



UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Departamento de Matemática

Tesis de Licenciatura

Estimadores de mínima distancia para datos circulares

Mercedes Fernández Sau

Directora: Dra. Daniela Rodríguez

21 de Agosto de 2012

# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Datos Circulares</b>	<b>4</b>
2.1. Medidas de Posición . . . . .	5
2.2. Medidas de Concentración y Dispersión . . . . .	8
2.3. Modelos Circulares . . . . .	9
2.3.1. Distribución Uniforme . . . . .	11
2.3.2. Distribución Von Mises . . . . .	11
2.3.3. Distribución Cardioide . . . . .	12
2.3.4. Distribución Normal Proyectada . . . . .	14
2.3.5. Distribución Wrapped . . . . .	15
2.4. Estimación por máxima verosimilitud . . . . .	16
<b>3. Estimación no paramétrica de la densidad circular</b>	<b>21</b>
3.1. Histograma . . . . .	21
3.2. Histograma móvil . . . . .	22
3.3. Estimación por núcleos de la densidad para datos lineales . . . . .	24
3.4. Estimación por núcleos de la densidad para datos circulares . . . . .	25
<b>4. Estimadores de mínima distancia para datos circulares</b>	<b>29</b>
4.1. Consistencia . . . . .	30
4.2. Estudio de Monte Carlo . . . . .	31
4.3. Resultados Numéricos . . . . .	36
4.4. Conclusiones . . . . .	40

# Capítulo 1

## Introducción

En diversos campos surgen problemas estadísticos donde las variables aleatorias involucradas, en lugar de tomar valores en todo  $\mathbb{R}^d$ , toman valores en un espacio de dimensión menor, por ejemplo en una determinada superficie. Tal es el caso en donde, por la particularidad del problema, las variables se encuentran concentradas en la esfera unidad. Este tipo de datos recibe el nombre de datos direccionales. Los datos circulares constituyen el caso más simple de esta categoría y están asociados directamente a ángulos en el plano. Si, en cambio, consideramos direcciones en el espacio, a este tipo de variables se las conoce como datos esféricos. Los mismos se encuentran comúnmente en biología, geología, medicina, meteorología, ecología, oceanografía y muchas otras áreas. Los ejemplos típicos incluyen direcciones de vuelo de aves, movimientos direccionales de animales en respuesta a ciertos estímulos, o las direcciones del viento y de las corrientes marinas.

Los datos direccionales tienen una serie de características que los hacen distintos de los datos lineales y por tanto, el análisis direccional es sustancialmente diferente del análisis estadístico “*lineal*” estándar ya que la propia naturaleza de los datos obliga a replantear aspectos tan básicos como la medición de distancias o la inclusión de su naturaleza periódica. El análisis de los datos direccionales dio lugar a una serie de nuevos problemas estadísticos que propulsaron, en los últimos 20 años, el desarrollo de métodos específicos. Varios modelos paramétricos y no paramétricos han sido ampliamente discutidos en este contexto. Para una revisión sobre este tema ver Mardia y Jupp [10], Batschelet [3], Watson [15], Fisher [5], Rao [13], He [7], Hall *et al.* [6], Bai *et al.* [2], entre otros.

Una práctica muy común en estadística es suponer que los datos han sido generados por un mecanismo aleatorio y que éste puede ser representado por un miembro de una familia paramétrica de funciones de distribución. La característica más sobresaliente de estos modelos matemáticos es la suposición de que el mecanismo aleatorio que generó las observaciones es totalmente conocido a excepción de un parámetro.

Naturalmente, el principal problema en estos casos es la estimación del parámetro desconocido con buenas propiedades estadísticas como sesgo pequeño o nulo y varianza pequeña. Un método que, en general, satisface estos requerimientos es el método de máxima verosimilitud.

Sin embargo, en algunos casos el estimador de máxima verosimilitud no resulta adecuado. Un ejemplo sencillo donde falla dicho método consiste en considerar una muestra distribuída como una mezcla de dos poblaciones normales. Éste y otros ejemplos no tan sencillos son expuestos por Le Cam [9].

Una alternativa a la estimación por máxima verosimilitud la constituyen los estimadores de mínima distancia que a su vez, bajo condiciones bastante amplias, tienen propiedades de robustez frente a desvíos del modelo paramétrico considerado. Algunas referencias clásicas sobre estimadores de mínima distancia pueden encontrarse en el trabajo seminal de Wolfowitz [16] y en el trabajo de Parr y Schucany [11], mientras que una versión suavizada fue estudiada por Cao *et al.* [4], siempre para datos en espacios euclídeos.

En esta tesis, definiremos estimadores de mínima distancia para datos direccionales bajo el supuesto de que las variables poseen una densidad perteneciente a una familia paramétrica cuyos parámetros se desconocen. La propuesta es considerar una distancia entre un estimador no paramétrico de la densidad para datos direccionales y la densidad paramétrica correspondiente. Aquellos parámetros que hagan mínima esta distancia es lo que definiremos como el estimador de los parámetros desconocidos. Para el estimador definido evaluaremos su comportamiento a través de un estudio de simulación y estudiaremos su consistencia.

El trabajo se organiza en tres capítulos. El Capítulo 2 contiene un resumen sobre conceptos definidos en variables aleatorias que toman valores en el círculo unitario, así como también algunos modelos circulares importantes. En el Capítulo 3, daremos un breve resumen de los estimadores no paramétricos de la función de densidad de variables aleatorias reales y exhibiremos su extensión al caso de variables aleatorias circulares. Por último, en el Capítulo 4 presentaremos los estimadores de mínima distancia para el caso de datos reales y estudiaremos una propuesta para datos circulares finalizando nuestro trabajo con un estudio de simulación que permita evaluar el comportamiento de la misma.



# Capítulo 2

## Datos Circulares

Los datos circulares aparecen en muchos contextos y se obtienen de diversas formas. Las principales corresponden a los instrumentos de medición circular clásicos: la brújula y el reloj. Entre las observaciones típicas medidas por la brújula se encuentran las direcciones del viento o las direcciones del vuelo migratorio de las aves, como así también, la dirección del polo magnético de la tierra. En el caso del reloj podrían ser de interés, por ejemplo, los horarios más frecuentes de llegada de los pacientes a la guardia de un hospital.

Las direcciones en el plano pueden representarse a través de vectores unitarios (i.e. direcciones) o como puntos en el círculo unidad. También podríamos pensar una dirección como un ángulo o bien como un número complejo unitario. En estos dos últimos casos es preciso fijar una dirección inicial y un sentido de rotación para evitar ambigüedades en la representación. De esta manera, dado un ángulo, un punto en el plano puede representarse como,

$$\mathbf{x} = (\cos \theta, \sin \theta)^T$$
$$z = e^{i\theta} = \cos \theta + i \sin \theta$$

Como en el caso de observaciones correspondientes a puntos sobre la recta real, una primera forma de obtener algún tipo de información sobre ellos es a través de un gráfico. Para el caso de datos circulares, el gráfico más simple es un *circular raw data plot*, en el cual cada observación está representada por un punto en el círculo unidad. La Figura 2.1 ilustra este método para el siguiente ejemplo:

**Ejemplo 2.0.1.** *Se hace girar una ruleta 9 veces y se registran las posiciones en las que se detuvo. Las mismas, medidas en grados, resultaron ser: 43°, 45°, 52°, 61°, 75°, 88°, 88°, 279° y 357°. El circular raw data plot de la Figura 2.1 sugiere la tendencia de los datos hacia una dirección particular.*

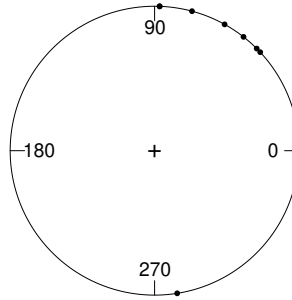


Figura 2.1: Circular plot para los datos del Ejemplo 2.0.1

Una vez obtenido el gráfico será de utilidad resumir la información mediante medidas estadísticas descriptivas. Resulta tentador *cortar* el círculo en algún punto y hacer uso de las medidas estadísticas conocidas para el caso lineal. El problema que trae tratar a los datos circulares de esta manera es que las medidas estadísticas resultantes van a depender fuertemente del punto en que hayamos decidido cortar al círculo. Como un ejemplo de esta situación consideremos una muestra de sólo 2 datos, formada por los ángulos  $1^\circ$  y  $359^\circ$ . Si cortamos el círculo en el  $0^\circ$  obtendremos una media igual a  $180^\circ$  y un desvío estándar de  $179^\circ$ . En cambio, si cortamos el círculo en  $180^\circ$ , es decir consideramos valores en el intervalo  $(-180^\circ, 180^\circ)$ , tendremos una media igual a  $0^\circ$  y un desvío estándar de  $1^\circ$ . Surge entonces la necesidad de definir medidas estadísticas que sean independientes de la posición que hayamos elegido como inicial.

En este capítulo, en las Secciones 2.1 y 2.2, presentaremos una descripción de las medidas de posición, concentración y dispersión para datos circulares. En la Sección 2.3, nos centraremos en las principales distribuciones circulares y finalmente, en la Sección 2.4, haremos referencia a la utilización del método de estimación de parámetros desconocidos por máxima verosimilitud en el caso de este tipo de distribuciones.

## 2.1. Medidas de Posición

Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  vectores unitarios que se corresponden con los ángulos  $\theta_j$ ,  $j=1, \dots, n$ . Se define la *dirección media*  $\bar{\theta}$  de  $\theta_1, \dots, \theta_n$  como la dirección de la resultante  $\mathbf{x}_1 + \dots + \mathbf{x}_n$ , que coincide con la dirección del centro de masa  $\bar{\mathbf{x}}$  de  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Dado que las coordenadas cartesianas de  $\mathbf{x}_j$  son  $(\cos \theta_j, \sin \theta_j)$  para  $j=1, \dots, n$ , se tiene que las coordenadas del centro de masa serán  $(\bar{C}, \bar{S})$ , donde

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos \theta_j, \quad \bar{S} = \frac{1}{n} \sum_{j=1}^n \sin \theta_j$$

Si definimos la *longitud media resultante*  $\bar{R}$  como

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}$$

resulta que  $\bar{\theta}$  es solución de las ecuaciones

$$\bar{C} = \bar{R} \cos \bar{\theta}, \quad \bar{S} = \bar{R} \sin \bar{\theta} \quad (2.1)$$

Esto es válido siempre y cuando  $\bar{R} > 0$ . En el caso en que  $\bar{R} = 0$ ,  $\bar{\theta}$  no está definida.

Teniendo en cuenta estas nuevas definiciones, queda claro que en el contexto de datos circulares  $\bar{\theta}$  *no* es lo mismo que  $(\theta_1 + \dots + \theta_n)/n$ , y que  $\bar{\theta}$  constituye una medida más apropiada que  $(\theta_1 + \dots + \theta_n)/n$  ya que no depende de la elección de la dirección inicial.

**Ejemplo 2.1.1.** *Para los datos del Ejemplo 2.0.1 se tiene que  $\bar{C} = 0.447$  y  $\bar{S} = 0.553$ , por lo tanto, la dirección media  $\bar{\theta}$  y la longitud media resultante son:*

$$\bar{\theta} = 51^\circ \text{ y } \bar{R} = 0.711$$

La Figura 2.2 muestra los resultados obtenidos para dicho conjunto de datos.

Consideremos ahora el efecto de las rotaciones sobre la dirección media. Supongamos que se elige una nueva dirección inicial que forma un ángulo  $\alpha$  con la dirección original. Entonces, ahora, los ángulos correspondientes a los datos en el nuevo sistema de referencia serán

$$\theta'_j = \theta_j - \alpha, \quad j = 1, \dots, n.$$

donde

$$\bar{C}' = \frac{1}{n} \sum_{j=1}^n \cos \theta'_j, \quad \bar{S}' = \frac{1}{n} \sum_{j=1}^n \sin \theta'_j.$$

Si reescribimos  $\bar{C}'$  y  $\bar{S}'$ , obtenemos lo siguiente:

$$\bar{C}' = \frac{1}{n} \sum_{j=1}^n \cos \theta'_j = \frac{1}{n} \sum_{j=1}^n \cos(\theta_j - \alpha) = \frac{1}{n} \sum_{j=1}^n [\cos \theta_j \cos \alpha + \sin \theta_j \sin \alpha] = \bar{C} \cos \alpha + \bar{S} \sin \alpha,$$



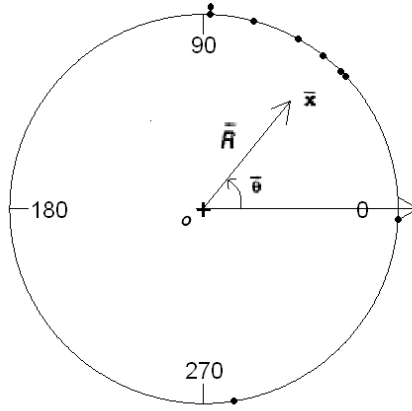


Figura 2.2: Dirección media  $\bar{\theta}$  y longitud media resultante  $\bar{R}$  para los datos del Ejemplo 2.1.1

$$\bar{S}' = \frac{1}{n} \sum_{j=1}^n \sin \theta'_j = \frac{1}{n} \sum_{j=1}^n \sin(\theta_j - \alpha) = \frac{1}{n} \sum_{j=1}^n [\sin \theta_j \cos \alpha - \cos \theta_j \sin \alpha] = \bar{S} \cos \alpha - \bar{C} \sin \alpha.$$

Usando la definición de longitud media resultante  $\bar{R}'$  para el nuevo sistema de referencia tenemos que

$$\begin{aligned} \bar{R}' &= (\bar{C}'^2 + \bar{S}'^2)^{1/2} = [(\bar{S} \cos \alpha - \bar{C} \sin \alpha)^2 + (\bar{C} \cos \alpha + \bar{S} \sin \alpha)^2]^{1/2} \\ &= [\bar{S}^2 \cos^2 \alpha + \bar{C}^2 \sin^2 \alpha - 2\bar{S} \cos \alpha \bar{C} \sin \alpha + \bar{C}^2 \cos^2 \alpha + \bar{S}^2 \sin^2 \alpha + 2\bar{C} \cos \alpha \bar{S} \sin \alpha]^{1/2} \\ &= [(\bar{S}^2 + \bar{C}^2)(\cos^2 \alpha + \sin^2 \alpha)]^{1/2} = [(\bar{S}^2 + \bar{C}^2)]^{1/2} = \bar{R}. \end{aligned}$$

Luego,

$$\bar{R}' = \bar{R}. \quad (2.2)$$

Si las coordenadas polares de  $(\bar{C}', \bar{S}')$  son  $(\bar{R}', \bar{\theta}')$ , tenemos que

$$\bar{C}' = \bar{R}' \cos \bar{\theta}', \quad \bar{S}' = \bar{R}' \sin \bar{\theta}' \quad (2.3)$$

Por otro lado, si reemplazamos las ecuaciones (2.1) en las expresiones obtenidas anteriormente para  $\bar{C}'$  y  $\bar{S}'$  nos queda que

$$\begin{aligned}\bar{C}' &= \bar{R} \cos \bar{\theta} \cos \alpha + \bar{R} \sin \bar{\theta} \sin \alpha = \bar{R} \cos(\bar{\theta} - \alpha) \\ \bar{S}' &= \bar{R} \sin \bar{\theta} \cos \alpha - \bar{R} \cos \bar{\theta} \sin \alpha = \bar{R} \sin(\bar{\theta} - \alpha).\end{aligned}$$

y por lo tanto de (2.2) y (2.3) se deduce que

$$\bar{\theta}' = \bar{\theta} - \alpha.$$

De esta forma, la dirección media de  $\theta_1 - \alpha, \dots, \theta_n - \alpha$  es  $\bar{\theta} - \alpha$ , es decir, la dirección media de la muestra es equivariante por rotaciones. Esta propiedad es análoga a la equivariancia por traslación de la media de un conjunto de datos lineales. La importancia de haber definido en el círculo (como en la recta) una medida que cumpla con esta propiedad, es lograr un acuerdo acerca de la ubicación de la media muestral, independientemente del sistema de referencia escogido.

Para ciertos casos (por ejemplo, la estimación robusta) es útil tener una versión de la mediana muestral adaptada al caso de datos circulares. Se define la *dirección mediana muestral*  $\tilde{\theta}$  de  $\theta_1, \dots, \theta_n$  como un ángulo  $\phi$  que verifique:

- la mitad de los datos caen dentro del arco  $[\phi, \phi + \pi]$  y,
- la mayoría de los datos están más cerca de  $\phi$  que de  $\phi + \pi$

Cuando el tamaño de la muestra es un número impar, la mediana muestral coincide con uno de los datos. Cuando es par, es conveniente tomar como mediana muestral el punto medio entre dos datos adyacentes apropiados.

**Ejemplo 2.1.2.** *Volvamos al ejemplo de la ruleta. Considerando las medidas obtenidas, se puede calcular la dirección de la mediana muestral, que resulta,  $\tilde{\theta} = 52^\circ$ , valor muy cercano a la dirección media  $\bar{\theta} = 51^\circ$  que calculamos en el Ejemplo 2.1.1*

## 2.2. Medidas de Concentración y Dispersión

En la Sección 2.1 definimos la longitud media resultante  $\bar{R}$  como la longitud del centro de masa  $\bar{\mathbf{x}}$ , dada por

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}.$$

Dado que  $\mathbf{x}_1, \dots, \mathbf{x}_n$  son vectores unitarios, se tiene que

$$0 \leq \bar{R} \leq 1.$$

Si las direcciones  $\theta_1, \dots, \theta_n$  están estrechamente agrupadas, el valor de  $\bar{R}$  será muy cercano a 1. Por otro lado, si  $\theta_1, \dots, \theta_n$  están muy dispersos,  $\bar{R}$  tendrá un valor cercano a 0. Por lo tanto,  $\bar{R}$  es una medida de *concentración* del conjunto de datos. La ecuación (2.2) muestra que  $\bar{R}$  es invariante por rotaciones.

Si bien la longitud media resultante  $\bar{R}$  es una medida útil, muchas veces se utilizan otras medidas de dispersión que resultan análogas a las utilizadas para datos en la recta real. La más simple de estas medidas es la *varianza muestral circular* definida por

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

que también satisface

$$0 \leq V \leq 1.$$

También podemos definir la *desviación circular standard* como

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

Notar que  $v$  toma valores en  $[0, \infty]$  mientras que  $V$  lo hace en  $[0, 1]$ .

Una manera de medir la distancia entre dos ángulos  $\theta$  y  $\xi$  es considerar

$$1 - \cos(\theta - \xi).$$

Luego, una forma de medir la dispersión de ángulos  $\theta_1, \dots, \theta_n$  con respecto a un  $\alpha$  dado es con

$$D(\alpha) = \frac{1}{n} \sum_{i=1}^n [1 - \cos(\theta_i - \alpha)]$$

y resulta que  $D(\bar{\theta}) = V$ .

## 2.3. Modelos Circulares

Una forma de especificar una distribución en el círculo unidad es por medio de su función de distribución. Suponemos que han sido escogidas una dirección y orientación

iniciales. Luego, podemos considerar un ángulo aleatorio  $\Theta$  cuya *función de distribución*  $F$  es una función definida en toda la recta real dada por

$$F(x) = P(0 < \Theta \leq x) \quad 0 \leq x \leq 2\pi,$$

y

$$F(x + 2\pi) - F(x) = 1 \quad -\infty < x < \infty.$$

Esta última ecuación afirma que todo arco de longitud  $2\pi$  en el círculo unidad tiene probabilidad 1. Este arco es exactamente el perímetro de todo el círculo unidad.

Para  $\alpha \leq \beta \leq \alpha + 2\pi$ ,

$$P(\alpha < \Theta \leq \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} dF(\theta). \quad (2.4)$$

La función de distribución es continua a derecha. A diferencia de las funciones de distribución en la recta se tiene que

$$\lim_{x \rightarrow \infty} F(x) = \infty, \quad \lim_{x \rightarrow -\infty} F(x) = -\infty.$$

Por definición,

$$F(0) = 0, \quad F(2\pi) = 1.$$

Observemos que, aunque la función  $F$  depende de la elección de la dirección inicial, la ecuación (2.4) muestra que  $F(\beta) - F(\alpha)$  es independiente de dicha elección. Luego, cambiar la dirección inicial sólo agrega una constante a  $F$ .

Si la función de distribución  $F$  es absolutamente continua,  $F$  tiene una función de densidad  $f$  dada por

$$\int_{\alpha}^{\beta} f(\theta) d\theta = F(\beta) - F(\alpha), \quad -\infty < \alpha \leq \beta < \infty.$$

Una función  $f$  es la función de densidad de una distribución absolutamente continua si y sólo si

1.  $f(\theta) \geq 0$  en casi todo punto de  $(-\infty, \infty)$ ,
2.  $f(\theta + 2\pi) = f(\theta)$  en casi todo punto de  $(-\infty, \infty)$ ,
3.  $\int_0^{2\pi} f(\theta) d\theta = 1$ .

En lo que sigue describiremos algunas de las familias más importantes de distribuciones en el círculo.

### 2.3.1. Distribución Uniforme

La *distribución uniforme* es la distribución más básica en el círculo y a menudo se utiliza como modelo nulo. Esta es la única distribución en el círculo que es invariante bajo rotaciones y reflexiones. Su función de densidad es

$$f(\theta) = \frac{1}{2\pi}.$$

Luego, para  $\alpha \leq \beta \leq \alpha + 2\pi$

$$P(\alpha < \theta \leq \beta) = \frac{\beta - \alpha}{2\pi},$$

es decir, la probabilidad es proporcional a la longitud de arco.

### 2.3.2. Distribución Von Mises

Desde el punto de vista de la inferencia estadística, una de las distribuciones de probabilidad circulares más utilizada es la *distribución Von Mises*  $VM(\mu, \kappa)$ , análoga a las distribuciones normales en la recta. Su función de densidad está dada por

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)},$$

donde  $I_0$  denota la función de Bessel modificada de primer tipo y orden 0, definida por

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \theta} d\theta.$$

El parámetro  $\mu \in [0, 2\pi)$  es la dirección media y el parámetro  $\kappa \geq 0$  recibe el nombre de *parámetro de concentración*. Esta distribución es unimodal y simétrica con respecto a  $\theta = \mu$ . La moda se encuentra en  $\theta = \mu$  y la antimoda en  $\theta = \mu + \pi$ . El cociente entre el valor de la densidad en la moda y el valor de la densidad en la antimoda es  $e^{2\kappa}$ , así que cuanto mayor sea el valor de  $\kappa$ , mayor será la concentración alrededor de la moda.

En la Figura 2.3 se muestran los gráficos de varias distribuciones Von Mises para distintos valores del parámetro  $\kappa$ .

### Mezcla de Von Mises

Al igual que en el caso de datos lineales, las mezclas finitas de distribuciones Von Mises,  $VM(\mu_i, \kappa_i)$ , con parámetros de mezcla  $p_i$ ,  $i = 1, \dots, n$ , proporcionan una clase mucho más rica de modelos circulares.

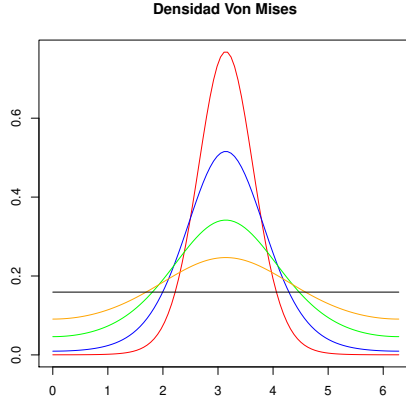


Figura 2.3: Densidad Von Mises  $VM(\pi, \kappa)$  para  $\kappa = 4$  (rojo),  $\kappa = 2$  (azul),  $\kappa = 1$  (verde),  $\kappa = 0.5$  (naranja),  $\kappa = 0$  (negro). Para el valor  $\kappa = 0$  esta distribución coincide con la uniforme.

Su función de densidad es

$$f(\theta) = \sum_{i=1}^n p_i \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(\theta - \mu_i)}$$

donde  $p_i \geq 0 \forall i$  y  $p_1 + \dots + p_n = 1$ .

Utilizaremos la notación  $VM(\mu_1, \mu_2, \kappa_1, \kappa_2, p)$  para denotar la mezcla de dos Von Mises,  $VM(\mu_1, \kappa_1)$  y  $VM(\mu_2, \kappa_2)$ , con proporciones de mezcla  $p_1 = p$  y  $p_2 = 1 - p$ .

En la Figura 2.4 se muestran los gráficos de algunas mezclas de distribuciones Von Mises.

### 2.3.3. Distribución Cardioide

La perturbación de la densidad uniforme por la función coseno da lugar a una distribución llamada *cardioide*  $C(\mu, \rho)$ , cuya función de densidad está dada por

$$f(\theta) = \frac{1}{2\pi} [1 + 2\rho \cos(\theta - \mu)], \quad |\rho| < \frac{1}{2}.$$

La longitud media resultante de una distribución  $C(\mu, \rho)$  es  $\rho$  y si  $\rho > 0$ , la dirección media es  $\mu$ . Esta es una distribución simétrica y unimodal con moda en  $\mu$  cuando  $\rho > 0$ . Si  $\rho = 0$ , la distribución cardioide se reduce a la uniforme.

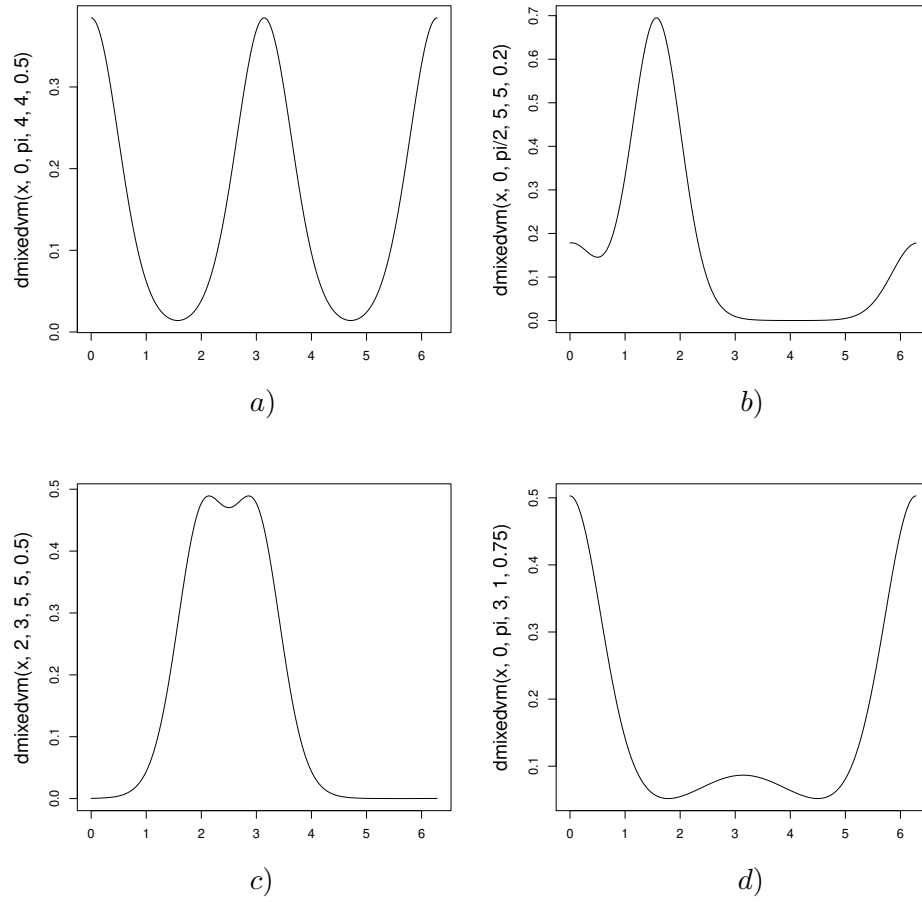


Figura 2.4: Mezclas de distribuciones Von Mises: a)  $VM(0, \pi, 4, 4, 0.5)$ , b)  $VM(0, \pi/2, 5, 5, 0.2)$ , c)  $VM(2, 3, 5, 5, 0.5)$  y d)  $VM(0, \pi, 3, 1, 0.75)$

En la Figura 2.5 se muestran los gráficos de varias distribuciones Cardioide para distintos valores del parámetro  $\rho$ .

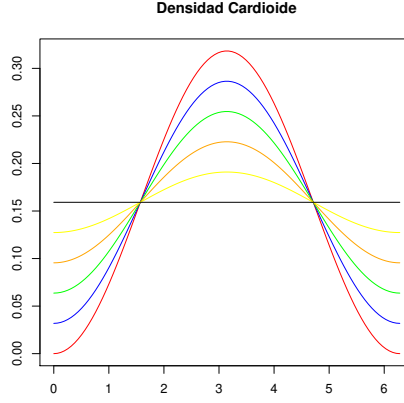


Figura 2.5: Densidad Cardioide  $C(\pi, \rho)$ , para  $\rho = 0.5$  (rojo),  $\rho = 0.4$  (azul),  $\rho = 0.3$  (verde),  $\rho = 0.2$  (naranja),  $\rho = 0.1$  (amarillo),  $\rho = 0$  (negro). Este último valor de  $\rho$  da lugar a la distribución uniforme.

### 2.3.4. Distribución Normal Proyectada

Una manera de obtener distribuciones en el círculo es proyectando radialmente distribuciones en el plano. Sea  $\mathbf{x}$  un vector aleatorio bidimensional tal que  $P(\mathbf{x} = \mathbf{0}) = 0$ . Entonces  $\|\mathbf{x}\|^{-1}\mathbf{x}$  es un punto aleatorio sobre el círculo unitario. Un ejemplo interesante es cuando el vector  $\mathbf{x}$  tiene una distribución normal bivariada  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , en cuyo caso se dice que  $\|\mathbf{x}\|^{-1}\mathbf{x}$  tiene una *distribución normal proyectada*  $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Su función de densidad está dada por

$$f(\theta; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\phi_2(\theta; \mathbf{0}, \boldsymbol{\Sigma}) + |\boldsymbol{\Sigma}|^{-1/2} D(\theta) \Phi(D(\theta)) \phi_1(|\boldsymbol{\Sigma}|^{-1/2} (\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1/2} \boldsymbol{\mu} \wedge \mathbf{x})}{\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}}$$

donde

- $\phi_2(\cdot; \mathbf{0}, \boldsymbol{\Sigma})$  denota la función de densidad de  $N_2(\mathbf{0}, \boldsymbol{\Sigma})$ ,
- $\phi_1$  y  $\Phi$  se refieren a la función de densidad y de distribución de  $N(0, 1)$ , respectivamente,



- $\mathbf{x} = (\cos \theta, \sin \theta)^T$ ,
- $D(\theta) = \frac{\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}}{(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})^{1/2}}$ ,
- $\boldsymbol{\mu} \wedge \mathbf{x} = \mu_1 \sin \theta - \mu_2 \cos \theta$ , con  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$

### 2.3.5. Distribución Wrapped

Dada una distribución cualquiera en la recta real, es posible *envolverla* alrededor de la circunferencia del círculo unitario. Es decir, si  $X$  es una variable aleatoria en la recta, la correspondiente variable aleatoria  $X_\omega$  de la *distribución wrapped* viene dada por

$$X_\omega = X \pmod{2\pi}.$$

Si el círculo es identificado con el conjunto de números complejos de módulo 1, entonces la función que *envuelve* se puede escribir como

$$X \mapsto e^{iX}.$$

Si  $X$  tiene función de distribución  $F$ , la función de distribución  $F_\omega$  de  $X_\omega$  está dada por

$$F_\omega(\theta) = \sum_{k=-\infty}^{\infty} [F(\theta + 2\pi k) - F(2\pi k)], \quad 0 \leq \theta \leq 2\pi$$

En particular, si la variable aleatoria  $X$  tiene función de densidad  $f$ , la correspondiente función de densidad  $f_\omega$  de  $X_\omega$  es

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

### Distribución Wrapped Normal

Como ejemplo de este último tipo de distribuciones podemos considerar la *distribución wrapped normal*  $WN(\mu, \rho)$ , que se obtiene *envolviendo* en el círculo una distribución  $N(\mu, \sigma^2)$ , donde

$$\sigma^2 = -2 \log \rho$$

es decir,

$$\rho = e^{-\sigma^2/2}.$$

Resulta entonces, que la función de densidad de una distribución  $WN(\mu, \rho)$  es

$$\phi_\omega(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-\frac{(\theta-\mu+2\pi k)^2}{2\sigma^2}}.$$

## 2.4. Estimación por máxima verosimilitud

Supongamos que se observa un vector muestra  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  discreto o continuo cuya función de densidad discreta o continua pertenece a una familia  $p(\mathbf{X}, \lambda)$ ,  $\lambda \in \Lambda$  y se quiere estimar  $\lambda$ . Diremos que  $\hat{\lambda}(\mathbf{X})$  es un *estimador de máxima verosimilitud* de  $\lambda$  si se cumple que

$$p(\mathbf{X}, \hat{\lambda}(\mathbf{X})) = \max_{\lambda \in \Lambda} p(\mathbf{X}, \lambda).$$

Por lo tanto, para hallar  $\hat{\lambda}(\mathbf{X})$  se debe maximizar la llamada *función de verosimilitud* dada por

$$L(\lambda; \mathbf{X}) = p(\mathbf{X}, \lambda) = p(X_1, X_2, \dots, X_n, \lambda) = \prod_{j=1}^n p(X_j, \lambda).$$

Cuando el soporte de  $p(\mathbf{X}, \lambda)$  no depende de  $\lambda$ , como la función  $\ln(x)$  es monótona creciente, maximizar  $L(\lambda; \mathbf{X})$  resulta equivalente a maximizar  $\ln L(\lambda; \mathbf{X})$ , que gracias a las propiedades del logaritmo lleva a una expresión más sencilla para hacer cálculos, esto es

$$\ln L(\lambda; \mathbf{X}) = \ln p(\mathbf{X}, \lambda) = \sum_{j=1}^n \ln p(X_j, \lambda).$$

Veamos, por ejemplo el caso de una distribución  $VM(\mu, \kappa)$ . La función de verosimilitud viene dada por

$$L(\mu, \kappa; \boldsymbol{\theta}) = \prod_{j=1}^n \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta_j - \mu)} = \frac{1}{(2\pi I_0(\kappa))^n} e^{\sum_{j=1}^n \kappa \cos(\theta_j - \mu)}.$$

Observemos que aquí estamos notando con  $\boldsymbol{\theta}$  la muestra correspondiente a los ángulos

de  $\mathbf{X}$ . Luego, al tomar logaritmo nos queda

$$\begin{aligned}\ln L(\mu, \kappa; \boldsymbol{\theta}) &= -n \ln(2\pi I_0(\kappa)) + \kappa \sum_{j=1}^n \cos(\theta_j - \mu) \\ &= -n \ln(2\pi I_0(\kappa)) + \kappa \sum_{j=1}^n [\cos \theta_j \cos \mu + \sin \theta_j \sin \mu] \\ &= -n \ln(2\pi I_0(\kappa)) + \kappa [\cos \mu \sum_{j=1}^n \cos \theta_j + \sin \mu \sum_{j=1}^n \sin \theta_j].\end{aligned}$$

Usando las definiciones de  $\bar{C}$  y  $\bar{S}$  dadas en la Sección 2.1 y la ecuación (2.1) se tiene

$$\ln L(\mu, \kappa; \boldsymbol{\theta}) = -n \ln(2\pi I_0(\kappa)) + n\kappa \bar{R} \cos(\bar{\theta} - \mu).$$

Como la función  $\cos(x)$  alcanza su valor máximo cuando  $x = 0$ , el estimador de máxima verosimilitud de  $\mu$  será

$$\hat{\mu} = \bar{\theta}.$$

Ahora para hallar el estimador  $\hat{\kappa}$  derivamos la función anterior con respecto a  $\kappa$  e igualamos a cero, obteniendo

$$\frac{\partial \ln L(\mu, \kappa; \boldsymbol{\theta})}{\partial \kappa} = -\frac{n}{I_0(\kappa)} \frac{\partial I_0(\kappa)}{\partial \kappa} + n\bar{R} \cos(\bar{\theta} - \mu).$$

Teniendo en cuenta que  $\frac{\partial I_0(\kappa)}{\partial \kappa} = I_1(\kappa)$ , donde

$$I_1(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos \theta e^{\kappa \cos \theta} d\theta$$

es la función de Bessel modificada de orden 1, y llamando  $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ , obtenemos la ecuación

$$-nA(\hat{\kappa}) + n\bar{R} \cos(\bar{\theta} - \hat{\mu}) = 0.$$

Si ahora reemplazamos  $\hat{\mu}$  por  $\bar{\theta}$ , nos queda que el estimador de máxima verosimilitud  $\hat{\kappa}$  de  $\kappa$  es el valor solución de

$$A(\hat{\kappa}) = \bar{R}.$$

Para el caso de una variable aleatoria cuya distribución corresponde a una mezcla de  $k$  Von Mises, donde su función de densidad está dada por

$$f(\theta) = \sum_{i=1}^k p_i \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(\theta - \mu_i)} \quad p_i \geq 0 \quad \forall i, \quad p_1 + \dots + p_k = 1,$$

existen algoritmos que permiten estimar los parámetros  $(\mu_1, \kappa_1, p_1), \dots, (\mu_k, \kappa_k, p_k)$  por máxima verosimilitud, así como también una alternativa a este método para el caso de una mezcla de solamente dos Von Mises (ver Mardia y Jupp [10]).

Como último ejemplo consideremos una variable aleatoria con distribución *Wrapped Cauchy*  $WC(\mu, \rho)$  con función de densidad

$$f(\theta; \mu, \rho) = \frac{\sqrt{1 - \|\boldsymbol{\mu}\|^2}}{2\pi(1 - \boldsymbol{\mu}^T \mathbf{x})},$$

donde

$$\mathbf{x} = (\cos \theta, \sin \theta)^T, \quad \boldsymbol{\mu} = \frac{2\rho}{1 + \rho^2} (\cos \mu, \sin \mu)^T,$$

de manera tal que  $\|\boldsymbol{\mu}\| < 1$ .

La función de verosimilitud, basada en observaciones independientes  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$  está dada por

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \prod_{j=1}^n \frac{\sqrt{1 - \|\boldsymbol{\mu}\|^2}}{2\pi(1 - \boldsymbol{\mu}^T \mathbf{x}_j)} = \left( \frac{\sqrt{1 - \|\boldsymbol{\mu}\|^2}}{2\pi} \right)^n \prod_{j=1}^n \frac{1}{1 - \boldsymbol{\mu}^T \mathbf{x}_j},$$

donde  $\mathbf{x}_j = (\cos \theta_j, \sin \theta_j)^T$ .

Si ahora tomamos logaritmo nos queda

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\theta}) = -n \ln(2\pi) + \frac{n}{2} \ln(1 - \|\boldsymbol{\mu}\|^2) - \sum_{j=1}^n \ln(1 - \boldsymbol{\mu}^T \mathbf{x}_j).$$

El estimador  $\hat{\boldsymbol{\mu}}$  se puede calcular usando un algoritmo iterativo (ver Mardia y Jupp [10]) que bajo ciertas condiciones converge al estimador de máxima verosimilitud  $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2)^T$ . Los estimadores de los parámetros originales  $\mu$  y  $\rho$  están dados por

$$\hat{\mu} = \tan^{-1} \left( \frac{\hat{\mu}_2}{\hat{\mu}_1} \right), \quad \hat{\rho} = \frac{1 - \sqrt{1 - \|\hat{\boldsymbol{\mu}}\|^2}}{\|\hat{\boldsymbol{\mu}}\|}.$$

Para el caso de variables aleatorias con distribuciones en la recta real sabemos que, bajo condiciones de regularidad, los estimadores de máxima verosimilitud son fuertemente consistentes y asintóticamente normales y eficientes. Recordemos que esto último quiere decir que si  $\hat{\lambda}(X_1, \dots, X_n)$  es el estimador de máxima verosimilitud de  $\lambda$ , para  $n$  *grande* el estimador se comporta aproximadamente como si tuviese distribución  $N(\lambda, 1/nI_1(\lambda))$ , donde  $I_1(\lambda)$  es el número de información de Fisher para  $X_1$ . Es decir, que su comportamiento es como si fuera insesgado con varianza  $1/nI_1(\lambda)$ , que es la menor varianza posible de acuerdo con el Teorema de Rao-Cramer.

Resultados análogos se obtienen para el caso en que la variable aleatoria tenga distribución en el círculo unitario. (Ver Mardia y Jupp [10]).



## Capítulo 3

# Estimación no paramétrica de la densidad circular

Una característica básica que describe el comportamiento de una variable aleatoria  $X$  es su función de densidad, pero en la mayoría de los estudios prácticos no se la conoce directamente. En su lugar sólo se cuenta con un conjunto de observaciones  $X_1, \dots, X_n$ , que suponemos independientes, idénticamente distribuidas y con función de densidad desconocida. Un problema fundamental de la estadística es la estimación de dicha función de densidad a partir de la información proporcionada por la muestra. Dicho problema se puede enfocar de dos formas. Un enfoque consiste en considerar que la función de densidad que deseamos estimar pertenece a una determinada familia, por ejemplo, Normal, Exponencial, etc. Bajo este enfoque, la estimación se reduce a determinar el valor de los parámetros desconocidos del modelo a partir de la muestra. Este tipo de estimación se denomina *estimación paramétrica* de la densidad. Una segunda alternativa consiste en no predeterminedar a priori ninguna condición para la función de densidad, salvo las condiciones impuestas a las funciones de densidad para ser consideradas como tales. Este enfoque se denomina *estimación no paramétrica* de la densidad y es en el que nos centraremos en este capítulo.

### 3.1. Histograma

El método clásico para la estimación de la densidad a partir de un conjunto de observaciones es el *histograma*. Si se supone  $X \in \mathbb{R}$  y se desea estimar su densidad en el intervalo  $I = (a, b]$ , el método consiste en dividir el intervalo en  $k$  subintervalos disjuntos, del mismo largo de forma tal que la unión de los mismos sea igual a  $I$  y contar las ocurrencias de los  $X_i$  en los intervalos.

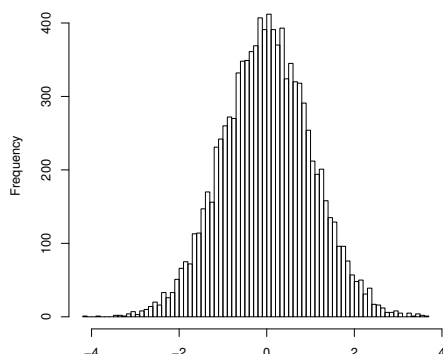


Figura 3.1: Histograma para una muestra de una variable  $X \sim N(0, 1)$  (100 particiones)

Si llamamos  $I_j$  al  $j$ -ésimo intervalo

$$I_j = \left( a + (j-1) \frac{(b-a)}{k}, a + j \frac{(b-a)}{k} \right],$$

entonces el histograma puede definirse como

$$\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{I_j}(X_i) \quad \text{si } x \in I_j.$$

Este estimador tiene una serie de problemas, el más notable es que  $\hat{g}_n$  es continua a trozos cuando  $g$ , por lo general, es continua. Por otra parte, es sensible tanto a la cantidad de elementos de la partición como al punto inicial de la partición. (Ver Figura 3.1).

## 3.2. Histograma móvil

Un método que intenta suavizar la estimación obtenida por el histograma es el llamado *histograma móvil*. En este caso, en lugar de dividir el intervalo de interés en subintervalos fijos, se considera, para un valor  $h > 0$  dado, un intervalo o ventana de ancho  $2h$  centrado en el punto en el que se desea realizar la estimación de la densidad. De esta forma la estimación resulta



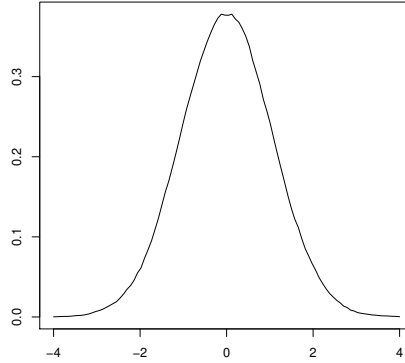


Figura 3.2: Histograma móvil para la muestra de la Figura 3.1 con  $h=0.5$

$$\hat{g}_n(x) = \frac{1}{2nh} \sum_{i=1}^n I_{(x-h, x+h]}(X_i). \quad (3.1)$$

Si bien esta estimación también es continua a trozos, es más suave que el histograma (ver Figura 3.2) y es quien va a dar lugar a la estimación por núcleos.

El estimador dado en (3.1) puede reescribirse de la siguiente manera:

$$\hat{g}_n(x) = \frac{1}{2nh} \sum_{i=1}^n I_{(x-h, x+h]}(X_i) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} I_{[-1,1)}\left(\frac{x - X_i}{h}\right).$$

Si llamamos  $\omega$  a la densidad de una variable aleatoria con distribución uniforme en  $[-1,1)$ , es decir

$$\omega(t) = \frac{1}{2} I_{[-1,1)}(t),$$

y la reemplazamos en la expresión de  $\hat{g}_n(x)$  nos queda que

$$\hat{g}_n(x) = \frac{1}{nh} \sum_{i=1}^n \omega\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \omega\left(\frac{x - X_i}{h}\right). \quad (3.2)$$

Notemos que

$$\blacksquare \quad \omega \geq 0$$

$$\blacksquare \int \omega(t) dt = 1.$$

Además, para cada  $1 \leq i \leq n$ , tenemos que

$$\omega\left(\frac{x - X_i}{h}\right) = \frac{1}{2} \Leftrightarrow X_i \in (x - h, x + h],$$

es decir, la función  $\omega$  le otorga un peso uniforme a todas las observaciones del intervalo  $(x - h, x + h]$ .

### 3.3. Estimación por núcleos de la densidad para datos lineales

En la Sección anterior, obtuvimos una expresión para el estimador por histograma móvil que resultará de gran utilidad para definir el estimador por núcleos de la densidad.

Una manera de generalizar (3.2) sería reemplazar la función  $\omega$  por una función continua  $K$ , no negativa y que verifique la condición  $\int K(t) dt = 1$ .

De esta manera obtenemos el estimador de la densidad *por núcleos* definido por Rosenblatt [14] y Parzen [12] dado por

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right), \quad (3.3)$$

donde  $h = h_n$  es llamado *parámetro de suavizado* o *ancho de ventana* y satisface  $h_n \rightarrow 0$  si  $n \rightarrow \infty$ .

Por lo tanto, la estimación de la densidad por núcleos no es otra cosa que un promedio ponderado por la distancia de las observaciones al punto que deseamos estimar. Cuanto mayor sea la distancia de la observación al punto, menor será su peso en la estimación. El peso lo determinarán el núcleo elegido y el valor de  $h$ . Cuanto mayor sea el valor de  $h$ , se le dará más peso a las observaciones que se encuentren alejadas del punto, dando resultados con un sesgo alto. Valores de  $h$  demasiado pequeños tienen como efecto que la varianza del estimador aumente, pues son pocas las observaciones consideradas por punto. Por lo tanto, determinar el valor de  $h$  constituye un punto muy importante en el proceso de estimación. (Ver Figura 3.3).

Se puede demostrar que si  $K$  cumple las siguientes condiciones

$$\blacksquare K(-t) = K(t) \forall t$$

- $\int |K(t)|dt < \infty$
- $|tK(t)| \rightarrow 0$  si  $|t| \rightarrow \infty$

se tiene que

$$E(\hat{f}_n(x)) \rightarrow f(x) \text{ si } h \rightarrow 0 \quad \text{y} \quad \text{Var}(\hat{f}_n(x)) \rightarrow 0 \text{ si } h \rightarrow 0, nh \rightarrow \infty,$$

es decir,  $\hat{f}_n(x)$  es *débilmente consistente* a  $f(x)$  para cada  $x$ .

A continuación, mencionaremos algunos ejemplos de posibles núcleos.

- **Núcleo Rectangular:** es el utilizado en el caso de los histogramas móviles

$$K(t) = \frac{1}{2}I_{(-1,1]}(t).$$

- **Núcleo triangular:** llamado así por su gráfica

$$K(t) = \begin{cases} 1 - |t| & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

- **Núcleo Gaussiano:** asociado a la densidad de una variable aleatoria con distribución  $N(0, 1)$

$$K(t) = e^{-\frac{1}{2}t^2}.$$

- **Núcleo de Epanechnikov:** con este núcleo se obtienen los mejores resultados en términos de error cuadrático medio (ECM)

$$K(t) = \begin{cases} \frac{3}{4}(1 - t^2) & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

### 3.4. Estimación por núcleos de la densidad para datos circulares

En la Sección anterior, obtuvimos el estimador de la densidad por núcleos para el caso de datos lineales. Si en lugar de tener una muestra de variables aleatorias  $X_1, \dots, X_n$ , tenemos vectores aleatorios  $(q+1)$ -dimensionales  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , el estimador (3.3) se escribe como

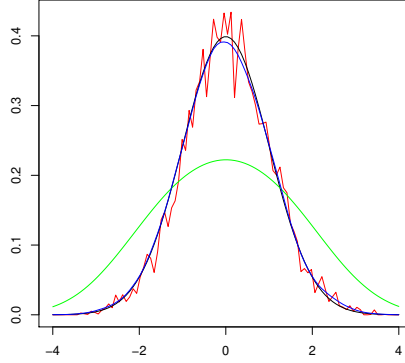


Figura 3.3: Estimación de la densidad de una variable  $N(0,1)$  (negro) por núcleo de Epanechnikov para  $h = 0.02$  (rojo),  $h = 0.5$  (azul) y  $h = 3$  (verde).

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh^{q+1}} \sum_{i=1}^n K[(\mathbf{x} - \mathbf{X}_i)/h], \quad \mathbf{x} \in \mathbb{R}^{q+1}. \quad (3.4)$$

Este estimador puede extenderse fácilmente a datos circulares. Al trabajar con puntos en el círculo, o con datos direccionales en general, podemos *medir* la distancia entre dos puntos por

$$\begin{aligned} \|\mathbf{x} - \mathbf{X}_i\|^2 &= (\mathbf{x} - \mathbf{X}_i)^T (\mathbf{x} - \mathbf{X}_i) = \mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{X}_i + \mathbf{X}_i^T \mathbf{X}_i \\ &= \|\mathbf{x}\|^2 - 2\mathbf{x}^T \mathbf{X}_i + \|\mathbf{X}_i\|^2. \end{aligned}$$

Como estamos considerando el círculo unitario resulta que

$$\|\mathbf{x} - \mathbf{X}_i\|^2 = 2 - 2\mathbf{x}^T \mathbf{X}_i = 2(1 - \mathbf{x}^T \mathbf{X}_i). \quad (3.5)$$

Dada una muestra aleatoria de ángulos  $\theta_1, \dots, \theta_n$  en  $[0, 2\pi)$ , consideramos los vectores aleatorios  $\mathbf{X}_i^T = (\cos \theta_i, \sin \theta_i)$ . Para estimar la densidad circular en un ángulo  $\theta$ , Hall *et al.* [6] proponen el estimador por núcleos dado por

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n} c_0(\kappa) \sum_{i=1}^n K(\kappa \mathbf{x}^T \mathbf{X}_i), \quad (3.6)$$

donde  $\mathbf{x}^T = (\cos \theta, \sin \theta)$ , el nuevo parámetro de suavizado  $\kappa$  reemplaza a  $h^{-1}$  y  $c_0(\kappa)$  se elige de manera tal que el estimador  $\hat{f}_n$  integre 1, es decir resulte también una función de densidad. En el citado trabajo, se demuestra, mediante un cambio de variables, que la integral de  $K(\kappa \mathbf{x}^T \mathbf{X}_i)$  no depende de  $i$ .

De (3.5) se deduce que si  $\mathbf{x}$  está cerca de  $\mathbf{X}_i$ ,  $\mathbf{x} - \mathbf{X}_i$  es cercano a 0, pero  $\mathbf{x}^T \mathbf{X}_i$  es cercano a 1. Luego, los estimadores (3.4) y (3.6) tienen características distintas lo que lleva a tener que usar núcleos diferentes. Una alternativa más semejante al caso del estimador de Rosenblatt y Parzen dado en (3.4) es

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n} d_0(\kappa) \sum_{i=1}^n K[\kappa(1 - \mathbf{x}^T \mathbf{X}_i)], \quad (3.7)$$

donde nuevamente  $d_0(\kappa)$  se elige de manera que el estimador integre 1.

Otro estimador posible, y con el cual vamos a trabajar en el Capítulo siguiente, es el propuesto por Bai *et al.* [2] dado por

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n} C(h) \sum_{i=1}^n K[(1 - \mathbf{x}^T \mathbf{X}_i)/h^2], \quad (3.8)$$

eligiendo  $C(h)$  de la misma manera que en los casos anteriores.

En los trabajos citados en [6] y [2] se estudian las propiedades asintóticas de los estimadores expuestos. En particular, Bai *et al.* obtienen resultados de consistencia fuerte puntual, consistencia fuerte uniforme y consistencia fuerte en norma  $L^1$ .



## Capítulo 4

# Estimadores de mínima distancia para datos circulares

Los estimadores de mínima distancia ocupan un lugar destacado entre las alternativas robustas al método clásico de estimación puntual por máxima verosimilitud descrito en el Capítulo 2. Esto se debe a que proporcionan el valor del parámetro buscado que más se acerca a la información suministrada por la muestra, aún en el caso en que el modelo que hayamos asumido no represente exactamente la situación real.

En el caso de datos lineales, la manera usual de definir un *estimador de mínima distancia*  $\hat{\lambda}_n$ , es

$$\hat{\lambda}_n = \operatorname{argmin} \delta(F_n, F_\lambda),$$

donde  $F_n$  es la distribución empírica,  $F_\lambda$  es la distribución del modelo asumido,  $\lambda \in \mathbb{R}^p$  es el parámetro que deseamos estimar y  $\delta$  es alguna discrepancia entre distribuciones. Como en la mayoría de los casos  $F_\lambda$  corresponde a una distribución absolutamente continua dada en términos de su función de densidad  $f_\lambda$ , una idea natural sería *suavizar* la distribución empírica  $F_n$  y definir así un estimador de mínima distancia de la siguiente manera

$$\hat{\lambda}_n = \operatorname{argmin} D(\hat{f}_n, f_\lambda) \quad (4.1)$$

donde  $\hat{f}_n$  es un estimador no paramétrico de la densidad, por ejemplo el estimador por núcleos definido en la Sección 3.3, y  $D$  es una distancia entre funciones de densidad. Las elecciones estándar para  $D$  corresponden a las métricas  $L^1$ ,  $L^2$  y  $L^\infty$ , siendo  $D_1(\hat{f}_n, f) = \int |\hat{f}_n - f|$ ,  $D_2(\hat{f}_n, f) = (\int (\hat{f}_n - f)^2)^{1/2}$  y  $D_\infty(\hat{f}_n, f) = \sup_t |\hat{f}_n(t) - f(t)|$ .

Para evitar problemas de existencia en la definición (4.1), podemos definir el estimador  $\hat{\lambda}_n$  como cualquier valor que satisfaga

$$D(\hat{\lambda}_n, f_{\hat{\lambda}_n}) \leq \inf_{\lambda} D(\hat{f}_n, f_\lambda) + \varepsilon_n \quad (4.2)$$

donde  $\varepsilon_n \rightarrow 0$  cuando  $n \rightarrow \infty$

En Cao *et al.* [4] se estudia la consistencia fuerte de los estimadores definidos en (4.1), así como también su distribución asintótica. Por otro lado, los autores muestran que el funcional que define los estimadores propuestos es continuo respecto de la métrica de Prokhorov, obteniendo así la robustez cualitativa de los mismos. Se describen, además, algunas técnicas de selección del ancho de ventana  $h$ , por ejemplo el que usaremos en la Sección 4.2 llamado *ancho de ventana de mínima distancia*  $h_{MD}$ . Este valor  $h_{MD}$  es aquel que minimiza, junto con  $\hat{\lambda}_n$ , la distancia entre la densidad del modelo que hayamos asumido y su estimación por núcleos, es decir

$$(\hat{\lambda}_n, h_{MD}) = \operatorname{argmin} D(\hat{f}_n(\cdot, h), f_\lambda). \quad (4.3)$$

### Nuestra propuesta

Para el caso de datos circulares, podemos definir el estimador de mínima distancia de la misma manera que lo hicimos en (4.1), es decir

$$\hat{\lambda}_n = \operatorname{argmin} D(\hat{f}_n, f_\lambda), \quad (4.4)$$

donde ahora  $\hat{f}_n$  es un estimador no paramétrico de la densidad circular, definido en la Sección 3.4.

El análisis de las propiedades y comportamiento asintótico de este tipo de estimadores resulta análogo al estudiado en el caso lineal por Cao *et al.* [4]. En la siguiente Sección analizaremos su consistencia.

## 4.1. Consistencia

Vamos a considerar las siguientes hipótesis bajo las cuales es posible obtener la consistencia fuerte del estimador propuesto.

**(H1)** La distancia  $D$  y el estimador no paramétrico de la densidad circular  $\hat{f}_n$  verifican que  $D(\hat{f}_n, f_\lambda) \rightarrow 0$  *c.t.p.*

**(H2)**  $\forall \lambda_0 \in \Lambda$  y  $\{\lambda_r\} \in \Lambda$  se tiene que  $\lim_{r \rightarrow \infty} D(f_\lambda, f_{\lambda_r}) = D(f_\lambda, f_{\lambda_0}) \Rightarrow \lim_{r \rightarrow \infty} \lambda_r = \lambda_0$ .

La hipótesis **(H1)** fue estudiada por Bai *et al.* [2] para  $D_1$  y  $D_\infty$ . Más precisamente, si  $K$  es acotado en  $\mathbb{R}_{>0}$  y verifica

$$\int_0^\infty \sup\{K(u) : |\sqrt{u} - \sqrt{v}| < 1\} v^{(q-2)/2} dv < \infty,$$



y si  $h_n \rightarrow 0$  y  $\frac{nh_n^q}{\log n} \rightarrow \infty$  cuando  $n \rightarrow \infty$ , en el citado trabajo se demuestra que

$$D_\infty(\hat{f}_n, f_\lambda) \rightarrow 0 \quad c.t.p.$$

Además, se demuestra que

$$D_1(\hat{f}_n, f_\lambda) \rightarrow 0 \quad c.t.p.,$$

suponiendo que  $h_n \rightarrow 0$  y  $nh_n^q \rightarrow \infty$  cuando  $n \rightarrow \infty$  y que  $\int_0^\infty v^{(q-2)/2} K(v) dv < \infty$ .

También es fácil ver que  $D_2(\hat{f}_n, f_\lambda) \rightarrow 0$  c.t.p., ya que, en nuestro caso

$$\begin{aligned} D_2(\hat{f}_n, f_\lambda)^2 &= \int_0^{2\pi} (\hat{f}_n(\theta) - f_\lambda(\theta))^2 d\theta = \int_0^{2\pi} |\hat{f}_n(\theta) - f_\lambda(\theta)|^2 d\theta \\ &\leq \int_0^{2\pi} D_\infty(\hat{f}_n, f_\lambda)^2 d\theta = 2\pi D_\infty(\hat{f}_n, f_\lambda)^2 \rightarrow 0 \quad c.t.p. \end{aligned}$$

Por otro lado, si la familia de densidades  $f_\lambda$  admite una *parametrización razonable* la hipótesis **(H2)** queda garantizada.

Ahora estamos en condiciones de enunciar el siguiente Teorema cuya demostración es análoga a la obtenida en Cao *et al.* [4].

**Teorema 4.1.1.** *Sea  $\{\hat{\lambda}_n\}$  una sucesión de estimadores de mínima distancia de  $\lambda$  como los definidos en (4.2). Bajo las hipótesis **H1** y **H2**,  $\{\hat{\lambda}_n\}$  resulta fuertemente consistente, es decir*

$$\lim_{n \rightarrow \infty} \hat{\lambda}_n = \lambda \quad c.t.p.$$

## 4.2. Estudio de Monte Carlo

En esta Sección, mostraremos los resultados de un estudio de simulación realizado con el objeto de evaluar el estimador propuesto y de compararlo con el de máxima verosimilitud. Usaremos como estimador de la densidad circular el dado en la ecuación (3.8), es decir

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n} C(h) \sum_{i=1}^n K[(1 - \mathbf{x}^T \mathbf{X}_i)/h^2],$$

escogiendo el núcleo de Epanechnikov, de manera que el estimador  $\hat{f}_n$  resulta

$$\hat{f}_n(\mathbf{x}) = \frac{1}{n} C(h) \sum_{i=1}^n \frac{3}{4} [1 - [(1 - \mathbf{x}^T \mathbf{X}_i)/h^2]^2] I_{(-1,1)}((1 - \mathbf{x}^T \mathbf{X}_i)/h^2).$$

Para obtener el valor de  $C(h)$ , que recordemos debía ser escogido de manera tal que el estimador  $\hat{f}_n$  integre 1, recurrimos al trabajo de Zhao y Wu [17]. Allí se establece que, suponiendo que  $h = h_n \rightarrow 0$  si  $n \rightarrow \infty$ ,

$$C(h)^{-1} \sim \lambda h^q, \quad \text{con } \lambda = 2^{q/2-1} \omega_{q-1} \int_0^\infty K(t) t^{q/2-1} dt,$$

donde  $C(h)^{-1} \sim \lambda h^q$  significa que  $C(h)^{-1}/\lambda h^q \rightarrow 1$  si  $n \rightarrow \infty$  y

$$\omega_{q-1} = 2\pi^{q/2} / \Gamma(q/2), \quad q \geq 1,$$

siendo  $\Omega_q \in \mathbb{R}^{q+1}$  una esfera unitaria  $q$ -dimensional en la cual toma valores la muestra aleatoria que se está analizando.

En nuestro caso estamos considerando datos circulares, por lo tanto  $q = 1$ . Luego,

$$\omega_0 = 2\pi^{1/2} / \Gamma(1/2) = 2\sqrt{\pi} / \sqrt{\pi} = 2,$$

$$\int_0^\infty \frac{3}{4} (1 - t^2) I(|t| < 1) t^{1/2-1} dt = \int_0^1 \frac{3}{4} (1 - t^2) t^{-1/2} dt = \frac{6}{5}$$

y

$$\lambda = 2^{1/2-1} 2 \frac{6}{5} = \frac{6}{5} \sqrt{2}.$$

Usando esta aproximación para  $C(h)$  nuestro estimador resulta

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh} \left( \frac{6}{5} \sqrt{2} \right)^{-1} \sum_{i=1}^n \frac{3}{4} [1 - [(1 - \mathbf{x}^T \mathbf{X}_i)/h^2]^2] I_{(-1,1)}((1 - \mathbf{x}^T \mathbf{X}_i)/h^2). \quad (4.5)$$

Para nuestro estudio hemos considerado los siguientes modelos:

### Modelo 1

En cada escenario se generaron 1000 muestras de tamaño  $n = 100$  de una variable aleatoria con distribución  $VM(\pi/2, 2)$ . Primero consideramos 10 valores distintos para el parámetro de suavizado  $h$  (ver Cuadro 4.1). Luego, repetimos el estudio considerando el ancho de ventana mínimo  $h_{MD}$  definido en (4.3). Además, con el propósito de evaluar el comportamiento del estimador propuesto bajo distintas contaminaciones se generaron 1000 muestras de tamaño  $n = 100$  para cada uno de los siguientes esquemas de contaminación del modelo original (ver Figura 4.1):

- **Contaminación 1:**  $0.9VM(\pi/2, 2) + 0.1WC(5, e^{-0.001})$
- **Contaminación 2:**  $0.9VM(\pi/2, 2) + 0.1WC(\pi/2, e^{-100})$
- **Contaminación 3:**  $0.9VM(\pi/2, 2) + 0.1WC(3\pi/2, e^{-0.05})$ .

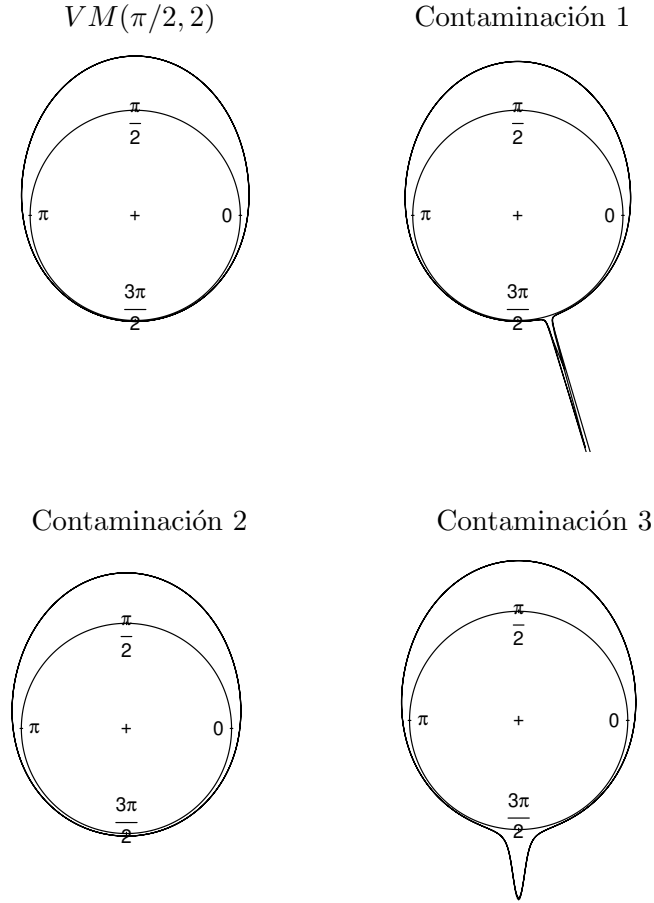


Figura 4.1: Gráficos correspondientes a las distribuciones del Modelo 1.

Los resultados obtenidos pueden verse en el Cuadro 4.2. Los Cuadros 4.3 y 4.4 reflejan distintas relaciones, en términos del  $ECM$ , entre los estimadores obtenidos. En el Cuadro 4.3, puede verse la relación entre los estimadores de mínima distancia y el obtenido mediante el método de máxima verosimilitud, para cada una de las distribuciones consideradas; mientras que en el Cuadro 4.4, se manifiesta, para cada uno de los estimadores, la relación entre el obtenido en el modelo contaminado y el obtenido bajo el modelo sin contaminar.

## Modelo 2

Se generaron 1000 muestras de tamaño  $n = 100$  de una variable aleatoria cuya distribución es una mezcla de dos Von Mises  $VM(\pi/2, 3\pi/2, 1, 2, 0.5)$  (ver Figura 4.2), considerando también en este caso distintos valores de  $h$ . Los resultados obtenidos en la estimación pueden verse en el Cuadro 4.5. Análogamente a lo realizado en el Modelo 1, en este caso también consideramos la relación existente entre los estimadores de mínima distancia y los correspondientes obtenidos por máxima verosimilitud. Dicha relación, en términos del ECM, está reflejada en el Cuadro 4.6.

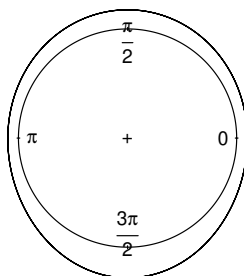


Figura 4.2: Gráfico correspondiente a la distribución del Modelo 2

Para analizar la efectividad del estimador propuesto, se consideraron las siguientes medidas de resumen:

$$\begin{aligned}
 media(\hat{\lambda}) &= \bar{\hat{\lambda}} = \frac{1}{N} \sum_{j=1}^N \hat{\lambda}_j \\
 sd(\hat{\lambda}) &= \frac{1}{N} \sum_{j=1}^N (\hat{\lambda}_j - \bar{\hat{\lambda}})^2 \\
 ECM(\hat{\lambda}) &= \frac{1}{N} \sum_{j=1}^N (\hat{\lambda}_j - \lambda_0)^2,
 \end{aligned}$$

donde  $\lambda_0$  representa el verdadero valor del parámetro y  $N = 1000$  en todos los casos.

En todos los casos la distancia "D" que hemos considerado para calcular el estimador de mínima distancia es la correspondiente a la métrica  $L^2$  y obtuvimos el

estimador definido en (4.4) minimizando el *cuadrado* de dicha distancia, es decir la función

$$\int_0^{2\pi} (\hat{f}_n(\theta) - f(\theta))^2 d\theta,$$

donde, haciendo un abuso de notación, llamamos  $\hat{f}_n(\theta)$  al estimador de la densidad circular definido en (4.5) con  $\mathbf{x} = (\cos \theta, \sin \theta)^T$  y  $f$  representa la densidad correspondiente a la distribución considerada en cada uno de los modelos. Con el fin de evaluar diferencias entre distintas métricas, en el Modelo 1 se consideró, además de la ya mencionada, la métrica  $L^1$ . Es decir que obtuvimos el estimador de mínima distancia, pero en este caso minimizando la función

$$\int_0^{2\pi} |\hat{f}_n(\theta) - f(\theta)| d\theta,$$

con  $\hat{f}_n$  y  $f$  iguales al caso anterior. En cualquiera de los dos casos, las integrales se obtuvieron promediando el valor de la función sobre una grilla de 1000 valores equidistantes entre 0 y  $2\pi$ .

Cabe destacar que, en nuestro estudio, los parámetros estimados fueron la dirección media  $\mu$  en el caso del Modelo 1 y  $\mu_1, \mu_2$  y la proporción de mezcla  $p$  en el caso del Modelo 2. En ningún modelo se estimó el parámetro de concentración  $\kappa$ . Debido a su estrecha relación con el ancho de ventana  $h$ , los resultados en la estimación del mismo no son adecuados cuando se utilizan este tipo de estimadores, ya que se produce un problema de identificación entre el parámetro de suavizado y el parámetro de concentración.

Para el Modelo 1, hemos incluido también una comparación con los estimadores introducidos por Agostinelli [1]. En el mismo, se proponen estimadores robustos basados en máxima verosimilitud pesada para datos circulares, que se encuentra implementada en la librería `wle` de R. El Cuadro 4.2 incluye un resumen de dichos estimadores que denotamos con  $\hat{\mu}_{Agos}$ . Por último, hemos comparado nuestra propuesta con una opción intermedia que consistió en aplicar nuestro estimador de mínima distancia, con distancia  $L^1$ , pero tomando como parámetro de concentración  $\kappa$  al obtenido mediante la estimación propuesta por Agostinelli. Los estimadores obtenidos bajo esta alternativa los hemos denotado por  $\hat{\mu}_{md,Agos}$ . También consideramos apropiado incluir, además de los resultados numéricos, boxplots correspondientes a cada una de las diferentes estimaciones, con el fin de apreciar las diferencias obtenidas en cada caso. Dichos gráficos corresponden a las Figuras 4.3, 4.4 y 4.5.

### 4.3. Resultados Numéricos

	$h$									
	0.01	0.706	1.403	2.1	2.796	3.493	4.19	4.886	5.583	6.28
$media(\hat{\mu})$	1.5769	1.5773	1.5759	1.5650	1.5422	1.5124	1.4942	1.5022	1.5312	1.5704
$sd(\hat{\mu})$	0.0955	0.0867	0.0842	0.0849	0.0841	0.0762	0.0587	0.0401	0.0271	0.0192
$ECM(\hat{\mu})$	0.0091	0.0075	0.0071	0.0072	0.0078	0.0092	0.0093	0.0063	0.0023	0.0003

Cuadro 4.1: **Modelo 1.** Resultados de la estimación de la dirección media  $\mu$  de una distribución  $VM(\pi/2, 2)$  para distintos valores de  $h$ .

$VM(\pi/2, 2)$	$\hat{\mu}_{md,L^2}$	$\hat{\mu}_{md,L^1}$	$\hat{\mu}_{mv}$	$\hat{\mu}_{Agos}$	$\hat{\mu}_{md,Agos}$
$media$	1.5772	1.5769	1.5777	1.5743	1.5736
$sd$	0.0907	0.0930	0.0820	0.0853	0.0964
$ECM$	0.0082	0.0086	0.0067	0.0072	0.0093
Contaminación 1	$\hat{\mu}_{md,L^2}$	$\hat{\mu}_{md,L^1}$	$\hat{\mu}_{mv}$	$\hat{\mu}_{Agos}$	$\hat{\mu}_{md,Agos}$
$media$	1.5690	1.5716	1.5209	1.5481	1.5629
$sd$	0.0981	0.0958	0.1098	0.0972	0.0948
$ECM$	0.0096	0.0091	0.0145	0.0099	0.0090
Contaminación 2	$\hat{\mu}_{md,L^2}$	$\hat{\mu}_{md,L^1}$	$\hat{\mu}_{mv}$	$\hat{\mu}_{Agos}$	$\hat{\mu}_{md,Agos}$
$media$	1.5714	1.5707	1.5743	1.5722	1.5688
$sd$	0.0997	0.1005	0.0921	0.0940	0.1001
$ECM$	0.0099	0.0100	0.0084	0.0088	0.0100
Contaminación 3	$\hat{\mu}_{md,L^2}$	$\hat{\mu}_{md,L^1}$	$\hat{\mu}_{mv}$	$\hat{\mu}_{Agos}$	$\hat{\mu}_{md,Agos}$
$media$	1.5726	1.5740	1.5758	1.5716	1.5703
$sd$	0.0981	0.0963	0.1065	0.0984	0.0957
$ECM$	0.0096	0.0092	0.0113	0.0096	0.0091

Cuadro 4.2: **Modelo 1.** Resultados de la estimación del parámetro  $\mu$  utilizando el ancho de ventana mínimo  $h_{MD}$ .

	$\frac{ECM(\hat{\mu}_{md,L^2})}{ECM(\hat{\mu}_{mv})}$	$\frac{ECM(\hat{\mu}_{md,L^1})}{ECM(\hat{\mu}_{mv})}$	$\frac{ECM(\hat{\mu}_{Agos})}{ECM(\hat{\mu}_{mv})}$	$\frac{ECM(\hat{\mu}_{md,Agos})}{ECM(\hat{\mu}_{mv})}$
$VM(\pi/2, 2)$	1.2206	1.2806	1.0755	1.3731
Contaminación 1	0.6618	0.6316	0.6855	0.6222
Contaminación 2	1.1703	1.1890	1.0403	1.1796
Contaminación 3	0.8478	0.8169	0.8522	0.8051

Cuadro 4.3: **Modelo 1.** Cocientes de los ECM del estimador propuesto y del obtenido por máxima verosimilitud.

	$\hat{\mu}_{md,L^2}$	$\hat{\mu}_{md,L^1}$	$\hat{\mu}_{mv}$	$\hat{\mu}_{Agos}$	$\hat{\mu}_{md,Agos}$
Contaminación 1	1.1707	1.0581	2.1641	1.3750	0.9677
Contaminación 2	1.2073	1.1627	1.2537	1.2222	1.0752
Contaminación 3	1.1707	1.0697	1.6865	1.3333	0.9784

Cuadro 4.4: **Modelo 1.** Cada columna representa el cociente entre los ECM del estimador obtenido en cada una de las distintas contaminaciones y el obtenido bajo el modelo sin contaminar.

	$\tilde{\mu}_1$	$sd(\hat{\mu}_1)$	$ECM(\hat{\mu}_1)$	$\tilde{\mu}_2$	$sd(\hat{\mu}_2)$	$ECM(\hat{\mu}_2)$	$\tilde{p}$	$sd(\hat{p})$	$ECM(\hat{p})$
$h$	Estimación por mínima distancia								
0.01	1.583	0.407	0.166	4.691	0.301	0.091	0.502	0.073	0.005
0.04	1.579	0.394	0.155	4.695	0.282	0.080	0.503	0.073	0.005
0.706	1.679	0.668	0.458	4.660	0.545	0.300	0.539	0.076	0.007
1	1.871	0.926	0.948	4.891	0.543	0.326	0.586	0.070	0.012
	Estimación por máxima verosimilitud								
	1.771	0.819	0.711	4.505	0.812	0.702	0.523	0.096	0.009

Cuadro 4.5: **Modelo 2.** Resultados de la estimación de las direcciones medias  $\mu_1$ ,  $\mu_2$  y del parámetro  $p$  para la distribución  $VM(\pi/2, 3\pi/2, 1, 2, 0.5)$ .

$h$	$ECM(\hat{\mu}_{1,md})$	$ECM(\hat{\mu}_{2,md})$	$ECM(\hat{p}_{md})$
	$ECM(\hat{\mu}_{1,mv})$	$ECM(\hat{\mu}_{2,mv})$	$ECM(\hat{p}_{mv})$
0.01	0.233	0.129	0.553
0.04	0.218	0.114	0.545
0.706	0.645	0.427	0.753
1	1.333	0.465	1.264

Cuadro 4.6: **Modelo 2.** Relación entre los estimadores de mínima distancia y los obtenidos mediante el método de máxima verosimilitud.

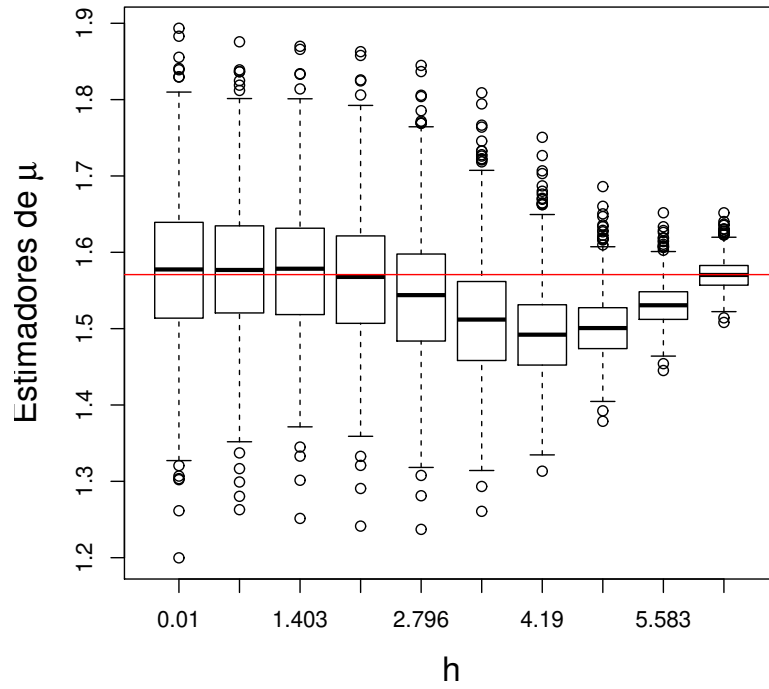


Figura 4.3: **Modelo 1.** Boxplots de los estimadores de  $\mu$  obtenidos en el Cuadro 4.1.



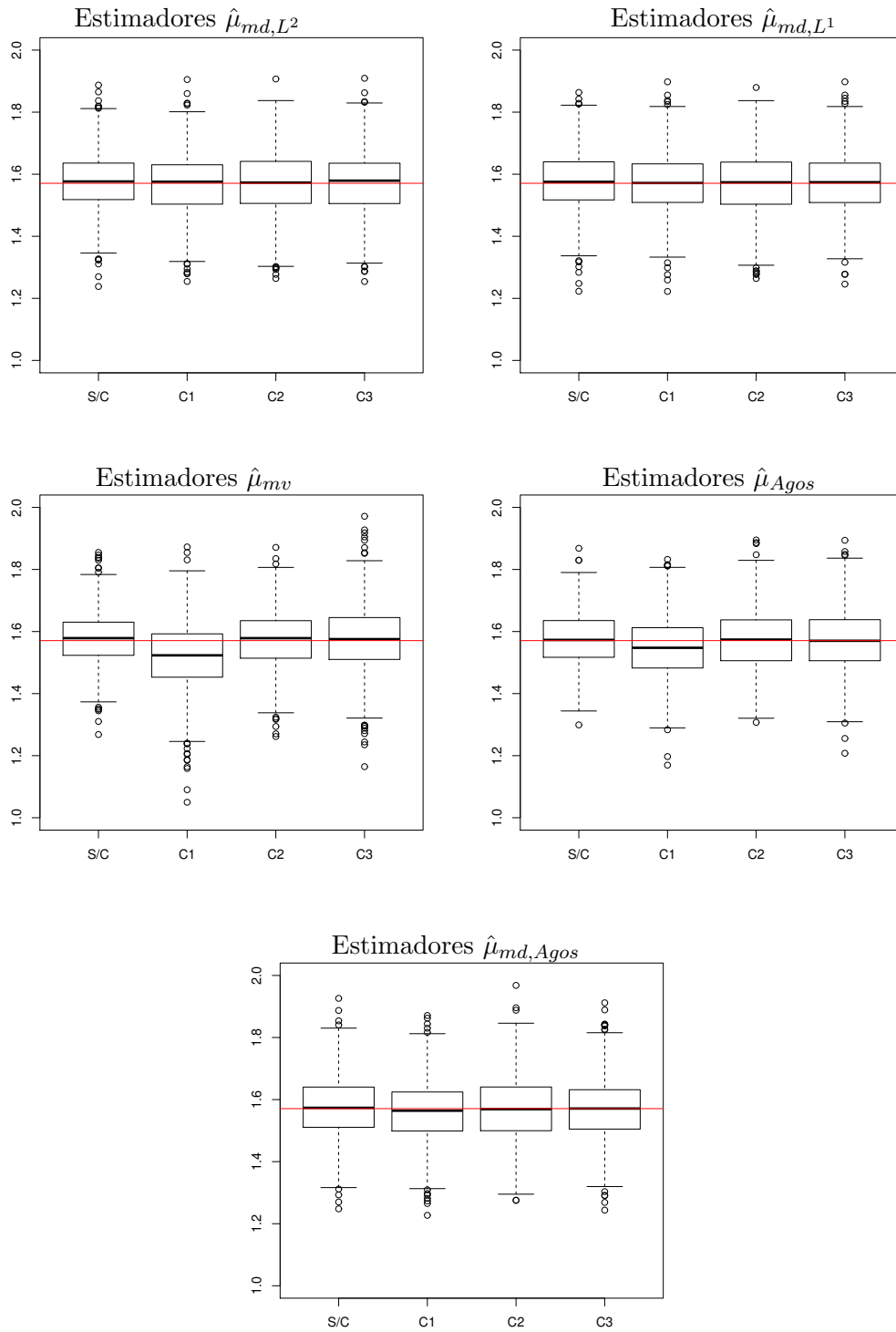


Figura 4.4: **Modelo 1.** Boxplots de los estimadores obtenidos en el Cuadro 4.2.

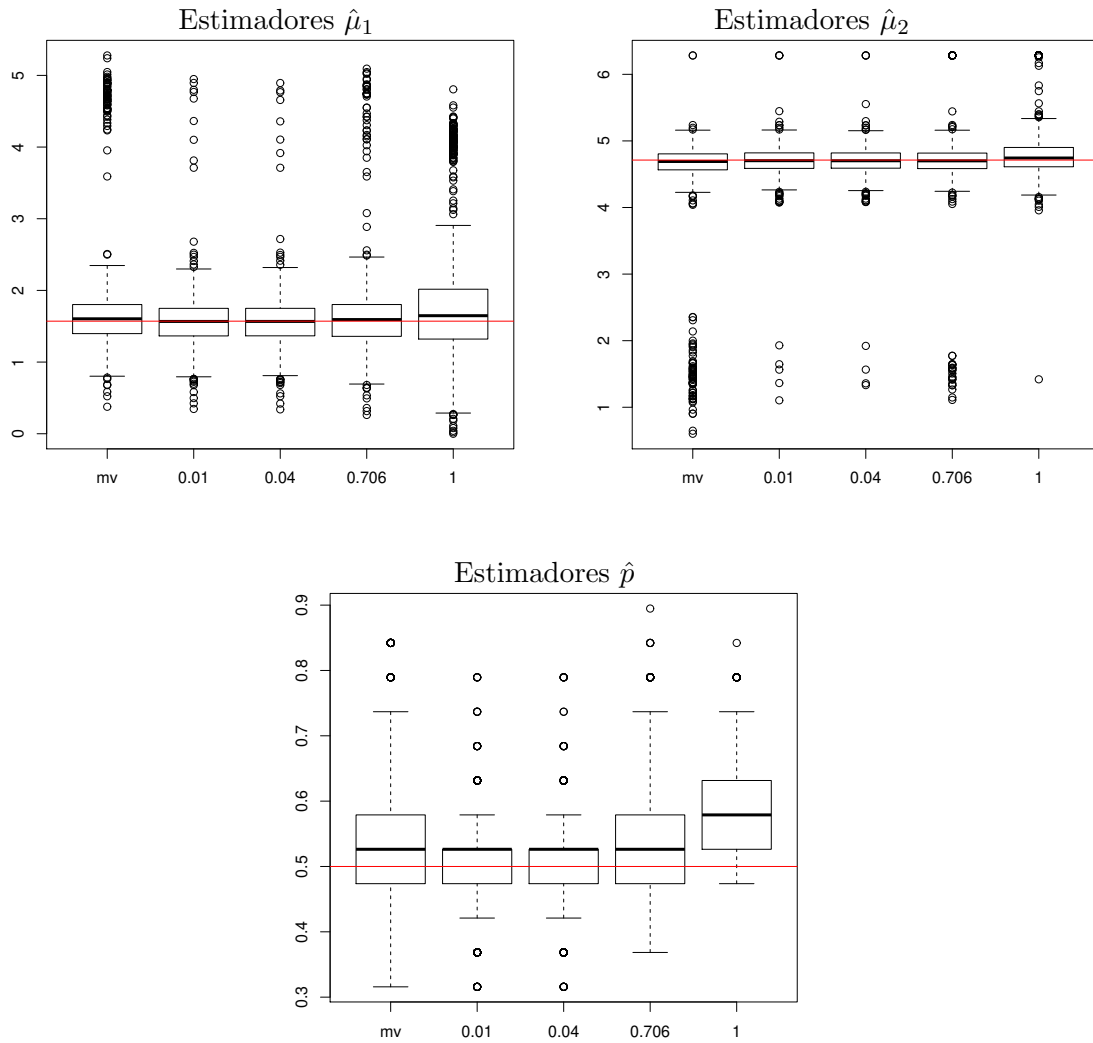


Figura 4.5: **Modelo 2.** Boxplots de los estimadores obtenidos en el Cuadro 4.4.

## 4.4. Conclusiones

El estudio de simulación llevado a cabo, aparte de evaluar el estimador propuesto frente a las alternativas existentes, fue diseñado para estudiar tres diferentes aspectos. Por un lado, la sensibilidad de la elección del ancho de ventana, como así también el comportamiento frente a diferentes desvíos del modelo original y el uso de distintas métricas.

Respecto a la elección del parámetro de suavizado necesario en el procedimiento de estimación no paramétrico, se puede observar que el estimador de mínima distancia resulta sensible a los diferentes valores del ancho de ventana. En el Modelo 1 (Cuadro 4.1 y Figura 4.3) la estimación fue adecuada para la mayoría de los valores de  $h$  considerados, pero

algunos producen estimadores sesgados. Los resultados obtenidos para el Modelo 2 (Cuadro 4.5 y Figura 4.5) producen resultados similares. Estos resultados muestran la necesidad de utilizar un método de selección de ventana.

Para el Modelo 1, hemos implementado los estimadores introducidos en (4.3), que estiman los parámetros desconocidos junto con el parámetro de suavizado. Este estimador, aparte de automatizar la elección del ancho de ventana, arroja mejores resultados (ver Cuadro 4.2 y Figura 4.4). También hemos conservado los valores de ventana seleccionados aunque no han sido reportados en los cuadros y figuras. En general, los valores obtenidos se mantuvieron siempre entre 0.4 y 0.5 siendo consistentes con lo observado en la Figura 4.3. Debido al tiempo que demanda su cálculo no hemos podido obtener resultados de selección de ventana para el Modelo 2.

Para evaluar las propiedades de robustez del estimador propuesto se han considerado tres contaminaciones distintas. Antes de continuar, es importante destacar que, en todos los casos, los parámetros estimados corresponden a ángulos, tomando por lo tanto sus valores en un intervalo acotado. Como podemos apreciar en el trabajo de Agostinelli [1] o Kato y Eguchi [8], el principal motivo de la falta de robustez en modelos de datos circulares se debe a perturbaciones que afectan al parámetro de concentración más que al de posición.

Por una cuestión de tiempo de cálculo nos hemos concentrado solo en el Modelo 1. En los resultados podemos apreciar que las contaminaciones 1 y 3 fueron las que más afectaron la estimación obtenida por máxima verosimilitud. En el caso de la Contaminación 1, en la cual se considera prácticamente una masa puntual, el buen desempeño de nuestra propuesta queda claramente evidenciada tanto en los valores numéricos obtenidos en el Cuadro 4.2, como en el Boxplot correspondiente representado en la Figura 4.4. En el caso de la Contaminación 3, si bien los Boxplots no muestran una significativa diferencia entre las distintas opciones, con respecto a la media, el efecto que causa esta contaminación en la estimación por máxima verosimilitud queda evidenciada en el Cuadro 4.4. Allí puede observarse que los cocientes que corresponden al estimador de máxima verosimilitud se alejan bastante de 1, en relación al resto, evidenciando el crecimiento del ECM en el caso del modelo contaminado. En la Contaminación 2, si bien el desempeño de nuestro estimador fue satisfactorio, el Cuadro 4.3 pone de manifiesto que el estimador de máxima verosimilitud no se ve fuertemente afectado por dicha contaminación.

Lógicamente y como se esperaba, en el Cuadro 4.3 se puede apreciar que el estimador de mínima distancia pierde eficiencia frente al estimador de máxima verosimilitud bajo el modelo sin contaminación. Lo mismo ocurre con el estimador de máxima verosimilitud pesado propuesto por Agostinelli, que tiene un comportamiento adecuado bajo el modelo original y bajo las Contaminaciones 2 y 3, pero que se ve afectado bajo la Contaminación 1. Como alternativa intermedia, y en caso de ser necesaria una estimación del parámetro de concentración, hemos combinado el estimador propuesto por Agostinelli del parámetro de concentración con nuestra propuesta, obteniendo un buen resultado como podemos apreciar en el Cuadro 4.2 y la Figura 4.4. También es importante destacar que los estimadores de mínima distancia resultan mucho más estables, en términos de ECM, que la opción por máxima verosimilitud. Esto puede verse en el Cuadro 4.4. Los cocientes correspondientes a

nuestra propuesta permanecen todos cercanos a 1, mientras que esto no sucede en el caso de máxima verosimilitud, presentando mayor variabilidad. Otro de los datos a observar en el Cuadro 4.4, es el hecho de que para cualquiera de las 3 contaminaciones, resulta más eficiente estimar la dirección media  $\mu$  utilizando un estimador del parámetro de concentración de  $\kappa$ , en lugar del verdadero valor del mismo.

En lo referente al Modelo 2, podemos concluir que, análogamente a lo expuesto por Le Cam, el estimador de máxima verosimilitud no resulta efectivo en el caso de una mezcla de dos Von Mises (ver Cuadro 4.4 y Figura 4.5). En cambio, nuestra propuesta arroja resultados satisfactorios, siempre y cuando se consideren valores de ventana pequeños. Resulta conveniente aclarar que el notorio sesgo reflejado en la Figura 4.5 para la estimación del parámetro de mezcla, se debe fuertemente a la necesidad computacional de calcular nuestro estimador minimizando la distancia para distintos valores de dicho parámetro escogidos en este caso sobre una grilla de longitud 20 equidistantes entre 0 y 1.

Finalmente, respecto a la selección de la métrica, sólo hemos comparado en el Modelo 1 la distancia  $L^2$  con la distancia  $L^1$ . Si bien los Boxplots de la Figura 4.4 no arrojaron diferencias significativas, el Cuadro 4.4 sugiere una leve ventaja de la distancia  $L^1$ .



# Bibliografía

- [1] Agostinelli, C. (2007) Robust Estimation for Circular Data. *Computacional Statistics & Data Analysis*. 5867-5875.
- [2] Bai, Z. D., Rao, R. and Zhao, L. C. (1988) Kernel Estimators of Density Function of Directional Data. *J. Multivariate Analysis*. **27**, 24-39.
- [3] Batschelet, E. (1981) Circular Statistics in Biology. *London: Academic Press*.
- [4] Cao, R., Cuevas, A. and Fraiman, R. (1995) Minimum distance density-based estimation. *Computacional Statistics & Data Analysis*. **20**, 611-631.
- [5] Fisher, N. I. (1993) Statistical Analysis of Circular Data. *Cambridge: Cambridge University Press*.
- [6] Hall, P. , Watson, G. S. and Cabrera, J. (1987) Kernel density estimation with spherical data. *Biometrika*. **74**, 751-762.
- [7] He X. (1992) Robust statistics of directional data: a survey. Nonparametric Statistics and Related Topics. *Amsterdam: North-Holand*. 87-95.
- [8] Kato, S. and Eguchi, S. (2012) Robust estimation of location and concentration parameters for the von Mises-Fisher distribution. Preprint.
- [9] Le Cam, L. (1990): Maximum Likelihood: An Introduction. *International Statistical Review*, 58, 2, 153-171.
- [10] Mardia, K. V. and Jupp, P. E. (2000) Directional Statistics. John Wiley.
- [11] Parr, W. C. and Schucany, W. R. (1982) Minimum distance estimation and componentes of fit. *J. Roy. Statistics. Soc B.* , 178-189.
- [12] Parzen, E. (1962) On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.
- [13] Rao, J. S. (1984) Nonparametric methods in directional data analysis. Handbook of Statistics, Vol. 4. *Amsterdam/New York: Elsevier*. 757-770.
- [14] Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-837

- [15] Watson, G. S. (1983) Statistics on Spheres. *New York: Wiley*.
- [16] Wolfowitz, J. (1957) The minimum distance method. *Ann. Math. Statist.* **28** . 75-88.
- [17] Zhao, L. and Wu, C. (2001) Central limit theorem for integrated square error of kernel estimators of spherical density. *Science in China (Serie A)*, **44**, 474-483.