0000	1.0000	1.0000	1.0000	0.9995	14	6	0.9998	0.9884	0.9067	0.6925	0.3953
11	11	1.0000	1.0000	1.0000	1.0000	1.0000	14	7	1.0000	0.9976	0.9685	0.8499	0.6047
12	0	0.2824	0.0687	0.0138	0.0022	0.0002	14	8	1.0000	0.9996	0.9917	0.9417	0.7880
12	1	0.6590	0.2749	0.0850	0.0196	0.0032	14	9	1.0000	1.0000	0.9983	0.9825	0.9102
12	2	0.8891	0.5583	0.2528	0.0834	0.0193	14	10	1.0000	1.0000	0.9998	0.9961	0.9713
12	3	0.9744	0.7946	0.4925	0.2253	0.0730	14	11	1.0000	1.0000	1.0000	0.9994	0.9935
12	4	0.9957	0.9274	0.7237	0.4382	0.1938	14	12	1.0000	1.0000	1.0000	0.9999	0.9991
12	5	0.9995	0.9806	0.8822	0.6652	0.3872	14	13	1.0000	1.0000	1.0000	1.0000	0.9999
12	6	0.9999	0.9961	0.9614	0.8418	0.6128	14	14	1.0000	1.0000	1.0000	1.0000	1.0000
12	7	1.0000	0.9994	0.9905	0.9427	0.8062	15	0	0.2059	0.0352	0.0047	0.0005	0.0000
12	8	1.0000	0.9999	0.9983	0.9847	0.9270	15	1	0.5490	0.1671	0.0353	0.0052	0.0005
12	9	1.0000	1.0000	0.9998	0.9972	0.9807	15	2	0.8159	0.3980	0.1268	0.0271	0.0037
12	10	1.0000	1.0000	1.0000	0.9997	0.9968	15	3	0.9444	0.6482	0.2969	0.0905	0.0176
12	11	1.0000	1.0000	1.0000	1.0000	0.9998	15	4	0.9873	0.8358	0.5155	0.2173	0.0592
12	12	1.0000	1.0000	1.0000	1.0000	1.0000	15	5	0.9978	0.9389	0.7216	0.4032	0.1509
13	0	0.2542	0.0550	0.0097	0.0013	0.0001	15	6	0.9997	0.9819	0.8689	0.6098	0.3036
13	1	0.6213	0.2336	0.0637	0.0126	0.0017	15	7	1.0000	0.9958	0.9500	0.7869	0.5000
13	2	0.8661	0.5017	0.2025	0.0579	0.0112	15	8	1.0000	0.9992	0.9848	0.9050	0.6964
13	3	0.9658	0.7473	0.4206	0.1686	0.0461	15	9	1.0000	0.9999	0.9963	0.9662	0.8491
13	4	0.9935	0.9009	0.6543	0.3530	0.1334	15	10	1.0000	1.0000	0.9993	0.9907	0.9408
13	5	0.9991	0.9700	0.8346	0.5744	0.2905	15	11	1.0000	1.0000	0.9999	0.9981	0.9824
13	6	0.9999	0.9930	0.9376	0.7712	0.5000	15	12	1.0000	1.0000	1.0000	0.9997	0.9963
13	7	1.0000	0.9988	0.9818	0.9023	0.7095	15	13	1.0000	1.0000	1.0000	1.0000	0.9995
13	8	1.0000	0.9998	0.9960	0.9679	0.8666	15	14	1.0000	1.0000	1.0000	1.0000	1.0000
13	9	1.0000	1.0000	0.9993	0.9922	0.9539	15	15	1.0000	1.0000	1.0000	1.0000	1.0000
13	10	1.0000	1.0000	0.9999	0.9987	0.9888							
13	11	1.0000	1.0000	1.0000	0.9999	0.9983							
13	12	1.0000	1.0000	1.0000	1.0000	0.9999							
13	13	1.0000	1.0000	1.0000	1.0000	1.0000							
14	0	0.2288	0.0440	0.0068	0.0008	0.0001							

n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5	n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5
16	0	0.1853	0.0281	0.0033	0.0003	0.0000	17	8	1.0000	0.9974	0.9597	0.8011	0.5000
16	1	0.5147	0.1407	0.0261	0.0033	0.0003	17	9	1.0000	0.9995	0.9873	0.9081	0.6855
16	2	0.7892	0.3518	0.0994	0.0183	0.0021	17	10	1.0000	0.9999	0.9968	0.9652	0.8338
16	3	0.9316	0.5981	0.2459	0.0651	0.0106	17	11	1.0000	1.0000	0.9993	0.9894	0.9283
16	4	0.9830	0.7982	0.4499	0.1666	0.0384	17	12	1.0000	1.0000	0.9999	0.9975	0.9755
16	5	0.9967	0.9183	0.6598	0.3288	0.1051	17	13	1.0000	1.0000	1.0000	0.9995	0.9936
16	6	0.9995	0.9733	0.8247	0.5272	0.2272	17	14	1.0000	1.0000	1.0000	0.9999	0.9988
16	7	0.9999	0.9930	0.9256	0.7161	0.4018	17	15	1.0000	1.0000	1.0000	1.0000	0.9999
16	8	1.0000	0.9985	0.9743	0.8577	0.5982	17	16	1.0000	1.0000	1.0000	1.0000	1.0000
16	9	1.0000	0.9998	0.9929	0.9417	0.7728	17	17	1.0000	1.0000	1.0000	1.0000	1.0000
16	10	1.0000	1.0000	0.9984	0.9809	0.8949	18	0	0.1501	0.0180	0.0016	0.0001	0.0000
16	11	1.0000	1.0000	0.9997	0.9951	0.9616	18	1	0.4503	0.0991	0.0142	0.0013	0.0001
16	12	1.0000	1.0000	1.0000	0.9991	0.9894	18	2	0.7338	0.2713	0.0600	0.0082	0.0007
16	13	1.0000	1.0000	1.0000	0.9999	0.9979	18	3	0.9018	0.5010	0.1646	0.0328	0.0038
16	14	1.0000	1.0000	1.0000	1.0000	0.9997	18	4	0.9718	0.7164	0.3327	0.0942	0.0154
16	15	1.0000	1.0000	1.0000	1.0000	1.0000	18	5	0.9936	0.8671	0.5344	0.2088	0.0481
16	16	1.0000	1.0000	1.0000	1.0000	1.0000	18	6	0.9988	0.9487	0.7217	0.3743	0.1189
17	0	0.1668	0.0225	0.0023	0.0002	0.0000	18	7	0.9998	0.9837	0.8593	0.5634	0.2403
17	1	0.4818	0.1182	0.0193	0.0021	0.0001	18	8	1.0000	0.9957	0.9404	0.7368	0.4073
17	2	0.7618	0.3096	0.0774	0.0123	0.0012	18	9	1.0000	0.9991	0.9790	0.8653	0.5927
17	3	0.9174	0.5489	0.2019	0.0464	0.0064	18	10	1.0000	0.9998	0.9939	0.9424	0.7597
17	4	0.9779	0.7582	0.3887	0.1260	0.0245	18	11	1.0000	1.0000	0.9986	0.9797	0.8811
17	5	0.9953	0.8943	0.5968	0.2639	0.0717	18	12	1.0000	1.0000	0.9997	0.9942	0.9519
17	6	0.9992	0.9623	0.7752	0.4478	0.1662	18	13	1.0000	1.0000	1.0000	0.9987	0.9846
17	7	0.9999	0.9891	0.8954	0.6405	0.3145	18	14	1.0000	1.0000	1.0000	0.9998	0.9962

n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5	n	t	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5
18	15	1.0000	1.0000	1.0000	1.0000	0.9993	20	1	0.3917	0.0692	0.0076	0.0005	0.0000
18	16	1.0000	1.0000	1.0000	1.0000	0.9999	20	2	0.6769	0.2061	0.0355	0.0036	0.0002
18	17	1.0000	1.0000	1.0000	1.0000	1.0000	20	3	0.8670	0.4114	0.1071	0.0160	0.0013
18	18	1.0000	1.0000	1.0000	1.0000	1.0000	20	4	0.9568	0.6296	0.2375	0.0510	0.0059
19	0	0.1351	0.0144	0.0011	0.0001	0.0000	20	5	0.9887	0.8042	0.4164	0.1256	0.0207
19	1	0.4203	0.0829	0.0104	0.0008	0.0000	20	6	0.9976	0.9133	0.6080	0.2500	0.0577
19	2	0.7054	0.2369	0.0462	0.0055	0.0004	20	7	0.9996	0.9679	0.7723	0.4159	0.1316
19	3	0.8850	0.4551	0.1332	0.0230	0.0022	20	8	0.9999	0.9900	0.8867	0.5956	0.2517
19	4	0.9648	0.6733	0.2822	0.0696	0.0096	20	9	1.0000	0.9974	0.9520	0.7553	0.4119
19	5	0.9914	0.8369	0.4739	0.1629	0.0318	20	10	1.0000	0.9994	0.9829	0.8725	0.5881
19	6	0.9983	0.9324	0.6655	0.3081	0.0835	20	11	1.0000	0.9999	0.9949	0.9435	0.7483
19	7	0.9997	0.9767	0.8180	0.4878	0.1796	20	12	1.0000	1.0000	0.9987	0.9790	0.8684
19	8	1.0000	0.9933	0.9161	0.6675	0.3238	20	13	1.0000	1.0000	0.9997	0.9935	0.9423
19	9	1.0000	0.9984	0.9674	0.8139	0.5000	20	14	1.0000	1.0000	1.0000	0.9984	0.9793
19	10	1.0000	0.9997	0.9895	0.9115	0.6762	20	15	1.0000	1.0000	1.0000	0.9997	0.9941
19	11	1.0000	1.0000	0.9972	0.9648	0.8204	20	16	1.0000	1.0000	1.0000	1.0000	0.9987
19	12	1.0000	1.0000	0.9994	0.9884	0.9165	20	17	1.0000	1.0000	1.0000	1.0000	0.9998
19	13	1.0000	1.0000	0.9999	0.9969	0.9682	20	18	1.0000	1.0000	1.0000	1.0000	1.0000
19	14	1.0000	1.0000	1.0000	0.9994	0.9904	20	19	1.0000	1.0000	1.0000	1.0000	1.0000
19	15	1.0000	1.0000	1.0000	0.9999	0.9978	20	20	1.0000	1.0000	1.0000	1.0000	1.0000
19	16	1.0000	1.0000	1.0000	1.0000	0.9996							
19	17	1.0000	1.0000	1.0000	1.0000	1.0000							
19	18	1.0000	1.0000	1.0000	1.0000	1.0000							
19	19	1.0000	1.0000	1.0000	1.0000	1.0000							
20	0	0.1216	0.0115	0.0008	0.0000	0.0000							

¹² Desarrollo propio con Microsoft Office Excel 7.

Apéndice C1

Programas en R: Son los códigos utilizados para la realización de las pruebas de simetría dentro de la paquetería de R.

— Ejemplo de los truenos

```
library(circular)

datosG<-read.table("C:/codigo/thunder.txt") ###Datos Agrupados
datosA=0 ###Datos Desagrupados en Grados
h=1
for(i in 1:length(datosG$V1))
  {for(j in 1:(datosG$V1[i]))
    {datosA[h]=15*(2*i-1)
     h=h+1
    }
  }
datos<-circular(datosA,units = "degrees",zero=1.570796,rotation="clock")
#####Mediana#####
d=0
n=length(datosA)
for(i in 1:360) d[i]=180-sum(abs(180-abs(datosA-i)))/n
median=which(d==min(d)) ###Mediana como valor circular #median(datos)

#####
##### PRUEBA GRAFICA #####
#####
Z=acos(cos((datosA-median)*pi/180))/2
aux=0
for(j in 1:(n-1))
  {for(i in 1:(n-j))
    {if(Z[i]>Z[i+1])
      {aux=Z[i]
       Z[i]=Z[i+1]
       Z[i+1]=aux
      }
    }
  }
x=sin(Z)
y=Z
for(i in 1:n)
  {y[i]=sin((Z[i]+Z[n+1-i])/2)
  }
```

```

plot(x,y)

#####
##### WILCOXON #####
#####

#####D's#####
D=acos(cos((datosA-median)*pi/180))
D=round(D*180/pi)
#####L(semicirculo)#####
Linf=median
Lsup=median+180
if(median>=180)
  {Linf=median-180
  Lsup=median
  }
#####R(rangos)#####
B=D
R=c(0*1:n)
m=n-length(which(D==0))
while(m>0)
  {for(i in 1:n)
    {rep=length(which(D==max(B)))
    if(D[i]==max(B))
      {if(rep==1)
        {R[i]=m
        m=m-1
        }
      if(rep>1)
        {R[i]=m-(rep-1)/2
        aux=length(which(B==max(B)))
        if(aux==1) m=m-rep
        }
      B[i]=0
    }
  }
}
for(i in 1:n)
  {aux=1
  if(datosA[i]>=Linf) B[i]=1
  if(datosA[i]>Lsup) aux=0
  B[i]=B[i]*aux
  }
R=R*B
#####estadistico W#####
W=sum(R)

```

```

m=n-length(which(D==0))
if(m>=16) wp=(W-m*(m+1)/4+0.5)/sqrt(m*(m+1)*(2*m+1)/24)

sig=pnorm(wp)
sig

#####
##### SIGNO #####
#####

#####L(semicirculo)#####
Linf=median
Lsup=median+180
if(median>=180)
{Linf=median-180
Lsup=median
}
#####estadistico N#####
B=c(0*1:n)
aux=1
for(i in 1:n)
{aux=1
if(datosA[i]>=Linf) B[i]=1
if(datosA[i]>Lsup) aux=0
B[i]=B[i]*aux
}
N=sum(B)
#####decision de rechazo#####
t=(n+qnorm(.025)*sqrt(n))/2
aux=c(1,1)
if(N>=n-t) aux[1]=0
if(N<=t) aux[2]=0
if(aux[1]*aux[2]==1) print("aceptar")

#####
##### General #####
#####
RefSim<-function(muestra)
{n=length(muestra)
R=rho.circular(muestra)
M2=trigonometric.moment(muestra,p=2,center=TRUE)
M3=trigonometric.moment(muestra,p=3,center=TRUE)
M4=trigonometric.moment(muestra,p=4,center=TRUE)
B2=M2$sin
A2=M2$cos
A3=M3$cos
}

```

```

A4=M4$cos
var=((1-A4)/2-2*A2+(2*A2/R)*(A3+A2*(1-A2)/R))/n
Z=abs(B2/sqrt(var))
return(Z)
}
Z=RefSim(datos)
pval=2*pnorm(Z,mean=0,sd=1,lower=FALSE)
pval

```

```

LSSim<-function(muestra,L)
{n=length(muestra)
Z=RefSim(muestra)
M=mean(muestra)
N=1
for(l in 2:(L+1))
{I=rbinom(n,1,0.5)
muestraA=muestra*I
muestraS=(2*M-muestra)*(I-1)
muestraLS=muestraA+muestraS
Z[l]=RefSim(muestraLS)
if(Z[l]>=Z[1]) N=N+1
}
pval=N/(L+1)
return(pval)
}

```

```

L=3000
pval=LSSim(datos,L)
pval

```

```

BootSim<-function(muestra,B)
{n=length(muestra)
Z=RefSim(muestra)
M=mean(muestra)
muestraS=2*M-muestra
muestra2n=c(muestra,muestraS)
N=1
for(b in 2:(B+1))
{muestraBoot=sample(muestra2n,size=n,replace=TRUE)
Z[b]=RefSim(muestraBoot)
if(Z[b]>=Z[1]) N=N+1
}
pval=N/(B+1)
return(pval)
}

```

```

B=3000
pval=BootSim(datos,B)
pval

```

— Ejemplo de las tortugas

```

library(circular)

datos<-circular(read.table("C:/codigo/t3.txt"),units = "degrees",zero=1.570796,rotation="clock")
datosA<-read.table("C:/codigo/t3.txt")
#####Mediana#####
d=0
n=length(datosA$V1)
for(i in 1:360) d[i]=180-sum(abs(180-abs(datosA-i)))/n
median=which(d==min(d)) ####Mediana como valor circular #median(datos)

#####
##### PRUEBA GRAFICA #####
#####
Z=acos(cos((datosA-median)*pi/180))/2
aux=0
Z=Z$V1
for(j in 1:(n-1))
  {for(i in 1:(n-j))
    {if(Z[i]>Z[i+1])
      {aux=Z[i]
        Z[i]=Z[i+1]
        Z[i+1]=aux
      }
    }
  }
x=sin(Z)
y=Z
for(i in 1:n)
  {y[i]=sin((Z[i]+Z[n+1-i])/2)
  }
plot(x,y)

#####
##### WILCOXON #####
#####
datosA=datosA-median
for(i in 1:n)
  {if(datosA$V1[i]<0) datosA$V1[i]=datosA$V1[i]+360
    if(datosA$V1[i]==360) datosA$V1[i]=0
  }

```

```

}
d=0
n=length(datosA$V1)
for(i in 1:360) d[i]=180-sum(abs(180-abs(datosA-i)))/n
median=which(d==min(d))
if(median==360) median=0
#####conjugado#####
D=acos(cos((datosA$V1)*pi/180))
D=round(D*180/pi)
#####L(semicirculo)#####
Linf=0
Lsup=180
#####R(rangos)#####
B=D
R=c(0*1:n)
m=n-length(which(D==0))
while(m>0)
  {for(i in 1:n)
    {rep=length(which(D==max(B)))
      if(D[i]==max(B))
        {if(rep==1)
          {R[i]=m
            m=m-1
          }
          if(rep>1)
            {R[i]=m-(rep-1)/2
              aux=length(which(B==max(B)))
              if(aux==1) m=m-rep
            }
          B[i]=0
        }
      }
    }
  }
for(i in 1:n)
  {aux=1
    if(datosA$V1[i]>=Linf) B[i]=1
    if(datosA$V1[i]>Lsup) aux=0
    B[i]=B[i]*aux
  }
R=R*B
#####estadistico W#####
T=sum(R)
m=n-length(which(D==0))
if(m>20) wp=(T-m*(m+1)/4)/sqrt(m*(m+1)*(2*m+1)/24)
sig=pnorm(wp)
sig

```

```

#####
##### SIGNO #####
#####
datosA=datosA-median
for(i in 1:n)
  {if(datosA$V1[i]<0) datosA$V1[i]=datosA$V1[i]+360
  if(datosA$V1[i]==360) datosA$V1[i]=0
  }
#####L(semicirculo)#####
Linf=0
Lsup=180
#####estadistico N#####
B=c(0*1:n)
aux=1
for(i in 1:n)
  {aux=1
  if(datosA$V1[i]>=Linf) B[i]=1
  if(datosA$V1[i]>Lsup) aux=0
  B[i]=B[i]*aux
  }
N=sum(B)
#####decision de rechazo#####
t=(n+qnorm(.025)*sqrt(n))/2
aux=c(1,1)
if(N>=n-t) aux[1]=0
if(N<=t) aux[2]=0
if(aux[1]*aux[2]==1) print("aceptar")

#####
##### General #####
#####
RefSim<-function(muestra)
  {n=length(muestra)
  R=rho.circular(muestra)
  M2=trigonometric.moment(muestra,p=2,center=TRUE)
  M3=trigonometric.moment(muestra,p=3,center=TRUE)
  M4=trigonometric.moment(muestra,p=4,center=TRUE)
  B2=M2$sin
  A2=M2$cos
  A3=M3$cos
  A4=M4$cos
  var=((1-A4)/2-2*A2+(2*A2/R)*(A3+A2*(1-A2)/R))/n
  Z=abs(B2/sqrt(var))
  return(Z)
  }
Z=RefSim(datos)

```

```
pval=2*pnorm(Z,mean=0,sd=1,lower=FALSE)
pval
```

```
LSSim<-function(muestra,L)
  {n=length(muestra)
  Z=RefSim(muestra)
  M=mean(muestra)
  N=1
  for(l in 2:(L+1))
    {I=rbinom(n,1,0.5)
    muestraA=muestra*I
    muestraS=(2*M-muestra)*(I-1)
    muestraLS=muestraA+muestraS
    Z[l]=RefSim(muestraLS)
    if(Z[l]>=Z[1]) N=N+1
    }
  pval=N/(L+1)
  return(pval)
}
```

```
L=3000
pval=LSSim(datos,L)
pval
```

```
BootSim<-function(muestra,B)
  {n=length(muestra)
  Z=RefSim(muestra)
  M=mean(muestra)
  muestraS=2*M-muestra
  muestra2n=c(muestra,muestraS)
  N=1
  for(b in 2:(B+1))
    {muestraBoot=sample(muestra2n,size=n,replace=TRUE)
    Z[b]=RefSim(muestraBoot)
    if(Z[b]>=Z[1]) N=N+1
    }
  pval=N/(B+1)
  return(pval)
}
```

```
B=3000
pval=BootSim(datos,B)
pval
```


Bibliografía

- Batschelet E. (1981). *Circular Statistics In Biology*. London: Academic Press.
- Conover W. J. (1971). *Practical Nonparametric Statistics*. John Wiley & Sons, Inc.
- Conover W. J. (1999). *Practical Nonparametric Statistics*. Third Edition. John Wiley & Sons, Inc.
- Fisher N. I. (1993). *Statistical Analysis Of Circular Data*. Cambridge University Press.
- Gupta M. K. (1967). *An Asymptotically Nonparametric Test Of Symmetry*. The Annals of Mathematical Statistics, 38, 849-866.
- György D. (2007). *Symmetry : Cultural-Historical And Ontological Aspects Of Science-Arts Relations : The Natural And Man-Made World In An Interdisciplinary Approach*. Basel : Birkhäuser.
- Jammalamadaka S. R. (2001). *Topics In Circular Statistics*. World Scientific.
- Mardia K. V. & Jupp P. E. (2000). *Directional Statistics*. John Wiley & Sons, LTD.
- Mardia K. V (1975). *Statistics Of Directional Data*. London: Academic Press.
- Pewsey A. (2002). *Testing circular symmetry*. The Canadian Journal of Statistics, Vol. 30 No. 4, p. 591-600.
- Schach S. (1969). *Nonparametric Symmetry Tests For Circular Distributions*. Biometrika, 56,3, p. 571-577.