



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**INFERENCIA BAYESIANA PARA DATOS  
CIRCULARES**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**MATEMÁTICO**

**P R E S E N T A:**

**UZIEL JOSÉ GONZÁLEZ GONZÁLEZ**

**DIRECTOR DE TESIS:  
DR. SERGIO JUÁREZ CERRILLO  
2012**



Universidad Nacional  
Autónoma de México



**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.



Universidad Nacional  
Autónoma de México



**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.

## Resumen

En esta tesis proponemos un enfoque de modelación estadística para una serie de tiempo de direcciones de viento observadas en la central núcleo-eléctrica Laguna Verde (CNLV). El objetivo de la tesis es proporcionar uno de los componentes de un modelo espacio-temporal para las series de tiempo de viento observadas en ocho estaciones meteorológicas ubicadas en los alrededores de la CNLV.

El modelo que se ajusta toma en cuenta que la dirección del viento es una magnitud angular. En estadística este tipo de datos, direcciones en el espacio, se conocen como datos circulares o direccionales y requieren de un tratamiento especial. Existen diversos enfoques para analizar datos circulares y en este trabajo seguimos el enfoque de *envolver*. La idea básica de este enfoque es enroscar una distribución con soporte en la línea real alrededor de un círculo unitario. En el modelo para las direcciones de viento consideramos un proceso autorregresivo para una variable normal envuelta. El ajuste del modelo lo hicimos con el algoritmo Metropolis-Hastings.

La efectividad de nuestro enfoque la evaluamos con diversas simulaciones. En la tesis presentamos los resultados de un par de estas simulaciones. En la primer simulación estimamos los parámetros de una distribución normal univariada envuelta, en la segunda ajustamos un proceso autorregresivo normal de orden 2 envuelto. En ambas simulaciones las estimaciones de los parámetros son considerablemente buenas. En este sentido, tenemos una pauta alentadora en cuanto al logro del objetivo de la tesis.

# Indice

<b>1. Introducción</b>	<b>3</b>
1.1. Datos Circulares . . . . .	3
1.2. Motivación de la Tesis . . . . .	4
1.3. Preliminares . . . . .	4
1.3.1. Inferencia Bayesiana . . . . .	4
1.3.2. Método de Monte Carlo . . . . .	6
1.3.3. Simulación . . . . .	6
1.3.4. Monte Carlo vía cadenas de Markov . . . . .	7
1.3.5. Variables latentes . . . . .	7
1.4. Estructura de la Tesis . . . . .	8
<b>2. Distribución Normal Envuelta</b>	<b>9</b>
2.1. Variable Envuelta . . . . .	9
2.2. Distribución Normal Envuelta . . . . .	11
<b>3. Inferencia Bayesiana para la Distribución Normal Envuelta</b>	<b>13</b>
3.1. Algoritmo Metropolis Hastings para la Distribución Normal Envuelta . . . . .	16
3.2. Simulaciones . . . . .	19
3.2.1. Distribución Normal Univariada Envuelta . . . . .	19
3.2.2. Proceso Autorregresivo Envuelto . . . . .	23
<b>4. Modelación de la Serie de Tiempo de Direcciones de Viento de la CNLV</b>	<b>30</b>
4.1. Datos de Direcciones de Viento . . . . .	31

4.2. El Modelo WAR(2) Ajustado . . . . .	31
<b>5. Conclusiones e Investigación Futura</b>	<b>34</b>
<b>Bibliografía</b>	<b>35</b>
<b>Apéndice</b>	<b>37</b>
<b>A. Monte Carlo vía Cadenas de Markov</b>	<b>38</b>
<b>B. Programa en R</b>	<b>41</b>

# Capítulo 1

## Introducción

### 1.1. Datos Circulares

Los datos circulares son observaciones de direcciones en el espacio. Un dato circular es un ángulo, por lo que se puede considerar como un punto del círculo unitario. En este trabajo nos referiremos a los datos definidos sobre  $\mathbb{R}$  o un subconjunto de  $\mathbb{R}$  como datos sobre la recta, mientras que los datos circulares están definidos en  $[0, 2\pi)$ .

Para analizar datos circulares se deben hacer diversas consideraciones. Por ejemplo, la dependencia de los resultados descriptivos e inferenciales con respecto al punto de referencia en el círculo unitario, hace que la naturaleza de los datos circulares sea diferente a la de los datos sobre la recta. Por lo que el análisis de datos circulares requiere de métodos estadísticos y modelos especiales.

Los datos circulares surgen en cualquier contexto donde se observen magnitudes angulares. Por ejemplo, aparecen en ciencias de la tierra, meteorología, (análisis de la dirección de viento), biología (estudios de la dirección de movimientos de especies), física y, en general, en cualquier contexto donde se registren direcciones.

La literatura sobre datos circulares no es extensa comparada a la de los datos sobre  $\mathbb{R}$ . Breckling (1989), Mardia (1972), Mardia y Jupp (1999), y Fisher (1994) son algunas referencias básicas para el análisis estadístico de datos circulares. Estos trabajos exponen procedimientos de estadística descriptiva, inferencia y modelación con regresión y modelos para series de tiempo. El enfoque de inferencia que utilizan estos autores es el frecuentista.

Sin embargo, no es difícil encontrar situaciones en las cuales los procedimientos de inferencia clásica tengan una implementación más compleja. Por ejemplo, la verosimilitud para una distribución normal envuelta no tiene forma cerrada (la densidad de la normal envuelta es una serie infinita) y por lo tanto su maximización es prácticamente intratable. Esto conduce a la necesidad de desarrollar procedimientos de inferencia estadística que permitan el ajuste de modelos para datos circulares. En este sentido, encontramos en el paradigma de la inferencia bayesiana un enfoque de análisis lo suficientemente poderoso y flexible para el tratamiento de datos circulares.

## **1.2. Motivación de la Tesis**

La motivación de esta tesis nace de la necesidad de modelar mediciones espacio-temporales de viento registradas en ocho estaciones meteorológicas localizadas en los alrededores de la CNLV. El objetivo del modelo es hacer pronóstico a corto plazo de la dirección y velocidad del viento.

El viento se registra en un vector de dos componentes, uno es la dirección y el otro es la velocidad. La dirección es una variable circular (un ángulo) y la velocidad es una variable sobre  $\mathbb{R}$ . De este modo, un modelo para viento debe considerar a una variable circular.

## **1.3. Preliminares**

### **1.3.1. Inferencia Bayesiana**

Cuando planteamos un problema estadístico, nuestra finalidad es poder hacer conclusiones acerca de fenómenos inciertos a partir de un conocimiento previo y de observaciones que se realizan sobre los mismos.

Dentro de la estadística clásica los planteamientos más comunes son la estimación puntual, la estimación por intervalos y el contraste de hipótesis. Pero debido a la naturaleza de algunos problemas, resulta imposible abordarlos mediante este enfoque.

A diferencia de lo que ocurre con los métodos clásicos; la flexibilidad de la inferencia bayesiana nos permite hacer un planteamiento más natural de una amplia variedad de



problemas. Otra de las ventajas del enfoque bayesiano es el diseño que hace de una teoría estadística, basada en una serie de principios básicos, que teóricamente nos permitiría dar una solución a cualquier problema de inferencia.

En la inferencia bayesiana en general tenemos los datos,  $\mathbf{X}$ , y unas cantidades desconocidas,  $\boldsymbol{\theta}$ , cuyo valor es de nuestro interés. Estas cantidades pueden ser de diversos tipos, tales como parámetros de un modelo, observaciones faltantes o mediciones que no podemos observar directamente. Si quisieramos resumir los pasos a seguir dentro del enfoque bayesiano podríamos hacerlo del siguiente modo:

- 1) Especificación de un modelo. Se debe de elegir un modelo para describir el proceso que generó los datos. El modelo elegido,  $P(\mathbf{x}|\boldsymbol{\theta})$ , dependerá del tipo de problema y siempre debe de tomarse en cuenta sus limitaciones.
- 2) Especificación de una distribución inicial,  $P(\boldsymbol{\theta})$ . Este punto es fundamental dentro del enfoque bayesiano, en conjunto con la interpretación subjetiva de la probabilidad distinguen la inferencia bayesiana de los métodos clásicos. La asignación de esta distribución inicial, en general, no es algo sencillo. En cualquier caso siempre debe asignarse un distribución inicial consistente con nuestra información previa. También debemos considerar que en caso de carecer de información inicial acerca del parámetro  $\boldsymbol{\theta}$ , debemos ser capaces de reflejar este desconocimiento. Esto se hace por medio de las llamadas distribuciones iniciales no informativas.
- 3) Cálculo de la distribución final. Si consideramos que  $\mathbf{X}$  es conocido, podemos condicionar en su valor observado  $\mathbf{x}$ . Por lo tanto nuestro conocimiento acerca de  $\boldsymbol{\theta}$  queda descrito por medio de la distribución final  $P(\boldsymbol{\theta}|\mathbf{x})$ . Y el teorema de Bayes nos ayuda a calcular la distribución final

$$P(\boldsymbol{\theta}|\mathbf{x}) = \frac{P(\boldsymbol{\theta})P(\mathbf{x}|\boldsymbol{\theta})}{\int_{\Theta} P(\tilde{\boldsymbol{\theta}})P(\mathbf{x}|\tilde{\boldsymbol{\theta}})d\tilde{\boldsymbol{\theta}}}.$$

El denominador  $P(\mathbf{x}) = \int_{\Theta} P(\tilde{\boldsymbol{\theta}})P(\mathbf{x}|\tilde{\boldsymbol{\theta}})d\tilde{\boldsymbol{\theta}}$ , no depende de  $\boldsymbol{\theta}$ , por lo que es común escribir

$$P(\boldsymbol{\theta}|x) \propto P(\boldsymbol{\theta})P(\mathbf{x}|\boldsymbol{\theta}).$$

En la práctica, el cálculo de la distribución final puede ser un asunto complicado, especialmente si la dimensión del parámetro no es pequeña. Y en otros casos se requieren técnicas computacionales relativamente sofisticadas.

- 4) Resumen de la información. Habiendo encontrado  $P(\boldsymbol{\theta}|\mathbf{x})$ , podemos decir que hemos resuelto el problema de inferencia; ya que toda la información disponible sobre el valor de  $\boldsymbol{\theta}$  queda representada por medio de la distribución final.

### 1.3.2. Método de Monte Carlo

La implementación de los métodos bayesianos involucra un esfuerzo computacional considerablemente alto. Este esfuerzo se concentra en el cálculo de ciertas características de la distribución final del parámetro de interés.

Por ejemplo, para pasar de una distribución conjunta a una colección de distribuciones y momentos marginales que sean útiles para hacer inferencias sobre subconjuntos de parámetros, es necesario integrar. Dentro de los métodos más usados para el cálculo de estas integrales se encuentran los métodos de Monte Carlo. La idea básica del método consiste en escribir la integral requerida

$$I = \int h(\theta) d\theta$$

como valor esperado de una función con respecto a alguna distribución de probabilidad, lo cual sugiere una solución estadística al problema de integración. Escribimos a  $I$  como un valor esperado

$$I = \int \left\{ \frac{h(\theta)}{s(\theta)} \right\} s(\theta) d\theta = E_{s(\theta)} \left\{ \frac{h(\theta)}{s(\theta)} \right\},$$

donde  $s(\theta)$  es una función fácil de simular. Después se genera una muestra  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}$  de  $s(\theta)$  con  $M$  grande. Finalmente se aproxima  $I$  a partir de la media muestral

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M \left\{ \frac{h(\theta^{(j)})}{s(\theta^{(j)})} \right\} s(\theta^{(j)}).$$

### 1.3.3. Simulación

El método de Monte Carlo permite calcular algunos resúmenes inferenciales elementales tales como valores esperados de ciertas funciones del parámetro respecto a la distribución

final. Es posible, además, aprovechar la muestra simulada con el fin de aproximar cualquier otra característica de interés de la distribución final. De hecho, existe una dualidad muy interesante entre una muestra dada y la distribución que la genera. Por un lado, la distribución permite generar la muestra; por otro lado, dada la muestra siempre es posible recrear la distribución que la generó (al menos aproximadamente).

### 1.3.4. Monte Carlo vía cadenas de Markov

Las técnicas de Monte Carlo vía cadenas de Markov permiten generar, de manera iterativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse usando métodos directos. La idea básica es simple: construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución final que nos interesa. Los dos algoritmos usados en la tesis son: Metropolis-Hastings y el muestreo de Gibbs. (Ver apéndice A).

En general no es fácil determinar en que momento la cadena ha convergido, un método empírico comúnmente utilizado, consiste en graficar las funciones de autocorrelación para ver el comportamiento de las cadenas; frecuentemente se omiten los primeros valores de la cadena para calcular estas funciones o los promedios ergódicos. La finalidad de este periodo de calentamiento es permitir que la cadena salga de una primera fase de inestabilidad.

### 1.3.5. Variables latentes

Existe una técnica general conocida como el método de las variables latentes. La idea consiste en observar que en ocasiones es más fácil y eficiente simular de una distribución  $P(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{x})$  que de la distribución  $P(\boldsymbol{\theta}|\mathbf{x})$ , donde el nuevo parámetro  $\boldsymbol{\lambda}$  puede ser esencialmente cualquier variable aleatoria. Es claro que al generar una muestra con distribución conjunta  $P(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{x})$  podemos obtener una muestra con distribución marginal  $P(\boldsymbol{\theta}|\mathbf{x})$  que nos interesaba originalmente.

## 1.4. Estructura de la Tesis

El resto de esta tesis está estructurado de la siguiente forma. En el Capítulo 2 proporcionamos una breve introducción a la distribución normal para datos circulares. En particular presentamos el enfoque de envolver para el manejo de datos circulares. En el Capítulo 3 revisamos con detalle el proceso de envolver. También se discuten las razones por las que consideramos que este enfoque es el más adecuado para analizar direcciones de viento. Siguiendo el trabajo de Coles (1998) describimos la estimación bayesiana de los parámetros de una distribución normal univariada envuelta. Siguiendo el mismo enfoque, aplicamos el método para hacer el ajuste de un proceso autorregresivo de orden 2 para una normal envuelta. Los modelos y procedimientos de inferencia del Capítulo 3 los ponemos en práctica en el Capítulo 4 al analizar la serie de tiempo de direcciones de viento de la CNLV. En el Capítulo 5 hacemos una reflexión sobre los resultados obtenidos e indicamos posibles líneas y direcciones de investigación. En el Apéndice hacemos una revisión breve del algoritmo Metropolis-Hastings, así como de algunas de sus variantes. Finalmente, incluimos el código en R para ajustar el modelo autorregresivo envuelto de orden 2.

# Capítulo 2

## Distribución Normal Envuelta

Una de las aproximaciones para el estudio de los datos circulares consiste en envolver una distribución conocida en la línea real, alrededor de la circunferencia de un círculo unitario. Su principal característica es la flexibilidad: una amplia clase de funciones sobre el círculo puede ser obtenida usando dicha técnica. En principio este procedimiento puede ser usado para envolver cualquier distribución.

### 2.1. Variable Envuelta

Sea  $Y$  una variable aleatoria sobre la línea con función de densidad  $f(y)$ . La correspondiente variable aleatoria circular  $X_w$  se obtiene al envolver de la siguiente manera:

$$X_w = Y(\text{mod } 2\pi).$$

La función de densidad de probabilidad  $f_w(\theta)$  de  $X_w$  se obtiene al envolver  $f(y)$ , definida en  $\mathbb{R}$ , alrededor de la circunferencia de un círculo unitario, esto es

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2k\pi), \quad 0 \leq \theta < 2\pi. \quad (2.1)$$

La correspondiente función de distribución de  $X_w$  es

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} F(\theta + 2k\pi) - F(2k\pi), \quad 0 \leq \theta < 2\pi, \quad (2.2)$$

donde  $F$  es la función de distribución de  $Y$ .

Algunas de sus propiedades como se pueden ver en Mardia (1972) son:

a) El proceso de envolver es un homomorfismo de  $\mathbb{R}$  al círculo, es decir  $(y_1 + y_2)_w = y_{1w} + y_{2w}$ .

b) Si la función característica de  $Y$  es  $\phi$ , entonces la función característica  $\{\phi_p : p = 0, \pm 1, \dots\}$  de  $X_w$  está dada por

$$\phi_p = \phi(p). \quad (2.3)$$

Para poder ver esto, veamos que:

$$\phi_p = \int_0^{2\pi} e^{ip\theta} F_w(\theta) d\theta = \sum_{k=-\infty}^{\infty} \int_{2k\pi}^{2\pi(k+1)} e^{ip\theta} F(\theta) d\theta = \int_{-\infty}^{\infty} e^{ipy} F(y) dy = \phi(p).$$

c) Si  $\phi$  es integrable entonces  $Y$  tiene una densidad y

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2k\pi) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right\} \quad (2.4)$$

donde  $\phi(p) = \alpha_p + i\beta_p$ . Para ver esto, notemos que  $\phi$  es integrable y

$$\sum_{p=-\infty}^{\infty} |\phi_p|^2 \leq \sum_{p=-\infty}^{\infty} |\phi_p| \leq \int_{-\infty}^{\infty} |\phi(p)| dp.$$

Por lo tanto,  $\sum_{p=1}^{\infty} (\alpha_p^2 + \beta_p^2)$  es convergente y el resultado se sigue de (2.1) y (2.3).

d) Si  $Y$  es infinitamente divisible, entonces  $X_w$  es infinitamente divisible.

e) Existen una infinidad de distribuciones definidas sobre la línea real, las cuales pueden ser envueltas sobre cualquier distribución en el círculo. Para ver esto, sea  $g$  la función de densidad de probabilidad de una distribución en el círculo y defina una función de densidad de probabilidad como

$$f_r(x) = p_r g(x), \quad 2\pi r \leq x \leq 2\pi(r+1), \quad r = 0, \pm 1, \pm 2, \dots,$$

donde  $p_r$  son números no negativos cualesquiera tales que  $\sum_{r=-\infty}^{\infty} p_r = 1$ . Si consideramos  $f(x) = \sum_{r=-\infty}^{\infty} f_r(x)$ , entonces  $f_w = g$ .

## 2.2. Distribución Normal Envuelta

La *distribución normal envuelta* es una función con dos parámetros, la cual se obtiene al envolver la línea de una distribución normal  $N(\tilde{\mu}, \sigma)$  alrededor del círculo.

De (2.1) la función de densidad de probabilidad de una distribución normal envuelta es

$$f_w(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{\frac{-(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta < 2\pi. \quad (2.5)$$

Dado que la función característica de  $N(\tilde{\mu}, \sigma^2)$  está dada por  $\phi(p) = \exp(i\mu p - p^2\sigma^2/2)$ , de la propiedad (b) de las distribuciones envueltas se tiene

$$\phi_p = e^{i\mu p - p^2\sigma^2/2}, \quad \alpha_p = e^{-p^2\sigma^2/2} \cos p\mu, \quad \beta_p = e^{-p^2\sigma^2/2} \sen p\mu.$$

Usando (2.4) obtenemos una útil representación de la función de densidad (2.6) como lo es

$$f_w(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right\}, \quad (2.6)$$

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

La abreviatura  $WN(\tilde{\mu}, \sigma)$  que usaremos, se debe a que tomamos las primeras letras del nombre en inglés de esta distribución, *Wrapped Normal distribution*.

Para fines prácticos, la función de densidad de probabilidad (2.7) se puede aproximar adecuadamente por los primeros tres términos cuando  $\sigma^2 > 2\pi$ , mientras que para  $\sigma^2 \leq 2\pi$  el término con  $k = 0$  de (2.6) nos da una aproximación razonable, (Mardia y Jupp, 1999, p.50)

Los momentos y propiedades de una distribución normal envuelta son:

- i) La dirección media es  $\mu = \tilde{\mu}(\text{mod } 2\pi)$ .
- ii) La longitud resultante media es  $\rho = e^{-\sigma^2/2}$ . Las propiedades *i*) y *ii*) indican una correspondencia entre los valores en la línea y en la distribución circular.
- iii) Dispersión circular:  $\delta = (1 - \rho^4)/(2\rho^2)$ .
- iv) Conforme  $\rho \rightarrow 0$ , la distribución converge a la distribución Uniforme  $U_c$ .
- v) Conforme  $\rho \rightarrow 1$ , la distribución tiende a una distribución degenerada concentrada en la dirección  $\mu$ .

- vi) La moda está en  $\theta = \mu$ .
- vii) Si  $\theta \sim WN(\mu, \rho)$  entonces  $(\theta - \psi) \sim WN(\mu - \psi, \rho)$ .
- viii) De la propiedad a) de las distribuciones envueltas y de la ecuación (2.5) se sigue que si  $\Theta_1$  y  $\Theta_2$  son dos variables Normales envueltas,  $\Theta_i \sim WN(\mu_i, \rho_i), (i = 1, 2)$ , entonces  $(\Theta_1 + \Theta_2) \sim WN(\mu_1 + \mu_2, \rho_1\rho_2)$ .

Aquí sólo presentamos la *distribución normal envuelta*, pero en teoría es posible generar un número infinito de distribuciones circulares, tantas como hay definidas en la línea. En la mayoría de las distribuciones envueltas existe una correspondencia entre los parámetros originales y sus respectivos parámetros envueltos. Esto nos permite una fácil interpretación de los parámetros circulares y por lo tanto de la inferencia.



# Capítulo 3

## Inferencia Bayesiana para la Distribución Normal Envuelta

En el capítulo anterior anticipábamos el modo en que el envolver una distribución nos puede ayudar a construir modelos para datos circulares. Este proceso lo aplicamos a una función  $f(y)$  definida sobre  $\mathbb{R}$  para obtener la función de densidad envuelta en el círculo unitario  $f_w(x)$  dada por

$$f_w(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi), \quad 0 \leq x < 2\pi. \quad (3.1)$$

Así, obtenemos una variable aleatoria circular  $X$  definida en  $[0, 2\pi)$  a través de la transformación  $X + 2k\pi$ , con  $k \in \mathbb{Z}$  una variable no observable. Esto es, obtenemos  $f_w$  aplicando una función de densidad en la línea,  $f$ , al argumento  $X + 2k\pi$ . La variable aleatoria  $X$  es una versión envuelta de la variable aleatoria en la línea,  $Y$ , la cual tiene función de densidad de probabilidad  $f(y|\Psi)$  que depende de un vector de parámetros desconocido  $\Psi$ . La ventaja de esta clase de modelos, y en particular del modelo normal envuelto, es la sencillez para extenderlo al caso multivariado circular. Por ejemplo, supongamos que  $Y$  es un vector aleatorio  $d$ -dimensional con distribución normal multivariada y función de densidad  $f(y|\mu, \Sigma)$ . El proceso de envolver  $X_w = Y \pmod{2\pi}$  se aplica componente a componente para obtener una distribución multivariada  $d$ -dimensional sobre el círculo, con densidad

$$f_w(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_d=-\infty}^{\infty} f(x_1 + 2k_1\pi, \dots, x_d + 2k_d\pi). \quad (3.2)$$

Marginalmente, cada variable  $X_i$  tiene una distribución normal envuelta univariada, mientras la estructura de correlación está definida por la matriz de varianzas-covarianzas  $\Sigma$  de la distribución en la línea  $Y$ . Supongamos que tenemos  $n$  repeticiones de la variable Normal  $d$ -dimensional  $Y$ ; en forma matricial podemos escribir:

$$\mathbf{Y} = \mathbf{X} + 2\mathbf{k}\pi, \quad (3.3)$$

donde  $\mathbf{k} = \{k^{(1)}, \dots, k^{(n)}\}'$  y  $\mathbf{X} = \{X^{(1)}, \dots, X^{(n)}\}'$  denota un conjunto de  $n$  variables aleatorias circulares  $d$ -dimensionales. Cada  $X^{(i)}, i = 1, \dots, n$  tiene función de densidad sobre el dominio  $[0, 2\pi)^d$  y tal que  $X^{(i)}, X^{(j)}, i \neq j$  se suponen independientes e idénticamente distribuido. Siendo más precisos, la ecuación (3.3) se puede reescribir como:

$$\begin{pmatrix} Y^{(1)} \\ \vdots \\ Y^{(i)} \\ \vdots \\ Y^{(n)} \end{pmatrix} = \begin{pmatrix} X^{(1)} + 2k^{(1)}\pi \\ \vdots \\ X^{(i)} + 2k^{(i)}\pi \\ \vdots \\ X^{(n)} + 2k^{(n)}\pi \end{pmatrix} = \begin{pmatrix} X_1^{(1)} + 2k_1^{(1)}\pi & \cdots & X_d^{(1)} + 2k_d^{(1)}\pi \\ \vdots & & \vdots \\ X_1^{(i)} + 2k_1^{(i)}\pi & \cdots & X_d^{(i)} + 2k_d^{(i)}\pi \\ \vdots & & \vdots \\ X_1^{(n)} + 2k_1^{(n)}\pi & \cdots & X_d^{(n)} + 2k_d^{(n)}\pi \end{pmatrix}.$$

Entonces, el vector observado  $X^{(i)}$  es un componente del par  $(X^{(i)}, k^{(i)})$ , donde el componente  $k^{(i)}$ , al que nos referiremos como el vector de coeficientes de envoltura, no es observable. Esta construcción se puede extender fácilmente a procesos estocásticos circulares. De hecho, considerando los  $d$ -componentes de (3.2) como variables de un proceso observado en  $d$  puntos temporales o  $d$  puntos espaciales  $(p_1, \dots, p_d)$  y denotando por  $f_d$  la densidad conjunta marginal de este proceso, la correspondiente densidad envuelta está dada por:

$$f_d(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_d=-\infty}^{\infty} f(p_1 + 2k_1\pi, \dots, p_d + 2k_d\pi). \quad (3.4)$$

Estos modelos se han estudiado para series de tiempo por Breckling (1989), mientras que la teoría general de distribuciones circulares envueltas se discute en Fisher (1993) y Mardia y Jupp(1999).

La extensión para el caso multivariado dado por (3.2) y (3.4) es una de las mayores ventajas de este enfoque. Debido a la complejidad de los datos, la posibilidad de modelar

bases más grandes, a menudo en un marco multivariado, es uno de los puntos medulares de las nuevas metodologías. Por esta razón, en los últimos años se han hecho esfuerzos para extender las distribuciones circulares más usadas al caso multivariado, entre ellas la von Mises. Sólo en la literatura más reciente podemos encontrar algunos resultados al respecto. Las distribuciones y procesos envueltos nos proporcionan una conveniente e intuitiva opción para generar modelos de datos circulares. Más aun, dichos modelos tienen grandes ventajas como la facilidad de interpretación debido a la correspondencia entre los parámetros circulares y envueltos; ver propiedades *i*) y *ii*) de la distribución normal envuelta. Todo esto debido a que las propiedades de la distribución definida en la línea se preservan bajo el proceso.

La dificultad al usar estos modelos se debe a la complejidad en la inferencia: trabajar directamente con la densidad (3.1) o (3.2) hace la estimación de parámetros intratable. Una de las soluciones fue desarrollada por Fisher y Lee (1994) quienes usaron el algoritmo EM para hacer la estimación, considerando los coeficientes de envoltura, los  $k$ 's, como datos perdidos. A pesar de la elegancia del método, éste presenta una elevada complejidad computacional: el paso E involucra una suma infinita de cocientes, los cuales deben de ser aproximados en cada paso, haciendo el algoritmo computacionalmente ineficiente.

En este trabajo seguiremos el procedimiento propuesto por Coles (1998). En particular, profundizaremos en el estudio del papel que desempeñan los coeficientes de envoltura  $k$ 's en la estimación de parámetros, con la finalidad de aclarar algunas dificultades que menciona Coles.

La principal dificultad al trabajar bajo este enfoque es que la forma de la función de densidad se construye a partir de sumas grandes, y no puede calcularse de manera más simple. Para poder resolver este inconveniente, supondremos que obtenemos una distribución circular al envolver sobre el círculo alguna distribución sobre la línea. En ese caso es posible desenvolver la distribución sobre el círculo y obtener una distribución sobre la línea real, llegando a un punto en que será posible usar todas las técnicas estadísticas estándar para las variables definidas sobre la línea real. En otras palabras dada una variable aleatoria circular,  $X$ , siempre es posible obtener la correspondiente variable aleatoria sobre la línea,  $Y$ , como  $Y = X + 2k\pi$ , donde  $X$  es definida en  $[0, 2\pi)$  y  $k$  es una variable no observable que representa el número de veces que  $X$  se desenvuelve para obtener  $Y$ . Por lo tanto, es

posible agregar la información de  $k$ , y entonces desenvolver  $X$ ; de esta manera estaríamos trabajando directamente con  $Y$ .

En este sentido, hacer inferencia sobre  $X$ , cuya función de densidad es  $f_w(y)$  escrita en (3.1), resulta imposible. En su lugar la función de densidad en la línea es más simple y describe la variable  $Y$  o, equivalentemente el par  $(X, k)$  del cual sólo observamos  $X$ . La equivalencia

$$Y \equiv (X, k) \tag{3.5}$$

se tiene que interpretar como la distribución conjunta de  $(X, k)$  con argumento  $X + 2k\pi$  y vector de parámetros  $\Psi$ .

Como  $k$  es una variable no observable, una aproximación natural (pero sobre todo conveniente) que nos permite superar la falta de certeza sobre  $k$  escribiéndola como distribución condicional, es el enfoque bayesiano. En particular, los métodos de Monte Carlo para cadenas markovianas (MCMC por sus siglas en inglés). La base de dicho método es simular una cadena de Markov cuya distribución de equilibrio es proporcional a la función objetivo, la cual, en nuestro caso, es la distribución final de  $(\mathbf{k}, \Psi)$ .

### 3.1. Algoritmo Metropolis Hastings para la Distribución Normal Envuelta

El método MCMC más general es el algoritmo Metropolis-Hastings (M-H) el cual utilizaremos en este caso (la versión general del algoritmo puede consultarse en el Apéndice A). Los vectores aleatorios circulares  $X^{(1)}, \dots, X^{(n)}$  se suponen independientes e idénticamente distribuidos con función de probabilidad  $f_w$  con dominio  $[0, 2\pi)^d$ . En nuestro modelo,  $X^{(i)}$  de hecho representa un componente del vector  $Y^{(i)} = X^{(i)} + 2k^{(i)}\pi = (X_1^{(i)} + 2k_1^{(i)}\pi, \dots, X_d^{(i)} + 2k_d^{(i)}\pi)$ , cuya función de densidad de probabilidad es  $g$  con dominio  $\mathbb{R}^d$  y parámetros desconocidos  $\Psi$ . Entonces, como se puede ver, el vector observado  $X^{(i)}$  es un componente del vector  $(X^{(i)}, k^{(i)})$ . El componente  $k^{(i)}$  no es observable. El modelo seleccionado es bastante simple (normal univariado o multivariado) y el objetivo es hacer inferencia para la distribución en ausencia de los coeficientes de envoltura. Esto a su vez, caracteriza el comportamiento

estocástico para el vector circular  $X^{(i)}$ .

En lo siguiente será conveniente adoptar las siguientes convenciones de notación. La muestra observada de tamaño  $n$  se denota como

$$\mathbf{X} = \{X^{(1)}, \dots, X^{(n)}\};$$

de manera similar

$$\mathbf{k} = \{k^{(1)}, \dots, k^{(n)}\}.$$

Los vectores o componentes perdidos los denotaremos con subíndices o superíndices negativos, por ejemplo:

$$\mathbf{k}^{(-i)} = \{k^{(1)}, \dots, k^{(i-1)}, k^{(i+1)}, \dots, k^{(n)}\}$$

$$k_{-j}^{(i)} = \{k_1^{(i)}, \dots, k_{j-1}^{(i)}, k_{j+1}^{(i)}, \dots, k_d^{(i)}\}.$$

Donde no presente ninguna ambigüedad, denotaremos un elemento de  $X^{(i)}$  como  $X$ .

Por razones de viabilidad es conveniente hacer inferencia desde un punto de vista bayesiano, el cual nos permite trabajar dentro de la incertidumbre provocada por los valores de  $k$ . El enfoque bayesiano clásico continua siendo intratable, pero el ajuste del modelo es particularmente manejable con los métodos MCMC.

Tal como se mencionó antes, el algoritmo M-H es el método más general: dado el valor actual  $t^{(i)} = t$ , un valor propuesto  $t'$  se simula de una densidad de transición  $q(t'|t)$ . Entonces  $t^{(i+1)}$  se actualiza de acuerdo a la siguiente regla:

$$t^{(i+1)} = \begin{cases} t' & \text{con probabilidad } \Delta(t^{(i)}, t') \\ t^{(i)} & \text{con probabilidad } 1 - \Delta(t^{(i)}, t') \end{cases} \quad (3.6)$$

donde

$$\Delta(t, t') = \min \left\{ 1, \frac{\pi(t') q(t|t')}{\pi(t) q(t'|t)} \right\}. \quad (3.7)$$

Bajo ciertas condiciones sobre  $q$ , la sucesión  $\{t^{(i)}\}$  tiene una distribución de equilibrio proporcional a  $\pi$ . En la mayoría de las situaciones, incluyendo nuestro caso,  $t$  es multivariado,  $t = \{\Psi, \mathbf{k}\}$ . En este caso el algoritmo se puede aplicar en ciclos a subconjuntos de  $t$ , simplemente reemplazando  $\pi(t)$  en (3.7) con la densidad condicional adecuada,  $\pi_m(t_m|t_{-m})$ , y repitiendo el ciclo a través de  $m$ , con la correspondiente transición,  $q_m$ , para cada componente; ver Smith y Roberts (1993).

Nuestra implementación del algoritmo Metropolis-Hastings se lleva cabo actualizando sucesivamente los componentes de  $\Psi$ , el vector de parámetros de la distribución en la línea, y el vector de coeficientes de envoltura,  $k$ , un componente a la vez. Notemos que en el caso multivariado, cada vector  $X^{(i)}$  tiene su propio vector  $k^{(i)} = (k_1^{(i)}, \dots, k_d^{(i)})$  de componentes perdidos; entonces el número de parámetros desconocidos en este modelo es bastante grande. Sin embargo, el algoritmo se simplifica con las siguientes identidades:

$$P(\Psi|X, k) = \frac{P(X, k)P(\Psi)}{P(X, k)} \propto P(X, k|\Psi)P(\Psi) \quad (3.8)$$

y

$$\begin{aligned} P(k_j^{(i)}|X, k^{(-i)}, k_{-j}^{(i)}, \Psi) &= P(k_j^{(i)}|X^{(i)}, k_{-j}^{(i)}, \Psi) = \frac{P(X^{(i)}, k^{(i)}|\Psi)}{P(X^{(i)}|\Psi)} \\ &\propto P(X^{(i)}, k^{(i)}|\Psi). \end{aligned} \quad (3.9)$$

Además,  $P(X^{(i)}, k^{(i)}|\Psi)$  es una función Normal (univariada o multivariada según sea el caso) mientras que  $P(X, k|\Psi)P(\Psi)$  es el producto de dichas funciones dada la independencia de  $X^{(i)}$ .

Sustituyendo las densidades posteriores (3.8) y (3.9) dentro de la ecuación (3.7) nos facilita las expresiones de probabilidad de aceptación de las transiciones propuestas en una cadena de Markov simulada.

Además de la especificación de la distribución inicial  $P(\Psi)$ , el algoritmo requiere hacer una elección de la función de transición,  $q_m$ . Generalmente esta elección puede afectar la eficiencia del algoritmo radicalmente. Para este modelo en particular, las transiciones basadas simplemente en caminatas aleatorias funcionan adecuadamente. En este caso, para cada componente de  $\Psi$ , la transición de la forma

$$t' = t + \epsilon$$

es la implementada, donde  $\epsilon \sim \text{Uniforme}[-a, a]$ . Por su parte, el modelo de transición para cada coeficiente de envoltura  $k_j^{(i)}$  es una caminata aleatoria dada por

$$k' = k + \epsilon$$

donde  $\epsilon = \{-1, 0, 1\}$  con probabilidades  $\{p, 1 - 2p, p\}$  respectivamente, para alguna elección apropiada de  $p$ . Las elecciones de  $a$  y  $p$  necesitan ser hechas con cierto juicio para obtener una convergencia razonable de la cadena simulada.

## 3.2. Simulaciones

A continuación presentamos unas simulaciones realizadas con el fin de ver cómo funciona el algoritmo M-H para la estimación de parámetros tal y como se describió en la sección anterior.

### 3.2.1. Distribución Normal Univariada Envuelta

En esta sección consideremos a las distribuciones normales  $N(0, \pi/4)$  y  $N(1, \pi)$ . De cada distribución se generaron 300 observaciones las cuales después se envolvieron en el círculo unitario. Cada una de las simulaciones es sencilla:  $Y$  se genera de una distribución normal y se envuelve para obtener  $X = Y(\text{mod } 2\pi)$ . Las gráficas 3.1 y 3.2 muestran los datos simulados resultantes.

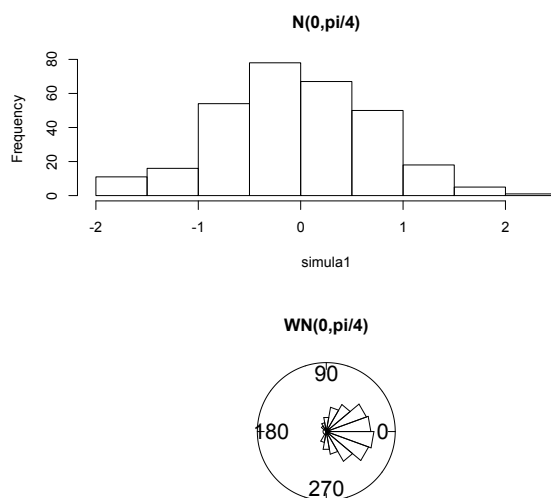


Figura 3.1: Histograma de los datos  $Y$  sobre la línea simulados de  $N(0, \pi/4)$  y el diagrama de rosa correspondiente a los datos envueltos.

Suponiendo que los coeficientes de envoltura son desconocidos, aplicamos el algoritmo M-H para estimar los parámetros. Las Figuras 3.3 y 3.4 muestran la salida del algoritmo MCMC para  $\mu$  y  $\sigma^2$  de la distribuciones  $N(0, \pi/4)$  y  $N(1, \pi)$ , respectivamente. En el Cuadro 3.1 pueden verse los resultados.

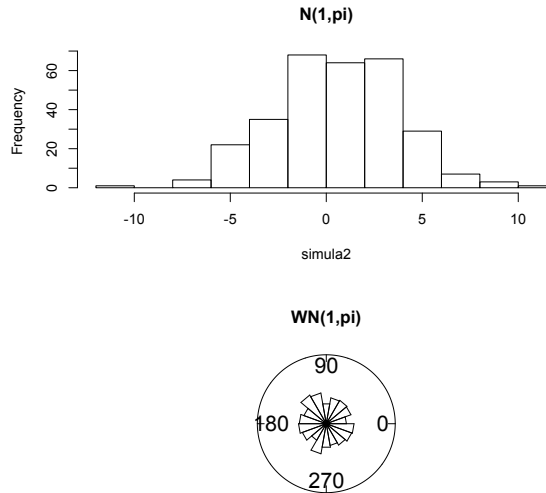


Figura 3.2: Histograma de los datos  $Y$  sobre la línea simulados de  $N(1, \pi)$  y el diagrama de rosa correspondiente a los datos envueltos.

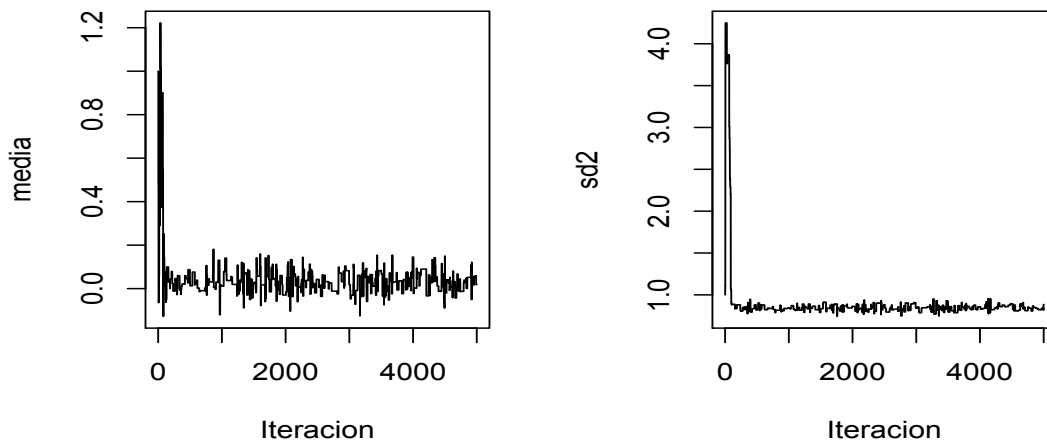


Figura 3.3: Trayectorias del algoritmo MCMC para  $\mu$ ,  $\sigma^2$  de una densidad  $N(0, \pi/4)$ .

A continuación discutimos con detalle el algoritmo MCMC para la función Normal Univariada. Antes de comenzar tenemos que aclarar de qué manera utilizaremos el algoritmo. Usando la notación de la sección anterior,  $t = (\mathbf{k}, \Psi)$  es multivariado y no sólo eso, a su vez  $\Psi = (\mu, \sigma)$  donde  $\mu$  y  $\sigma$  son la media y la desviación estándar de la distribución Normal



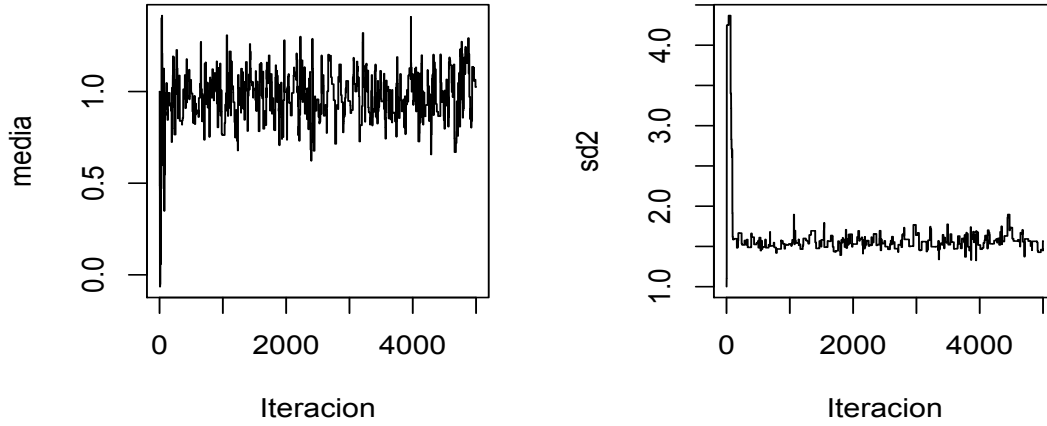


Figura 3.4: Trayectorias del algoritmo MCMC para  $\mu$ ,  $\sigma^2$  de una densidad  $N(1, \pi)$ .

Parámetros	Ejemplo 1			Ejemplo 2		
	Valor verdadero	Media final	Intervalo final 95 %	Valor verdadero	Media final	Intervalo final 95 %
$\mu$	0	0.0399	[-0.050,0.145]	1	0.9720	[0.719,1.205]
$\sigma^2$	$\pi/4$	0.7957	[0.787,0.945]	$\pi$	2.5496	[2.062,3.459]

Cuadro 3.1: Inferencia sobre los parámetros de una distribución normal univariada envuelta respectivamente. En el artículo de Coles, no se sugiere con cuál de los parámetros comenzar, pero nosotros decidimos hacerlo con los coeficientes  $k'$ s. Dado el vector  $k^{(j)}$  y siguiendo la recomendación de Coles generaremos un vector  $k^{(j+1)}$  a partir de una caminata aleatoria (ver Apéndice A ) con  $p = 1/3$  para cada uno de los elementos del vector. Utilizamos el criterio de decisión (3.7) con (3.9) para obtener:

$$\Delta(k_j, k'_j) = \min \left\{ 1, \frac{\exp \left[ -\frac{1}{2} \left( \frac{Y'_j - \mu}{\sigma} \right)^2 \right]}{\exp \left[ -\frac{1}{2} \left( \frac{Y_j - \mu}{\sigma} \right)^2 \right]} \right\},$$

donde  $Y'_j = X_j + 2k'_j\pi$  para toda  $j = 1, \dots, 300$ . Habiendo seleccionado cada uno de los componentes del vector  $k^{(j+1)}$  continuamos con el M-H en una de sus variantes (ver muestreo

de Gibbs en Apéndice A). Tenemos el vector  $k^{(j+1)}$  actualizado y valor dado de  $\mu_{(j)}$ . Tal como procedimos antes ahora generamos un valor de  $\mu^{(j+1)}$  a partir de una caminata aleatoria, que podrá o no ser seleccionado. Ahora utilizamos (3.7) con (3.8) obteniendo:

$$\Delta(\mu, \mu') = \min \left\{ 1, \frac{\prod_{j=1}^n \exp \left[ -\frac{1}{2} \left( \frac{Y_j - \mu'}{\sigma} \right)^2 \right] N(\mu', \sigma)}{\prod_{j=1}^n \exp \left[ -\frac{1}{2} \left( \frac{Y_j - \mu}{\sigma} \right)^2 \right] N(\mu, \sigma)} \right\}.$$

La expresión anterior nos permite actualizar el valor de  $\mu_{(j+1)}$ . Para terminar con un ciclo del M-H, procedemos de manera análoga, partimos del valor dado para después calcular la probabilidad de aceptación del valor propuesto de  $\sigma'$ , el cual también se obtuvo a partir de una caminata aleatoria. Como nuevo posible valor de la cadena, utilizando (3.7) con (3.8) tenemos:

$$\Delta(\sigma, \sigma') = \min \left\{ 1, \frac{\prod_{j=1}^n (2\pi\sigma'^2)^{-N/2} \frac{1}{\sigma' \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{Y_j - \mu}{\sigma'} \right)^2 \right] g(\sigma'^2, a, b)}{\prod_{i=1}^n (2\pi\sigma^2)^{-N/2} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{Y_j - \mu}{\sigma} \right)^2 \right] g(\sigma^2, a, b)} \right\}$$

donde  $a$  y  $b$  son los valores de forma y escala de la densidad gamma,  $g$  respectivamente.

Por medio de las simulaciones hechas para el caso univariado de una distribución normal, tratamos de señalar algunos problemas de inferencia que ocurren en el tratamiento de datos envueltos. Encontramos una fuerte relación entre una variabilidad grande de los datos y una inferencia pobre debido al problema de identificabilidad de los coeficientes.

Tendremos que destacar que cuando la variabilidad de los datos es pequeña, la densidad Normal envuelta

$$f(x) = (2\pi\sigma^2)^{-1/2} \sum_{k=-\infty}^{\infty} \exp \left\{ \frac{-(x - \mu + 2k\pi)^2}{2\sigma^2} \right\}$$

se aproxima adecuadamente por sólo tres valores  $\{-1, 0, 1\}$ . Mientras en el caso en que tenemos una variabilidad grande los procedimientos de inferencia fallan Ferrari (2009). Esta evidencia nos ayuda a resolver el problema de identificabilidad, al menos en caso de poca variabilidad; lo que nos permite hacer más eficiente la inferencia en los modelos de distribuciones envueltas.

Por lo tanto, el poder superar esta dificultad inicial nos permite ver la ventajas de este enfoque: la facilidad de extenderlo a casos multivariados, así como la fácil interpretación de los parámetros en términos del comportamiento de fenómenos circulares.

### 3.2.2. Proceso Autorregresivo Envuelto

El proceso de envolver puede ser aplicado a cada componente marginal de un proceso estocástico para obtener procesos estocásticos circulares. En este caso, (3.4) determina la función de densidad de probabilidad de subcomponente  $d$ -dimensional del proceso. Este argumento puede ser utilizado tanto para un proceso discreto como para uno continuo e incluso espacial.

Ilustraremos el método en el contexto del modelado de series de tiempo, restringido por simplicidad, a los procesos autoregresivos envueltos (Breckling, 1989),

$$Y_t - \mu = \sum_{j=1}^p \beta_j (Y_{t-j} - \mu) + \epsilon_t \quad (3.10)$$

donde  $\epsilon_t$  son variables independientes  $N(0, \sigma^2)$ , seguidas de la transformación componente a componente

$$X_t = Y_t \pmod{2\pi}. \quad (3.11)$$

Se puede manipular otra clase de procesos de manera similar. Nuestra presentación consiste en el análisis de un proceso autoregresivo envuelto de orden 2 que denotaremos como WAR(2). En la figura 3.5 se observa la serie simulada original (a) con su correspondiente versión envuelta (b). Una vez más, el objetivo es hacer inferencia sobre (a) a partir de (b). En principio, la distribución conjunta del proceso completo es un caso especial del modelo (3.2), con matriz de covarianzas determinada por la estructura del proceso autoregresivo, por lo que se puede ajustar usando las técnicas de la sección anterior. En la práctica, es más eficiente hacer uso de la propiedad Markoviana del modelo con la finalidad de que los cocientes requeridos para la evaluación de la probabilidad de aceptación del M-H se simplifiquen considerablemente. Primero, tal como en el ejemplo anterior, empezaremos por determinar cuáles serán los coeficientes de envoltura; para los primeros dos valores de los coeficientes tenemos una expresión diferente al resto, las expresiones se darán con detalle más adelante para evitar ambigüedades con la notación.

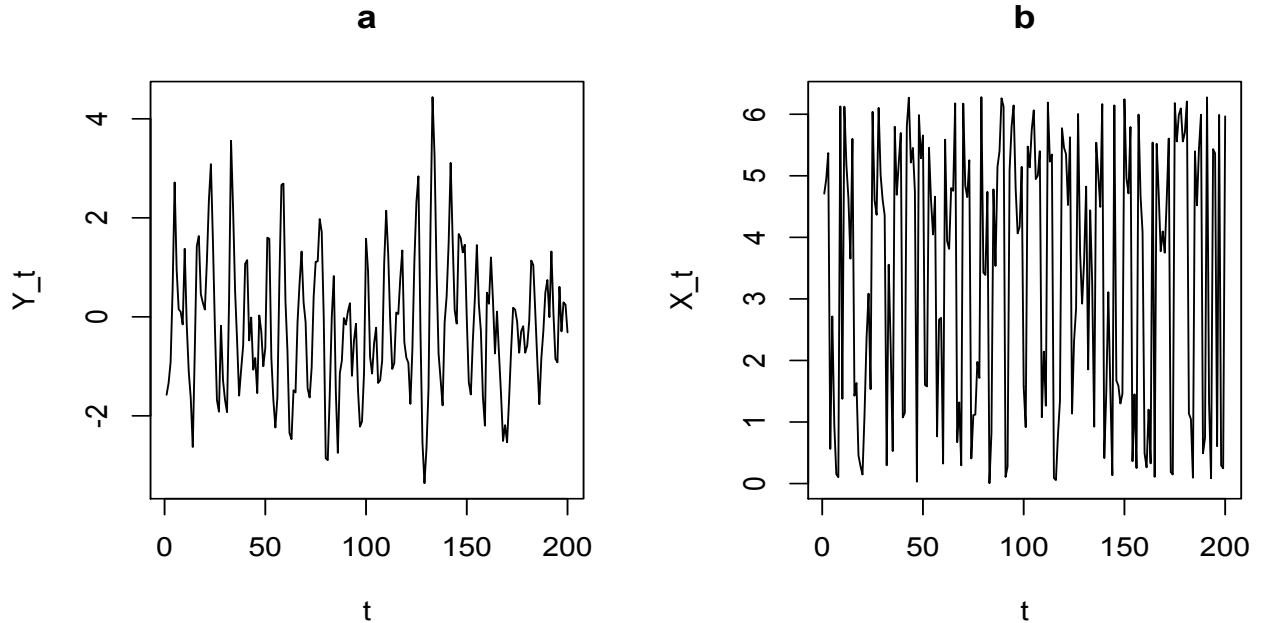


Figura 3.5: Proceso autorregresivo simulado con  $\mu = 0$ ,  $\phi_1 = 1$ ,  $\phi_2 = -0.5$  y  $\sigma^2 = 1$ . a) Antes de envolver. b) Después de envolver.

Después de haber hecho la selección de los coeficientes, podemos desenvolver los datos y así continuar con el algoritmo, tomando dos bloques más; esta parte del proceso puede verse a detalle en Chib (1994). El primer bloque corresponde al vector de parámetros de regresión y el segundo a la varianza de un proceso AR(2) que suponemos de la forma

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \quad t = 1, \dots, 200 \quad (3.12)$$

Reportamos los resultados correspondientes a un proceso con  $\mu = 0$ ,  $\phi_1 = 1$ ,  $\phi_2 = -0.5$  y  $\sigma^2 = 1$ . Para facilitar la notación, entenderemos a  $Y_n$  como el vector de datos desenvueltos componente a componente con la correspondiente  $k_i$ ; notemos que no hay necesidad de un superíndice en este caso. Los valores de  $\phi = (\phi_1, \phi_2)$  se encuentran en la región  $S \subset \mathbb{R}^2$  y satisfacen las restricciones de estacionariedad

$$\phi_1 + \phi_2 < 1; \quad -\phi_1 + \phi_2 < 1; \quad \phi_2 > -1.$$

Tomando la función de verosimilitud exacta y dados los  $n = 200$  datos  $Y_n = (y_1, y_2, \dots, y_n)'$

tenemos

$$l(\phi, \sigma^2) = \Psi(\psi, \sigma^2) \times (\sigma^2)^{-(n-2)/2} \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=3}^n (y_t - w_t' \phi)^2 \right], \quad (3.13)$$

donde  $w_t = (y_{t-1}, y_{t-2})'$ ,

$$\Psi(\psi, \sigma^2) = (\sigma^2)^{|V^{-1}|^{1/2}} \exp \left[ -\frac{1}{2\sigma^2} Y_2' V^{-1} Y_2 \right] \quad (3.14)$$

es la densidad de  $Y_2 = (y_1, y_2)'$ ,

$$V^{-1} = \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix},$$

y el tercer término de (3.13) es proporcional a la densidad de observaciones  $(y_3, \dots, y_n)$  dado  $Y_2$ . Si la única información inicial disponible es que el proceso es estacionario, entonces la distribución final de los parámetros es

$$P(\phi, \sigma^2 | Y_n) \propto l(\phi, \sigma^2) I[\phi \in S]$$

donde  $I[\phi \in S]$  es 1 si  $\phi \in S$  y 0 en otro caso. Para poder simular la distribución final debemos hacer primero dos observaciones. Primero, tal como mencionamos antes, la estrategia por bloques nos resultará útil; usaremos un bloque para  $\phi$  y otro para  $\sigma^2$ . Segundo, de la descomposición ANOVA, el término exponencial de (3.13) es proporcional a

$$\exp \left[ -\frac{1}{2\sigma^2} (\phi - \hat{\phi})' G (\phi - \hat{\phi}) \right],$$

donde  $\hat{\phi} = G^{-1} \sum_{t=3}^n (w_t y_t)$  y  $G = \sum_{t=3}^n (w_t w_t')$ . Esto es el *kernel* de una densidad normal con media  $\hat{\phi}$  matrix de covarianzas  $\sigma^2 G^{-1}$ . Estas observaciones inmediatamente nos llevan a las siguientes densidades condicionales para  $\phi$  y  $\sigma^2$ :

La densidad de  $\phi$  dado los  $k$ 's,  $X_n$  y  $\sigma^2$  es

$$P(\phi | Y_n, \sigma^2) \propto \Psi(\phi, \sigma^2) \times \{f_{nor}(\phi | \hat{\phi}, \sigma^2 G^{-1}) I[\phi \in S]\} \quad (3.15)$$

donde  $f_{nor}$  es una función de densidad normal.

La densidad de  $\sigma^2$  dado  $\phi$  los  $k$ 's y  $X_n$  es una gamma inversa con parámetros  $n/2$  y  $[Y_2 V^{-1} Y_2 + \sum_{t=3}^n (y_t - w_t' \phi)^2]/2$ .

Una muestra de  $P(\sigma^2, \phi | Y_n)$  se obtiene muestreando sucesivamente  $P(\phi | Y_n, \sigma^2)$ , y dado el valor de  $\phi$ , simular  $\sigma^2$  de  $P(\sigma^2 | Y_n, \phi)$ . Dichas expresiones se dan más adelante.

Tal como se mencionó antes, la actualización de los primeros dos coeficientes de envoltura difiere del resto. El valor simulado en la iteración  $j$  corresponde al valor sugerido por

$$\min \left\{ 1, \frac{\Psi(Y_2', \phi^{(j)}, \sigma^{2(j)})}{\Psi(Y_2, \phi^{(j)}, \sigma^{2(j)})} \right\}$$

donde  $Y_2'$  es el vector  $Y_2$  con una caminata aleatoria pertinente. Para el resto de las  $k$ 's tendríamos

$$\min \left\{ 1, \frac{\frac{1}{2\sigma^{2(j)}}(y_j' - \mu)^2}{\frac{1}{2\sigma^{2(j)}}(y_j - \mu)^2} \right\}$$

donde  $\mu = \phi_1(y_{i-1}) + \phi_2(y_{i-2})$  para  $i = 3, \dots, n$ .

El valor para  $\phi$  se simula del siguiente modo: En la iteración  $j$  (dado el valor actualizado  $\sigma^{2(j)}$ ), generamos un candidato  $\phi^{(j+1)}$  de una densidad normal con media  $\hat{\phi}$  y covarianza  $\sigma^{2(j)}G^{-1}$ ; si satisface la estacionariedad cambiamos a este valor con probabilidad

$$\min \left\{ 1, \frac{\Psi(\phi^{(j+1)}, \sigma^{2(j)})}{\Psi(\phi^{(j)}, \sigma^{2(j)})} \right\}$$

y en otro caso tomamos  $\phi^{(j+1)} = \phi^{(j)}$ , donde  $\Psi(\cdot, \cdot)$  se definió en (3.14).

Para terminar,  $\sigma^2$  simplemente se simula a partir de un función gamma inversa cuyos parámetros se dieron antes.

La identificabilidad del modelo es generalmente un problema cuando se trabaja con series de tiempo; en nuestro caso mostramos los resultados correspondientes a un proceso WAR(2). La Cadena de Markov simulada con la inferencia de dicho proceso se muestra en la Figura 3.6. La convergencia se presenta más lentamente que en las simulaciones anteriores, de hecho en la imagen sólo presentamos las últimas 10000 iteraciones de las 20000 que se realizaron, mientras que en las simulaciones anteriores no fue necesario un número tan grande. En la Figura 3.7 podemos observar tanto los promedios ergódicos como las funciones de autocorrelación para cada uno de los parámetros del modelo. La convergencia lenta surge debido a que los datos simulados causan dependencia en los coeficientes de envoltura asociados: por ejemplo, si  $Y_t$  excede  $2\pi$ , existe una posibilidad mayor de que  $Y_{t+1}$  sea mayor que  $2\pi$ ;

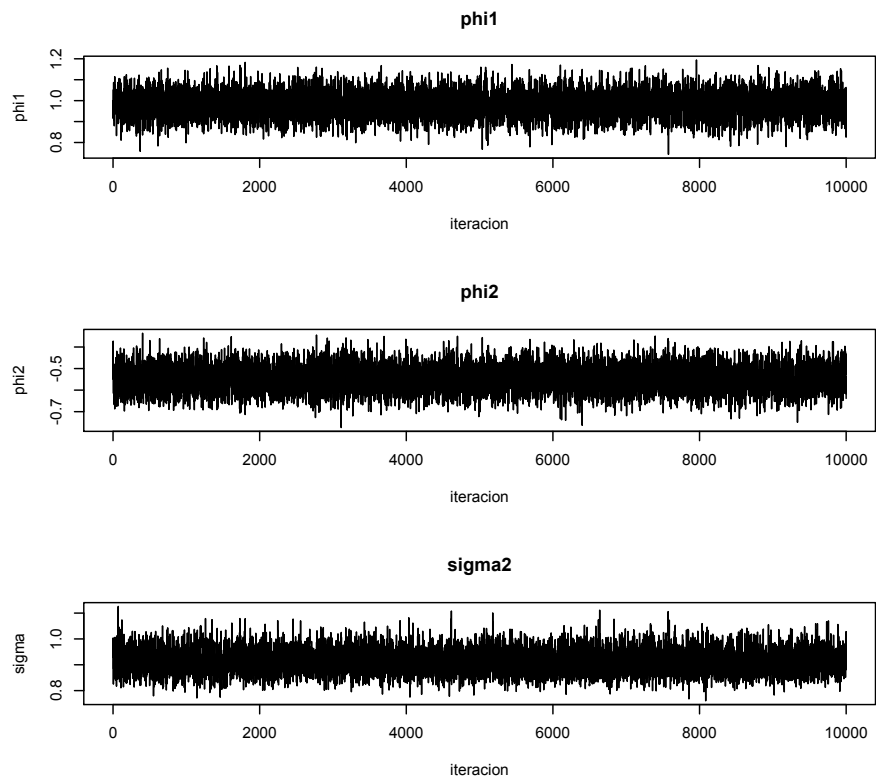


Figura 3.6: Trayectorias del algoritmo MCMC para los parámetros  $\phi_1$ ,  $\phi_2$  y  $\sigma^2$  del modelo WAR(2) cuyas observaciones se muestran en la Figura 3.5 b).

Parámetros	Valor	Media	Intervalo
	Verdadero	final	final 95 %
$\phi_1$	1	0.98	[0.638,1.096]
$\phi_2$	-0.5	-0.54	[-0.661,-0.299]
$\sigma^2$	1	0.91	[0.830,1.23]

Cuadro 3.2: Resumen de la distribución final del proceso WAR(2) simulado.

entonces tanto  $k_j$  como  $k_{j+1}$  son no cero. Sin embargo, la actualización en la rutina para cada coeficiente es aplicada independientemente del valor de los otros. Esto resulta en una rutina bastante lenta para obtener la cadena de coeficientes, los cuales son consistentes (en algunos casos totalmente) con los datos generados a partir del proceso AR(2). Por lo anterior tenemos un tiempo de convergencia muy lento.

En principio, un algoritmo modificado se podría aplicar, en el cual la regla de transición para cada coeficiente fuese dependiente de los coeficientes cercanos, en la práctica, el problema no es lo suficientemente serio como para justificar aumentar la complejidad del algoritmo, comparado con la simplicidad de correrlo por un periodo más largo.

Los parámetros son estimados razonablemente bien, aunque con un aparente sesgo. De hecho, la media final de cada parámetro se acerca a los valores de la muestra, es decir, a los valores de la serie sin envolver. Las medias finales de  $\hat{\phi}_1$ ,  $\hat{\phi}_2$  y  $\hat{\sigma}^2$  fueron 0.98, -0.54 y 0.91 respectivamente, mientras los valores muestrales son 1.0, -0.5 y 1.0; esta información se resume en el Cuadro 3.2. Lo obtenido nos sugiere que el algoritmo una vez más está estimando bien a pesar de la incertidumbre inducida por los coeficientes de envoltura desconocidos.



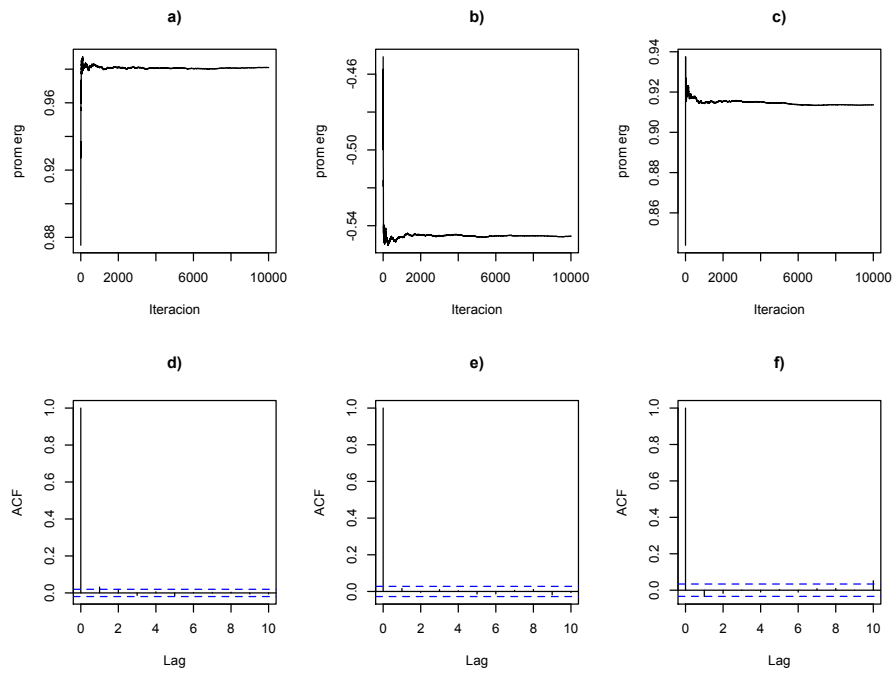


Figura 3.7: Promedios ergódicos a),b),c) de  $\phi_1$ ,  $\phi_2$  y  $\sigma^2$  respectivamente; así como sus funciones de autocorrelación d) e) y f).

## Capítulo 4

# Modelación de la Serie de Tiempo de Direcciones de Viento de la CNLV

La central núcleo-eléctrica Laguna Verde (CNLV) está ubicada en la zona centro-este del estado de Veracruz e inició operaciones en 1989. LA CNLV provee en la actualidad aproximadamente el 5% de la energía eléctrica en México.

Protocolos internacionales de seguridad requieren que en la CNLV se cuente con un modelo de dispersión de contaminantes. Este modelo tendría la finalidad de anticipar los alcances geográficos de los contaminantes expulsados a la atmósfera durante una posible contingencia. El modelo de dispersión de contaminantes requiere de la simulación del campo de viento. En la CNLV ya se cuenta con un modelo Gaussiano de viento recomendado por el Organismo Internacional de Energía Atómica. Sin embargo, se ha determinado en diversos estudios (Tejeda et al. (1994), Tejeda et al. (1989)), que este modelo no es adecuado debido a las complejas características climáticas y orográficas de la región próxima a la CNLV. También se demuestra en estos estudios que el desarrollo de un modelo dinámico no es posible, por lo que se debe abordar el problema desde un enfoque estadístico. En este capítulo aplicamos los procedimientos del Capítulo 3 para modelar la dirección del viento registrada en la CNLV.

## 4.1. Datos de Direcciones de Viento

Aplicaremos el procedimiento desarrollado en el capítulo anterior para modelar una serie de tiempo de 400 observaciones de dirección de viento registradas por hora a una altura de 10m; tomamos las primeras de 1980. En períodos largos de tiempo se aprecia una marcada no estacionariedad, pero en intervalos de tiempo cortos, como el que analizamos, se puede considerar que el proceso es estacionario. La Figura 4.1 muestra la serie de tiempo analizada.

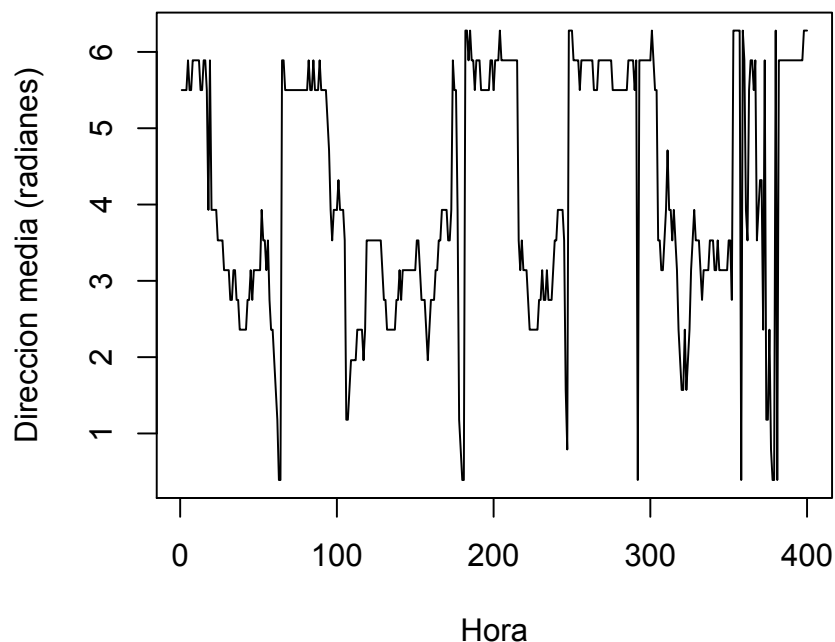


Figura 4.1: Serie de tiempo de la dirección media de viento registrada cada hora en la CNLV a partir de las 0 horas del 01/01/1980.

## 4.2. El Modelo WAR(2) Ajustado

El proceso autoregresivo envuelto definido en el capítulo anterior lo ajustaremos a esta serie. Las trayectorias del MCMC así como las distribuciones finales de cada uno de los

parámetros del modelo autoregresivo de orden dos WAR(2) se muestran en la Figura 4.2. Tenemos las siguientes estimaciones:  $\hat{\phi}_1 = 0.928$ ,  $\hat{\phi}_2 = 0.002$  y  $\hat{\sigma}^2 = 1.169$ ; lo que nos sugiere que con retraso de una hora la dependencia es bastante grande mientras que con dos retrasos la dependencia es casi nula.

Tal como ocurrió en el caso de del proceso autorregresivo envuelto, la convergencia se presenta de manera bastante lenta, en la Figura 4.3 podemos ver los promedios ergódicos.

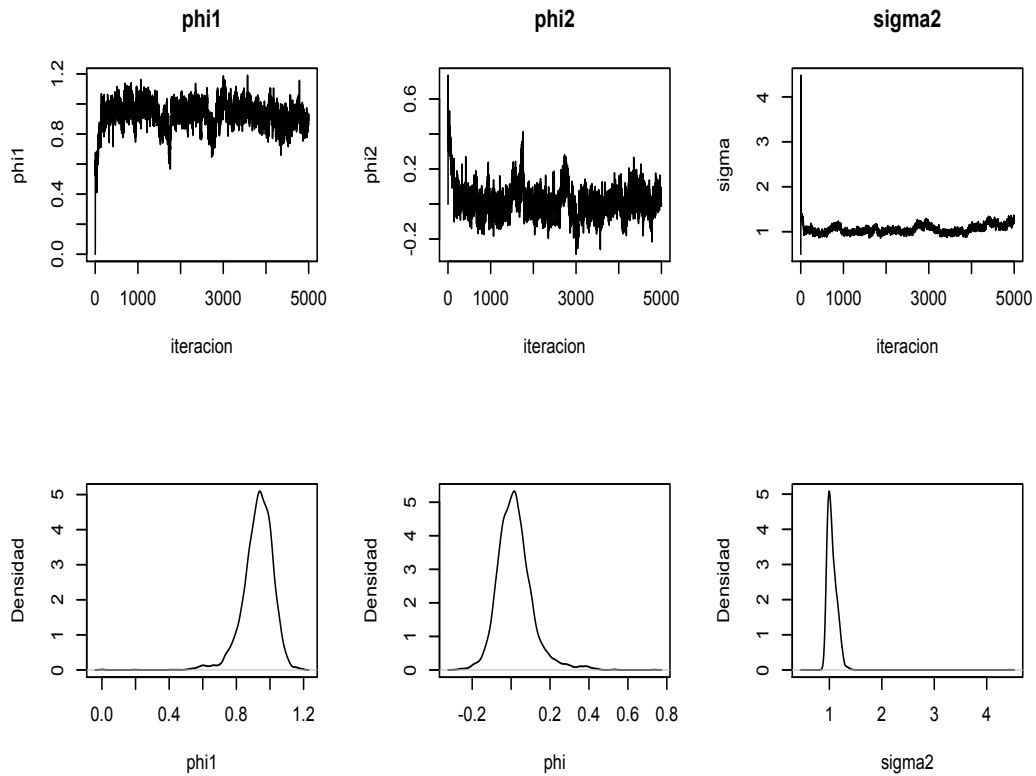


Figura 4.2: Trayectorias del MCMC y distribuciones finales suavizadas (por columnas) de las estimaciones hechas para  $\phi_1$ ,  $\phi_2$  y  $\sigma^2$  correspondientes al proceso WAR(2) ajustado para la serie que se presenta en la Figura 4.1.

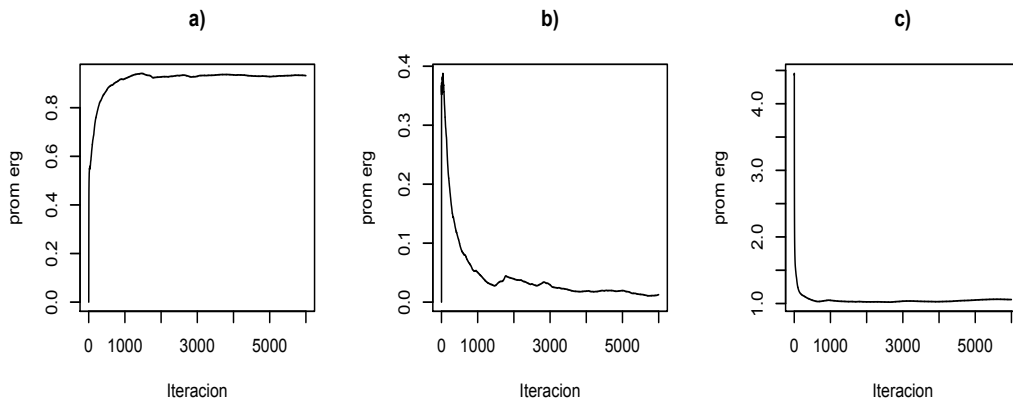


Figura 4.3: Promedios ergódicos a), b) y c) para  $\phi_1$ ,  $\phi_2$  y  $\sigma^2$ .

# Capítulo 5

## Conclusiones e Investigación Futura

El proceso de envolver y alrededor del círculo es un modo natural de construcción modelos para el análisis de datos circulares. En las dos simulaciones que se realizaron podemos ver que la aplicación de los métodos estándar MCMC nos dan un medio efectivo y eficiente para hacer inferencia en dichos modelos. Esto a pesar del desconocimiento de un número importante de coeficientes. Los ejemplos simulados evidencian la precisión del procedimiento; en el caso de los datos reales de dirección de viento revelan información interesante desde el punto de vista meteorológico; por ejemplo, podemos apreciar que existe una marcada dependencia de una observación con respecto a la dirección media de la hora anterior.

También podemos decir que en ambas simulaciones, es especial en el proceso WAR, el método es sumamente sensible a la elección de la caminata aleatoria, una caminata propuesta no adecuada impide la convergencia. En todas las estimaciones hechas para los ejemplos simulados se puede apreciar un error mayor cuanto más grande es la varianza del proceso que genera los datos.

En todos los casos la información inicial no informativa nos permite hacer inferencia en cada uno de los parámetros, aunque por supuesto, si se cuenta con información inicial informativa puede ser implementada fácilmente. Tenemos que hacer notar que en los casos cuya varianza en los datos es mayor nos obliga a utilizar más información inicial con la finalidad de no tener problemas con el procedimiento. Una de las ventajas de este enfoque que toma como base el MCMC es la simplicidad. Modelos de dimensión mayor y estructura más compleja se puede implementar de manera relativamente sencilla.

Una de las motivaciones de este trabajo fue precisamente la construcción de un modelo de viento en la región cercana a Laguna Verde. Después de una amplia revisión bibliográfica el enfoque bayesiano parece ser el que respeta la estructura intrínseca de los datos circulares. A partir de este trabajo, queda abierta la posibilidad de continuar desarrollando el modelo. Uno de los primeros puntos a tratar es el manejo de series más largas, para lo cual necesitaríamos remover los componentes del viento para contar con una serie estacionaria, y así poder emplear lo que se ha hecho hasta el momento. También sería interesante analizar dos series diferentes, una con mediciones a 10m y otra a 60m; claro, ambas del mismo punto para saber cómo se correlacionan, lo que nos daría información meteorológica importante. Otro posible dirección que puede tomar la investigación es extender a un modelo espacio temporal, ya que se cuenta con registros de más de una estación de medición.

# Bibliografía

- [1] Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, London: Springer.
- [2] Chib, S. and Greenberg, E. (1994). Bayes Inference for Regression Models with ARMA( $p,q$ ) Errors. *Journal of Econometrics*, 64, 183-206
- [3] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 4, 327-335.
- [4] Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, 8, 105-113.
- [5] Ferrari, C. (2009). *The Wrapping Approach for Circular Data Bayesian Modeling*. Ph.D. thesis, Università di Bologna.
- [6] Fisher, N. I. (1993). *The Analysis of Circular Data*, Cambridge: Cambridge University Press.
- [7] Fisher, N. I. and Lee, A. J. (1994). Time Series Analysis of Circular Data. *Journal of the Royal Statistical Society, B*, 56, 327-339.
- [8] Gutiérrez, P. E. (1997). *Métodos Computacionales en la Inferencia Bayesiana*, Serie Monografías, Vol. 6, No. 15. IIMAS, UNAM.
- [9] Mardia, K. V. (1972). *Statistics of Directional Data*, Cambridge: Cambridge University Press.
- [10] Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley.



- [11] Robert, C. P. (2001). *The Bayesian Choice*, Springer
- [12] Smith, A. F. M., and Roberts, G. O. (1993). Bayesian Computation Via Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 55, 3-23.
- [13] Tejeda, A., Ojeda, M. M. and Contreras, A. D. (1994). Local Winds South of Sierra de Misantla (México), and their statistical relationships with 500 mb circulation. *Geofísica Internacional*, Vol. 33, 2, 303-309.
- [14] Tejeda, A., Morales, T., and Cervantes, J. (1989). The Surface Wind Direction as a Function of The Wind at 850 mb in Laguna Verde zone, Veracruz, México. *Atmósfera*, 2, 181-187.

# Apéndice A

## Monte Carlo vía Cadenas de Markov

En este apéndice revisamos brevemente el algoritmo de Metropolis-Hastings; ver Robert (2001). Consideremos el problema de estimar la integral  $I = \int h(x)f(x)dx$ . Ahora introduciremos los métodos de Monte Carlo para cadenas de Markov (MCMC). La idea es construir una cadena de Markov  $X_1, X_2, \dots$ , cuya distribución estacionaria es  $f$ . Bajo ciertas condiciones se puede demostrar una ley de los grande números para cadenas de Markov, de tal manera que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N h(X_i) E_f(h(X)) = I.$$

El algoritmo Metropolis-Hastings es un método MCMC específico que trabaja de la siguiente manera. Sea  $q(y|x)$  una distribución arbitraria y conocida (i.e., sabemos cómo generar muestras de  $q(y|x)$ ). La densidad condicional  $q(y|x)$  es conocida como distribución propuesta. El algoritmo genera una secuencia de observaciones  $X_0, X_1, \dots$ , de la siguiente manera

Elige  $X_0$  arbitrariamente. Supongamos que hemos generado  $X_0, X_1, \dots, X_i$ . Para generar  $X_{i+1}$  hacemos lo siguiente:

1. Generamos un candidato  $Y \sim q(y|X_i)$ .

2. Evaluamos  $r \equiv r(X_i, Y)$  donde

$$r(x, y) = \min \left\{ 1, \frac{f(y) q(x|y)}{f(x) q(y|x)} \right\}$$

3. Hagamos

$$X_{i+1} = \begin{cases} Y & \text{con probabilidad } r \\ X_i & \text{con probabilidad } 1 - r. \end{cases}$$

Existen algunas variantes del algoritmo MCMC. Aquí consideraremos algunas de las versiones más conocidas.

METROPOLIS-HASTINGS CON CAMINATA ALEATORIA. Si consideramos la distribución propuesta  $Y$  de la forma

$$Y = X_i + \epsilon_i$$

donde  $\epsilon_i$  se genera a partir de una densidad con distribución  $g$ . En otras palabras,  $q(y|x) = g(y - x)$ . Por lo que podemos ver que, para este caso, tenemos,

$$r(x, y) = \min \left\{ 1, \frac{f(x)}{f(y)} \right\}.$$

Éste es el llamado método de Metropolis-Hastings con caminata aleatoria. La razón para el nombre es que, si no hicieramos el paso de aceptación-rechazo, estaríamos simulando una caminata aleatoria.

Este método no tiene sentido a menos que  $X$  tome valores sobre toda la línea real. Si  $X$  está restringido en algún intervalo entonces lo más conveniente es hacer una transformación a  $X$ . Por ejemplo, si  $X \in (0, \infty)$  entonces podemos tomar  $Y = \log(X)$  y entonces simulamos la distribución para  $Y$  en lugar de  $X$ .

METROPOLIS HASTINGS CON INDEPENDENCIA. Esta es una versión con muestreo de importancia de MCMC. Nosotros proponemos una distribución fija  $g$ . Generalmente,  $g$  se selecciona de tal manera que sea una aproximación de  $f$ . Como probabilidad de aceptación tenemos

$$r(x, y) = \min \left\{ 1, \frac{f(y) g(x)}{f(x) g(y)} \right\}.$$

MUESTREO DE GIBBS. Los dos métodos previos pueden ser fácilmente adaptados, en principio, para trabajar en dimensiones mayores. En la práctica, ajustar las cadenas para hacerlas converger bien es difícil. El muestreo de Gibbs es una manera de simplificar un

problema de dimension mayor a varios unidimensionales.

Aquí presentamos cómo trabaja para un problema bivariado. Supongamos que  $(X, Y)$  tiene densidad  $f_{X,Y}(x, y)$ . Primero, supongamos que es posible simular de las distribuciones condicionales  $f_{X|Y}(x|y)$  y  $f_{Y|X}(Y|X)$ . Sea  $(X_0, Y_0)$  los valores iniciales. Asumamos que hemos generado  $(X_0, Y_0), \dots, (X_n, Y_n)$ . Entonces el muestreo de Gibbs para generar  $(X_{n+1}, Y_{n+1})$  es:

$$X_{n+1} \sim f_{X|Y}(x|Y_n)$$

$$Y_{n+1} \sim f_{Y|X}(y|X_{n+1})$$

*iterar*

Lo anterior claramente puede ser generalizado a dimensiones mayores.

# Apéndice B

## Programa en R

```
# METROPOLIS-HASTINGS PARA UN WAR(2)

# USAMOS LA LIBRERIA MASS Y EL PAQUETE MCMCpack.
library(MASS)
library("MCMCpack")

# FUNCION QUE ENVUELVE
w <- function(x)
{
  y <- 0
  y <- x %% (2*pi)
  return(y)
}

# FUNCION QUE ENVUELVE Y REGRESA LOS K'S
w2 <- function(x)
{
  y <- 0
  k <- 0
  y <- x %% (2*pi)
  k <- x %/% (2*pi)
  out <- matrix(c(y,k),nrow=length(y),ncol=2)
  return(out)
}

# FUNCION QUE DESENVUELVE
dw <- function(y,k)
{
  x <- y + 2*k*pi
  return(x)
}
```

```

# *****
# AQUI ESTIMAMOS TANTO PARAMETROS COMO K'S
# *****

# GENERACION DE UN AR(2)
set.seed(190)
phi1 <- 1
phi2 <- -0.5
desv <- 1
n <- 200
x <- arima.sim(n=n,model=list(ar=c(phi1,phi2)),sd=desv)
z <- w(x)

par(mfrow=c(1,2))
plot(x,main="a",xlab="t",ylab="Y_t")
plot(z,main="b",xlab="t",ylab="X_t")

# ENVOLVEMOS EL AR(2) GENERADO
# Y CALCULO DE LOS COEFICIENTES DE ENVOLTURA
a <- w2(x)
y <- a[,1]
ks <- a[,2]

# ALGORITMO MHARW2
set.seed(160)
niter <- 20000
phi1.ini <- 0
phi2.ini <- 0
sigma.ini <- .5
p1 <- 0.025
p2 <- 0.95
p3 <- 0.025

Psi <- function(phi1,phi2,sigma,Y2)
{
  Vinv <- matrix(c(1 - phi2^2,-phi1*(1 + phi2),-phi1*(1 + phi2), 1 - phi2^2),nrow=2,ncol=2)
  u <- (1/sigma^2)*sqrt(det(Vinv))*exp(-(0.5*t(Y2)%*%Vinv%*%Y2)/sigma^2)
  return(u)
}

MHARW2 <- function(y,niter,phi1.ini,phi2.ini,sigma.ini,p1,p2,p3)
{
  n <- length(y)
  mk <- matrix(0,n,niter)
  k <-rep(0,n)

  phi1.vec <- rep(0,niter)

```

```

phi2.vec <- rep(0,niter)
sigma.vec <- rep(0,niter)

phi1.vec[1] <- phi1.ini
phi2.vec[1] <- phi2.ini
sigma.vec[1] <- sigma.ini

phi1 <- phi1.vec[1]
phi2 <- phi2.vec[1]
phi <- matrix(c(phi1,phi2),nrow=2,ncol=1)
sigma <- sigma.vec[1]

Y2 <- matrix(c(y[1]+2*k[1]*pi,y[2]+2*k[2]*pi),nrow=2,ncol=1)
B <- matrix(c(1,-1,0,1,1,1),nrow=3,ncol=2)

# Inicia iteracion M-H
for (t in 1:(niter-1))
{
#print(t)
mk[1,t] <- k[1]
mk[2,t] <- k[2]
k1cand <- 999
k2cand <- 999
tri <- rmultinom(1,1,c(p1,p2,p3))
if(tri[1,]=="1") k1cand <- -1
if(tri[2,]=="1") k1cand <- 0
if(tri[3,]=="1") k1cand <- 1
tri <- rmultinom(1,1,c(p1,p2,p3))
if(tri[1,]=="1") k2cand <- -1
if(tri[2,]=="1") k2cand <- 0
if(tri[3,]=="1") k2cand <- 1
k1_cand <- k[1] + k1cand
k2_cand <- k[2] + k2cand

# calcula probabilidad de aceptacion
Y2.cand <- matrix(c(y[1]+2*k1_cand*pi,y[2]+2*k2_cand*pi),nrow=2,ncol=1)
lograzon <- log(Psi(phi1,phi2,sigma,Y2.cand)) - log(Psi(phi1,phi2,sigma,Y2))

if( log(runif(1)) < min(lograzon,0) )
{
mk[1,t+1] <- k1_cand
mk[2,t+1] <- k2_cand
}
else
{
mk[1,t+1] <- k[1]
mk[2,t+1] <- k[2]
}
}

```

```

}

for(i in 3:n)
{
  mk[i,t] <- k[i]
  kicand <- 999
  tri <- rmultinom(1,1,c(p1,p2,p3))
  if(tri[1,]=="1") kicand <- -1
  if(tri[2,]=="1") kicand <- 0
  if(tri[3,]=="1") kicand <- 1
  ki_cand <- k[i] + kicand
  # calcula probabilidad de aceptacion para ki
  mu <- phi1*(y[i-1]+2*mk[i-1,t+1]*pi) + phi2*(y[i-2]+2*mk[i-2,t+1]*pi)
  log.num.k <- (-1/(2*sigma^2))*(((y[i] + 2*ki_cand*pi) - mu)^2)
  log.den.k <- (-1/(2*sigma^2))*(((y[i] + 2*k[i]*pi) - mu)^2)
  lr.k <- (log.num.k - log.den.k)

  if( log(runif(1)) < min(lr.k,0) )
  { mk[i,t+1] <- ki_cand }
  else
  { mk[i,t+1] <- k[i] }
}
# termina el ciclo i
k <- mk[,t+1]

suma <- 0

for(i in 3:n)
{
  wt <- matrix(c(y[i-1] + 2*k[i-1]*pi,y[i-2] + 2*k[i-2]*pi),nrow=2,ncol=1)
  suma <- suma + ((y[i]+2*k[i]*pi - t(wt)%*%phi)^2)
}

A <- 0
G <- matrix(0,nrow=2,ncol=2)
for(i in 3:n)
{
  wt <- matrix(c(y[i-1]+2*k[i-1]*pi,y[i-2]+2*k[i-2]*pi),nrow=2,ncol=1)
  A <- A + wt%*(y[i]+2*k[i]*pi)
  G <- G + wt%*t(wt)
}
GInv <- solve(G)
phi.hat <- GInv%*A

Y2 <- matrix(c(y[1]+2*k[1]*pi,y[2]+2*k[2]*pi),nrow=2,ncol=1)
Vinv <- matrix(c(1 - phi2^2,-phi1*(1 + phi2),-phi1*(1 + phi2), 1 - phi2^2),nrow=2,ncol=2)
a.sigma <- n/2

```



```

b.sigma <- ((t(Y2)%*%Vinv%*%Y2) + suma)/2

# Se genera un nuevo valor de  $\sigma^2$  de una distribucion gamma inversa
sigma.vec[t+1] <- sqrt(rinvgamma(1,a.sigma,b.sigma))
sigma <- sigma.vec[t+1]

# Se propone un nuevo valor para  $\phi = (\phi_1, \phi_2)$  a partir de una normal bivariada con
# media =  $\phi.hat$ , var =  $\sigma^2 \cdot \text{inv}(G)$ 

sigma.phi <- sigma^2*GInv
phi.cand <- mvrnorm(1, mu=phi.hat, Sigma=sigma.phi)
phi.cand <- matrix(phi.cand,nrow=2,ncol=1)

# print(phi.cand)

# condicion de estacionariedad
a <- B%*%phi.cand
if((a[1] < 1) & (a[2] < 1) & (a[3] > -1))
{
  lograzon <- log(Psi(phi.cand[1],phi.cand[2],sigma,Y2)) - log(Psi(phi1,phi2,sigma,Y2))
  if (log(runif(1)) < min(lograzon,0))
  {
    phi1.vec[t+1] <- phi.cand[1]
    phi2.vec[t+1] <- phi.cand[2]
    phi[1] <- phi1.vec[t+1]
    phi[2] <- phi2.vec[t+2]
  } else{
    phi1.vec[t+1] <- phi1
    phi2.vec[t+1] <- phi2
    phi[1] <- phi1.vec[t+1]
    phi[2] <- phi2.vec[t+2]
  }
}
else{
  phi1.vec[t+1] <- phi1
  phi2.vec[t+1] <- phi2
  phi[1] <- phi1.vec[t+1]
  phi[2] <- phi2.vec[t+2]
}
phi1 <- phi1.vec[t+1]
phi2 <- phi2.vec[t+1]
phi[1] <- phi1
phi[2] <- phi2
} # Termina la iteracion M-H
print(k)
out <- matrix(c(phi1.vec,phi2.vec,sigma.vec),nrow=niter,ncol=3)
return(out)
#return(k)

```

```

}

run1 <- MHARW2(y,niter,phi1.ini,phi2.ini,sigma.ini,p1,p2,p3)

# RESULTADOS

par(mfrow=c(3,2))
hist(run1[,1],main = paste("Phi1"),xlab="phi1", ylab="Frec")
ts.plot(run1[,1],xlab="iteracion",ylab="phi1")
hist(run1[,2],main = paste("Phi2"),xlab="phi2", ylab="Frec")
ts.plot(run1[,2],xlab="iteracion",ylab="phi2")
hist(run1[,3],main = paste("Sigma"),xlab="sigma", ylab="Frec")
ts.plot(run1[,3],xlab="iteracion",ylab="sigma")

mean(run1[,1])
mean(run1[,2])
mean(run1[,3])

# quito los primeros 10000
par(mfrow=c(3,1))
ts.plot(run1[-seq(1,10000),1],main="phi1",xlab="iteracion",ylab="phi1")
ts.plot(run1[-seq(1,10000),2],main="phi2",xlab="iteracion",ylab="phi2")
ts.plot(run1[-seq(1,10000),3],main="sigma2",xlab="iteracion",ylab="sigma")

mean(run1[-seq(1,10000),1])
mean(run1[-seq(1,10000),2])
mean(run1[-seq(1,10000),3])
#-----

```