



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA.

INFERENCIA BAYESIANA EN MODELOS LINEALES DINÁMICOS PARA DATOS
CIRCULARES

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:
UZIEL JOSÉ GONZÁLEZ GONZÁLEZ

DIRECTOR:
DR. EDUARDO GUTIÉRREZ PEÑA
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN
SISTEMAS

MÉXICO, D. F. FEBRERO DE 2015.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

TESIS DE MAESTRÍA

**Inferencia bayesiana en modelos lineales
dinámicos para datos circulares**

Presenta:

Uziel González González

Director:

Eduardo Gutiérrez Peña

*Tesis que para obtener
el grado de Maestro en Ciencias Matemáticas*

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y EN SISTEMAS

Febrero de 2015

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Resumen

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y
EN SISTEMAS

Departamento de Probabilidad y Estadística

Inferencia bayesiana en modelos lineales dinámicos para datos circulares

por Uziel González

En esta tesis proponemos un modelo estadístico para una serie de tiempo de direcciones de viento observadas en la central núcleo-eléctrica de 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 la CNLV.

El modelo contemplado toma en cuenta que la dirección de viento es una medida 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; en este trabajo decidimos seguir el enfoque de *envolver* o *enroscar* una distribución. La idea consiste en *envolver* una distribución con soporte en la recta real alrededor del círculo unitario. El proceso para las direcciones de viento lo consideramos como un proceso autorregresivo para una variable envuelta. La especificación del modelo la hacemos considerando el proceso como un caso particular de los modelos lineales dinámicos, mientras que la estimación de los parámetros la hacemos por medio de inferencia bayesiana basada en simulación.

La efectividad del modelo la evaluamos por medio de un ejemplo simulado para después aplicarlo a datos de reales. Los resultados tanto para el ejemplo simulado como para el caso real son satisfactorios, por lo que tenemos una pauta alentadora en cuanto al objetivo de la tesis.

Agradecimientos

Este trabajo no habría podido concluirlo sin el apoyo y la paciencia de mis directores Dr. Sergio Juárez y Dr. Eduardo Gutiérrez. A ambos les agradezco el aprendizaje a largo de este tiempo.

También quisiera agradecer a todos los sinodales, Dr. Raúl Rueda, Dr. Ramsés Mena, Dr. Alberto Contreras y M. en C. Antonio Flores por sus comentarios, los cuales enriquecieron este trabajo.

Índice general

Resumen	I
Agradecimientos	II
Contenido	III
Lista de figuras	v
1. Introducción	1
2. Preliminares	5
2.1. Inferencia bayesiana	5
2.1.1. Diagnóstico de modelos por medio de la distribución de pronóstico	7
2.1.2. Método de Monte Carlo	8
2.1.3. Monte Carlo vía cadenas de Markov	9
2.1.4. Variables latentes	12
2.2. Datos circulares	12
2.2.1. Variable envuelta	13
2.2.2. Distribución normal envuelta	14
2.3. Modelo de espacio de estados	16
2.3.1. Modelos lineales dinámicos	18
2.3.2. Estimación de estados y pronóstico	20
2.3.3. Filtrado	22
2.3.4. Filtro de Kalman para modelos lineales dinámicos	23
2.3.5. Suavizamiento	24
2.3.6. Pronóstico	25
3. Especificación del modelo	28
3.1. MLD's univariados para análisis de series de tiempo	28
3.1.1. Representación de los modelos ARMA y ARIMA por medio de un MLD	30
3.2. Modelos con parámetros desconocidos	34

3.2.1.	Estimación por máxima verosimilitud	34
3.2.2.	Inferencia bayesiana basada en simulación	35
3.2.3.	Simulando los estados dadas $y_{1:T}$: <i>forward filtering backward sampling</i>	37
3.2.4.	Varianza desconocida constante: gamma inversa como distribución inicial	39
4.	Análisis bayesiano de datos circulares	42
4.1.	Proceso autorregresivo envuelto con parámetros desconocidos: ejemplo simulado	45
5.	Análisis de los datos de la estación Laguna Verde	57
6.	Conclusiones	66
A.	Código en R	69
	Bibliografía	82

Índice de figuras

2.1.	Estructura de dependencia de un modelo de espacio de estados.	18
2.2.	Flujo de información de $Y_{1:t}$ a Y_{t+k}	25
4.1.	a) Proceso AR(2), b) Proceso WAR(2) con $\mu = 0$, $\phi_1 = 0.6$, $\phi_2 = -0.2$ y $\sigma^2 = 1$	47
4.2.	Trayectorias del algoritmo MCMC de los parámetros ϕ_1 , ϕ_2 y σ^2 respectivamente.	51
4.3.	Promedios ergódicos y funciones de autocorrelación de ϕ_1 , ϕ_2 y σ^2 respectivamente.	52
4.4.	Densidades finales suavizadas de los parámetros ϕ_1 , ϕ_2 y σ^2 del ejemplo simulado.	53
4.5.	Pronóstico a 24 pasos de la serie correspondiente al proceso WAR(2).	54
4.6.	Evaluación de valor omitido en la distribución de pronóstico a un paso.	55
4.7.	Evaluación de valor omitido en la distribución de pronóstico circular a un paso.	56
5.1.	Rosa de vientos correspondiente el mes de noviembre de 2007, la velocidad está medida en m/s y la dirección en grados.	58
5.2.	Serie de tiempo de direcciones medidas en radianes correspondientes al mes de noviembre de 2007.	59
5.3.	Funciones de autocorrelación circulares de la serie de tiempo original y la serie con una diferencia.	60
5.4.	Densidades finales suavizadas de los parámetros ϕ_1 , ϕ_2 y σ^2	61
5.5.	Trayectorias del algoritmo MCMC de los parámetros ϕ_1 , ϕ_2 y σ^2	62
5.6.	Promedios ergódicos y funciones de autocorrelación correspondientes a los parámetros ϕ_1 , ϕ_2 y σ^2	63
5.7.	Pronóstico de las direcciones a 24 horas.	64
5.8.	Evaluación de valor omitido en la distribución de pronóstico.	65
5.9.	Evaluación de valor omitido en la distribución de pronóstico circular.	65

Dedico este trabajo a mi familia y amigos quienes tenazmente me han apoyado en todo momento. Y sobre todo a mi madre, a ella le pertenece todo mi esfuerzo.

Capítulo 1

Introducción

*A circle is the reflection of
eternity. It has no beginning and
it has no end, and if you put
several circles over each other,
then you get a spiral.*

MAYNARD JAMES KEENAN

Esta tesis nace de la necesidad de modelar mediciones espacio-temporales de viento registradas en ocho estaciones meteorológicas localizadas en los alrededores de la central núcleo-eléctrica de Laguna Verde (CNLV). El objetivo del modelo es hacer pronósticos a corto plazo de la dirección y velocidad¹ 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 datos de viento debe considerar una variable circular. La bibliografía relacionada con el estudio de series de tiempo con datos direccionales no es tan extensa, ver Breckling (1989) y Craig (1988).

En el primero de estos trabajos se hace una revisión extensa de los componentes meteorológicos del viento y se propone modelar cada uno de los componentes

¹En estricto sentido, el término velocidad alude a la magnitud física de carácter vectorial que expresa el desplazamiento por unidad de tiempo. Los componentes de este vector son la dirección y la rapidez. Por el uso coloquial del término velocidad y para facilitar la interpretación haremos un abuso del lenguaje. En adelante nos referiremos a la rapidez como velocidad, dejando de lado su carácter vectorial.

de manera independiente. Después de haber modelado cada uno de los componentes se plantea el análisis del componente aleatorio mediante una variante de los ampliamente conocidos modelos autorregresivos. Es notable lo exhaustivo del tratamiento de los componentes meteorológicos que hace el autor. Si quisiéramos modelar series de tiempo más largas debemos considerar cada uno de estos factores, los cuales están determinados por las características geográficas de la zona. En el caso de la zona de Laguna Verde tenemos condiciones que no pueden ser incorporadas en la construcción de un modelo físico de viento por lo restrictivo de los supuestos, ver Morales et al. (1989). Por esta razón un tratamiento estadístico de la información puede ser más adecuado. El segundo trabajo plantea el mismo tipo de análisis, pero en este caso se hace una revisión técnica del componente autorregresivo del modelo. En ambos se concluye la dificultad que involucra la inferencia basada en la verosimilitud. Un enfoque alternativo basado en transformaciones se puede ver en Fisher and Lee (1994), que no resuelve el problema de la implementación para modelos complejos ni resuelve el problema de pronóstico.

En Coles (1998) tenemos unos de los primeros intentos por modelar datos circulares por medio del paradigma bayesiano. En esta misma línea se encuentra el trabajo de Ferrari (2009), en donde se propone la introducción de los modelos lineales dinámicos para el análisis de procesos circulares. En ambos casos se prueban los modelos tanto para datos simulados como para datos observados con resultados alentadores pero sin abordar el problema del pronóstico. Un modelo bayesiano para datos de viento en la región de interés se puede ver en González (2012). El inconveniente en este caso es lo limitado del modelo para incluir un eventual análisis simultáneo de la velocidad y dirección, así como la dificultad para extenderlo a un modelo espacio-temporal.

Después de una extensa revisión de los diferentes modelos para datos circulares, pudimos concluir que el enfoque de los modelos lineales dinámicos es el más adecuado para trabajar con series de tiempo de mediciones de viento. Son varias las ventajas que tiene esta propuesta respecto a las antes mencionadas. Entre ellas se encuentra la flexibilidad para incorporar diferentes componentes en el modelo así como la relativa facilidad para poder pronosticar. En conjunto con estos modelos lineales dinámicos, emplearemos el enfoque de envolver un proceso estocástico para generar una nueva clase de modelos. Para poder hacer inferencia decidimos seguir el paradigma bayesiano, realizando el análisis por medio de algoritmos computacionales modernos.

Ahora daremos una descripción general de la organización de la tesis. En el capítulo 2 presentamos los resultados preliminares que nos ayudarán a hacer el planteamiento del modelo. Comenzaremos con una descripción general de la inferencia bayesiana, además de algunas propuestas para poder evaluar modelos por medio de la distribución de pronóstico o predictiva. Continuaremos con una descripción breve de algunos métodos de Monte Carlo vía cadenas de Markov (MCMC)² que usaremos en el análisis final. Es importante destacar el uso de variables latentes dentro del modelo que plantearemos, por lo que hacemos mención de ellas. A pesar de existir varios planteamientos de las diferentes distribuciones para analizar datos circulares, escogimos el enfoque de *envolver* o *enroscar*³. Si bien este enfoque es en realidad una aproximación, nos permite plantear modelos generales. Otro enfoque consiste en estudiar distribuciones intrínsecamente definidas en el círculo, como la distribución von Mises, también llamada distribución normal circular. Uno de los principales problemas de la inferencia sobre estas distribuciones, incluso en modelos sencillos, es que su tratamiento es complicado. Dedicaremos una sección para dar algunas generalidades de las distribuciones envueltas y algunas características de la distribución normal envuelta, que es una aproximación de la distribución von Mises. También presentaremos los modelos lineales dinámicos, los cuales son un caso especial de una clase de modelos más generales conocidos como modelos de espacio de estados. Una de las referencias básicas para este tema es West and Harrison (1997), donde se puede ver que son varios los problemas en estadística que se pueden abordar por medio de esta clase de modelos. En esta sección trataremos los principales cuestiones de estudio en esta clase de modelo, tales como el *filtrado*, *suavizamiento* y *pronóstico* por medio del filtro de Kalman. Es importante mencionar que el enfoque bayesiano tiene ventajas sobre el enfoque frecuentista y que haremos la inferencia por medio de simulación.

En el capítulo 3 daremos las representaciones como modelos lineales dinámicos de los modelos ARMA y ARIMA. Además veremos el planteamiento general de la inferencia basada en simulación cuando tenemos parámetros desconocidos, tal como se propone en Migon et al. (2005) y Petris et al. (2009)

En el capítulo 4 conjuntamos los resultados preliminares con los modelos lineales dinámicos para hacer un estudio de un proceso circular. En este caso simulamos una serie de tiempo para después envolverla y finalmente hacer inferencia

²Por sus siglas en inglés, *Markov chain Monte Carlo Methods*.

³Traducción literal del inglés de *wrapping*.

suponiendo desconocidos los parámetros de envoltura o variables latentes. Los resultados son satisfactorios en cuanto al pronóstico, lo cual verificamos por medio de la distribución predictiva. Todo el análisis lo hacemos por medio del software estadístico R, R Core Team (2014).

En el capítulo 5 haremos una análisis de las direcciones de viento medidas a 10 metros en una de las estaciones meteorológicas de Laguna Verde siguiendo el modelo propuesto en capítulos anteriores. Finalmente, en el último capítulo discutimos algunos problemas tanto de implementación como los resultados obtenidos para los datos observados.

Capítulo 2

Preliminares

2.1. Inferencia bayesiana

Cuando planteamos un problema estadístico, nuestra finalidad es generar 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. 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. Una de las referencias básicas para el estudio de la estadística bayesiana es Robert (2001), y en Gutiérrez-Peña (1997) se discuten algunas de las ideas de esta sección de preliminares.

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, X , y unas cantidades desconocidas, θ , cuyo valor es de nuestro interés. Estas cantidades pueden ser de diversos tipos, tales como parámetros de un modelo, observaciones futuras o mediciones que no podemos observar directamente; tanto X como θ pueden ser vectores. Los pasos a seguir dentro del enfoque bayesiano se pueden resumir del siguiente modo:

- Especificación de un modelo. Se debe de elegir un modelo para describir el proceso que generó los datos. El modelo elegido, $p(x|\theta)$, dependerá del tipo de problema y siempre debe de tomarse en cuenta sus limitaciones.

- Especificación de una distribución inicial, $p(\theta)$. Este punto es fundamental dentro del enfoque bayesiano. En conjunto con la interpretación subjetiva de la probabilidad distingue 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 una distribución inicial consistente con nuestra información previa. También debemos considerar que en caso de carecer de información inicial acerca del parámetro θ , debemos ser capaces de reflejar este desconocimiento. Esto se hace por medio de las llamadas distribuciones iniciales no informativas.
- Cálculo de la distribución final. Si consideramos que X es conocido, podemos condicionar en su valor observado x . Por lo tanto, nuestro conocimiento acerca de θ queda descrito por medio de la distribución final $p(\theta|x)$. El teorema de Bayes nos ayuda a calcular la distribución final

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

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

$$p(\theta|x) \propto p(\theta)p(x|\theta).$$

En la práctica, el cálculo de la distribución final puede ser un asunto complicado, especialmente si la dimensión del parámetro es grande. Existen casos en los que se requieren técnicas computacionales sofisticadas.

- Resumen de la información. Habiendo encontrado $p(\theta|x)$, podemos decir que hemos resuelto el problema de inferencia; ya que toda la información disponible sobre el valor de θ queda contenida en la distribución final. Sin embargo, en general es necesario resumir esta información para poder interpretar los resultados y responder las preguntas de interés. Esto se puede hacer a través de procedimientos de estimación, prueba de hipótesis o pronósticos por dar algunos ejemplos.

2.1.1. Diagnóstico de modelos por medio de la distribución de pronóstico

En la teoría de estadística bayesiana, el pronóstico de observaciones futuras se basa en la distribución predictiva asociada al modelo, la cual podemos interpretar como la distribución de los datos promediados sobre todos los valores de los parámetros posibles. Por esta razón, cuando aún no hemos observado los datos, x , el pronóstico se basa en

$$p(x) = \int p(x|\theta)p(\theta)d\theta.$$

Ésta es precisamente la función de densidad de los datos promediada sobre los valores de los parámetros de acuerdo al conocimiento previo que tenemos de ellos. Por lo anterior, $p(x)$ se conoce como *distribución de pronóstico inicial*. Es más común e interesante pronosticar x' después de haber observado los datos, x . De manera análoga podemos calcular la distribución de pronóstico final

$$p(x'|x) = \int p(x'|\theta)p(\theta|x)d\theta.$$

En este caso la distribución es la densidad de los datos promediada sobre la distribución final $p(\theta|x)$. Las observaciones futuras, x' , se pueden ver de manera alternativa como parámetros adicionales a estimar dentro de un modelo, es decir, la distribución final conjunta es $p(x', \theta|x)$. En este caso, la inferencia de las observaciones futuras x' se pueden obtener marginalizando distribución con respecto a θ

$$\begin{aligned} p(x'|x) &= \int p(x', \theta|x)d\theta \\ &= \int p(x'|\theta, x)p(\theta|x)d\theta. \end{aligned}$$

La densidad de pronóstico no sólo es útil para cuantificar nuestro conocimiento acerca de las observaciones futuras sino además nos permite evaluar la calidad de nuestro modelo. Esta distribución podemos usarla para responder algunas preguntas acerca de la confiabilidad de nuestro pronóstico. Algunos aspectos de cómo justificar el diagnóstico de modelos basado en la distribución de pronóstico se discuten en Gelfand et al. (1992). En el caso de análisis de series de tiempo podemos utilizar esta idea para evaluar el pronóstico a un paso. Para poder verificar la confiabilidad del pronóstico, podemos omitir la última observación de la serie para

ajustar el modelo. Después predecimos el valor omitido usando la distribución de pronóstico.

Una propuesta simple es considerar la siguiente densidad de pronóstico

$$p(x_i|x_{(-i)}) = \int p(x_i|\theta)p(\theta|x_{(-i)})d\theta$$

donde x_i es la observación i de x y $x_{(-i)}$ es x omitiendo x_i . El valor de la integral anterior también se conoce como ordenada de pronóstico condicional o CPO¹. Se puede considerar como una medida cuantitativa para el efecto de la observación i sobre todas las densidades de pronóstico, $p(x)$, dado que

$$\text{CPO}_i = p(x_i|x_{(-i)}) = \frac{p(x)}{p(x_{(-i)})}$$

Valores pequeños de CPO's indican que no esperamos ver las observaciones omitidas bajo nuestro modelo. Podemos construir una medida que involucra todas las observaciones que omitimos. Si estamos en el contexto de series de tiempo, consideraremos la partición x_1 como la última observación, que será la que evaluaremos y que omitimos al ajustar el modelo.

2.1.2. 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 de una función arbitraria, h , como

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

¹Por sus siglas en inglés, *conditional predictive ordinate*.

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 de $s(\theta)$, $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}$, con M grande. Finalmente se aproxima I a partir de la media muestral

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

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.

2.1.3. 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, tenemos que construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución final que nos interesa. Más adelante presentamos algunos de los algoritmos de que usamos en este trabajo, los cuales son: muestreo de Gibbs, Metropolis-Hastings así como algunas variantes. Una amplia discusión y ejemplos de estas técnicas pueden verse en Robert and Casella (2004), Chib and Greenberg (1994) y Chib and Greenberg (1995).

En general no es fácil determinar en qué momento la cadena ha convergido, en general se utilizan métodos empíricos para verificarlo. Por otro lado, podemos usar las funciones de autocorrelación para determinar si la muestra que generamos es independiente. Con frecuencia 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.

Muestreo de Gibbs

Supongamos que el parámetro es multidimensional, $\psi = (\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(k)})$. Sea $\pi(\psi) = \pi(\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(k)})$ la densidad objetivo. El muestreo de Gibbs inicia en un valor arbitrario $\psi_0 = (\psi_0^{(1)}, \dots, \psi_0^{(k)})$ en el espacio parametral y actualiza un componente a la vez, $\psi^{(i)}$, $i = 1, \dots, k$, de la distribución condicional, de acuerdo al algoritmo 1. Un punto importante es que este algoritmo funciona incluso cuando uno o más componentes $\psi^{(i)}$ por sí mismos son multidimensionales. En cuyo caso, el algoritmo funciona por bloques de los componentes de ψ .

Algoritmo 1 Muestreo de Gibbs

0. Inicia: fijamos $\psi_0 = (\psi_0^{(1)}, \dots, \psi_0^{(k)})$.
 1. Para $j = 1, \dots, N$:
 - 1.1) genera $\psi_j^{(1)}$ de $\pi(\psi^{(1)} | \psi^{(2)} = \psi_{j-1}^{(2)}, \dots, \psi^{(k)} = \psi_{j-1}^{(k)})$;
 - 1.2) genera $\psi_j^{(2)}$ de $\pi(\psi^{(2)} | \psi^{(1)} = \psi_j^{(1)}, \psi^{(3)} = \psi_{j-1}^{(3)}, \dots, \psi^{(k)} = \psi_{j-1}^{(k)})$;
 - ⋮
 - 1.k) genera $\psi_j^{(k)}$ de $\pi(\psi^{(k)} | \psi^{(1)} = \psi_j^{(1)}, \dots, \psi^{(k-1)} = \psi_j^{(k-1)})$.
-

Metropolis-Hastings

Este método es bastante general, ya que nos permite generar el siguiente estado de la cadena de una distribución esencialmente arbitraria, π . La invarianza de la distribución objetivo se obtiene por medio de un paso de aceptación o rechazo. Supongamos que la cadena se encuentra actualmente en ψ . Entonces el valor propuesto es $\tilde{\psi}$ que se generó por medio de la densidad $q(\psi, \cdot)$. En este caso q es una densidad en su segundo argumento, pero es parametrizada por el primero. Es decir, la densidad propuesta puede depender del estado actual, ψ . El valor propuesto $\tilde{\psi}$ se acepta como un nuevo estado de la cadena con probabilidad

$$\alpha(\psi, \tilde{\psi}) = \min \left\{ 1, \frac{\pi(\tilde{\psi})q(\tilde{\psi}, \psi)}{\pi(\psi)q(\psi, \tilde{\psi})} \right\}. \quad (2.1)$$

Si el valor propuesto se rechaza, la cadena permanece en el estado ψ . El algoritmo 2 nos muestra los pasos necesarios. La selección de la densidad q es una cuestión importante. En particular, cuando la elegimos simétrica, la expresión (2.1) se simplifica y sólo tenemos

$$\alpha(\psi, \tilde{\psi}) = \min \left\{ 1, \frac{\pi(\tilde{\psi})}{\pi(\psi)} \right\}.$$

Algoritmo 2 Metropolis-Hastings

-
0. Inicia: fijamos ψ_0 .
 1. Para $j = 1, \dots, N$:
 - 1.1) genera $\tilde{\psi}_j$ de $q(\psi_{j-1}, \cdot)$;
 - 1.2) calcula $\alpha = \alpha(\psi_{j-1}, \tilde{\psi}_j)$ de acuerdo a (2.1);
 - 1.3) genera una variable aleatoria independiente $U_j \sim \mathcal{Bernoulli}(\alpha)$;
 - 1.4) si $U_j = 1$ hacemos $\psi_j = \tilde{\psi}_j$, en otro caso hacemos $\psi_j = \psi_{j-1}$.
-

El muestreo de Gibbs y el algoritmo de Metropolis-Hastings pueden emplearse de manera combinada de acuerdo al problema en cuestión. Supongamos que dentro de un muestreo de Gibbs queremos simular de la distribución condicional $\pi(\psi^{(1)}|\psi^{(2)})$, pero ésta no tiene una forma conocida o es demasiado complicado simularla. En este caso, en lugar de generar $\psi^{(1)}$ a partir de $\pi(\psi^{(1)}|\psi^{(2)})$, podemos hacerlo por medio de un paso del algoritmo de Metropolis-Hastings sin que la distribución de la cadena de Markov deje de ser invariante.

Metropolis-Hastings con adaptación y rechazo

El muestreo por rechazo es un algoritmo sencillo que nos permite generar una variable aleatoria de una distribución objetivo, π , simulando de una distribución propuesta, f , y aceptando con una probabilidad específica. Supongamos que existe una constante C tal que $\pi(\psi) \leq Cf(\psi)$ para cada ψ y definimos $r(\psi) = \pi(\psi)/Cf(\psi)$, por lo que $0 \leq r(\psi) \leq 1$. Generamos dos variables independientes U y V , con U distribuida uniforme en $(0, 1)$ y $V \sim f$. Si $U \leq r(V)$ hacemos $\psi = V$, en otro caso repetimos el proceso. En otras palabras, simulamos V de $f(\psi)$ y aceptamos V como muestra de $\pi(\psi)$ con probabilidad $r(V)$. En caso de rechazo, reiniciamos el proceso. Se puede demostrar que si el soporte de f incluye el soporte de π , el algoritmo eventualmente genera un valor V que se acepta. Si la distribución objetivo univariada, π , no es log-cóncava² podemos combinar el muestreo de adaptación y rechazo con el algoritmo de Metropolis-Hastings. Este algoritmo se conoce como ARMS³. También existe una variante que nos permite simular de una distribución objetivo multivariada.

²Una función g es cóncava si $g(\alpha x + (1 - \alpha)y) \geq \alpha g(x) + (1 - \alpha)g(y)$ para cada $\alpha \in (0, 1)$. π es log-cóncava si $\log \pi(\psi)$ es una función cóncava.

³Por sus siglas en inglés, *Adaptive rejection Metropolis sampling*.

2.1.4. Variables latentes

Existe una técnica general muy útil 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 $p(\theta, \lambda|x)$ que de la distribución $p(\theta|x)$, donde el nuevo parámetro λ puede ser esencialmente cualquier variable aleatoria. Es claro que al generar una muestra con distribución conjunta $p(\theta, \lambda|x)$ podemos obtener una muestra con distribución marginal $p(\theta|x)$ que nos interesaba originalmente.

2.2. Datos circulares

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

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 tanto, 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 and Jupp (1999), Jammalamadaka and SenGupta (2001), y Fisher and Lee (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 frecuentista. También es común encontrar situaciones en las cuales los procedimientos de inferencia clásica tienen una implementación más compleja. Por ejemplo, la verosimilitud para una distribución normal envuelta, que se define más adelante, no tiene forma cerrada (la densidad de la normal envuelta es una serie

infinita) y por lo tanto su maximización es prácticamente intratable, ver Craig (1988). Por las razones anteriores es necesario 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.

2.2.1. Variable envuelta

Una de las aproximaciones para el estudio de los datos circulares consiste en envolver una distribución en la línea real, Y , alrededor del círculo unitario. Su principal característica es la flexibilidad: una amplia clase de funciones sobre el círculo se puede obtener usando dicho enfoque. En principio, este procedimiento puede ser usado para envolver cualquier distribución.

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

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

La correspondiente función de distribución de X_w es

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} F(\theta + 2k\pi) - F(2k\pi), \quad 0 \leq \theta < 2\pi, \quad (2.3)$$

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 ϕ , entonces la función característica $\{\phi_p : p = 0, \pm 1, \dots\}$ de X_w es

$$\phi_p = \phi(p). \quad (2.4)$$

Para poder ver esto, notemos 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 ϕ es integrable entonces Y tiene una densidad y

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2k\pi) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right\} \quad (2.5)$$

donde $\phi(p) = \alpha_p + i\beta_p$. El resultado se sigue de (2.2), de (2.4) y de considerar que la serie $\sum_{p=1}^{\infty} (\alpha_p^2 + \beta_p^2)$ converge. Para verificar el último punto, basta observar que si ϕ es integrable, tenemos

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

- 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. 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 $f_w = g$.

2.2.2. Distribución normal envuelta

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.

La *distribución normal envuelta* es una función con dos parámetros, la cual se obtiene al envolver la línea de una distribución normal $\mathcal{N}(\tilde{\mu}, \sigma)$ alrededor del círculo. De (2.2) la función de densidad de probabilidad de una distribución normal envuelta es

$$f_w(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta < 2\pi. \quad (2.6)$$

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

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

Usando (2.5) obtenemos una útil representación de la función de densidad (2.6) como lo es

$$f_w(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right\}, \quad (2.7)$$

donde $\rho = e^{-\sigma^2/2}$. La distribución normal envuelta con parámetros μ y σ la denotaremos como $\mathcal{WN}(\mu, \sigma)$ ⁴.

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

Los momentos y propiedades de una distribución normal envuelta son:

- i) La dirección circular media es $\mu = \tilde{\mu} \pmod{2\pi}$ donde $\tilde{\mu}$ es la media de una distribución normal.
- 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 circular, U_c .

⁴Por la traducción del inglés, *wrapped normal distribution*.

- v) Conforme $\rho \rightarrow 1$, la distribución tiende a una distribución degenerada concentrada en la dirección μ .
- vi) La moda está en $\theta = \mu$.
- vii) Si $\theta \sim \mathcal{WN}(\mu, \rho)$ entonces $(\theta - \psi) \sim \mathcal{WN}(\mu - \psi, \rho)$, donde $\psi \in [0, 2\pi)$.
- viii) De la propiedad a) de las distribuciones envueltas y de la ecuación (2.6) se sigue que si Θ_1 y Θ_2 son dos variables normales envueltas, $\Theta_i \sim \mathcal{WN}(\mu_i, \rho_i)$, ($i = 1, 2$), entonces $(\Theta_1 + \Theta_2) \sim \mathcal{WN}(\mu_1 + \mu_2, \rho_1\rho_2)$.

Tal como se mencionó, en principio a cualquier distribución definida en \mathbb{R} puede aplicarse el proceso de envolver. Este proceso no está restringido a distribuciones de probabilidad, también podemos hacerlo para diferentes procesos estocásticos.

2.3. Modelo de espacio de estados

*Nous devons envisager l'état
présent de l'Univers comme un
effet de l'état antérieur et comme
la cause de l'état suivant.*

LAPLACE (1795)

Recientemente ha habido un interés creciente en las aplicaciones de los modelos de espacio de estados en el análisis de series de tiempo. Dichos modelos consideran una serie de tiempo como un sistema dinámico que tiene una perturbación aleatoria. Este enfoque nos permite dar una interpretación natural de las series de tiempo como una combinación de varios componentes, tales como tendencia o estacionalidad. Además, nos proporcionan una estructura probabilística poderosa, que nos ofrece una amplia variedad de aplicaciones.

Consideremos a la serie de tiempo $(Y_t)_{t \geq 1}$. En general, dada la estructura de dependencia de las Y_t 's, no es sencillo determinar la distribución conjunta de dimensión finita de (Y_1, \dots, Y_t) , para cualquier $t \geq 1$. Los supuestos de independencia o intercambiabilidad harían al tiempo irrelevante. Por lo tanto, estos supuestos no son razonables para la mayoría de las series de tiempo con datos reales. La

dependencia markoviana es la más simple de las formas de dependencia entre las Y_t 's, en las cuales el tiempo tiene una función determinada. Decimos que $(Y_t)_{t \geq 1}$ es una cadena de Markov si, para cada $t \geq 1$, $\pi(y_t|y_{1:t-1}) = \pi(y_t|y_{t-1})$, donde $y_{1:t} = (y_1, \dots, y_t)$. Esto significa que la información relativa a Y_t que se tiene a partir de las observaciones hasta el tiempo $t - 1$ es exactamente la misma que se tiene con y_{t-1} únicamente. Otra manera de decirlo es que Y_t y $Y_{1:t-2}$ son condicionalmente independientes dado y_{t-1} . Para una cadena de Markov las distribuciones conjuntas de dimensión finita se puede escribir de la siguiente manera

$$\pi(y_{1:t}) = \pi(y_1) \prod_{j=2}^t \pi(y_j|y_{j-1}).$$

Es conveniente aclarar que suponer una estructura markoviana en las observaciones no es apropiada en muchas de las aplicaciones. Los modelos de espacio de estados poseen una estructura relativamente simple que permite construir modelos más complejos. En este modelo suponemos que existe una cadena de Markov que no es observable (θ_t), conocida como proceso de estados, y que Y_t depende de θ_t . En las aplicaciones en ingeniería, θ_t generalmente es observable físicamente y describe un sistema que tiene como salida Y_t . Por otro lado, en las aplicaciones en econometría y estadística, θ_t es una variable latente, la cual además puede tener una interpretación útil. En cualquier caso, podemos considerar a θ_t como una serie de tiempo auxiliar que nos facilita la tarea de determinar la distribución de probabilidad de la serie de tiempo observable, Y_t . En términos formales, un modelo de espacio de estados consiste en una serie de tiempo con valores en \mathbb{R}^p , $(\theta_t : t = 0, 1, \dots)$, y una serie de tiempo con valores en \mathbb{R}^m , $(Y_t : t = 1, 2, \dots)$, que satisfacen las siguientes hipótesis.

- (H.1) (θ_t) es una cadena de Markov.
- (H.2) Condicionado en (θ_t) , las Y_t 's son independientes y Y_t depende únicamente de θ_t .

La consecuencia de (H.1)-(H.2) es que el modelo de espacio de estados queda determinado por la distribución inicial $\pi(\theta_0)$ y por las densidades $\pi(\theta_t|\theta_{t-1})$ y $\pi(y_t|\theta_t)$, $t \geq 1$. De hecho, para cada $t > 0$,

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \prod_{j=1}^t \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j). \quad (2.8)$$

De (2.8) podemos generar, por condicionamiento o marginalizando, algunas otras distribuciones de interés. Por ejemplo, la densidad conjunta de las observaciones de $Y_{1:t}$ se puede obtener integrando respecto a las θ_j 's en (2.8). Si suponemos un espacio de estados, el flujo de información queda expresado de manera simple en la figura 2.1. El punto clave de estos modelos es que, dado el presente, el futuro es independiente del pasado. Es precisamente esta propiedad estructural de independencia condicional la que nos permite tener una gran efectividad en los modelos dinámicos.

De manera similar, podemos observar que θ_t y $(\theta_{0:t-2}, Y_{1:t-1})$ son condicionalmente independientes dado θ_{t-1} , lo cual se puede expresar en términos de la distribución condicional como

$$\pi(\theta_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(\theta_t | \theta_{t-1}).$$

Los modelos de espacio de estados en los cuales las variables aleatorias toman valores discretos usualmente se conocen como *modelos de Markov ocultos*.

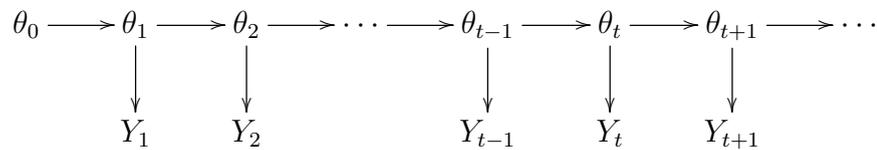


FIGURA 2.1: Estructura de dependencia de un modelo de espacio de estados.

2.3.1. Modelos lineales dinámicos

Los modelos lineales dinámicos son una amplia clase de modelos que nos permiten incorporar parámetros que varían a lo largo del tiempo. Una referencia básica sobre este tema es West and Harrison (1997).

Para ser precisos diremos que un *modelo lineal dinámico* (MLD) es un modelo de espacio de estados determinado por una distribución inicial normal para el vector de estados de dimensión p , al tiempo $t = 0$,

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0), \quad (2.9a)$$

y el par de ecuaciones para cada tiempo $t \geq 1$,

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim \mathcal{N}_m(0, V_t), \quad (2.9b)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}_p(0, W_t), \quad (2.9c)$$

donde G_t y F_t son matrices conocidas (de orden $p \times p$ y $m \times p$ respectivamente) y $(v_t)_{t \geq 1}$ y $(w_t)_{t \geq 1}$ son dos sucesiones independientes de vectores aleatorios gaussianos con medias cero y matrices de varianzas conocidas $(V_t)_{t \geq 1}$ y $(W_t)_{t \geq 1}$, respectivamente. La ecuación (2.9b) se conoce como *ecuación de observación*, mientras que (2.9c) es la *ecuación del sistema* o *ecuación de estados*. Además suponemos que θ_0 es independiente de (v_t) y (w_t) . Un MLD satisface las hipótesis (H.1) y (H.2) que se mencionaron en la sección anterior, con $Y_t | \theta_t \sim \mathcal{N}(F_t \theta_t, V_t)$ y $\theta_t | \theta_{t-1} \sim \mathcal{N}(G_t \theta_{t-1}, W_t)$. A diferencia de lo que ocurre con el modelo (2.9), el modelo general de espacio de estados queda totalmente definido con una distribución para θ_0 junto con las ecuaciones de observación y evolución

$$Y_t = h_t(\theta_t, v_t),$$

$$\theta_t = g_t(\theta_{t-1}, w_t),$$

donde g_t y h_t son funciones arbitrarias. Un espacio de estados *lineal* supone que g_t y h_t son funciones lineales en θ , mientras que un modelo lineal *gaussiano* supone que las distribuciones son gaussianas. En caso de relajar el supuesto de normalidad tendríamos que pagar un costo computacional considerable. A manera de ejemplo veremos cómo los MLD's pueden plantearse como una generalización de los modelos de regresión pero considerando variación a través del tiempo en los coeficientes de regresión. El caso simple, el modelo de regresión estático, considera la relación entre la variable Y y la variable explicativa no aleatoria x del siguiente modo

$$Y_t = \theta_1 + \theta_2 x_t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_t^2).$$

En nuestro contexto consideraremos a (Y_t, x_t) como observable a través del tiempo. Por el hecho de poder incorporar variación en el tiempo de los parámetros de regresión, en principio somos capaces de modelar la relación funcional no lineal entre x y y . Además podemos incluir en el modelo cambios en la estructura del proceso que analizamos. Un modelo de regresión lineal dinámico supone

$$Y_t = \theta_{t,1} + \theta_{t,2} x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2),$$

con la ecuación adicional que describe la evolución del sistema

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}_2(0, W_t).$$

Este es un MLD con $F_t = [1, x_t]$ y estados $\theta_t = (\theta_{t,1}, \theta_{t,2})'$. Como caso particular, si $G_t = I$, la matriz identidad, $\sigma_t^2 = \sigma^2$ y $w_t = 0$ para cada t , tenemos el modelo de regresión estático.

Análisis de los modelos lineales dinámicos en R

Sin entrar en detalles podemos decir que es considerable el análisis que se puede hacer de los MLD's por medio del software R, que además tiene disponible un paquete específico para ello, `d1m`. El paquete representa al modelo como un objeto de la clase `d1m`. Los objetos de este tipo pueden representar los DLM's constantes o variando a través del tiempo. Una vez que conocemos las matrices F, G, V, W, C_0 y el vector m_0 , el MLD constante queda totalmente determinado. Dentro de R, estos componentes se almacenan en `FF`, `G`, `VV`, `W`, `CO` y `m0` respectivamente.

Respecto los MLD's que varían en el tiempo, el paquete en lugar de guardar las matrices F_t, V_t, G_t, W_t para todos los valores de t , guarda un plantilla de cada una de ellas. En una matriz separada se guardan las entradas que varían en el tiempo. Se pueden consultar los detalles del paquete en Petris et al. (2009).

2.3.2. Estimación de estados y pronóstico

Una de las razones por las cuales los MLD's tienen una extensa aplicación en diversos problemas es por su flexibilidad. Sin embargo, a pesar de las ventajas de los MLD's en las aplicaciones estadísticas, éstos no quedan exentos de una de las principales dificultades como lo es la especificación del modelo. Idealmente, al momento de plantear un problema el experto en el tema y el estadístico podrían construir un espacio de estados con una interpretación intuitiva. Sin embargo, las complicaciones son variadas; podríamos no tener una interpretación clara de los estados, una identificabilidad pobre del espacio de estados o simplemente el modelo podría ser demasiado complejo. De momento supondremos que conocemos el modelo, es decir, que conocemos las densidades $\pi(y_t|\theta_t)$ y $\pi(\theta_t|\theta_{t-1})$. Daremos

algunas nociones de la estimación y el pronóstico, y más adelante abordaremos los modelos con parámetros desconocidos, ψ .

Para un modelo de espacio de estados conocido, algunos de los principales objetivos son la inferencia de los estados no observables y el pronóstico de observaciones futuras. Tanto la estimación como el pronóstico pueden abordarse calculando las distribuciones condicionales de las cantidades de interés, dada la información disponible.

Para calcular el vector de estados tenemos que calcular la densidad condicional $\pi(\theta_s|y_{1:t})$. Debemos distinguir entre los problemas de *filtrado* (cuando $s = t$), *pronóstico de estados* (cuando $s > t$) y *suavizamiento* (cuando $s < t$). Es importante enfatizar la diferencia que existe entre el filtrado y el suavizamiento. En el filtrado suponemos que los datos los obtenemos secuencialmente en el tiempo. En otras palabras, hacemos una estimación del vector de estados actual hasta el tiempo t , para ello necesitamos la densidad condicional $\pi(\theta_t|y_{1:t})$.

Por otro lado, el suavizamiento o análisis retrospectivo consiste en detectar el comportamiento periódico que está detrás de las observaciones de alguna serie de tiempo. Para poder abordar este problema debemos calcular la distribución condicional $\pi(\theta_{1:t}|y_{1:t})$. Tanto para el filtrado como para el suavizamiento existen algoritmos iterativos que pueden implementarse.

Uno de los principales objetivos en el estudio de series de tiempo es el pronóstico. Es por ello que usualmente la estimación de los estados se convierte en un paso intermedio para pronosticar el valor de las observaciones futuras. Para hacer pronóstico a un paso, es decir, obtener la observación Y_{t+1} habiendo observado $y_{1:t}$, primero estimamos θ_{t+1} y entonces podremos hacer pronóstico para Y_{t+1} . La densidad de pronóstico, $\pi(\theta_{t+1}|y_{1:t})$, se basa en la densidad de filtrado de θ . Es entonces que podemos calcular la densidad de pronóstico para las observaciones a un paso, $\pi(y_{t+1}|y_{1:t})$.

En algunos casos podríamos estar interesados en hacer pronóstico más a futuro, es decir, necesitaríamos calcular θ_{t+k} para después hacer pronóstico k pasos adelante, Y_{t+k} , con $k > 1$. Es importante aclarar que existe menos precisión conforme crece el número de pasos que deseamos pronosticar. Más adelante mostraremos cómo calcular de manera recursiva estas densidades. En particular, la media condicional $E(Y_{t+1}|y_{1:t})$ proporciona una estimación puntual a un paso de manera óptima para el valor de Y_{t+1} , minimizando el error cuadrado esperado condicional.

Como función de k , $E(y_{t+k}|y_{1:t})$ se conoce comunmente como *función de pronóstico*.

2.3.3. Filtrado

Primero describimos los pasos para calcular las densidades de filtrado, $\pi(\theta_t|y_{1:t})$, para el modelo general de espacio de estados. Incluso si no hacemos un uso extensivo de estas expresiones, es útil revisar estas recursiones para entender la importancia del supuesto de independencia condicional que ya hemos discutido antes. Más adelante, dentro del contexto de los MLD's, tenemos expresiones particulares conocidas como el filtro de Kalman.

Una de las ventajas de los modelos de espacio de estados, es precisamente que, debido a la estructura markoviana de los estados dinámicos (H.1) y la independencia condicional de las observaciones (H.2), las densidades de pronóstico pueden calcularse recursivamente. La siguiente proposición contiene una presentación formal de las recursiones de filtrado para el modelo general de espacio de estados. Las demostraciones de las siguientes proposiciones son resultados conocidos cuya demostración puede verse en West and Harrison (1997).

Proposición 1. (Recursiones de filtrado). *Para un modelo general de espacio de estados que definimos en (H.1)-(H.2), tenemos que se cumplen las siguientes afirmaciones:*

- (i) *La densidad de pronóstico a un paso para los estados se puede calcular de la densidad de filtrado $\pi(\theta_{t-1}|y_{1:t-1})$ de acuerdo a*

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})d\theta_{t-1}. \quad (2.10a)$$

- (ii) *La densidad de pronóstico a un paso para las observaciones se puede calcular a partir de la densidad de pronóstico de los estados del siguiente modo*

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})d\theta_t. \quad (2.10b)$$

- (iii) *La densidad de filtrado se puede calcular de las densidades anteriores del siguiente modo*

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})}. \quad (2.10c)$$

2.3.4. Filtro de Kalman para modelos lineales dinámicos

Los resultados anteriores en principio nos permitirían abordar cualquier problema de pronóstico o filtrado. Sin embargo, el cálculo de estas distribuciones por lo general es bastante complicado. Los MLD's son un caso importante en el cual las recursiones se simplifican considerablemente. En particular, considerando resultados conocidos de la distribución gaussiana multivariada, no es difícil verificar que el vector aleatorio $(\theta_0, \theta_1, \dots, \theta_t, Y_1, Y_2, \dots, Y_t)$ tiene una distribución gaussiana para cualquier $t \geq 1$. Por lo tanto, las distribuciones marginales y condicionales también son gaussianas. Dado que todas las distribuciones involucradas son gaussianas, éstas quedan totalmente determinadas por la media y la varianza. Este resultado se conoce como filtro de Kalman.

Proposición 2. (Filtro de Kalman). *Consideremos el MLD definido en (2.9). Sea*

$$\theta_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1}).$$

Entonces tenemos que:

- (i) *La distribución de pronóstico a un paso de θ_t condicionado a $y_{1:t-1}$ es gaussiana, con parámetros*

$$\begin{aligned} a_t &= \mathbb{E}(\theta_t|y_{1:t-1}) = G_t m_{t-1}, \\ R_t &= \text{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t. \end{aligned} \tag{2.11a}$$

- (ii) *La distribución de pronóstico a un paso Y_t condicionado a $y_{1:t-1}$ es gaussiana, con parámetros*

$$\begin{aligned} f_t &= \mathbb{E}(Y_t|y_{1:t-1}) = F_t a_t, \\ Q_t &= \text{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t. \end{aligned} \tag{2.11b}$$

- (iii) *La distribución de filtrado de θ_t condicionada a $y_{1:t}$ es gaussiana, con parámetros*

$$\begin{aligned} m_t &= \mathbb{E}(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t, \\ C_t &= \text{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t, \end{aligned} \tag{2.11c}$$

donde $e_t = Y_t - f_t$ es el error de pronóstico.

El filtro de Kalman nos permite calcular las distribuciones de pronóstico y de filtrado recursivamente. Iniciamos con $\theta_0 \sim \mathcal{N}(m_0, C_0)$, y continuamos con

$\pi(\theta_1|y_1)$, procedemos del mismo modo incorporando la información disponible recursivamente.

2.3.5. Suavizamiento

Una de las características útiles en los modelos de espacio de estados es que la estimación y el pronóstico se pueden aplicar recursivamente conforme tenemos información disponible. Sin embargo, en el análisis de series de tiempo usualmente las observaciones tienen cierto comportamiento periódico, el cual queremos reconstruir retrospectivamente. En este caso usaremos un algoritmo recursivo hacia atrás en el tiempo para calcular la distribución condicional de θ_t dado $y_{1:T}$, para $t < T$. Iniciaremos con las distribuciones $\pi(\theta_T|y_{1:T})$ y estimaremos los estados anteriores, veamos los detalles en la siguiente proposición.

Proposición 3. (Recursiones de suavizamiento). *Para un modelo general de espacio de estados que definimos en (H.1)-(H.2), tenemos que se cumplen las siguientes afirmaciones:*

(i) *Condicionado en $y_{1:T}$, la sucesión de estados $(\theta_0, \dots, \theta_T)$ tiene la siguiente transición de probabilidades hacia atrás*

$$\pi(\theta_t|\theta_{t+1}, y_{1:T}) = \frac{\pi(\theta_{t+1}|\theta_t)\pi(\theta_t|y_{1:t})}{\pi(\theta_{t+1}|y_{1:t})}.$$

(ii) *Las distribuciones de suavizamiento de θ_t dado $y_{1:T}$ se puede calcular de acuerdo a la siguiente recursión hacia atrás en t , iniciando en $\pi(\theta_T|y_{1:T})$*

$$\pi(\theta_t|y_{1:T}) = \pi(\theta_t|y_{1:t}) \int \frac{\pi(\theta_{t+1}|\theta_t)}{\pi(\theta_{t+1}|y_{1:t})} \pi(\theta_{t+1}|y_{1:T}) d\theta_{t+1}. \quad (2.12a)$$

En el contexto de los MLD's, las recursiones de suavizamiento tienen expresiones explícitas en términos de las medias y las varianzas de las distribuciones de suavizamiento.

Proposición 4. (Suavizador de Kalman). *Considera el MLD definido en (2.9), si $\theta_{t+1}|y_{1:t} \sim \mathcal{N}(s_{t+1}, S_{t+1})$, entonces $\theta_t|y_{1:T} \sim \mathcal{N}(s_t, S_t)$, donde*

$$\begin{aligned} s_t &= m_t + C_t G'_{t+1} R_{t+1}^{-1} (s_{t+1} - a_{t+1}), \\ S_t &= C_t - C_t G'_{t+1} R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

El suavizador de Kalman nos permite calcular la distribución de $\theta_t|y_{1:T}$, iniciando en $t = T - 1$, en cuyo caso $\theta_T|y_{1:T} \sim \mathcal{N}(s_T = m_T, S_T = C_T)$. Entonces continuamos con el cálculo hacia atrás de las distribuciones de $\theta_t|y_{1:T}$ para $t = T - 2$, $t = T - 3$ y así sucesivamente.

2.3.6. Pronóstico

Si contamos con $y_{1:t}$, usualmente estamos interesados en el pronóstico de las futuras observaciones, Y_{t+k} , o en los vectores de estado, θ_{t+k} . Para los modelos de espacio de estados, la forma recursiva del cálculo del pronóstico a un paso nos permite de manera natural actualizar la información disponible. Esto es de interés en problemas de aplicaciones en las que la información llega continuamente. Para los MLD's, el pronóstico a un paso para estados y observaciones se obtiene como consecuencia del filtro de Kalman que presentamos en la proposición 2.

Son muchas las aplicaciones en las cuales podemos estar interesados en hacer pronóstico más allá de un solo paso. Aquí presentamos expresiones recursivas para las medias y las varianzas de las distribuciones condicionales de los estados y las observaciones al tiempo $t + k$, conociendo toda la información hasta el tiempo t . Por la estructura markoviana del modelo, la distribución de filtrado funciona como una distribución inicial para la ecuación de evolución del modelo. Para los MLD's, sólo usamos los datos para obtener m_t , que es la media de $\pi(\theta_t|y_{1:t})$, ésta nos da la información suficiente para nuestra finalidad de pronosticar. Observemos que para pasar de $Y_{1:t}$ a Y_{t+k} , tal como nos muestra la figura 2.2, los datos $Y_{1:t}$ nos proporcionan la información concerniente a θ_t , la cual a su vez nos da información de la evolución del estado hasta θ_{t+k} y por lo tanto de Y_{t+k} . Es importante aclarar que mientras más grande sea k , la precisión del pronóstico será menor.

La siguiente proposición nos da expresiones recursivas para calcular las densidades de pronóstico para el modelo general de espacio de estados.

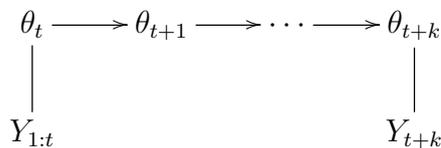


FIGURA 2.2: Flujo de información de $Y_{1:t}$ a Y_{t+k} .

Proposición 5. (Recursiones de pronóstico). *Para un modelo general de espacio de estados que definimos en (H.1)-(H.2), tenemos que para cualquier $k > 0$ se cumplen las siguientes afirmaciones:*

(i) *La densidad de pronóstico a k pasos para los estados es*

$$\pi(\theta_{t+k}|y_{1:t}) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|y_{1:t})d\theta_{t+k-1}.$$

(ii) *La densidad de pronóstico a k pasos para las observaciones es*

$$\pi(y_{t+k}|y_{1:t}) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|y_{1:t})d\theta_{t+k}.$$

Para los MLD's la proposición 5 tiene una forma específica ya que todas las integrales pueden calcularse explícitamente. Tal como ocurrió con las recursiones de filtrado, debido a que las distribuciones de pronóstico son gaussianas, es suficiente calcular las medias y las varianzas. En la proposición siguiente tenemos una manera recursiva para calcularlas, pero primero necesitamos algo de notación. Para $k \geq 1$, definamos

$$a_t(k) = E(\theta_{t+k}|y_{1:t}),$$

$$R_t(k) = \text{Var}(\theta_{t+k}|y_{1:t}),$$

$$f_t(k) = E(Y_{t+k}|y_{1:t}),$$

$$Q_t(k) = \text{Var}(Y_{t+k}|y_{1:t}).$$

Proposición 6. *Considera el MLD definido en (2.9), sea $a_t(0) = m_t$ y $R_t(0) = C_t$. Entonces, para cada $k \geq 1$, tenemos que se cumplen las siguientes afirmaciones:*

(i) *La distribución de pronóstico de θ_{t+k} condicionado a $y_{1:t-1}$ es gaussiana, con parámetros*

$$a_t(k) = G_{t+k}a_t(k-1),$$

$$R_t(k) = G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k}.$$

(ii) *La distribución de pronóstico de Y_{t+k} condicionado a $y_{1:t-1}$ es gaussiana, con parámetros*

$$f_t(k) = F_{t+k}a_t(k),$$

$$Q_t = F_{t+k}R_t(k)F'_{t+k} + V_{t+k}.$$

Capítulo 3

Especificación del modelo

3.1. MLD's univariados para análisis de series de tiempo

Cuando tenemos un MLD totalmente especificado, es decir, cuando conocemos las matrices F_t , G_t y las matrices de covarianza V_t , W_t estamos en condiciones para hacer pronóstico. Sin embargo en la práctica no es tan sencillo especificar el modelo. Un enfoque bastante general que ha mostrado ser útil en la práctica, consiste en imaginar la serie de tiempo como una combinación de componentes elementales simples. Cada uno de estos componentes captura diferentes características de la serie, tales como tendencia, estacionalidad y dependencia de covariables (en el caso de regresión). Cada componente se representa como un MLD individual, los cuales se combinan en un único MLD que nos permite analizar la serie de tiempo original. Para ser más precisos, los componentes se combinan en forma aditiva.

Consideremos una serie de tiempo univariada (Y_t). Podemos suponer que la serie se puede escribir como la suma de componentes independientes

$$Y_t = Y_{1,t} + \cdots + Y_{h,t}, \quad (3.1)$$

donde $Y_{i,t}$ puede representar el componente de tendencia, $Y_{2,t}$ un componente estacional y así sucesivamente. El componente i , $Y_{i,t}$, $i = 1, \dots, h$, se puede ver

como un MLD del siguiente modo:

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t}, & v_{i,t} &\sim \mathcal{N}(0, V_{i,t}), \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + w_{i,t}, & w_{i,t} &\sim \mathcal{N}(0, W_{i,t}), \end{aligned}$$

donde el vector de estados de dimensión p_i , $\theta_{i,t}$ son distintos y las series $(Y_{i,t}, \theta_{i,t})$ y $(Y_{j,t}, \theta_{j,t})$ son mutuamente independientes para toda $i \neq j$. Los componentes del MLD se combinan para obtener el MLD para Y_t . Por la hipótesis de independencia de los componentes, tenemos que $Y_t = \sum_{i=1}^h Y_{i,t}$ puede representarse como un MLD de la siguiente manera

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, & v_t &\sim \mathcal{N}(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t), \end{aligned}$$

donde

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{bmatrix}, \quad F_t = \begin{bmatrix} F_{1,t} & \cdots & F_{h,t} \end{bmatrix},$$

G_t y W_t son matrices diagonales de bloques

$$G_t = \begin{bmatrix} G_{1,t} & & \\ & \ddots & \\ & & G_{h,t} \end{bmatrix}, \quad W_t = \begin{bmatrix} W_{1,t} & & \\ & \ddots & \\ & & W_{h,t} \end{bmatrix},$$

con $V_t = \sum_{i=1}^h V_{i,t}$.

Esta clase de modelos nos permite analizar por bloques cada uno de los componentes de la serie de tiempo. Por lo anterior, podemos incorporar información de manera relativamente simple; ésta fue una de las razones por las que decidimos que este enfoque será el más adecuado para el estudio de nuestros datos. En el caso de datos direccionales, cuando contamos con muchos datos no podemos asegurar la estacionariedad de la serie.

3.1.1. Representación de los modelos ARMA y ARIMA por medio de un MLD

Dentro de los modelos más ampliamente usados para el estudio de series de tiempo se encuentran los modelos autorregresivos de promedios móviles o ARMA¹. Estos fueron popularizados por Box y Jenkins, ver Box et al. (2008). Como ya se mencionó, una de las ventajas del enfoque de modelo de espacio de estados es que nos permite hacer un análisis estructural del problema. Todos los componentes de la serie, tales como tendencia o periodicidad pueden ser modelados de manera independiente para después ser incorporados en un único modelo de espacio de estados. Esto nos permite incorporar características propias de nuestra información. Bajo el enfoque usual no podríamos ser capaces de incorporar de manera natural un análisis inicial de la estructura del sistema que generó los datos.

Un modelo ARMA(p, q) se define por

$$Y_t = \mu + \sum_{j=1}^p \phi_j(Y_{t-j} - \mu) + \sum_{j=1}^q \psi_j \epsilon_{t-j} + \epsilon_t, \quad (3.2)$$

donde (ϵ_t) es un ruido blanco gaussiano con varianza σ_ϵ^2 y $(Y_t)_{t \geq 1}$ es un proceso estacionario de segundo orden. Cuando los datos no son estacionarios, usualmente se toman diferencias de los datos hasta alcanzar la estacionariedad. Un modelo cuya diferencia d es un modelo ARMA(p, q) se conoce como proceso autorregresivo integrado de promedios móviles, ARIMA(p, d, q)². Cualquier modelo ARIMA puede expresarse como un MLD. Sin embargo, el espacio de estados no tiene una representación única. Existen varias propuestas de representación, pero aquí sólo veremos las más usadas. Primero analizaremos el caso estacionario. Consideremos el proceso ARMA(p, q) que se definió en (3.2), por simplicidad supondremos que μ es cero. Podemos escribir la relación como

$$Y_t = \sum_{j=1}^r \phi_j(Y_{t-j}) + \sum_{j=1}^{r-1} \psi_j \epsilon_{t-j} + \epsilon_t,$$

con $r = \max\{p, q + 1\}$, $\phi_j = 0$ para $j > p$ y $\psi_j = 0$ para $j > q$. Definimos las siguientes matrices

¹Por sus siglas en inglés, *autoregressive moving average*.

²Por sus siglas en inglés, *autoregressive integrated moving average*.

$$\begin{aligned}
F &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, \\
G &= \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \cdots & 0 & 1 \\ \phi_r & 0 & \cdots & 0 & 0 \end{bmatrix}, \\
R &= \begin{bmatrix} 1 & \psi_1 & \cdots & \psi_{r-2} & \psi_{r-1} \end{bmatrix}'.
\end{aligned} \tag{3.3}$$

Si introducimos un vector de estados de dimensión r , $\theta_t = (\theta_{1,t}, \dots, \theta_{r,t})'$, entonces el modelo ARMA puede verse como un MLD de la siguiente manera

$$\begin{aligned}
Y_t &= F\theta_t, \\
\theta_{t+1} &= G\theta_t + R\epsilon_t.
\end{aligned} \tag{3.4}$$

Este es un MLD con $V = 0$ y $W = RR'\sigma^2$, donde σ^2 es la varianza de la sucesión de error (ϵ_t). Para comprobar la equivalencia, notemos que la ecuación de observación es $y_t = \theta_{1,t}$ y las ecuaciones de estados son

$$\begin{aligned}
\theta_{1,t} &= \phi_1\theta_{1,t-1} + \theta_{2,t-1} + \epsilon_t \\
\theta_{2,t} &= \phi_2\theta_{1,t-1} + \theta_{3,t-1} + \psi_1\epsilon_t \\
&\vdots \\
\theta_{r-1,t} &= \phi_{r-1}\theta_{1,t-1} + \theta_{r,t-1} + \psi_{r-2}\epsilon_t \\
\theta_{r,t} &= \phi_r\theta_{1,t-1} + \psi_{r-1}\epsilon_t.
\end{aligned} \tag{3.5}$$

Sustituyendo la expresión $\theta_{2,t-1}$ que aparece en la segunda ecuación en la primera ecuación, tenemos

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \phi_2\theta_{1,t-2} + \theta_{3,t-2} + \psi_1\epsilon_{t-1} + \epsilon_t.$$

Haciendo sustituciones sucesivas del mismo modo, finalmente obtenemos

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \cdots + \phi_r\theta_{1,t-r} + \psi_1\epsilon_{t-1} + \cdots + \psi_{r-1}\epsilon_{t-r+1} + \epsilon_t.$$

Recordando que $r = \max\{p, q + 1\}$ y que $y_t = \theta_{1,t}$, basta observar que este modelo ARMA es de la forma de (3.2).

La representación de un modelo ARMA como un MLD (3.4) es bastante artificial.

Para poder entenderla mejor es conveniente ver algunos ejemplos que resulten ilustrativos. Consideremos un proceso autorregresivo de orden dos, es decir, un proceso AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

Un primer intento por tener una representación como MLD de un proceso AR(2) sería considerar la ecuación de observación como $F_t[Y_{t-1}, Y_{t-2}]$ y $\theta_t = [\phi_1, \phi_2]'$. Tendríamos la posibilidad de incluir la evolución temporal de los parámetros del proceso AR, o en otro caso podemos considerar $W = 0$. Pero tenemos un inconveniente, pues la matriz F_t del MLD no puede depender de los valores anteriores de las observaciones; en caso de elegir F_t del modo anterior implicaría que

$$Y_t | y_{t-1}, y_{t-2}, \theta \sim \mathcal{N}(\phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma^2).$$

Esto nos diría que Y_t no sería independiente de los valores anteriores y_{t-1}, y_{t-2} dado θ_t , en cuyo caso no cumpliríamos con la hipótesis (H.1) de la definición de modelo de espacio de estados.

Con la finalidad de mantenernos dentro de la definición, lo primero que consideramos en la representación del proceso (3.2) como MLDMLD, fue *trasladar* la dependencia del proceso autorregresivo, de Y_t al vector de estados, es decir, consideramos $Y_t = \theta_{1,t}$. Por otro lado, la segunda hipótesis de los modelos de espacio de estados considera que el proceso es markoviano, por ello necesitamos otra condición para representar la dependencia de segundo orden. Lo que haremos será aumentar el vector de estados por medio del segundo componente, $\theta_{2,t}$, y seleccionaremos G y W tales que

$$\begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} \theta_{1,t-1} \\ \theta_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}.$$

En este caso tenemos

$$\theta_{1,t} = \phi_1 \theta_{1,t-1} + \phi_2 \theta_{1,t-2} + \epsilon_t.$$

De este modo tenemos la representación de un AR(p) como MLD que cumple con las hipótesis (H.1) y (H.2) de la definición. Cuando tenemos un proceso AR(2), en

particular

$$F = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad V = 0,$$

$$G = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Este modelo nos servirá como ejemplo en el capítulo 6. El siguiente paso es expresar el componente MA(q). Observemos que para el modelo ARMA(1,1),

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t + \psi_1 \epsilon_{t-1}, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \quad (3.6)$$

con $r = q + 1 = 2$ y las matrices del MLD correspondientes son

$$F = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad V = 0,$$

$$G = \begin{bmatrix} \phi_1 & 1 \\ 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} 1 & \psi_1 \\ \psi_1 & \psi_1^2 \end{bmatrix} \sigma_y^2.$$

La representación de un proceso ARMA como un MLD es bastante útil por dos razones. La primera es porque un componente ARMA de un DLM puede explicar la autocorrelación que no explican los demás componentes estructurales como la tendencia y la estacionalidad. La segunda razón es porque la optimización de la función de verosimilitud de un modelo ARMA se puede realizar eficientemente aplicando la recursión general que se usa para calcular la verosimilitud de un MLD. La representación de un modelos ARIMA(p, d, q) como un MLD, con $d > 0$, se puede obtener como una extensión del caso estacionario. Si consideramos $Y_t^* = \Delta^d Y_t$, entonces Y_t^* es un modelo ARMA estacionario, cuya representación es la anterior. Con la finalidad de modelar la serie original (Y_t) debemos ser capaces de recuperarla a partir de Y_t^* y posiblemente de otros componentes del vector de estados. Por ejemplo, si $d = 1$, $Y_t^* = Y_t - Y_{t-1}$ y por lo tanto $Y_t = Y_t^* + Y_{t-1}$. Supongamos que Y_t^* satisface un modelo AR(2), entonces una representación por medio de un MLD para Y_t la tenemos por medio del sistema

$$Y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \theta_{t-1},$$

$$\theta_t = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \phi_1 & 1 \\ 0 & \phi_2 & 0 \end{bmatrix} \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, W), \quad (3.7)$$

con

$$\theta_t = \begin{bmatrix} Y_{t-1} \\ Y_t^* \\ \phi_2 Y_{t-1}^* \end{bmatrix} \quad (3.8)$$

y $W = \text{diag}(0, \sigma^2, 0)$. Para el caso general con $d > 1$ se puede consultar Petris et al. (2009) o Durbin and Koopman (2012). En ambas referencias se pueden encontrar las representaciones para los modelos ARMA y ARIMA. La primera de ellas sigue el enfoque bayesiano y la segunda se concentra en el análisis frecuentista de los MLD's.

3.2. Modelos con parámetros desconocidos

En las secciones anteriores presentamos algunos resultados para el análisis de series de tiempo, suponiendo que el sistema de matrices compuesto por F_t , G_t , V_t y W_t es conocido. Sin embargo, en la mayoría de las aplicaciones estas matrices son desconocidas. Las matrices asociadas a los modelos que analizaremos dependerán de un vector de parámetros desconocido, ψ . Aunque usualmente estos parámetros son constantes a través del tiempo, existe la posibilidad de incorporar variación en el tiempo.

3.2.1. Estimación por máxima verosimilitud

Supongamos que tenemos n vectores aleatorios, Y_1, \dots, Y_n , cuya distribución depende de un parámetro desconocido ψ . Denotaremos la densidad conjunta de las observaciones para valor particular del parámetro como $p(y_1, \dots, y_n; \psi)$. La función de verosimilitud se define como la densidad de probabilidad de los datos vista como función de ψ . Es decir, si denotamos la verosimilitud como L , podemos escribir $L(\psi) \propto \text{constante} \cdot p(y_1, \dots, y_n; \psi)$. En el contexto de los MLD's es conveniente escribir la densidad conjunta de las observaciones del siguiente modo

$$p(y_1, \dots, y_n; \psi) = \prod_{t=1}^n p(y_t | y_{1:t-1}; \psi), \quad (3.9)$$

donde $p(y_t|y_{1:t-1}; \psi)$ es la densidad condicional de y_t condicionada a la información hasta el tiempo $t - 1$ y el parámetro ψ es desconocido. Recordemos que las densidades del lado derecho de (3.9) son gaussianas con media f_t y varianza Q_t . Entonces podemos escribir la log-verosimilitud como

$$\ell(\psi) = -\frac{1}{2} \sum_{t=1}^n \log |Q_t| - \frac{1}{2} \sum_{t=1}^n (y_t - f_t)' Q_t^{-1} (y_t - f_t), \quad (3.10)$$

donde f_t y Q_t dependen de ψ . La ecuación (3.10) se puede maximizar numéricamente para obtener el estimador máximo verosímil (EMV) de ψ :

$$\hat{\psi} = \arg \max_{\psi} \ell(\psi). \quad (3.11)$$

En la práctica, utilizar el EMV, $\hat{\psi}$, para hacer inferencia no es conveniente cuando queremos utilizarlo en las recursiones de filtrado y suavizamiento. Una referencia para el análisis de los modelos de espacio de estados por medio de la verosimilitud es Durbin and Koopman (2012).

3.2.2. Inferencia bayesiana basada en simulación

En general, el sistema de matrices compuesto por F_t , G_t , V_t y W_t es desconocido. La inferencia se debe de hacer por medio de la distribución final asociada al modelo, la cual en la mayoría de los casos no tiene una forma explícita. Las técnicas de Monte Carlo vía cadenas de Markov nos ayudan a simular de estas distribuciones.

Tal como ya hemos hecho antes, supongamos que tenemos n vectores aleatorios, Y_1, \dots, Y_n , cuya distribución depende de un vector de parámetros desconocido, ψ . Denotaremos la densidad conjunta de las observaciones como $p(y_1, \dots, y_n|\psi)$. Debido a que cálculo del estimador máximo verosímil no nos permite abordar los problemas de filtrado ni suavizamiento, tenemos una razón más para creer que el enfoque bayesiano es el más adecuado para hacer inferencia en los MLD's tal como se propone en Migon et al. (2005) y Petris et al. (2009). Este enfoque nos ofrece una formulación más consistente del problema. El vector desconocido, ψ , puede considerarse como un vector aleatorio. Las hipótesis generales de los modelos de espacio de estados ((H.1) y (H.2)) para el proceso (Y_t) y (θ_t) se cumplen condicionados en los parámetros ψ . Observemos que $\pi(\psi)$ nos proporciona la información

inicial acerca de ψ . Entonces, para cualquier $n \geq 1$, supongamos que

$$(\theta_0, \theta_1, \dots, \theta_n, Y_1, \dots, Y_n, \psi) \sim \pi(\theta_0|\psi)\pi(\psi) \prod_{t=1}^n \pi(y_t|\theta_t, \psi)\pi(\theta_t|\theta_{t-1}, \psi) \quad (3.12)$$

(comparar con (2.8)).

Cuando contamos con los datos $y_{1:t}$, la inferencia sobre los estados θ_s al tiempo s y sobre los parámetros se resuelve calculando la distribución final conjunta

$$\pi(\theta_s, \psi|y_{1:t}) = \pi(\theta_s|\psi, y_{1:t})\pi(\psi|y_{1:t}). \quad (3.13)$$

Podemos estar interesados en el filtrado, cuando $s = t$; en el pronóstico de estados, cuando $s > t$; o en el suavizamiento, cuando $s < t$. La densidad marginal condicional de θ_s se obtiene de la ecuación (3.13); por ejemplo, la densidad de filtrado es

$$\pi(\theta_t, \psi|y_{1:t}) = \int \pi(\theta_t|\psi, y_{1:t})\pi(\psi|y_{1:t})d\psi.$$

Usualmente estamos interesados en reconstruir toda la historia de los estados hasta el tiempo t . La inferencia sobre los estados, $\theta_{0:t}$ y ψ condicionados en los datos $y_{1:t}$, se expresa por medio de la densidad final conjunta

$$\pi(\theta_{0:t}, \psi|y_{1:t}) = \pi(\theta_{0:t}|\psi, y_{1:t})\pi(\psi|y_{1:t}). \quad (3.14)$$

En este caso emplearemos los métodos MCMC para aproximar las distribuciones finales que requerimos porque no tenemos una forma explícita de la distribución. En general queremos hacer inferencia sobre MLD's que incluyen un vector de parámetros desconocidos, ψ , y las observaciones $y_{1:T}$. La incorporación de los estados en la distribución final usualmente simplifica el diseño para simular una muestra, incluso cuando sólo estamos interesados en la distribución final de los parámetros desconocidos, $\pi(\psi|y_{1:T})$. De hecho, simular de $\pi(\psi|\theta_{0:T}, y_{1:T})$ es casi invariablemente más fácil que hacerlo de $\pi(\psi|y_{1:T})$. Además existen algunos algoritmos eficientes para generar los estados condicionados en los datos y los parámetros conocidos. Esto nos sugiere que podemos muestrear de (3.14) alternando una muestra de $(\psi|\theta_{0:T}, y_{1:T})$ y $\pi(\theta_{0:T}|\psi, y_{1:T})$ por medio del algoritmo de Gibbs. Este algoritmo por bloques se propone en Migon et al. (2005). La muestra simulada nos servirá de entrada para calcular una muestra de la distribución de pronóstico de

los estados y las observaciones

$$\pi(\theta_{T+1:T+k}, y_{T+1:T+k}, \psi, \theta_T | y_{1:T}) = \pi(\theta_{T+1:T+k}, y_{T+1:T+k} | \psi, \theta_T) \pi(\psi, \theta_T | y_{1:T}).$$

Entonces, para cada par (ψ, θ_T) muestreado de $\pi(\psi, \theta_T | y_{1:T})$, podemos generar el pronóstico $\theta_{T+1:T+k}, y_{T+1:T+k}$, y así obtener la distribución de pronóstico. En principio ya somos capaces de simular de cada una de las distribuciones que necesitamos; el inconveniente es que hacerlo por medio del algoritmo de Gibbs resultaría demasiado ineficiente. Afortunadamente podemos emplear las técnicas de Monte Carlo secuenciales para evitar ese problema.

3.2.3. Simulando los estados dadas $y_{1:T}$: *forward filtering backward sampling*

Para simular de la densidad condicional $\pi(\theta_{0:T}, \psi | y_{1:T})$ necesitamos muestrear por medio del algoritmo de Gibbs de las densidades condicionales $\pi(\psi | \theta_{0:T}, y_{1:T})$ y de $\pi(\theta_{0:T} | \psi, y_{1:T})$. Mientras que para la primera de las densidades no tenemos una expresión general, para la segunda sí tenemos una manera de simular eficientemente.

Las ecuaciones de suavizamiento nos proporcionan un algoritmo para calcular $\pi(\theta_{0:T} | \psi, y_{1:T})$ para $(t = 0, 1, \dots, T)$. Para ello nos valdremos del algoritmo conocido como *forward filtering backward sampling*, FFBS³. Queremos calcular

$$\pi(\theta_{0:T} | \psi, y_{1:T}) = \prod_{t=0}^T \pi(\theta_t | \theta_{t+1}, y_{1:T}), \quad (3.15)$$

donde el último factor en el producto es simplemente $\pi(\theta_T | y_{1:T})$, es decir, la distribución de filtrado de θ_T , la cual es $\mathcal{N}(m_T, C_T)$. La ecuación (3.15) nos sugiere que si deseamos simular del lado izquierdo, podemos iniciar simulando θ_T de una $\mathcal{N}(m_T, C_T)$, después simular recursivamente θ_t de $\pi(\theta_t | \theta_{t+1:T}, y_{1:T})$ para $t = T - 1, T - 2, \dots, 0$. Por la proposición 4 sabemos que $\pi(\theta_t | \theta_{t+1:T}, y_{1:T}) = \pi(\theta_t | \theta_{t+1}, y_{1:T})$, cuya distribución es $\mathcal{N}(h_t, H_t)$ con

$$\begin{aligned} h_t &= m_t + C_t G'_{t+1} R_{t+1}^{-1} (\theta_{t+1} - a_{t+1}), \\ H_t &= C_t - C_t G'_{t+1} R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

³Por sus siglas en inglés.

Por lo tanto, después de obtener $(\theta_{t+1}, \dots, \theta_T)$, el siguiente paso consiste en simular θ_t de una $\mathcal{N}(h_t, H_t)$. Observemos que h_t depende del valor de θ_{t+1} que se generó. El muestreo que hacemos por FFBS puede verse en el algoritmo 3, tal como se propone en Migon et al. (2005).

Algoritmo 3 Forward filtering backward sampling

1. Aplica el filtro de Kalman.
 2. Simula $\theta_T \sim \mathcal{N}(m_T, C_T)$.
 3. Para $t = T - 1, \dots, 0$, simular $\theta_t \sim \mathcal{N}(h_t, H_t)$.
-

En la mayoría de las aplicaciones de los MLD's, comúnmente usamos los métodos MCMC para obtener la distribución final de ψ . Dependiendo del contexto, podemos estar interesados en hacer inferencia en los estados no observables, los parámetros desconocidos o en ambos. Para cada una de estas situaciones existen diferentes algoritmos, dos de ellos son el de Gibbs y un algoritmo híbrido, que generan una muestra de la densidad conjunta de los estados y los parámetros. Mientras que un tercer algoritmo, el muestreo marginal, sólo genera muestras de la distribución final de los parámetros. Sin embargo es útil observar que si podemos simular de la distribución final de parámetros, entonces hacerlo de la distribución conjunta de estados y parámetros se puede obtener fácilmente por medio de

$$\pi(\theta_{0:T}, \psi | y_{1:T}) = \pi(\theta_{0:T} | \psi, y_{1:T}) \pi(\psi | y_{1:T}).$$

Para ver cómo funciona exactamente el algoritmo de Gibbs en conjunto con el FFBS presentamos el algoritmo 4.

Algoritmo 4 Forward filtering backward sampling dentro del algoritmo de Gibbs

0. Inicia: fija $\psi = \psi^{(0)}$.
 1. Para $i = 1, \dots, N$:
 - a) Simula $\theta_{0:T}^{(i)}$ de $\pi(\theta_{0:T} | y_{1:T}, \psi = \psi^{(i-1)})$ usando FFBS.
 - b) Simula $\psi^{(i)}$ de $\pi(\psi | y_{1:T}, \theta_{0:T} = \theta_{0:T}^{(i)})$.
-

Cuando ψ es un vector de dimensión r , es más sencillo simular por medio de Gibbs cada uno de los componentes del vector en lugar de hacerlo todo de una sola vez. En caso de que la distribución final fuese más difícil de simular, podemos emplear el algoritmo de Metropolis-Hastings o el Metropolis-Hastings adaptativo (ARMS).

El segundo enfoque, muestreo marginal, es más simple conceptualmente. Consiste

en muestrear de $\pi(\psi|y_{0:T})$. La implementación concreta depende del modelo en cuestión y una vez más, ψ normalmente es multivariado. Tal como mencionamos antes, en este caso los algoritmos de Metropolis-Hastings y ARMS son los más adecuados.

Finalmente, existe un algoritmo híbrido cuando el parámetro ψ tiene dos componentes y se puede escribir como $\psi = (\psi_1, \psi_2)$, donde cada componente puede ser univariado o multivariado. Una descripción concisa se presenta en el algoritmo 5.

Algoritmo 5 Forward filtering backward sampling dentro de un algoritmo híbrido

0. Inicia: fija $\psi_2 = \psi_2^{(0)}$.
 1. Para $i = 1, \dots, N$:
 - a) Simula $\psi_1^{(i)}$ de $\pi(\psi_1|y_{1:T}, \psi_2 = \psi_2^{(i-1)})$.
 - b) Simula $\theta_{0:T}^{(i)}$ de $\pi(\theta_{0:T}|y_{1:T}, \psi_1 = \psi_1^{(i)}, \psi_2 = \psi_2^{(i-1)})$ usando FFBS.
 - c) Simula $\psi_2^{(i)}$ de $\pi(\psi_2|y_{1:T}, \theta_{0:T} = \theta_{0:T}^{(i)}, \psi_1 = \psi_1^{(i)})$.
-

Una vez más, tal como ocurrió en los esquemas anteriores, dentro de los pasos a) o c) es posible hacer la simulación por medio de Metropolis-Hastings o ARMS.

Cuando generamos una cadena de Markov por medio de simulación siempre debemos revisar que la trayectoria alcance la convergencia a la distribución estacionaria. Esto podemos verificarlo empíricamente a través de las trayectorias de la cadena y los promedios ergódicos. Idealmente, esperaríamos que las muestras tuvieran una correlación baja entre ellas. En caso de no ser así podemos *adelgazar* la cadena simulada, es decir, quitar ciertas simulaciones cada cierto número fijo de veces. También es recomendable incluir un periodo de *calentamiento* para permitir que la cadena alcance la estacionariedad.

3.2.4. Varianza desconocida constante: gamma inversa como distribución inicial

Este es uno de los modelos más simples y más usados cuando desconocemos la varianza. Supongamos que las observaciones son univariadas, de dimensión $m = 1$. Si la varianza de las observaciones y las matrices de covarianzas asociadas a la evolución son desconocidas, lo más simple es considerarlas invariantes en el tiempo, con W diagonal. Es decir, si trabajamos con las precisiones, la distribución inicial gamma inversa supone que

$$V_t = \phi_y^{-1}, \quad W_t = \text{diag}(\phi_{\theta,1}^{-1}, \dots, \phi_{\theta,p}^{-1})$$

y $\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}$ tienen distribuciones gamma independientes. Por lo tanto, la distribución inicial del vector de varianzas $(\phi_y^{-1}, \phi_{\theta,1}^{-1}, \dots, \phi_{\theta,p}^{-1})$ es el producto de $d = (p+1)$ densidades gamma inversas. Respecto a los hiperparámetros, es conveniente expresar la información inicial por medio de precisiones desconocidas, $E(\phi_y) = a_y$, y $E(\phi_{\theta,i}) = a_{\theta,i}$, con incertidumbre en términos de la varianza inicial $\text{Var}(\phi_y) = b_y$, $\text{Var}(\phi_{\theta,y}) = b_{\theta,i}$, con $i = 1, \dots, p$. En este caso, podemos parametrizar la distribución gamma inicial como

$$\phi_y \sim \mathcal{G}(\alpha_y, \beta_y), \quad \phi_{\theta,i} \sim \mathcal{G}(\alpha_{\theta,i}, \beta_{\theta,i}), \quad i = 1, \dots, p,$$

con

$$\alpha_y = \frac{a_y^2}{b_y}, \quad \beta_y = \frac{a_y}{b_y}, \quad \alpha_{\theta,i} = \frac{a_{\theta,i}^2}{b_{\theta,i}}, \quad \beta_{\theta,i} = \frac{a_{\theta,i}}{b_{\theta,i}}, \quad i = 1, \dots, p.$$

Cuando tenemos las observaciones, $y_{1:T}$, la distribución final conjunta de los estados $\theta_{0:T}$ y de los parámetros desconocidos $\psi = (\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p})$ es proporcional a la densidad conjunta

$$\begin{aligned} \pi(y_{1:T}, \theta_{0:T}, \psi) &= \pi(y_{1:T} | \theta_{0:T}, \psi) \pi(\theta_{0:T} | \psi) \pi(\psi) \\ &= \prod_{t=1}^T \pi(y_t | \theta_t, \phi_y) \cdot \prod_{t=1}^T \pi(\theta_t | \theta_{t-1}, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \cdot \pi(\theta_0) \cdot \pi(\phi_y) \cdot \prod_{i=1}^p \pi(\phi_{\theta,i}). \end{aligned}$$

Observemos que el segundo producto en la factorización también se puede escribir como un producto sobre $i = 1, \dots, p$, debido a que W es diagonal. Esta factorización alternativa es útil cuando queremos generar la distribución completa de las $\phi_{\theta,i}$'s. Podríamos usar el algoritmo de Gibbs para el modelo de la gamma inversa, así simularíamos de la distribución condicional completa de los estados y de la distribución de $\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}$. Para simular de los estados podríamos usar el algoritmo 4 (FFBS).

Ahora veamos cómo podemos generar la distribución condicional de ϕ_y ⁴

$$\begin{aligned} \pi(\phi_y | \dots) &\propto \prod_{t=1}^T \pi(y_t | \theta_t, \phi_y) \pi(\phi_y) \\ &\propto \phi_y^{\frac{T}{2} + \alpha_y - 1} \exp \left\{ \phi_y \left[\frac{1}{2} \sum_{t=1}^T (y_t - F_t \theta_t)^2 + \beta_y \right] \right\}. \end{aligned}$$

⁴Los puntos en las expresiones $\pi(\phi_y | \dots)$ y $\pi(\phi_{\theta,i} | \dots)$ sustituyen cualquier otra variable aleatoria en el modelo excepto ϕ_y , incluyendo los estados $\theta_{0:T}$.

La distribución condicional de ϕ_y es nuevamente una distribución gamma,

$$\pi(\phi_y | \dots) \sim \mathcal{G} \left(\alpha_y + \frac{T}{2}, \beta_y + \frac{1}{2}SS_y \right),$$

con $SS_y = \sum_{t=1}^T (y_t - F_t \theta_t)^2$. Del mismo modo, se puede comprobar que distribuciones de las $\phi_{\theta,i}$'s son

$$\pi(\phi_{\theta,i} | \dots) \sim \mathcal{G} \left(\alpha_{\theta,i} + \frac{T}{2}, \beta_{\theta,i} + \frac{1}{2}SS_{\theta,i} \right) \quad i = 1, \dots, p,$$

con $SS_{\theta,i} = \sum_{t=1}^T (\theta_{t,i} - (G_t \theta_{t-1})_i)^2$. Una discusión extensa con ejemplos de este tipo de planteamientos para MLD's puede verse en Petris et al. (2009).

Capítulo 4

Análisis bayesiano de datos circulares

En el capítulo 2 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 $f_w(x)$ del siguiente modo

$$f_w(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi), \quad 0 \leq x < 2\pi. \quad (4.1)$$

Así, obtenemos una variable aleatoria circular X definida en $[0, 2\pi)$ a través de la transformación $Y = X + 2k\pi$, con $k \in \mathbb{Z}$ una variable latente que además no es observable. Es decir, obtenemos f_w aplicando una función de densidad definida en \mathbb{R} , f , al argumento $X + 2k\pi$. La variable aleatoria X es una versión envuelta de la variable aleatoria en la recta, Y , la cual tiene función de densidad de probabilidad $f(y|\varphi)$ que depende de un vector de parámetros desconocido φ . 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 de dimensión d 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 de dimensión d sobre el círculo, con densidad

$$f_w(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_d=-\infty}^{\infty} f(x_1 + 2k_1\pi, \dots, x_d + 2k_d\pi). \quad (4.2)$$

Marginalmente, cada variable X_i tiene una distribución normal envuelta univariada, mientras la estructura de correlación está definida por la matriz de varianzas-covarianzas Σ de la distribución en la recta Y . Supongamos que tenemos n repeticiones de la variable normal Y , de dimensión d ; en forma matricial podemos escribir del siguiente modo

$$Y = X + 2k\pi, \quad (4.3)$$

donde $k = \{k^{(1)}, \dots, k^{(n)}\}'$ y $X = \{X^{(1)}, \dots, X^{(n)}\}'$ denota un conjunto de n variables aleatorias circulares de dimensión d . 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 distribuidos. Siendo más precisos, la ecuación (4.3) 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 (4.2) 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 de este proceso, la correspondiente densidad envuelta es

$$f_d(x_1, \dots, x_d) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_d=-\infty}^{\infty} f(p_1 + 2k_1\pi, \dots, p_d + 2k_d\pi). \quad (4.4)$$

Estos modelos para series de tiempo han sido estudiados por Breckling (1989) y Craig (1988), mientras que la teoría general de distribuciones circulares envueltas se discute en Fisher (1993) y Mardia and Jupp (1999).

La posible extensión para el caso multivariado que presentan (4.2) y (4.4) 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. Además, 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 recta se preservan bajo el proceso.

Por otro lado, la dificultad al usar estos modelos se debe a la complejidad en la inferencia. Trabajar directamente con la densidad (4.1) o (4.2) hace la estimación de parámetros casi intratable. Una de las soluciones fue desarrollada en Fisher and Lee (1994), quienes usaron el algoritmo EM para hacer la estimación, considerando los coeficientes de envoltura, los k 's, como datos faltantes. 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 un procedimiento motivado por el trabajo de Coles (1998), en conjunto con la teoría de los MLD's. 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 se menciona en Coles (1998). Este enfoque es más natural que un tratamiento de los datos por medio del paradigma frecuentista, además de la menor dificultad en cuanto a la implementación. Un trabajo basado en la verosimilitud de los procesos envueltos puede verse en Craig (1988). En éste se exponen los resultados correspondientes a un tipo específico de modelo con resultados poco alentadores en cuanto a su generalización, sobre todo por la dificultad para calcular el máximo de la función de verosimilitud.

La principal dificultad al trabajar bajo el enfoque de envolver 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 recta. En ese caso es posible desenvolver la distribución sobre

el círculo y obtener una distribución sobre la recta real, llegando a un punto en que será posible usar toda la teoría de los MLD's para las variables definidas sobre la recta real. En otras palabras, dada una variable aleatoria circular, X , siempre es posible obtener la correspondiente variable aleatoria sobre la recta, 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 .

Hacer inferencia sobre X , cuya función de densidad es $f_w(y)$, escrita en (4.1), resultará imposible. En su lugar la función de densidad en la recta es más simple y describe la variable Y , o equivalentemente el par (X, k) del cual sólo observamos X . La equivalencia

$$Y \equiv (X, k), \quad (4.5)$$

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

Como k es una variable no observable, una aproximación natural (pero sobre todo conveniente) que nos permite superar la incertidumbre sobre k escribiéndola como distribución condicional, es el enfoque bayesiano con el apoyo los métodos MCMC. La base de dichos métodos 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 (k, φ) .

4.1. Proceso autorregresivo envuelto con parámetros desconocidos: ejemplo simulado

El proceso de envolver puede ser aplicado a cada componente marginal de un proceso estocástico para obtener procesos estocásticos circulares. En este caso, (4.4) determina la función de densidad de probabilidad del subcomponente de dimensión d 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 a los procesos autoregresivos envueltos, ver Breckling (1989). Como caso particular de

(3.2), tenemos los procesos $\text{AR}(p)$

$$Y_t - \mu = \sum_{j=1}^p \phi_j (Y_{t-j} - \mu) + \epsilon_t, \quad (4.6)$$

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

$$X_t = Y_t \pmod{2\pi}. \quad (4.7)$$

Se puede manipular otra clase de procesos de manera similar. Como resultado de la transformación obtendremos una serie univariada de datos envueltos y sus respectivos coeficientes de envoltura. En esta sección analizaremos un proceso autorregresivo envuelto simulado de orden 2 que denotaremos como $\text{WAR}(2)$ ¹. En lo sucesivo será conveniente adoptar algunas convenciones de notación. La muestra observada de tamaño T , la denotaremos como

$$X = \{x_1, \dots, x_T\}.$$

De manera similar, denotamos los coeficientes de envoltura como

$$k = \{k_1, \dots, k_T\},$$

los cuales supondremos desconocidos. Los componentes faltantes los denotaremos con superíndices negativos del siguiente modo

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

Nuestra finalidad es obtener la distribución final $\pi(\varphi|X)$. Ya que hemos elegido trabajar en el contexto de los MLD's, en este caso φ es un vector que incluye todos los parámetros asociados al modelo. Es decir, incluye los parámetros de estado, los coeficientes autoregresivos y la varianza del proceso. Hacer inferencia directamente resulta imposible, por lo que incorporamos el vector k como variable latente. Ahora, nuestra intención es simular de $\pi(\varphi, k|X)$. Para ello necesitamos la siguiente expresión

$$\pi(\varphi|k, X) = \pi(\varphi|y_1, y_2, \dots, y_T), \quad (4.8)$$

¹Por la traducción de inglés, *wrapped autoregressive process*.

la igualdad se cumple por (4.7). Por otro lado observemos que

$$\pi(k_j | X, k^{(-j)} \varphi) = \frac{\pi(X, k | \varphi)}{\pi(X, k^{(-j)} | \varphi)} \propto \pi(X, k | \varphi). \quad (4.9)$$

El algoritmo que emplearemos será el muestreo de Gibbs, que funcionará por bloques. El primero de ellos está formado por todos los parámetros del modelo autorregresivo, φ , cuya simulación se hace con los datos desenvueltos, como vemos en la ecuación (4.8), en conjunto con la teoría de los MLD's. El segundo bloque es el de los coeficientes de envoltura y los simulamos por medio de la ecuación (4.9). Dentro del mismo muestreo de Gibbs principal, emplearemos el algoritmo Metropolis-Hastings para generar el vector k . Como ejemplo, simulamos un proceso AR(2) para generar un proceso WAR(2), cuyo vector k suponemos desconocido. En particular, $T = 350$, $\phi_1 = 0.6$, $\phi_2 = -0.2$ y $\sigma^2 = 1$. En la figura 4.1 se observa la serie simulada original a) con su correspondiente versión envuelta b). En este caso, el objetivo es hacer inferencia sobre a) a partir de b).

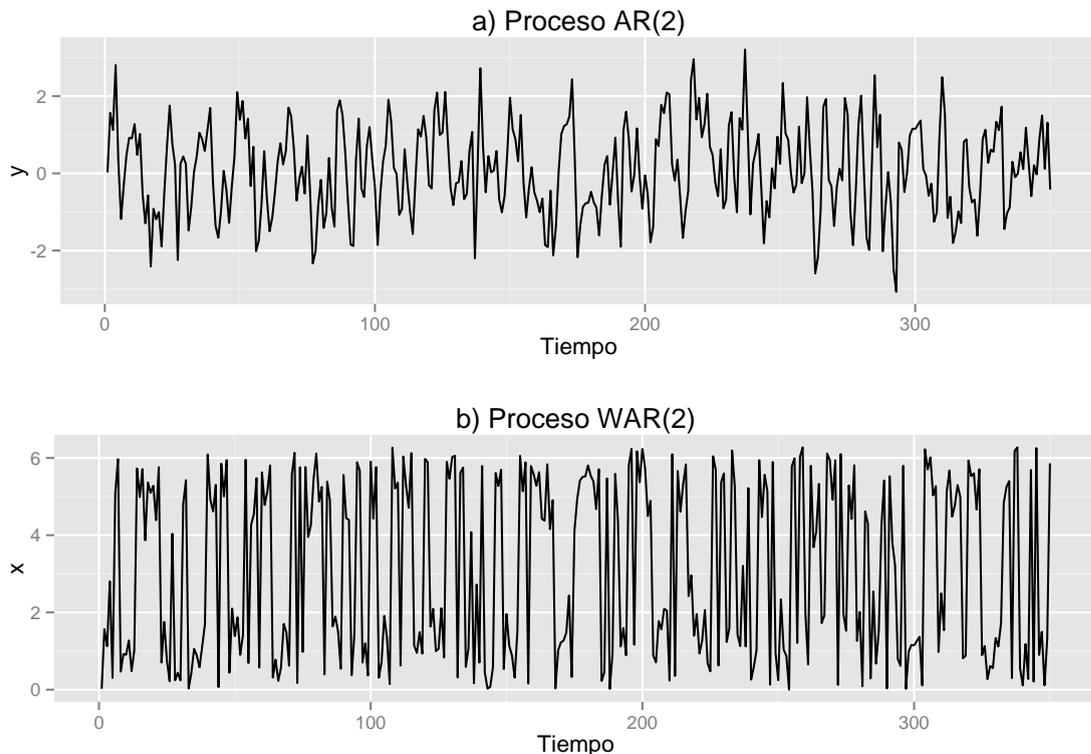


FIGURA 4.1: a) Proceso AR(2), b) Proceso WAR(2) con $\mu = 0$, $\phi_1 = 0.6$, $\phi_2 = -0.2$ y $\sigma^2 = 1$.

Los parámetros autorregresivos ϕ_1 y ϕ_2 deben encontrarse en la región de estacionariedad, $\mathcal{S} \subset \mathbb{R}^2$, que cumple

$$\phi_1 + \phi_2 < 1, \quad \phi_1 - \phi_2 > -1, \quad |\phi_2| < 1.$$

Aunque en principio no existe un orden para iniciar el algoritmo, nosotros decidimos simular primero el coeficiente k . Para tener una muestra de este bloque, necesitamos la verosimilitud exacta del proceso AR y los T datos. Si $Y = (y_1, y_2, \dots, y_T)'$ y $\phi = (\phi_1, \phi_2)$, tenemos entonces

$$\ell(\phi, \sigma^2) = \Psi(\phi, \sigma^2) \times (\sigma^2)^{-(n-2)/2} \times \exp \left[-\frac{1}{2\sigma^2} \sum_{t=3}^T (y_t - w_t' \phi)^2 \right], \quad (4.10)$$

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

$$\Psi(\phi, \sigma^2) = (\sigma^2)^{-1} |V^{-1}|^{1/2} \exp \left[-\frac{1}{2\sigma^2} Y_2' V^{-1} Y_2 \right] \quad (4.11)$$

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 (4.10) es proporcional a la densidad de observaciones (y_3, \dots, y_T) dado Y_2 . Si la única información inicial disponible es que el proceso es estacionario, entonces la distribución final de los parámetros es

$$\pi(\phi, \sigma^2 | Y) \propto \ell(\phi, \sigma^2) I[\phi \in \mathcal{S}]$$

donde $I[\phi \in \mathcal{S}]$ es 1 si $\phi \in \mathcal{S}$ y 0 en otro caso. En resumen, el tercer término de la ecuación (4.10) y la ecuación (4.11) son las expresiones explícitas de (4.9) para (y_1, y_2) y (y_3, \dots, y_T) respectivamente; recordemos que $y_i = (x_i, k_i)$ con $i = 1, \dots, T$. Las expresiones anteriores nos permiten calcular la probabilidad (2.1) del algoritmo Metropolis-Hastings, con la cual decidimos si rechazamos o no el valor propuesto para cada uno de los k 's.

Además de la especificación de la distribución objetivo, el algoritmo requiere hacer una elección de la función de transición, q . Generalmente esta elección puede afectar la eficiencia del algoritmo radicalmente. El modelo de transición elegido en este

trabajo para cada coeficiente de envoltura k_j es la siguiente caminata aleatoria

$$q(k_j, k'_j) = k_j + \epsilon, \quad j = 1, \dots, T,$$

donde $\epsilon = \{-1, 0, 1\}$ con probabilidades $\{p, 1-2p, p\}$ respectivamente, para alguna elección apropiada de p .

En lo concerniente al segundo bloque, usaremos la teoría de los MLD's para simular de (4.8), que es precisamente un modelo AR(2). El algoritmo híbrido 5 nos permite simular primero de los coeficientes autorregresivos por medio de $\pi(\phi_1, \phi_2 | \sigma_y^2, y_{1:T})$. De acuerdo con la notación del algoritmo, $\psi_1 = (\phi_1, \phi_2)$ y $\psi_2 = (\sigma_y^2)^{-1}$. La distribución inicial para (ϕ_1, ϕ_2) es el producto de dos normales, $\mathcal{N}(0, (2/3)^2)$ y $\mathcal{N}(0, (1/3)^2)$, restringido a \mathcal{S} . De este modo, penalizamos aquellos valores cercanos a la región de no estacionariedad. Los coeficientes autorregresivos condicionados a las precisiones, pero no a los estados, no tienen una distribución estándar. En este caso es recomendable usar el algoritmo ARMS para simular de distribución condicional completa. Por otro lado, y tal como lo indica dicho algoritmo, los estados los generamos por FFBS. Para la precisión, es decir, para el recíproco de la varianza, σ_y^2 , suponemos una distribución gamma como inicial con media a y varianza b como hiperparámetros, es decir,

$$\mathcal{G}\left(\frac{a^2}{b}, \frac{a}{b}\right).$$

En particular, elegimos $a = 1$ y $b = 1000$, la precisión condicionada a los estados y los coeficientes autorregresivos tiene distribución gamma,

$$(\sigma_y^2)^{-1} | \dots \sim \mathcal{G}\left(\frac{a^2}{b} + \frac{T}{2}, \frac{a}{b} + \frac{1}{2}SS_y\right),$$

con $SS_y = \sum_{t=1}^T (y_t - F_t \theta_t)^2$. Los pasos para poder hacer las simulaciones correspondientes al proceso WAR(2) se encuentran en el algoritmo 6.

Algoritmo 6 Muestreo para el proceso WAR(2)

0. Inicia: fija $k = k^{(0)}$ y $\psi_2 = \psi_2^{(0)}$.
 1. Para $i = 1, \dots, N$:
 - a) Simula $k^{(i)}$ de acuerdo a (4.9) usando Metropolis-Hastings.
 - b) Simula $\psi_1^{(i)}$ de $\pi(\psi_1 | y_{1:T}, \psi_2 = \psi_2^{(i-1)})$ usando ARMS.
 - c) Simula $\theta_{0:T}^{(i)}$ de $\pi(\theta_{0:T} | y_{1:T}, \psi_1 = \psi_1^{(i)}, \psi_2 = \psi_2^{(i-1)})$ usando FFBS.
 - d) Simula $\psi_2^{(i)}$ de $\pi(\psi_2 | y_{1:T}, \theta_{0:T} = \theta_{0:T}^{(i)}, \psi_1 = \psi_1^{(i)})$ de acuerdo a la sección 3.2.4.
-

La identificabilidad del modelo es generalmente un problema cuando se trabaja con series de tiempo; en este modelo es aún más complicado por la incorporación del vector k . Presentaremos los resultados correspondientes al proceso WAR(2) simulado. Las trayectorias de los parámetros simulados se muestran en la figura 4.2. En la imagen sólo presentamos las últimas 5000 iteraciones de las 10000 que se realizaron; también es importante mencionar que adelgazamos las cadenas para alcanzar la estacionariedad. En el caso del bloque para estimar el vector k , el adelgazamiento fue de 50, mientras que para el bloque del resto de los parámetros fue de 10. En la figura 4.3 podemos observar tanto los promedios ergódicos como las funciones de autocorrelación para cada uno de los parámetros del modelo. Basados en este diagnóstico simple, descartamos un comportamiento de no estacionariedad. Podemos observar las densidades suavizadas en la figura 4.4.

La convergencia no acelerada se debe a que los datos simulados causan dependencia en los coeficientes de envoltura asociados. Por ejemplo, si Y_t excede 2π , existe una posibilidad mayor de que Y_{t+1} sea mayor que 2π ; entonces tanto k_j como k_{j+1} no son cero. Sin embargo, la actualización en la rutina para cada coeficiente es aplicada independientemente del valor de los otros. Esto resulta en una rutina 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) simulado.

En principio, se podría aplicar un algoritmo modificado 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 que generan 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 0.57, -0.21 y 1.05 respectivamente, mientras los valores muestrales son 0.6, -0.2 y 1.0; esta información se resume en el cuadro 4.1. Los resultados sugieren que el algoritmo está estimando bien a pesar de la incertidumbre inducida por los coeficientes de envoltura desconocidos.

Dado que una de las ventajas de los MLD's es precisamente la relativa facilidad para implementar el pronóstico de un modelo, podemos incluir el pronóstico a uno o más pasos del proceso WAR(2). Este es un avance respecto a los trabajos

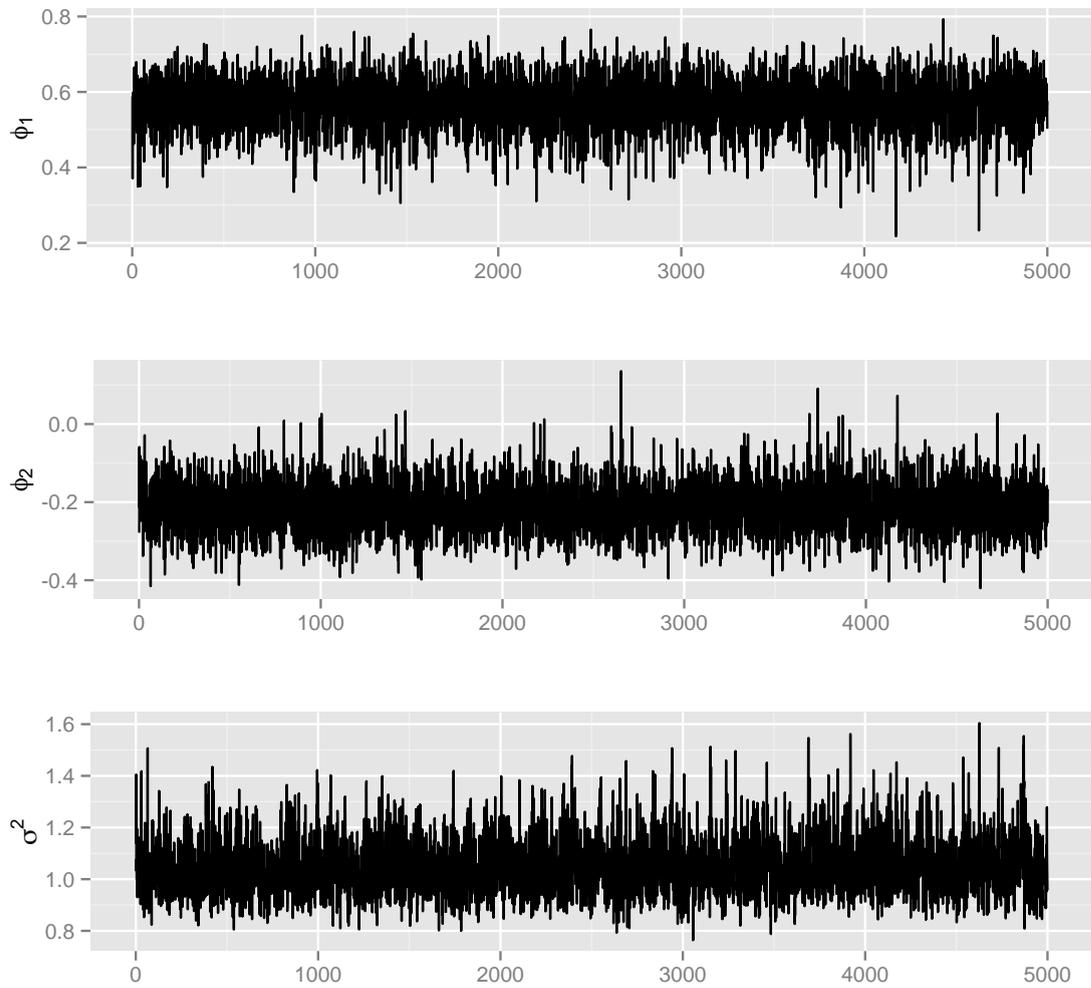


FIGURA 4.2: Trayectorias del algoritmo MCMC de los parámetros ϕ_1 , ϕ_2 y σ^2 respectivamente.

Parámetro	Valor verdadero	Media	Intervalo del 0.90 de probabilidad
ϕ_1	0.6	0.57	[0.45, 0.67]
ϕ_2	-0.2	-0.21	[-0.31, -0.11]
σ^2	1.0	1.05	[0.90, 1.25]

CUADRO 4.1: Resumen del proceso WAR(2) simulado.

previos de Coles (1998) y Ferrari (2009) que no abordaron el tema de pronóstico para los modelos autorregresivos. Por lo anterior, podemos hablar de la ventaja del planteamiento de este tipo de problemas por medio de un MLD respecto a modelos menos generales.

En cuanto al pronóstico a un paso, el enfoque bayesiano nos dice que tendríamos

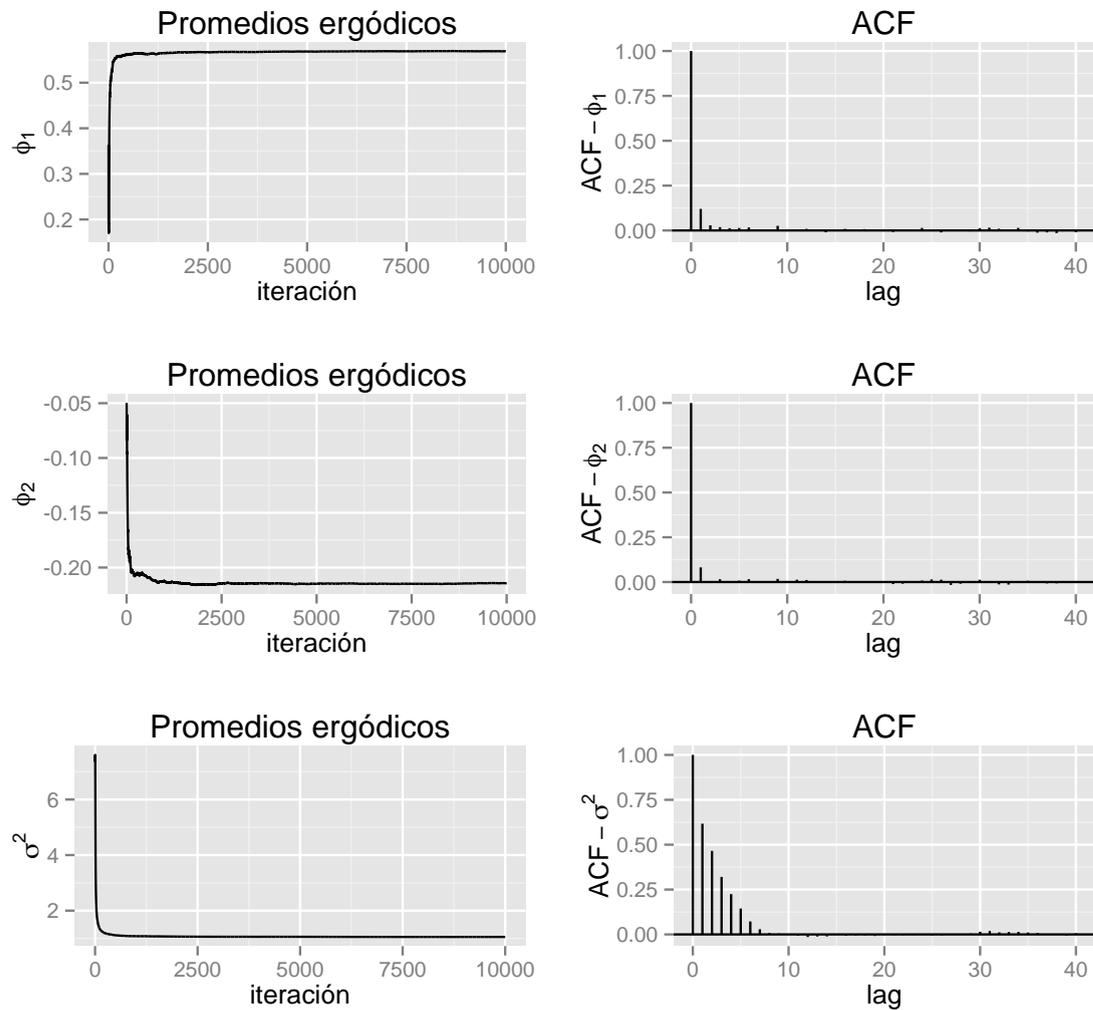


FIGURA 4.3: Promedios ergódicos y funciones de autocorrelación de ϕ_1 , ϕ_2 y σ^2 respectivamente.

que simular cada uno de los coeficientes de envoltura tal como hemos venido haciéndolo para inmediatamente desenvolver el valor obtenido. Después del paso anterior estamos en condiciones de usar el filtro de Kalman para obtener el pronóstico. En particular, para nuestro ejemplo simulado, después de analizar la estabilización de los coeficientes de envoltura decidimos quedarnos con la moda de cada uno de ellos. De manera similar procedemos cuando hacemos predicción a más pasos, simplemente para no incrementar los pasos de cómputo. En la figura 4.5² presentamos en azul una simulación del pronóstico a 24 pasos adelante del

²La representación en la recta de datos circulares presenta una dificultad importante. Para cálculos anteriores hemos considerado la medición de los ángulos a partir de 0 y en dirección contraria al movimiento del reloj. En adelante, para facilitar la interpretación de algunas gráficas consideraremos una escala alternativa. Cuando el ángulo medido en radianes se encuentre en $[0, \pi]$, lo mediremos a partir de 0 y en dirección contraria al reloj. Si el ángulo en cuestión

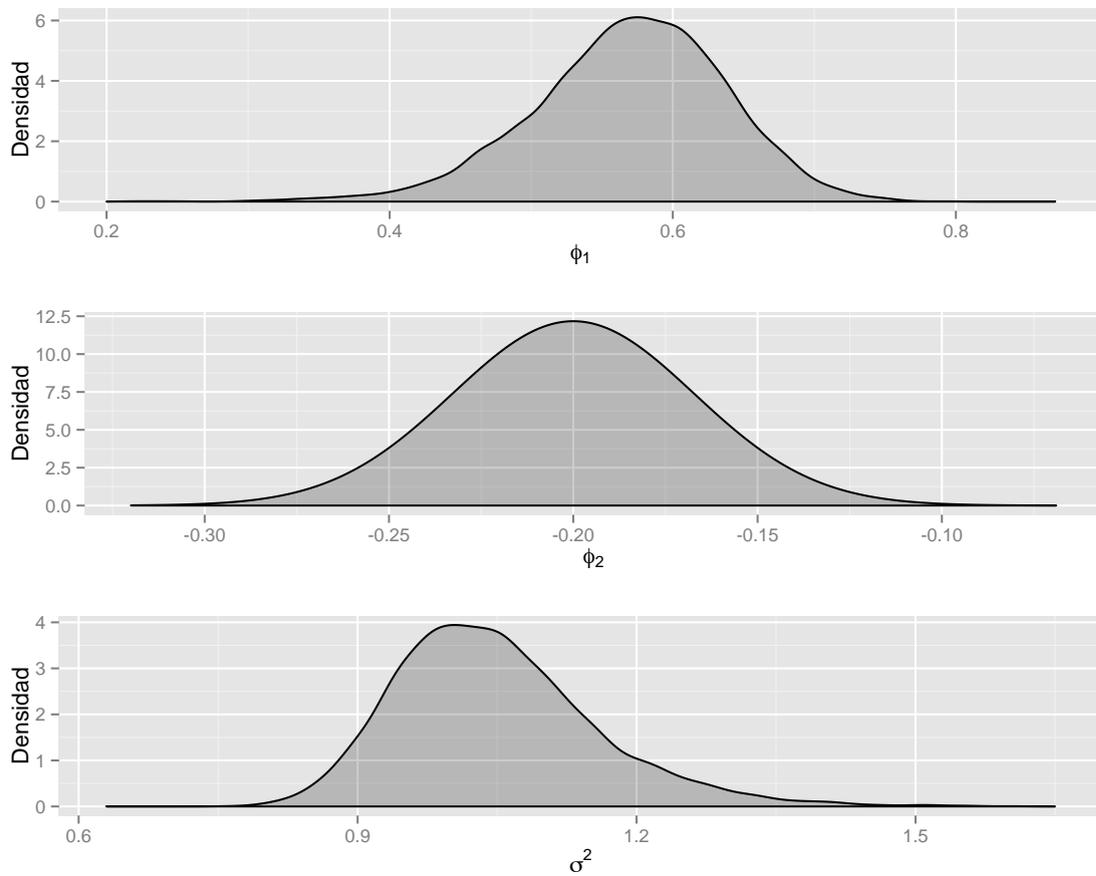


FIGURA 4.4: Densidades finales suavizadas de los parámetros ϕ_1 , ϕ_2 y σ^2 del ejemplo simulado.

proceso envuelto, después de haber estimado el vector k . También incluimos, en azul marino, la media del pronóstico; así como un intervalo de predicción para la media. En caso de que quisiéramos obtener el pronóstico para el proceso AR(2) sin envolver, simplemente tenemos que tomar en cuenta la igualdad (4.7) para recuperar el pronóstico del modelo AR(2)

El siguiente paso es evaluar el pronóstico de nuestro modelo. Una manera consistente con el enfoque bayesiano es utilizar la distribución de pronóstico tal como se sugiere en la subsección 2.1.1. La distribución de pronóstico en este caso está definida en el círculo.

se encuentra en $(\pi, 2\pi)$ lo mediremos en dirección del movimiento del reloj y lo registraremos con valor negativo con respecto a 0. Usaremos de manera libre una u otra escala, únicamente aclararemos si los datos son envueltos o no. Con esto, eliminaremos la falsa noción de una excesiva variabilidad alrededor del 0.

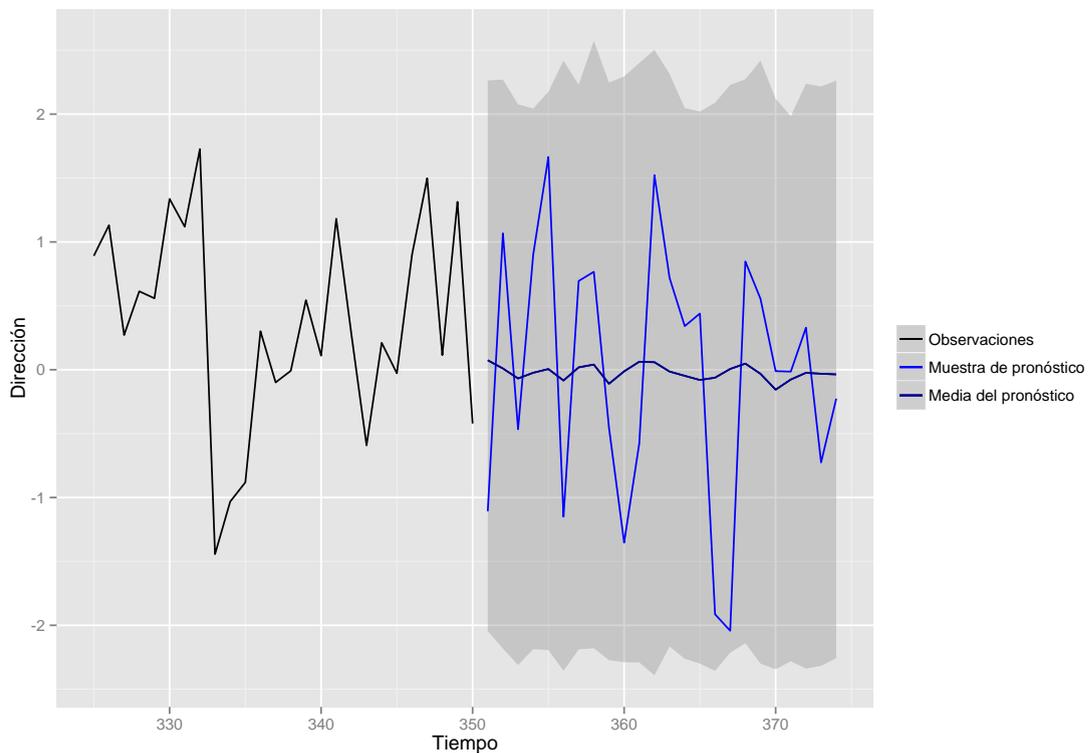


FIGURA 4.5: Pronóstico a 24 pasos de la serie correspondiente al proceso $WAR(2)$.

Para la validación de nuestro modelo de pronóstico omitimos la última observación de la serie simulada, en este caso la número 350. Ajustamos el modelo y calculamos la función predictiva, en la cual evaluamos el valor que omitimos. Tal como podemos ver en la figura 4.6, la observación que omitimos toma un valor elevado de la distribución de predictiva bimodal. En cuanto a la estimación puntual del pronóstico tenemos que ser cuidadosos considerando que no tenemos un único estimador puntual. En la figura 4.7 podemos ver la distribución de pronóstico circular a un paso con el respectivo valor que omitimos, el punto en color azul. El intervalo del 0.90 de probabilidad para el pronóstico se encuentra delimitado por las flechas delgadas; mientras que la media circular es marcada por medio de la tercera flecha. Es importante destacar que para dar una estimación puntual del pronóstico es importante analizar tanto la distribución de pronóstico en su representación en la recta, ver figura 4.6, como la representación circular, ver figura 4.7. Observemos que a pesar de que nuestro pronóstico no es tan cercano a la media circular sí es un valor cercano a una de las modas que se puede observar en la figura 4.6.

Para construir un intervalo de predicción para nuestro pronóstico; una propuesta sería incluir una serie de círculos consecutivos similares a la figura 4.7, uno por cada uno de los 24 valores pronosticados. De manera alternativa podemos plantear una banda de predicción para la media de los pronósticos no envueltos, tal como decidimos hacerlo en la figura 4.5.

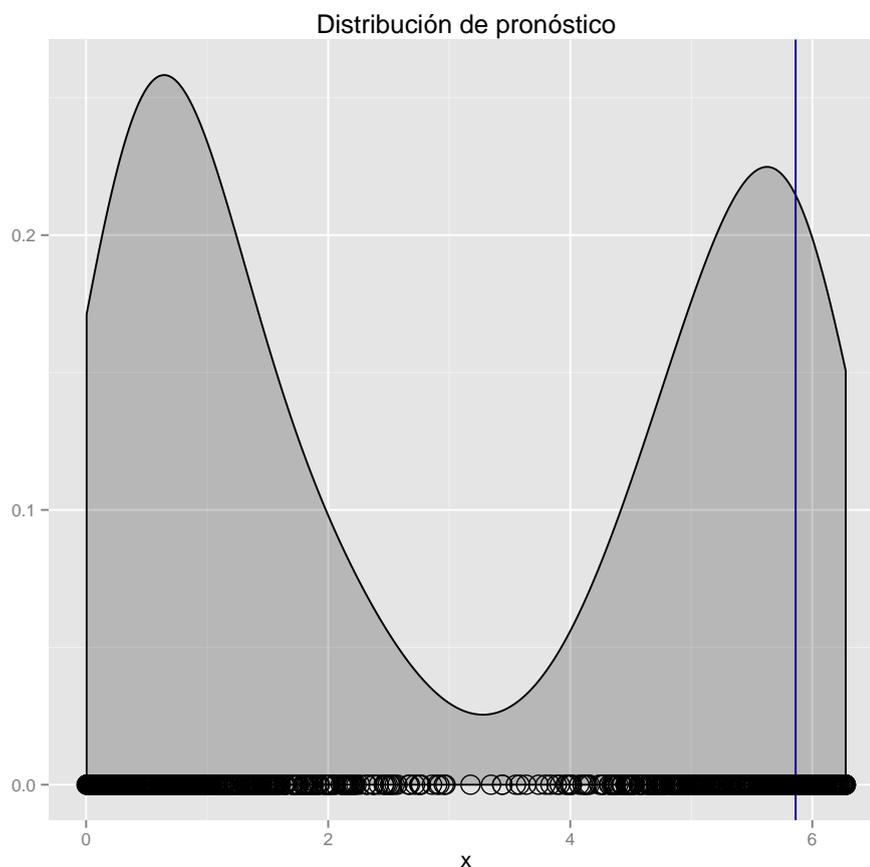


FIGURA 4.6: Evaluación de valor omitido en la distribución de pronóstico a un paso.

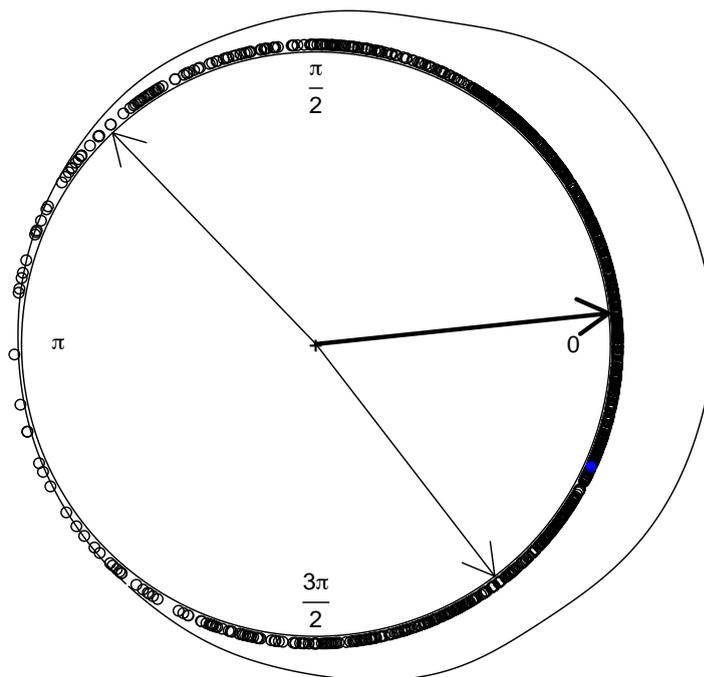
Distribución de pronóstico circular

FIGURA 4.7: Evaluación de valor omitido en la distribución de pronóstico circular a un paso.

Capítulo 5

Análisis de los datos de la estación Laguna Verde

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 que este modelo no es adecuado debido a las complejas características climáticas y orográficas de la región próxima a la CNLV. También se ha concluido que un modelo físico no es posible, por lo que se debe abordar el problema desde un enfoque estadístico. En este capítulo aplicamos los procedimientos mostrados en capítulos anteriores para modelar la dirección del viento registrada en la CNLV.

Datos de direcciones de viento

Contamos con información correspondiente a cuatro estaciones ananométricas ubicadas en los alrededores de la zona conocida como Laguna Verde. En cada una

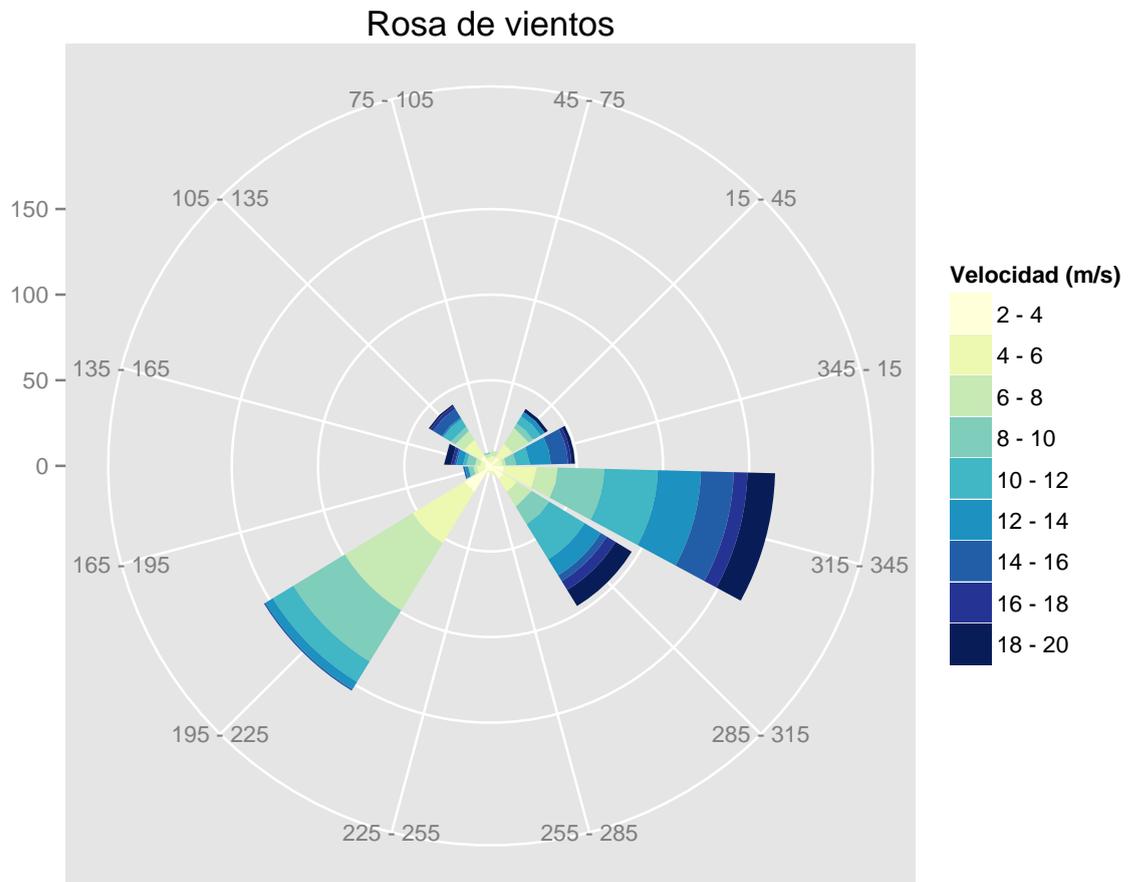


FIGURA 5.1: Rosa de vientos correspondiente el mes de noviembre de 2007, la velocidad está medida en m/s y la dirección en grados.

de estas estaciones se midió la dirección y velocidad del viento a 10 y 60 metros durante el mes de noviembre de 2007. Los datos con la información registrada a 10m en una de las estaciones podemos verla en figura 5.1, la cual se encuentra registrada en grados. En nuestro caso utilizaremos toda la teoría y metodologías que desarrollamos en el capítulo anterior para modelar una serie de tiempo de las direcciones de viento registradas cada hora a una altura de 10 metros durante el mes de noviembre de 2007. Esta serie de tiempo se presenta en la figura 5.2.

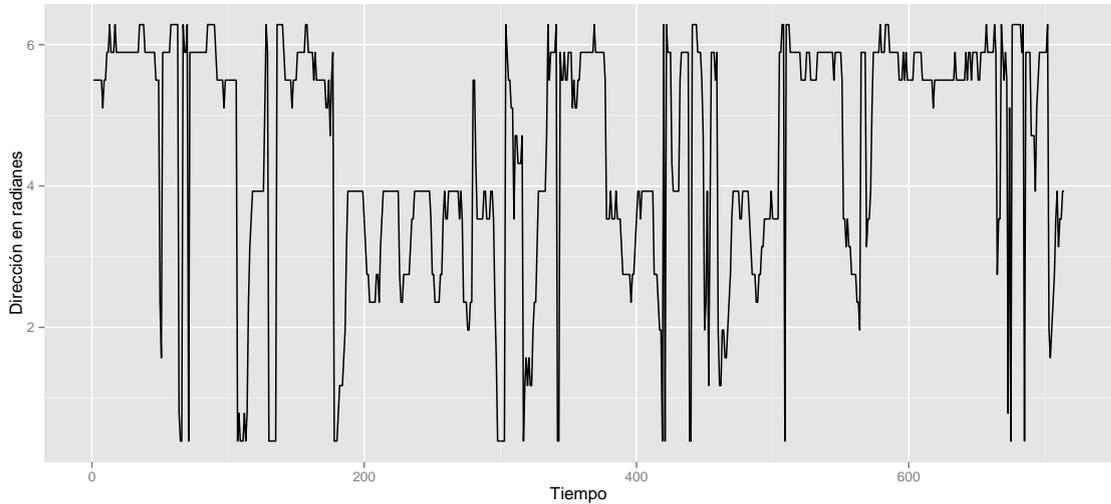


FIGURA 5.2: Serie de tiempo de direcciones medidas en radianes correspondientes al mes de noviembre de 2007.

En periodos largos de tiempo se aprecia una ausencia de estacionariedad, por lo que aplicaremos el operador diferencia tomando en cuenta que si

$$\begin{aligned}x_1 &= y_1 \pmod{2\pi}, \\x_2 &= y_2 \pmod{2\pi},\end{aligned}$$

tenemos

$$(x_1 - x_2) = (y_1 - y_2) \pmod{2\pi},$$

donde tal como antes, x representa un valor definido en $(0, 2\pi)$ y y se encuentra definido en \mathbb{R} , por lo que el operador diferencia también está bien definido.

Para poder verificar la falta de estacionariedad analizaremos las funciones de autocorrelación circulares¹. Para el análisis básico de datos circulares el software R

¹De acuerdo a Jammalamadaka and SenGupta (2001), si $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$ es una muestra aleatoria de datos direccionales, consideramos el coeficiente de correlación muestral

$$r_{c,n} = \frac{\sum_{i=1}^n \text{sen}(\alpha_i - \bar{\alpha}) \text{sen}(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^n \text{sen}^2(\alpha_i - \bar{\alpha}) \text{sen}^2(\beta_i - \bar{\beta})}},$$

donde $\bar{\alpha}$ y $\bar{\beta}$ son las direcciones medias muestrales.

cuenta con los paquetes `CircStats`, Lund and Agostinelli (2012) y `circular`, Agostinelli and Lund (2013); los cuáles empleamos en esta tesis. De acuerdo a la figura 5.3 podemos observar que mientras la serie original tiene una correlación persistente, en la serie de una diferencia la correlación decae rápidamente. Lo anterior nos indica que será conveniente modelar la diferencia de las observaciones; para ello tenemos que considerar la representación como modelo lineal dinámico de un modelo ARIMA.

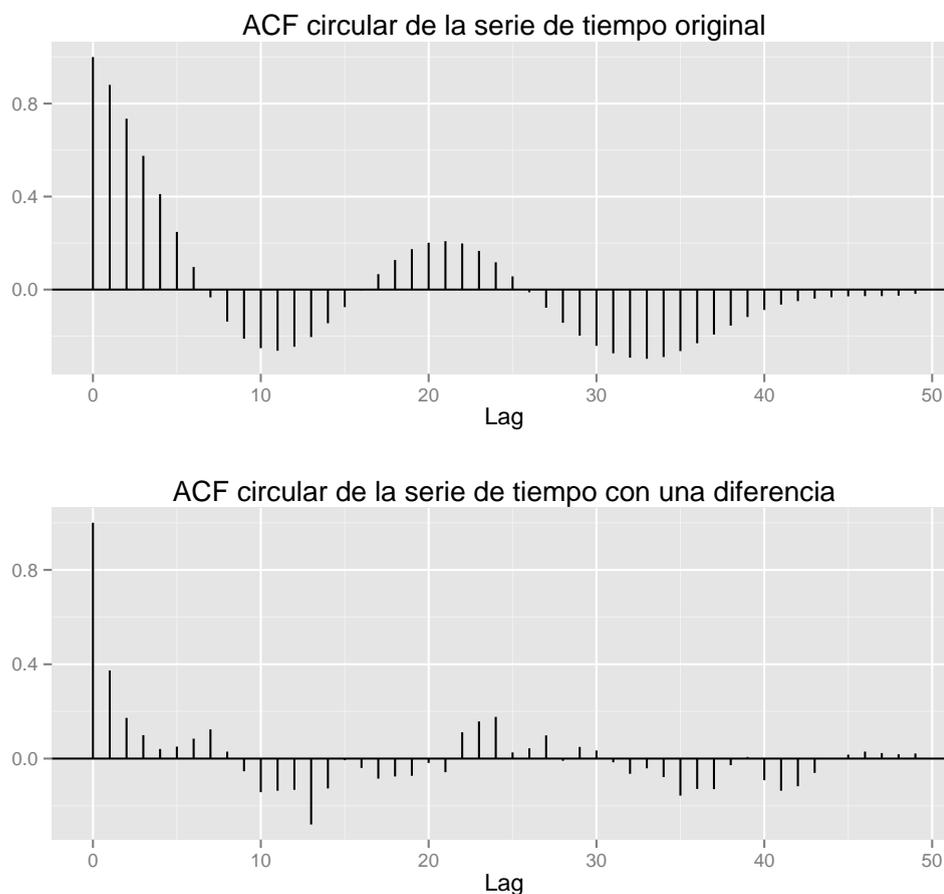


FIGURA 5.3: Funciones de autocorrelación circulares de la serie de tiempo original y la serie con una diferencia.

Resultados

Para generar los resultados usamos las mismas funciones del ejemplo simulado del capítulo anterior, incluso los hiperparámetros y valores iniciales para la simulación son los mismos. La única diferencia es que los datos analizados son los de la

serie de tiempo después de haber aplicado el operador diferencia una vez. La representación de este modelo se encuentra en las ecuaciones (3.7) y (3.8). El código para los datos reales es esencialmente el mismo salvo por la variantes que ya se mencionaron, además de que tuvimos que adelgazar más la cadena con respecto al ejemplo simulado. Para estos datos utilizamos 5000 iteraciones con un periodo de calentamiento de 1000.

Las densidades finales suavizadas de los parámetros estimados del modelo se pueden ver en la figura 5.4. En la figura 5.5 tenemos las trayectorias para cada

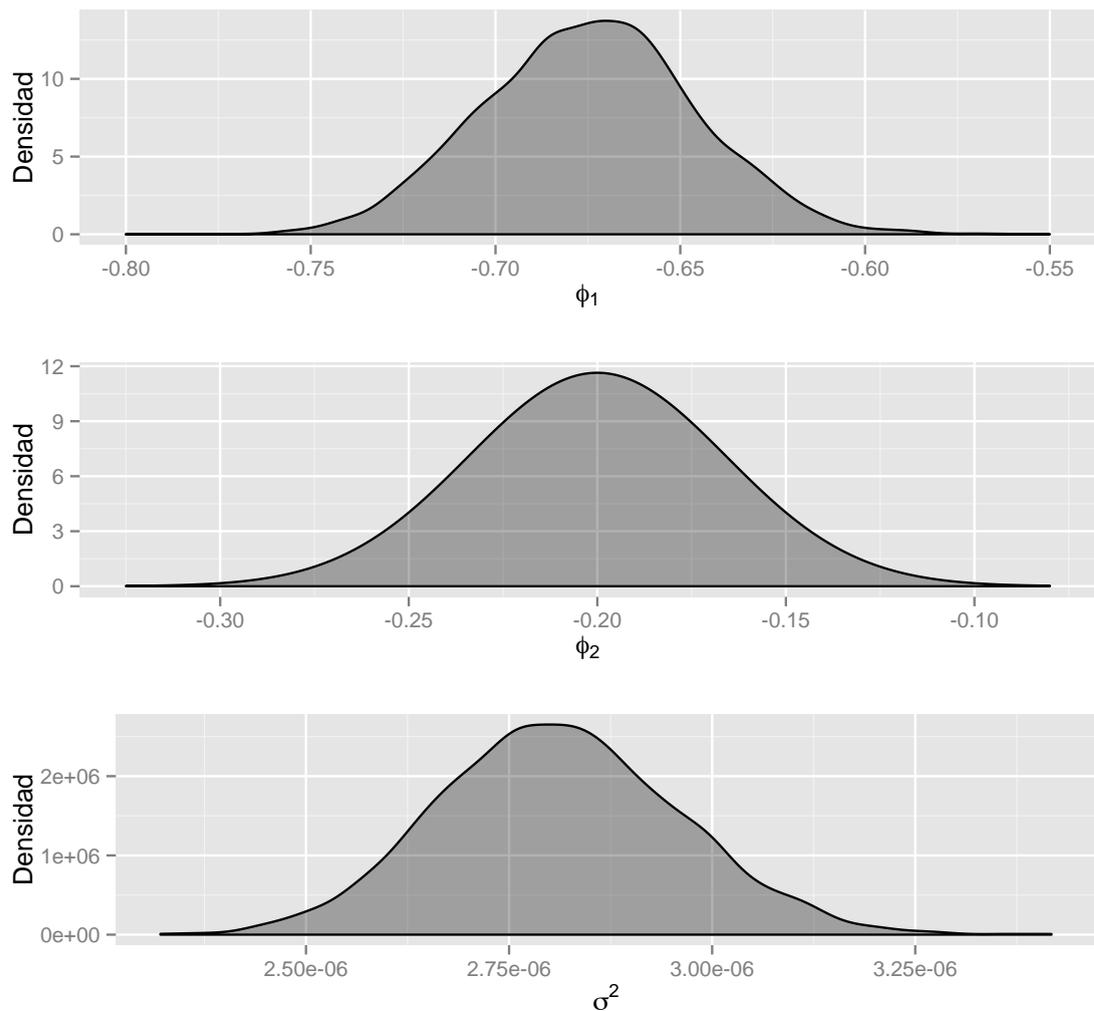


FIGURA 5.4: Densidades finales suavizadas de los parámetros ϕ_1 , ϕ_2 y σ^2 .

uno de los parámetros del modelo, las cuales nos permiten concluir empíricamente que tenemos convergencia. Del mismo modo, también tenemos evidencia empírica

de convergencia por medio de los promedios ergódicos y las funciones de autocorrelación que podemos ver en la figura 5.6.

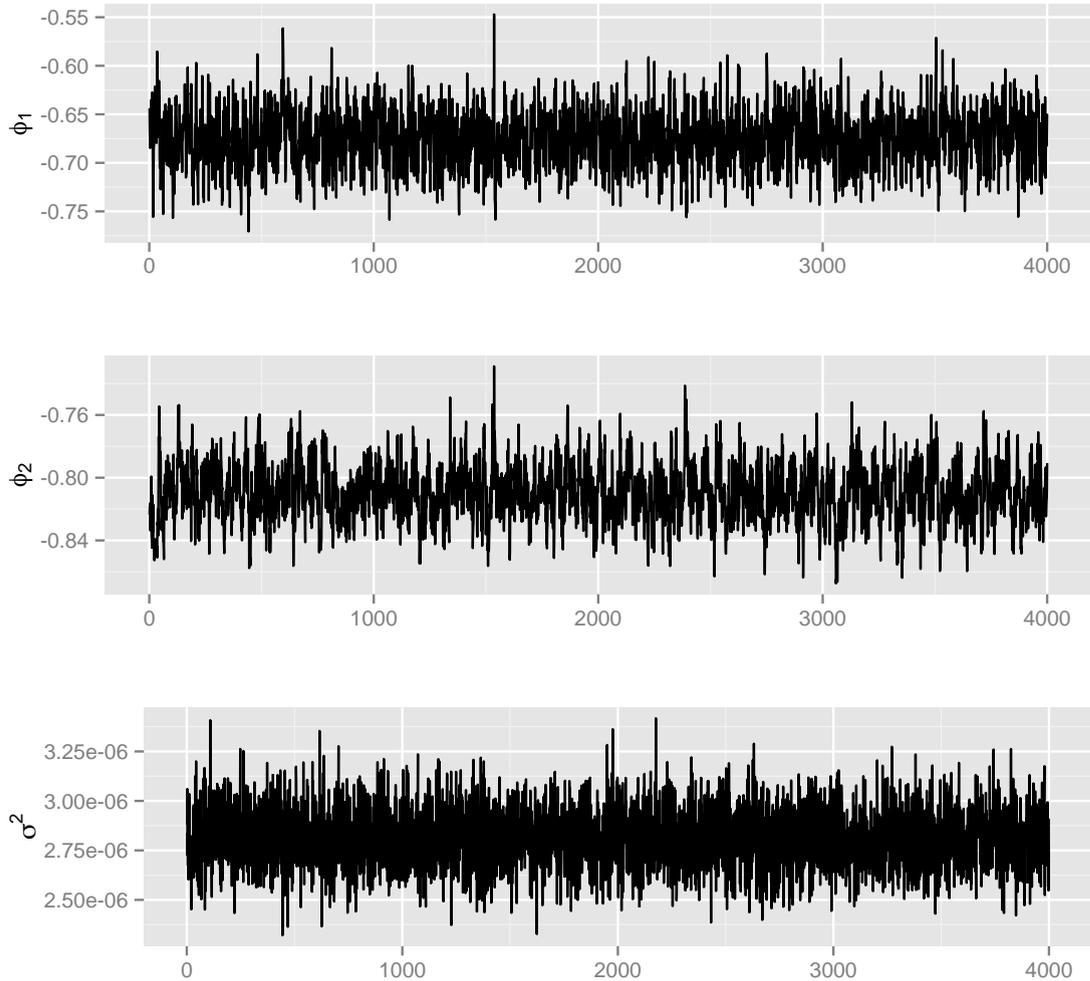


FIGURA 5.5: Trayectorias del algoritmo MCMC de los parámetros ϕ_1 , ϕ_2 y σ^2 .

Un resumen de las estimaciones del proceso se puede ver en el cuadro 5.1, las cuales sugieren que con retraso de dos horas la dependencia es considerable.

Parámetro	Media	Intervalo del 0.90 de probabilidad
ϕ_1	-0.67	$[-0.72, -0.63]$
ϕ_2	-0.81	$[-0.84, -0.78]$
σ^2	$2.82e - 06$	$[2.57e - 06, 3.1e - 06]$

CUADRO 5.1: Resumen de las estimaciones para los datos de la CNLV.

Una de las ventajas de considerar este modelo es la relativa facilidad con la que se puede implementar el pronóstico de las observaciones. En la figura 5.7

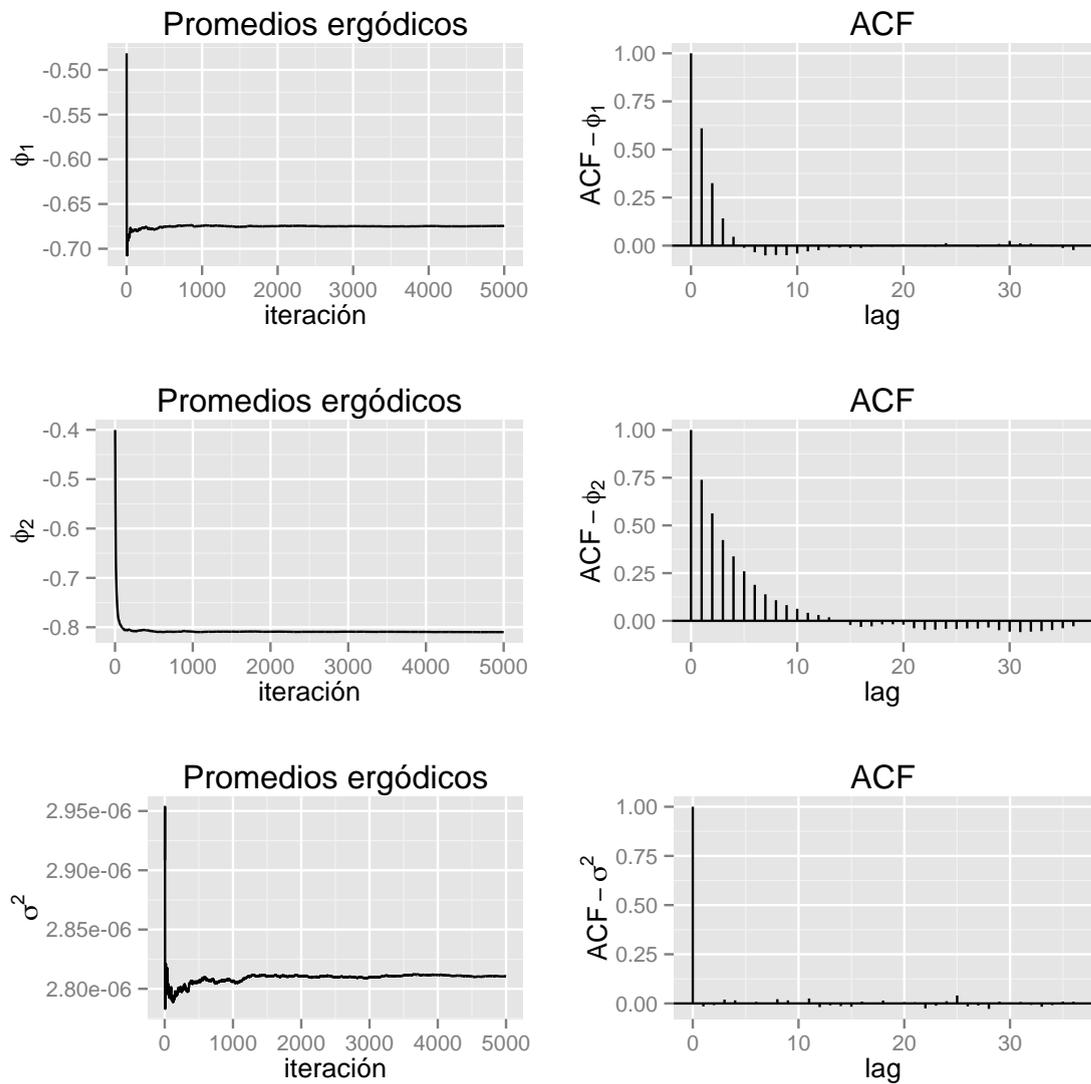


FIGURA 5.6: Promedios ergódicos y funciones de autocorrelación correspondientes a los parámetros ϕ_1 , ϕ_2 y σ^2 .

podemos observar el pronóstico a 24 horas correspondientes a los mismos datos que hemos analizado. Incluimos una muestra, en color azul; además de la media del pronóstico, en azul marino; así como la banda de predicción para la media.

El siguiente paso es la evaluación del pronóstico tal como procedimos en el ejemplo simulado. Es importante notar que, a diferencia del caso simulado, la variabilidad del vector k es mayor. Por lo tanto es importante considerar la moda de los coeficientes de envoltura, de otro modo el pronóstico es deficiente. En el ejemplo simulado la estabilización de los coeficientes nos permitía incluso considerar el último vector k de la muestra simulada con resultados aceptables.

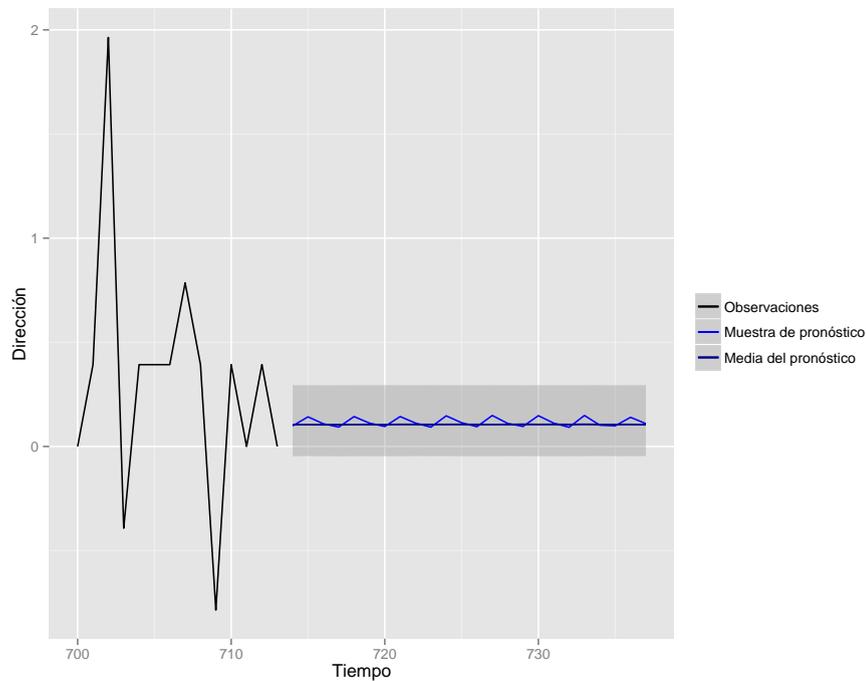


FIGURA 5.7: Pronóstico de las direcciones a 24 horas.

Es de destacar lo acertado del pronóstico; mientras que el valor la última observación es de 0, la media circular de la muestra de pronóstico es 0.11. Lo eficiente del pronóstico podemos corroborarlo por medio de la distribución de pronóstico que podemos ver en la figura 5.8. En la misma figura podemos ver la poca variabilidad del pronóstico, la cual se puede explicar por la varianza chica asociada al proceso. En la figura 5.9 podemos observar la distribución de pronóstico que confirma la concentración de la densidad que ya se mencionó. La flecha representa la media circular y el círculo azul el valor que pronosticamos. En ambos casos, el valor que omitimos toma un valor elevado en cada una de las densidades, lo que sugiere que el modelo es adecuado para pronosticar direcciones de viento. Como parte de las observaciones que podemos hacer a partir del modelo, al menos en en la temporada que corresponde al mes de noviembre, podemos concluir que los cambios de dirección no son drásticos y existe poca variabilidad.

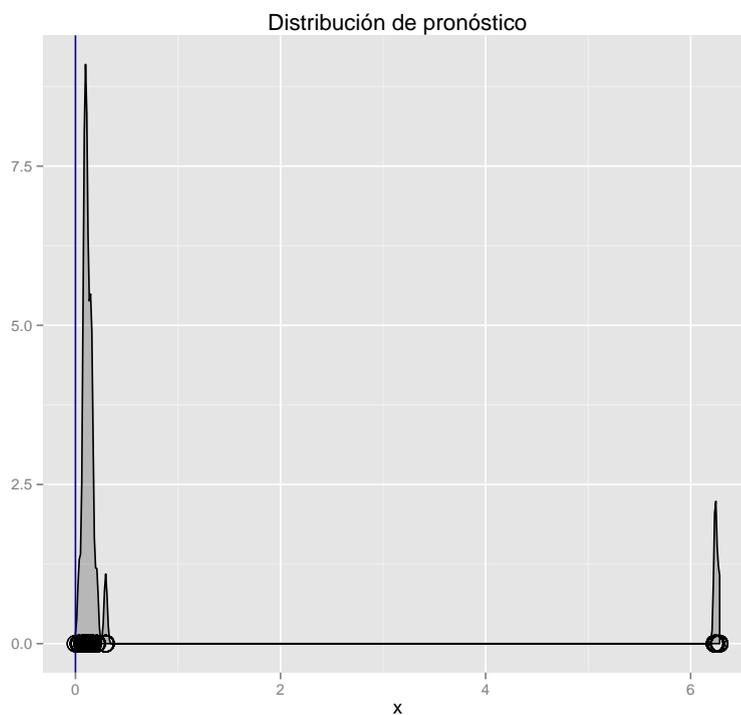


FIGURA 5.8: Evaluación de valor omitido en la distribución de pronóstico.

Distribución de pronóstico circular

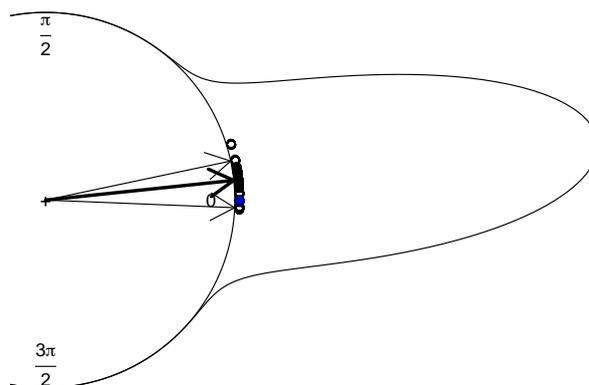


FIGURA 5.9: Evaluación de valor omitido en la distribución de pronóstico circular.

Capítulo 6

Conclusiones

El proceso de envolver distribuciones y procesos alrededor del círculo es un método natural para construir modelos con la finalidad de analizar datos circulares. Hemos podido ver que mediante la incorporación de los coeficientes de envoltura como variables latentes dentro de un modelo lineal dinámico tenemos un método confiable para poder hacer inferencia. Con el planteamiento propuesto eludimos el problema de la estimación de distribuciones complejas, sobre todo en el caso de dimensiones mayor a uno.

Tanto en el ejemplo simulado como con las observaciones reales hemos podido observar una alta sensibilidad del modelo respecto a la probabilidad de transición en la caminata aleatoria. Del mismo modo, en ambos casos fuimos capaces de generar un pronóstico razonable tomando en cuenta la variabilidad que inducen los coeficientes de envoltura. La posibilidad de incluir un pronóstico en el modelo es una de las principales virtudes de éste, ya que no existe un trabajo previo en el que este tipo de análisis se realizara para modelos autorregresivos.

Otra de las ventajas del modelo es que prácticamente no existe ninguna transformación o supuesto que no admita una interpretación natural en la estimación de los parámetros del modelo autorregresivo, dificultad que sí presentan algunas de las transformaciones que se proponen en los primeros intentos para modelar datos direccionales. En todo, caso el único inconveniente que se presenta es la interpretación del operador diferencia aplicado a datos de viento, el cual es inherente a cualquier modelo ARIMA. A pesar de esto, a favor del modelo podemos decir que la principal finalidad en nuestro caso es tener un pronóstico adecuado, dejando

en segundo término una interpretación natural de los supuestos y el modelo para obtenerlo.

El ejemplo simulado sustenta la precisión del procedimiento, mientras que su aplicación en el estudio de datos de viento nos ha dado cierta información de interés meteorológico. Con ayuda del modelo podemos inferir que existe una tendencia del viento de mantenerse en una dirección no necesariamente cercana a la media y que existe una marcada relación con las dos horas precedentes a cada una de las observaciones.

Respecto a la implementación, podemos decir de manera general que utilizamos distribuciones iniciales poco informativas para los parámetros autorregresivos pero de ser necesario podríamos incluir algunas que nos proporcionen más información. Sobre todo sería importante incluir más información en caso de que existiera mayor variabilidad en el componente direccional. En el caso de los coeficientes de envoltura, el modelo es sensible a los cambios en las probabilidades de transición, al grado que un cambio mínimo en éstas impide la convergencia. Un punto importante y que requiere de cuidado es la estimación del vector de coeficientes de envoltura. En el ejemplo simulado tenemos que estos se estabilizan relativamente rápido, mientras que en las observaciones de viento esta estabilización es más lenta. Para poder resolver esto tuvimos que considerar la moda de cada uno de ellos, con esto fuimos capaces de generar un pronóstico acertado. El modelo también podría admitir un cambio en la función de transición en caso de requerir un algoritmo más eficiente, pero para nuestra información no fue necesario a pesar de que requerimos de un número alto de simulaciones y un adelgazamiento considerable. Una de las ventajas de los métodos MCMC en este contexto es la flexibilidad en cuanto a la aplicación en estructuras complejas y de dimensión mayor.

Si bien los resultados en cuanto al pronóstico parecen razonables, por la importancia de los datos que se están modelando se debe proceder con cautela. Tal como vimos en el ejemplo simulado, la estimación puntual del pronóstico no necesariamente es adecuada. En este caso sería pertinente un intervalo de pronóstico, acorde a la variabilidad del modelo. El planteamiento de un intervalo no es inmediato por la complejidad de los datos de viento; la falta de un punto de referencia único impide una respuesta inmediata.

En cuanto a la investigación futura, los resultados son lo suficientemente alentadores para pensar que una manera adecuada para analizar simultáneamente los

componentes de velocidad y dirección es mediante los modelos lineales dinámicos. Es decir, podríamos considerar la representación como MLD de los llamados modelos VARMA¹, en el cual incluiríamos los coeficientes de envoltura como variables latentes. Además una extensión a un modelo espacio temporal que incluya varias estaciones de medición no sólo es posible sino adecuado por la flexibilidad de esta clase de modelos.

¹Por sus siglas en inglés, *vector autoregressive moving average*.

Apéndice A

Código en R

```
#####  
## Ejemplo simulado, capitulo 5  
#####  
  
## Intencionalmente sin acentos ##  
  
set.seed(321)  
  
##### Necesito paquetes  
#####  
  
library(dlm) # manejo general de mld's  
library(circular) # analisis de datos circulares  
library(CircStats) # analaisis de datos circulares  
library(lattice) # habilita algunos graficos  
library(ggplot2)  
library(grid)  
library(gridExtra)  
  
##### Funciones auxiliares  
#####  
w <- function(y) # envuelve sin regresar coeficientes de envoltura  
{  
  x <- 0  
  x <- y %% (2*pi)  
  return(x)  
}  
  
w2 <- function(x) # envuelve y devuelve coeficientes  
{
```

```

y <- 0
k <- 0
y <- x %% (2*pi)
k <- x %/% (2*pi)
out <- matrix(c(y,k),nrow=length(y),ncol=2)
return(out)
}

# Desenvuelve los datos, necesita K
dw <- function(y,k)
{
x <- y + 2*k*pi
return(x)
}

# Calcula la moda de un vector
# En caso de repetirse valores se queda con la primera moda
#por orden de aparicion
# La uso para calcular el la moda de los coeficientes de envoltura

Mode <-function(x)
{
ux<-unique(x)
ux[which.max(tabulate(match(x,ux)))]
}

##### genero el proceso WAR(2)
#####

# parametros del modelo AR
n <- 350
desv<-1
phi1 <- .6
phi2 <- -0.2

x <- arima.sim(n=n,model=list(ar=c(phi1,phi2)),sd=desv)
a <- w2(x)
y <- a[,1] # datos envueltos
ks <- a[,2] # coeficientes originales
n<-length(y)

## proceso ar(2) y war(2)
q1<-qplot(seq_along(x),x, geom="line", main="a) Proceso AR(2)",
xlab="Tiempo",ylab="y")
q2<-qplot(seq_along(y),y, geom="line", main="b) Proceso WAR(2)",

```

```
xlab="Tiempo",ylab="x")
grid.arrange(q1,q2,ncol=1)

## Calcula el vector k

kas<-function(y,niter,k,phi1,phi2,sigma)
{
n<-length(y)
mk <- matrix(0,n,niter)

p1<-0.025 # para la caminata aleatoria de las k'as
p2<-0.95
p3<-0.025

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

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)
}

# Inicia iteracion M-H
for (t in 1:(niter-1))
{
#print(t)
mk[1,t+1] <- k[1]
mk[2,t+1] <- k[2]

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+1] <- k[i]

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 <- 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 para i
k <- mk[,t+1]

} #termina ciclo para MH (t)
return(k)
}
#####
### Gibbs sampler, modelo AR(2)
## la referencia para el codigo es el libro
##"Dynamic Linear Models with R" de Giovanni Petris p. 186-191

ar2gdpGibbs <- function(y, a.theta, b.theta, shape.theta, rate.theta,
dV = 1e-7, m0 = rep(0,2),

```

```

                                C0 = diag(x=c(rep(1,2))),
                                n.sample = 1, thin = 0, Theta,save.states = TRUE)
{
  mod <- dlmModARMA(ar = rep(0,2), sigma2 = 1)
  mod$m0 <- m0
  mod$C0 <- C0
  p <- 2 # dimension del espacio de estados
  r <- 1 # numero de varianzas desconocidas
  nobs <- NROW(y)
  if ( is.numeric(thin) && (thin <- as.integer(thin)) >= 0 )
  {
    every <- thin + 1
    mcmc <- n.sample * every
  }

  shape.theta <- a.theta^2 / b.theta
  rate.theta <- a.theta / b.theta

  shape.theta <- shape.theta + 0.5 * nobs

  if ( save.states )
    #Theta <- matrix(0, nobs + 1, p)
    theta<-Theta
  gibbsPhi <- matrix(0, nrow = n.sample, ncol = 2)
  gibbsVars <- matrix(0, nrow = n.sample, ncol = 1)
  AR2support <- function(u)
  {
    ## region de estacionariedad para el proceso AR(2)
    (sum(u) < 1) && (diff(u) < 1) && (abs(u[2]) < 1)
  }
  ARfullCond <- function(u)
  {
    ## log-densidad condicional completa para parametros del AR(2)
    mod$GG[1:2,1] <- u
    -dlmLL(y, mod) + sum(dnorm(u, sd = c(0.66,0.33) , log=TRUE))
  }
  it.save <- 0
  for (it in 1:mcmc)
  {
    ## genera parametros del AR(2)
    mod$GG[1:2,1] <- arms(mod$GG[1:2,1],
                          ARfullCond, AR2support, 1)
# por eficiencia usa ARMS (adaptive rejection Metropolis-Hastings)
    ## genera los estados - FFBS (Forward Filtering-backward...)
    modFilt <- dlmFilter(y, mod, simplify=TRUE)
    theta[] <- dlmBSample(modFilt)
  }
}

```

```

## genera matriz W
theta.center <- theta[-1,-2,drop=FALSE] -
  (theta[-(nobs + 1),,drop=FALSE] %*% t(mod$GG))[-2]
SStheta <- drop(sapply( 1 : 1, function(i)
  crossprod(theta.center[,i])))
diag(mod$W)[1:1] <-
  1 / rgamma(r, shape = shape.theta,
    rate = rate.theta + 0.5 * SStheta)
## save current iteration, if appropriate
if ( !(it %% every) )
{
  it.save <- it.save + 1
  #print(it.save)
  if ( save.states )
    #gibbsTheta[, ,it.save] <- theta
    Theta<-theta
  gibbsPhi[it.save,] <- mod$GG[(p-1):p,r]
  gibbsVars[it.save,] <- diag(mod$W)[1:r]
}
}
if ( save.states )
  return(list(phi = gibbsPhi, vars = gibbsVars, theta = Theta))
else
  return(list(phi = gibbsPhi, vars = gibbsVars))
}

#####
### Funcion principal
#####

## Usa kas() y ar2gdpGibbs()

war2<-function(y,niter)
{
n<-length(y)

phi1 <- 0.5
phi2 <- -0.5
sigma<-0.5
k<-rep(0,n)
MK<-NULL

p <- 2 # dimension del espacio de estados
r <- 1 # numero de varianzas desconocidas
nobs <- NROW(y)
Theta <- matrix(0, nobs + 1, p)
phi1.vec<-NULL

```

```
phi2.vec<-NULL
var.vec<-NULL

for (t in 1:niter)
{
print(t)

k.vec<-kas(y,50,k,phi1,phi2,sigma) # uso adelgazamiento de 50 para k
  MK<-cbind(MK,k.vec)
k<-k.vec

ar2gibbsOut <- ar2gdpGibbs(dw(y,k), a=1,
b=1000, n=1, thin = 9,Theta=Theta)
## uso adelgazamiento de 10 para el mld
  phi1.vec<-c(phi1.vec,ar2gibbsOut$phi[,1])
  phi2.vec<-c(phi2.vec,ar2gibbsOut$phi[,2])
  var.vec<-c(var.vec,ar2gibbsOut$vars[,1])
  Theta<-ar2gibbsOut$theta

  phi1<-ar2gibbsOut$phi[,1]
  phi2<-ar2gibbsOut$phi[,2]
  sigma<-sqrt(ar2gibbsOut$vars[,1])

} # termina for t

  return(list(phi1 = phi1.vec, phi2 = phi2.vec, var = var.vec,
mk=MK,theta=Theta))
} #termina la funcion war2

prueba<-war2(y,10000) # genero 10'000 muestras

prueba$mk[,10000] ## ultimo vector k
prueba$mk[,10000]-ks ## comparo el ultimo vector k con el verdadero

burn<-5000 # calentamiento

## Estimadores puntuales
mcmcMeans(prueba$phi1[-(1:burn)])
mcmcMeans(prueba$phi2[-(1:burn)])
mcmcMeans(prueba$var[-(1:burn)])

## Intervalos de probabilidad
quantile(prueba$phi1[-(1:burn)] , probs=c(.05,.95))
quantile(prueba$phi2[-(1:burn)] , probs=c(.05,.95))
quantile(prueba$var[-(1:burn)] , probs=c(.05,.95))
```

```
## Trayectorias
q1<-qplot(seq_along(prueba$phi1[-(1:burn)]),prueba$phi1[-(1:burn)],
geom="line", xlab="", ylab=expression(phi[1]))
q2<-qplot(seq_along(prueba$phi2[-(1:burn)]),prueba$phi2[-(1:burn)],
geom="line", xlab="", ylab=expression(phi[2]))
q3<-qplot(seq_along(prueba$var[-(1:burn)]),prueba$var[-(1:burn)],
geom="line", xlab="", ylab=expression(sigma^2))
grid.arrange(q1,q2,q3,ncol=1)

## Promedios ergodicos
q1<-qplot(seq_along(ergMean(prueba$phi1)),ergMean(prueba$phi1),
geom="line", xlab="iteracion",ylab = expression(phi[1]),
main="Promedios ergodicos")

bacf1 <- acf(prueba$phi1, plot = FALSE)
bacfdf1 <- with(bacf1, data.frame(lag, acf))

q2 <- ggplot(data = bacfdf1, mapping = aes(x = lag, y = acf)) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = lag, yend = 0))+
  labs(title="ACF")+ylab(expression(ACF-phi[1]))

q3<-qplot(seq_along(ergMean(prueba$phi2)),ergMean(prueba$phi2),
geom="line", xlab="iteracion",ylab = expression(phi[2]),
main="Promedios ergodicos")

bacf2 <- acf(prueba$phi2, plot = FALSE)
bacfdf2 <- with(bacf2, data.frame(lag, acf))

q4 <- ggplot(data = bacfdf2, mapping = aes(x = lag, y = acf)) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = lag, yend = 0))+
  labs(title="ACF")+ylab(expression(ACF-phi[2]))

q5<-qplot(seq_along(ergMean(prueba$var)),ergMean(prueba$var),
geom="line", xlab="iteracion",ylab = expression(sigma^2),
main="Promedios ergodicos")

bacf <- acf(prueba$var, plot = FALSE)
bacfdf <- with(bacf ,data.frame(lag, acf))

q6 <- ggplot(data = bacfdf, mapping = aes(x = lag, y = acf)) +
  geom_hline(aes(yintercept = 0)) +
  geom_segment(mapping = aes(xend = lag, yend = 0))+
  labs(title="ACF")+ylab(expression(ACF-sigma^2))
```

```

grid.arrange(q1,q2,q3,q4,q5,q6,ncol=2)

## Densidades

x1=seq(prueba$phi1[-(1:burn)])
phi1=prueba$phi1[-(1:burn)]
data1<-data.frame(x1,phi1)
q1<-ggplot(data1, aes(x=phi1)) + geom_density(alpha=.2, fill="black")
+xlabs(expression(phi[1]))+
ylab("Densidad")+xlim(.3, .8)

x2=seq(prueba$phi2[-(1:burn)])
phi1=prueba$phi2[-(1:burn)]
data2<-data.frame(x2,phi2)
q2<-ggplot(data2, aes(x=phi2)) + geom_density(alpha=.2, fill="black")
+xlabs(expression(phi[2]))+
ylab("Densidad")+xlim(-.33,-.09)

x3=seq(prueba$var[-(1:burn)])
var=prueba$var[-(1:burn)]
data3<-data.frame(x3,var)
q3<-ggplot(data3, aes(x=var)) + geom_density(alpha=.2, fill="black")
+xlabs(expression(sigma^2))+
ylab("Densidad")+xlim(.8, 1.5)

grid.arrange(q1,q2,q3,ncol=1)

#####
### Pronostico
#####

## Para el pronostico uso el vector con
## modas de los coeficientes de envoltura
## generaro 1000 muestras de 24 pasos adelante para cada una (24x1000)

mode<-NULL
for (i in 1:length(y))
{
mod<-Mode(prueba$mk[i,])
mode<-c(mode,mod)
}

mode-ks # comparo con el vector real
        # en este caso coincide con la ultima muestra simulada k

X<-ts(dw(y,mode)) # desenvuelvo los datos con el vector de modas

```

```

X.dlm<-dlmModARMA(ar = c(.57,-.21), sigma2=1.01)
# especifico modelo
X.dlm$m0<-prueba$theta[351,]
# valor esperado del vector de estados anterior
X.dlm$C0<-matrix(c(prueba$var[10000],0,0,0),2,2)
# matriz de varianza del vector de estados anterior

plot(X, xlab = "", ylab = "Serie AR(2)", type = "l",
      col = "darkgray")

### Filtro
XFilt <- dlmFilter(X, X.dlm)
lines(dropFirst(XFilt$m[, 1]), lty = "dotdash")
### Suavizamiento
XSmooth <- dlmSmooth(XFilt)
lines(dropFirst(XSmooth$s[,1]), lty = "longdash")
legend("bottomright", col = c("darkgrey", rep("black", 2)),
      lty = c("solid", "dotdash", "longdash"),
      pch = c(1, NA, NA), bty = "n",
      legend = c("data", "filtered level", "smoothed level"))

#####
### Grafico de pronostico

### Muestra de 1000, 24 pasos adelante cada una
XFore <- dlmForecast(XFilt, nAhead =24, sampleNew =1000)

## Pronostico sin envolver

plot(window(X, start = 325), type = "o",
      xlim = c(325,375), ylim = c(-3.5, 5),
      xlab = "tiempo", ylab = "y")
names(XFore)
attach(XFore)
invisible(lapply(newObs, function(x)
  lines(x, col = "darkgray",
        type = "o", pch = 4)))
#lines(f, type = "o", lwd = 2, pch = 16)
#abline(v = mean(c(time(f)[1], time(x)[length(x)])))
detach()

### Pronostico envuelto
#XFore <- dlmForecast(XFilt, nAhead =24, sampleNew = 1)

```

```
plot(window(w(X), start = 325), type = "o",
      xlim = c(325,375), ylim = c(-0.5, 6.5),
      xlab = "tiempo", ylab = "Direccion",main="")
names(XFore)
attach(XFore)
invisible(lapply(newObs, function(x)
  lines(w(x), col = "darkgray",
        type = "o", pch = 4)))
#lines(f, type = "o", lwd = 2, pch = 16)
#abline(v = mean(c(time(f)[1], time(x)[length(x)])))
detach()

## Grafico muestra n evuelta vs media

sampleNew<-1000 # numero de muestras

NewObs<-unlist(XFore$newObs)
# muestra de la densidad de pronostico, 24 pasos

Matrix<-NULL

for (i in 1:24)
{
vector<-NewObs[((i-1)*sampleNew+1):(i*sampleNew)]
Matrix<-cbind(Matrix,vector)
}

mean.vec<-colMeans(Matrix) ## media de las predicciones a cada paso

upper.q<-NULL ## cuantil .025

for (i in 1:24)
{
upper.q<-c(upper.q,quantile(Matrix[,i],.975) )
}

lower.q<-NULL ## cuantile .975
for (i in 1:24)
{
lower.q<-c(lower.q,quantile(Matrix[,i],.025) )
}
```

```

## banda con media y muestra sin envolver
## (coincide con war2 es su representacion [-pi,pi])

df1 = data.frame(Tiempo = seq(325,350,length=26),
Direccion = X[325:350], isin = "Observaciones")
df2 = data.frame(Tiempo = seq(351,374,length=24),
Direccion= Matrix[100,1:24] , isin = "Pronostico")
df3 = data.frame(Tiempo = seq(351,374,length=24),
Direccion = upper.q , isin = "Intervalo_superior")
df4 = data.frame(Tiempo = seq(351,374,length=24),
Direccion = lower.q, isin = "Intervalo_inferior")
df5 = data.frame(Tiempo = seq(351,374,length=24),
Direccion = mean.vec, isin = "Media")
df = rbind(df1, df2, df3, df4,df5)

ggplot(df1, aes(x = Tiempo, y =
Direccion)) + geom_line(colour='black') +

    geom_line(aes(x=Tiempo, y=Direccion,color=isin),
              colour='darkblue', data=df5)+

    geom_smooth(aes(x=Tiempo, y=Direccion, ymax=upper.q, ymin=lower.q),
               colour='blue', data=df2, stat='identity')

## Evaluamos el pronostico
#####

x<-w(Matrix[,1])
x<-data.frame(seq(x),x)

g <- ggplot(x,aes(x=x))
g <- g + geom_density(alpha=.2, fill="black")
g <- g + geom_vline(xintercept =y[350],colour="darkblue")
g <- g + geom_point(aes(x=x,y=0),shape=1,size=5) +
labs(title="Distribucion de pronostico")+ylab(NULL)
g

# Distribucion de pronostico en el circulo

## asigno la clase de circular a los pronosticos
NewObs <- circular(Matrix[,1], type="angles", units="radians")

## intervalo del 90% de probabilidad
quantile.circular(NewObs,c(0.05,.95))
##      5%      95%

```

```
## 5.365111 2.333436

# Distribucion de pronostico en el circulo
fore <- mean(circular(Matrix[,1]))
# estimacion puntual del pronostico, media circular
res25x <- density(circular(Matrix[,1]), bw=27)
res<-plot(res25x, points.plot=TRUE, xlim=c(-.8,1.0),ylim=c(-1.1,1.1),
main="Distribucion de pronostico circular", ylab="", xlab="")
#points(data.2, plot.info=res, col="black",type="o")
points(circular(y[350]), plot.info=res, col="blue",type="o")
#points(0, plot.info=res, col="darkgray",type="o")
arrows.circular(c(5.365111, 2.333436  )) ## extremos de intervalos
arrows.circular(fore, lwd=3) ## media circular, ie pronostico
```

Bibliografía

- Agostinelli, C. and Lund, U. (2013). *R package `circular`: Circular Statistics (version 0.4-7)*.
URL: <https://r-forge.r-project.org/projects/circular/>
- Box, G., Jenkins, G. and Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*, 4th edn, Wiley, New York.
- Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, London: Springer.
- Chib, S. and Greenberg, E. (1994). Bayes inference for regression models with ARMA(p, q) errors, *Journal of Econometrics* **64**: 183–206.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm., *The American Statistician* **49**(4): 327–335.
- Coles, S. (1998). Inference for circular distributions and processes., *Statistics and Computing*. **8**: 105–113.
- Craig, P. (1988). *Time Series Analysis for Directional Data*, PhD thesis, Trinity College Dublin.
- Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods*, 2nd edn, Oxford University Press, Oxford.
- Ferrari, C. (2009). *The Wrapping Approach for circular data Bayesian modeling.*, PhD thesis, Università di Bologna.
- Fisher, N. I. (1993). *The Analysis of Circular data*, Cambridge University Press.
- Fisher, N. I. and Lee, A. J. (1994). Time series analysis of circular data., *Journal of the Royal Statistical Society* **B**(56): 327–339.

- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion), *Bayesian Statistics 4* pp. 147–167.
- González, U. (2012). *Inferencia bayesiana para datos circulares.*, Tesis de licenciatura en matemáticas, Facultad de Ciencias, UNAM.
- Gutiérrez-Peña, E. (1997). Métodos computacionales en la inferencia bayesiana, *Serie Monografías*, Vol. 6-15, IIMAS, UNAM.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific.
- Lund, U. and Agostinelli, C. (2012). *R package CircStats: Circular Statistics (version 0.2-4)*.
URL: <http://CRAN.R-project.org/package=CircStats>
- Mardia, K. V. (1972). *Statistics of Directional Data*, Cambridge University Press.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley.
- Migon, H., Gamerman, D., Lopez, H. and Ferreira, M. (2005). Bayesian dynamic models, in D. Day and C. Rao (eds), *Handbook of Statistics*, Vol. 25, Elsevier B.V, chapter 19, pp. 557–592.
- Morales, T., Tejeda, A. 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**(3): 181–187.
- Petris, G., Petrone, S. and Campagnoli, P. (2009). *Dynamic Linear Models with R, useR!*, Springer-Verlag, New York.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <http://www.R-project.org/>
- Robert, C. P. (2001). *The Bayesian Choice*, Springer Texts in Statistics, 2nd edn, Springer.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer Texts in Statistics, 2nd edn, Springer.

-
- 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* **B**(55): 3–23.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer.