



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

ANÁLISIS DE DATOS CIRCULARES  
CON ESTRUCTURA TEMPORAL

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A

CITLALLI ELIZABETH DORANTES BOLAÑOS

TUTORA DE TESIS:

MAT. MARGARITA ELVIRA CHÁVEZ CANO



Facultad de Ciencias  
UNAM

México, D.F., junio de 2009



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

1. Datos del alumno  
Dorantes  
Bolaños  
Citlalli Elizabeth  
56 16 16 04  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
301033734

---

2. Datos del tutor  
Mat.  
Margarita  
Chávez  
Cano

---

3. Datos del sinodal 1  
Dra.  
Guillermina  
Eslava  
Gómez

---

4. Datos del sinodal 2  
Dra.  
Ruth Selene  
Fuentes  
García

---

5. Datos del sinodal 3  
Act.  
Jaime  
Vázquez  
Alamilla

---

6. Datos del sinodal 4  
Act.  
Francisco  
Sánchez  
Villareal

---

7. Datos del trabajo escrito  
Análisis de Datos Circulares con Estructura Temporal  
71 p.  
2009

# Dedicatorias y Agradecimientos

*A mis padres: a ustedes que me dieron la vida, les agradezco todo el amor que me han dado desde entonces; gracias por el apoyo y la protección incondicional que siempre me han brindado; gracias por los valores y principios que me han inculcado; por no dejarme nunca a la deriva; gracias por educarme como lo han hecho. Sin ustedes este trabajo no hubiera sido posible. Mamá, papá, les dedico este gran logro personal. Los amo.*

*A Ricardo, mi hermano: contigo he compartido risas, alegrías, tristezas, peleas, y 21 años de mi vida; te quiero mucho hermano y te agradezco de todo corazón todo tu apoyo; sé que cuento contigo incondicionalmente y tu sabes que cuentas conmigo siempre. Con mucho cariño para ti.*

*A mi abuelita Consuelo: gracias por tu amor y tu apoyo. Eres parte importante de mi vida y también de este logro.*

*A Renato: te doy las gracias por caminar a mi lado, por tu apoyo incondicional, tu comprensión, tu lealtad, tus palabras de aliento que nunca me faltaron; gracias por tus comentarios y sugerencias para esta tesis que te dedico con mucho amor.*

*A toda mi familia y amigos que creyeron en mi desde siempre: con mucho afecto para todos ustedes.*

*Asimismo, a todos aquéllos que contribuyeron de una u otra manera a la realización de ésta tesis: gracias.*

*A mi Alma Máter, la Universidad Nacional Autónoma de México, a quién pondré siempre muy en alto: agradezco la formación profesional que me entregó. De igual forma agradezco a mi querida Facultad de Ciencias, de quien me siento muy orgullosa de formar parte.*

*Con mucho aprecio para mi Directora de Tesis, Mat. Margarita E. Chávez Cano: mi más sincero agradecimiento por brindarme la oportunidad de ser su tesista, así como permitirme recurrir a su capacidad y experiencia, fundamentales para este trabajo. Asimismo, le agradezco por dedicar su tiempo a este proyecto; por mostrarme su apoyo, su confianza y su amistad en todo momento. Siempre le estaré agradecida.*

*A Germán Valle, con especial cariño: te agradezco toda tu ayuda desinteresada, en especial por tu apoyo con el magnífico LaTeX, tus sugerencias, tus consejos...*

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Introducción al Análisis de Series de Tiempo</b>	<b>4</b>
1.1. Conceptos Fundamentales	4
1.1.1. Procesos Estocásticos	5
1.1.2. Series de Tiempo Continuas y Discretas	5
1.1.3. Series de Tiempo Determinísticas y Estocásticas	5
1.2. Propósitos del Análisis de Series de Tiempo	5
1.3. Tipos de Variación	6
1.3.1. Descomposición de la Serie de Tiempo	8
1.4. Series de Tiempo Estacionarias	8
1.5. Transformaciones	9
1.6. Análisis de Series con Tendencia	10
1.7. Variación Estacional	11
1.8. Autocorrelación	12
1.8.1. Autocorrelación Parcial	14
1.8.2. Correlograma	15
1.9. Representación Autoregresiva de un Proceso Lineal	20
1.9.1. Procesos de Ruido Blanco	20
1.9.2. Representaciones para los Procesos de Series de Tiempo	20
1.9.3. Proceso Autoregresivo: Proceso $AR$	21
1.9.4. Proceso de Media Móvil: Proceso $MA$	21
1.9.5. Proceso $ARMA(p, q)$	22
1.9.6. Proceso Circular $CAR(p)$	23
<b>2. Introducción a la Estadística Circular</b>	<b>26</b>
2.1. Datos Circulares	27
2.2. Momentos Trigonométricos	29
2.3. Distribuciones Circulares	30
2.3.1. Distribución Uniforme o Isotrópica	31
2.3.2. Distribución Cardioide	32
2.3.3. Distribución Envuelta	33

---

2.3.4. Distribución Cauchy Envuelta . . . . .	34
2.3.5. Distribución Normal Envuelta . . . . .	35
2.4. Distribución von Mises . . . . .	35
<b>3. Análisis de Datos Circulares con Estructura Temporal</b>	<b>38</b>
3.1. Algunos Modelos para Series de Tiempo de Datos Circulares . . . . .	40
3.1.1. Procesos Gaussianos Proyectados . . . . .	40
3.1.2. Procesos “Envueltos” . . . . .	41
3.1.3. Procesos Derivados Utilizando Funciones de Enlace . . . . .	42
3.2. Selección del Modelo e Identificación . . . . .	43
<b>4. Aplicación</b>	<b>47</b>
<b>Conclusiones</b>	<b>56</b>
<b>Apéndice A.</b>	<b>58</b>
<b>Apéndice B.</b>	<b>59</b>
<b>Apéndice C.</b>	<b>61</b>
<b>Apéndice D.</b>	<b>63</b>
<b>Bibliografía</b>	<b>65</b>

# Introducción

Lo poco que he aprendido carece de valor, comparado con lo que ignoro y no desespere en aprender.  
René Descartes (1596-1650)

Desde los comienzos de la civilización han existido formas sencillas de hacer estadística. Hacia el año 3000 a.C. los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y de las mercancías vendidas o cambiadas mediante trueque. Por su parte, los egipcios analizaban los datos de la población y la renta del país en el siglo XXXI a.C.

El Imperio romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. El registro de nacimientos y defunciones comenzó en Inglaterra a principios del siglo XVI, y en 1662 apareció el primer estudio estadístico notable de población titulado “Observations on the London Bills Of Mortality” (Comentarios sobre las actas de defunción en Londres). Durante el siglo XX, la creación de instrumentos precisos para asuntos de salud pública y propósitos económicos y sociales (tasa de desempleo, econometría, etc.) necesitó de avances sustanciales en las prácticas estadísticas.

Hoy en día, los métodos estadísticos constituyen una herramienta de trabajo cada vez más importante en los campos profesionales más diversos. La estadística se ha convertido en una metodología efectiva para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dicha información.

Difícilmente puede ser considerada completa la formación de economistas, físicos, ingenieros, políticos, sociólogos, psicólogos, químicos, biólogos o cualquier otro grupo profesional, sin la asimilación de algunos conceptos básicos de metodología estadística que les permitan analizar e interpretar adecuadamente los datos sobre los que trabajan.

El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchas poblaciones se pueden modelar con gran exactitud, utilizando determinadas distribuciones probabilísticas; los resultados de éstas

se pueden utilizar para el análisis de datos. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

Es de nuestro conocimiento que para los estadísticos resulta rutinario trabajar y manipular distintos tipos de datos; sin embargo, existe una familia de datos un tanto diferente a los que se está acostumbrado. Es a finales de los años 70's cuando comenzaron a desarrollarse rápidamente métodos estadísticos para analizar este tipo de datos muy particular, conocidos como Datos Circulares: una gran clase de datos direccionales (datos medidos como ángulos o bidimensionales); actualmente se han convertido en datos de interés para los científicos en muchos ámbitos: biológicos (los movimientos de migración de los animales), meteorológicos (vientos), geológicos (direcciones de las articulaciones y defectos), médicos, etc. Muchos ejemplos de este tipo de datos, son los registrados por fenómenos que son periódicos en el tiempo, incluidos los ritmos diarios y estacionales.

El desarrollo de la estadística circular comprende cientos de años de investigación, desde el siglo X se encuentran registros de análisis con carácter circular. La demanda y falta de metodología para investigaciones de orientaciones originó la búsqueda de métodos adecuados para estos estudios y, hasta la fecha, se siguen haciendo aportes novedosos a los ya desarrollados.

A finales del primer milenio se encontraron las primera gráficas con características circulares. Funlkhouser (Fisher, 1993) comenta que en los siglos X y XI se realizaban estudios astrológicos con base en el movimiento de los planetas a través del zodiaco en diagramas oscilatorios.

En 1767, el reverendo John Mitchell analizó las separaciones angulares entre las estrellas y encontró que el número de parejas cercanas era tan grande para ser consistente; formuló la siguiente hipótesis: las direcciones de las estrellas se distribuyen uniformemente. Mitchell infirió que tales direcciones se debían a fuerzas gravitacionales.

Durante la guerra de Crimea, en 1858, la enfermera de la Armada Británica, Florence Nightingale, utilizó una variedad de métodos que le ayudaron a registrar y analizar de manera más sencilla los registros sanitarios. Uno de estos métodos matemáticos, fue una nueva gráfica a la que llamó **CoxComb**, la cual consistía de un círculo dividido en doce partes iguales, con diferentes colores que representaban los meses del año. Este diagrama es una primera referencia de lo que actualmente conocemos como *diagrama de rosa*.

Para el año de 1917, surgieron en el área de biogeografía, gráficas estándar de datos circulares conocidas hoy en día como diagramas de rosa, utilizadas para representar datos axiales.

En 1918, von Mises introdujo la **distribución probabilística normal circular**, base de la inferencia estadística paramétrica de datos circulares.

Fue en la década de los 50's cuando comenzaron a desarrollarse las técnicas pro-

babilísticas apropiadas para distribuciones de datos circulares. En 1956, el desarrollo de la metodología para analizar datos angulares se vuelve razonablemente acertado, con los primeros trabajos de Watson<sup>1</sup> y Williams<sup>2</sup> (Fisher, 1993), quienes no solo unificaron los problemas inferenciales que planteaba von Mises, sino que también aportaron novedosas ideas y resultados. Watson, por ejemplo, introdujo el análisis de varianza, pruebas paramétricas, pruebas no paramétricas, entre otras.

En 1972, Kantilal (“Kanti”) Mardia<sup>3</sup> realizó avances en pruebas tales como la correlación y la regresión y, a fechas más recientes, en series de análisis temporal de datos circulares, este último, tema central del presente trabajo, en el cual se pretende mostrar de una manera concisa cómo modelar y analizar los datos circulares, se exponen diferentes modelos así como sus principales características y resultados. Asimismo, al tratar con este tipo de datos, se logran observar algunas diferencias y similitudes con los datos lineales y sus modelos correspondientes<sup>4</sup>.

En el capítulo I se describen los lineamientos generales iniciales para estudiar una serie de tiempo, así como los pasos a seguir para facilitar su análisis. Se detallan las características principales de los distintos tipos de series y se explican los procesos AR, MA, ARMA y LARMA.

En el capítulo II se presenta una breve introducción a la Estadística Circular, con el fin de comprender mejor el manejo de los datos direccionales. Se describen las distribuciones circulares, exponiendo la función de probabilidad de cada una de ellas y sus momentos.

En el tercer capítulo se describen algunos modelos para analizar series de tiempo de datos circulares, se explica cómo seleccionar el modelo más adecuado para el análisis.

Por último, se expone una aplicación con una serie de tiempo con datos direccionales.

---

<sup>1</sup>Geoffrey Stuart Watson (1921 – 1998) fue un estadístico australiano; su interés principal fue la estadística aplicada. Entre sus aportaciones más importantes a la estadística se encuentra la prueba de Durbin-Watson.

<sup>2</sup>Evan James Williams (1917–), estadístico australiano fue profesor de estadística en la Universidad de Melbourne de 1964 a 1982.

<sup>3</sup>Estadístico especializado en estadística direccional, análisis multivariado, geoestadística y estadística de bioinformática.

<sup>4</sup>Eloisa Reyes, *op. cit.*, pp. 5-15.

# Capítulo 1

## Introducción al Análisis de Series de Tiempo

Una serie de tiempo es una colección de observaciones realizadas secuencialmente a través del tiempo, generalmente a intervalos iguales. Su aplicación se extiende a una gran variedad de campos como lo son: economía, física, mercadotecnia, demografía, medicina, astronomía, agricultura, ingeniería, procesos de control, ciencias sociales, etc. La naturaleza fundamental de una serie de tiempo es que sus observaciones dependen entre sí (*i.e.*, están correlacionadas), por lo que el orden de las observaciones es muy importante.

Existen varios objetivos para estudiar una serie de tiempo. Estos incluyen la comprensión y descripción de los datos, la predicción de los futuros valores, y un control óptimo de un sistema; más adelante se exponen con más detalle cada uno.

Los métodos para analizar las series de tiempo constituyen una importante área de la estadística, por lo que es recomendable tener conocimientos previos de estadística para un manejo más adecuado de las series de tiempo. Mucha de la teoría estadística se refiere a muestras aleatorias con observaciones independientes. La característica especial del análisis de series de tiempo es el hecho de que las observaciones sucesivas generalmente no son independientes y el análisis toma en cuenta el orden de las observaciones. Cuando las observaciones sucesivas son dependientes, los valores futuros pueden predecirse a través de observaciones anteriores.

### 1.1. Conceptos Fundamentales

Para comprender mejor aquello que engloba una *serie de tiempo*, es necesario introducir algunos conceptos fundamentales. Se comenzará por estudiar la relación entre un proceso estocástico y una serie de tiempo, en seguida se estudiarán los diferentes tipos de series de tiempo: continuas o discretas, que a su vez pueden ser determinísticas o estocásticas.

### 1.1.1. Procesos Estocásticos

Un *proceso estocástico* es una familia de variables aleatorias indexadas  $Z(\omega, t)$ , donde  $\omega$  pertenece a un espacio muestral y  $t$  pertenece a un conjunto de índices (generalmente el tiempo). Para una  $t$  fija,  $Z(\omega, t)$  es una variable aleatoria. Para una  $\omega$  dada,  $Z(\omega, t)$  (que estará en función de  $t$ ) es llamada *realización*. El conjunto que se compone de todas las realizaciones posibles es conocido como *el conjunto de procesos estocásticos*. Por lo tanto, una serie de tiempo es una realización de un proceso estocástico determinado.

### 1.1.2. Series de Tiempo Continuas y Discretas

Las series de tiempo, tales como las señales eléctricas y de voltaje, que pueden ser registradas de manera continua en el tiempo, son llamadas *series de tiempo continuas*. Más ejemplos de este tipo de series son los detectores de mentiras y los electrocardiogramas. Por otro lado, las series de tiempo, tales como el tipo de interés (en la banca) o el volumen de ventas en una empresa, que son registradas solo en un intervalo de tiempo específico, son llamadas *series de tiempo discretas*.

Las series de tiempo discretas pueden originarse de varias maneras. Dada una serie de tiempo continua, se pueden tomar valores a intervalos iguales de tiempo y así se obtiene una serie de tiempo discreta.

### 1.1.3. Series de Tiempo Determinísticas y Estocásticas

Si una serie de tiempo puede predecirse de manera exacta, se dice que la serie es *determinística*. Sin embargo, la mayoría de las series de tiempo son estocásticas, en donde los valores futuros son determinados, sólo en parte, por valores pasados, por lo que en este tipo de series las predicciones exactas son imposibles y se debe considerar que los valores futuros siguen una distribución de probabilidad, la cual es condicionada por valores pasados conocidos.

## 1.2. Propósitos del Análisis de Series de Tiempo

Como ya se mencionó anteriormente, los objetivos del análisis son los siguientes:

*Descripción:* cuando se inicia el análisis de una serie de tiempo, generalmente el primer paso a seguir es graficar las observaciones contra el tiempo, esta gráfica es conocido como **gráfica de tiempo**. Este paso es importante para poder obtener medidas descriptivas de las principales propiedades de la serie en cuestión. Una gráfica no solo mostrará la tendencia y la variación estacional, sino también revelará puntos (observaciones) llamados “outliers” que son aquéllos que no son consistentes con el resto de los datos. Otros rasgos

que se pueden encontrar en una gráfica de tiempo son los cambios repentinos o graduales en las series, así como la presencia de puntos influyentes, éstos son aquéllos en donde si la tendencia era creciente se vuelve decreciente o viceversa. Cuando existe un punto influyente se pueden usar dos modelos ajustados para cada una de las dos partes de la serie.

*Explicación:* cuando las observaciones involucran dos o más variables, es posible explicar el cambio de una variable por medio de lo sucedido con la otra. Por ejemplo, se puede tener el caso en el que sea de interés estudiar cómo el nivel del mar se ve afectada por la temperatura y la presión, o cómo las ventas se ven afectadas por los precios y las condiciones económicas.

*Predicción:* por lo general, una serie de tiempo es analizada con el fin de poder predecir los valores futuros de la serie; por lo que un buen análisis permitirá una buena previsión. Esta es una tarea muy importante en el pronóstico de ventas y en el análisis económico. Cabe mencionar que muchos autores utilizan, de manera indistinta, los términos “predicción” y “pronóstico”.

*Control:* muchas veces las series de tiempo son utilizadas y/o analizadas para mejorar el control sobre algunos sistemas físicos o económicos. Por ejemplo, si uno puede predecir que un proceso de fabricación no llegará al objetivo, entonces se pueden tomar medidas correctivas. Existe una gran variedad de procesos de control, por ejemplo, en el control estadístico, las observaciones se grafican en cuadros de control y el encargado actúa con base en el estudio de estos cuadros.

### 1.3. Tipos de Variación

A continuación, se describen los tipos de variación utilizados en los métodos tradicionales de análisis de series de tiempo<sup>1</sup>:

1. *Variación estacional o efecto estacional:* se define como las oscilaciones periódicas a una distancia constante (año, mes, semana). En caso de que la estacionalidad no sea de interés, la variación estacional puede ser removida de los datos.
2. *Variación cíclica o cambios cíclicos:* la diferencia entre efecto estacional y cambios cíclicos, es que la estacionalidad se repite en intervalos fijos (y constantes: un año, un mes, etc.); mientras que los cambios cíclicos tienen una duración más larga que varía de ciclo a ciclo.

---

<sup>1</sup>Chris Chatfield, *op. cit.*, pp. 11-13.

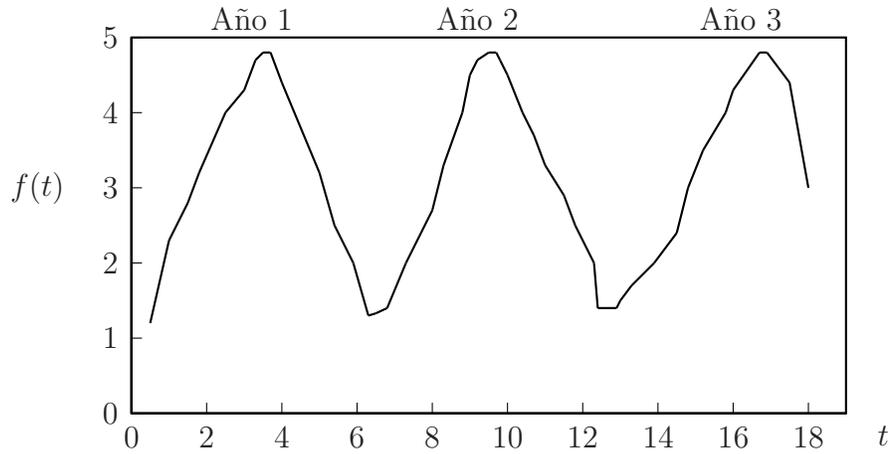


Figura 1.1: Gráfica de una serie de tiempo con efecto estacional.

3. *Tendencia*: es el comportamiento de los datos a largo plazo, que puede ser creciente o decreciente (Figura 1.2). Es muy importante tomar en cuenta el número de observaciones realizadas y con base en ellas definir lo que se entiende por periodo largo. En términos generales, la tendencia puede ser determinística (la tendencia es predecible) o estocástica (no hay forma de predecir su comportamiento).

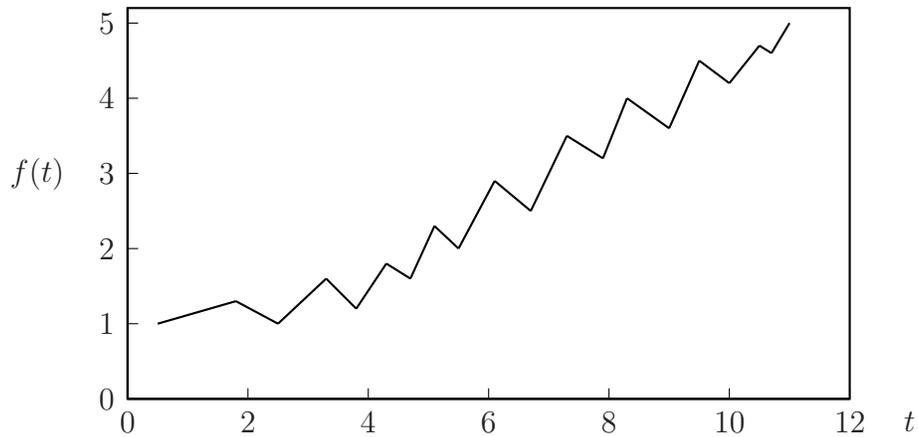


Figura 1.2: Gráfica de una serie de tiempo con tendencia.

4. *Otras fluctuaciones irregulares*: para analizar y/o modelar una serie de tiempo es conveniente que ésta tenga un comportamiento estacionario<sup>2</sup>, es decir, que la

---

<sup>2</sup>Veáse la Sección 1.4.

media no tenga un comportamiento dinámico (que no dependa del tiempo; que sea constante) por lo que es necesario remover o quitar los tipos de variación antes mencionados: la tendencia, periodicidad y ciclos, una vez hecho esto nos quedan series de residuales, que pueden ser o no aleatorios.

### 1.3.1. Descomposición de la Serie de Tiempo

Los métodos tradicionales para analizar las series de tiempo, generalmente involucran la descomposición de la serie. Convencionalmente, se ha pensado que las series de tiempo están constituidas por una “mezcla” de tendencia-cambios cíclicos ( $P_t$ ), estacionalidad ( $S_t$ ) e irregularidades ( $\varepsilon_t$ ) de los componentes (v.a.’s no correlacionadas). Si suponemos que estos componentes son independientes y aditivos, uno puede reescribir las series de tiempo como:

$$\begin{aligned} x_t = \text{t-ésima obs.} &= \text{patrón} + \text{error} \\ &= (\text{tendencia, cambios cíclicos}) + \text{estacionalidad} + \text{error} \\ &= P_t + S_t + \varepsilon_t \end{aligned}$$

Donde las componentes de tendencia y estacionalidad son la parte determinística o semideterminística de la serie, mientras que el componente irregular es la parte estocástica.

Recordando las definiciones de los tres componentes arriba mencionados:

- i) Tendencia: es el comportamiento de la serie a largo plazo.
- ii) Estacionalidad: representa los efectos producidos por fenómenos que se presentan con cierta frecuencia dentro del periodo de observación.
- iii) Irregularidades: se utiliza para caracterizar los movimientos imprevistos y considerados como aleatorios y no correlacionados entre sí.

## 1.4. Series de Tiempo Estacionarias

Se dice que una serie de tiempo es estacionaria si no tiene un cambio sistemático tanto en la media como en la varianza (es decir, la serie no presenta tendencia alguna o variaciones estrictamente periódicas). La estacionariedad tal como se ha definido anteriormente es referida frecuentemente en la literatura por *estacionariedad débil*, *estacionariedad en el sentido amplio* o *estacionariedad en segundo orden*.

Por otra parte, se dice que una serie es *estrictamente estacionaria* o *completamente estacionaria* si:

$$F(Y_{t1}, \dots, Y_{tn}) = F(Y_{t1+k}, \dots, Y_{tn+k})$$

para cualquier conjunto de observaciones en el tiempo  $t_1, \dots, t_n$  y  $t_{1+k}, \dots, t_{n+k}$ .

Una manera de describir un proceso estocástico con más de una variable aleatoria, consiste en especificar la distribución de probabilidad conjunta (o función de distribución conjunta) de los valores del proceso en varios puntos del tiempo:  $t_1, \dots, t_n$ . La función de distribución conjunta para un número finito de variables aleatorias  $Y_1, \dots, Y_n$  se define por  $F(Y_{t_1}, \dots, Y_{t_n})$ .

Por lo anterior, se dice que una serie es estrictamente estacionaria si la distribución de probabilidad conjunta permanece invariante con respecto a un desplazamiento en el tiempo.

Para efectos prácticos se utilizará el término de serie estacionaria (o proceso estacionario) como aquella que presenta una estacionariedad débil.

Mucha de la teoría de probabilidad está enfocada a las series de tiempo estacionarias, por lo que dentro del análisis de series de tiempo es conveniente convertir las series no estacionarias en estacionarias para poder aplicar esta teoría.

## 1.5. Transformaciones

Al analizar una serie de tiempo, es deseable transformar los valores observados. Las dos principales razones para hacer una transformación son:

1. *Estabilizar la varianza:* ésto será necesario si existe una tendencia en la serie y la varianza se incrementa junto con la media. En particular, si la desviación estándar es directamente proporcional a la media, se utiliza una transformación logarítmica.
2. *Producir el efecto estacional aditivo:* si existe tendencia en la serie y el tamaño del efecto estacional se incrementa con la media. Esta transformación es recomendable para lograr que el efecto estacional sea constante.

Comúnmente se utilizan tres modelos estacionales:

$$x_t = \mu_t + s_t + \varepsilon_t \quad (1.1)$$

$$x_t = \mu_t s_t + \varepsilon_t \quad (1.2)$$

$$x_t = \mu_t s_t \varepsilon_t \quad (1.3)$$

donde:

$x_t$  =observación al tiempo t

$\mu_t$  =media general

$s_t$  =efecto estacional

$\varepsilon_t$  =error aleatorio

El primer modelo es completamente aditivo, por lo que no requiere ninguna transformación. El segundo modelo tiene estacionalidad multiplicativa y error aditivo, por lo que el tamaño relativo de estos efectos serán los que determinarán la transformación adecuada para aplicar. Por último, se tiene que el tercer modelo es completamente multiplicativo, por lo que es apropiado aplicar una transformación logarítmica.

## 1.6. Análisis de Series con Tendencia

Al analizar una serie de este tipo, debemos saber si queremos medir la tendencia, o bien, removerla para así poder analizar fluctuaciones locales. Con datos estacionales, es conveniente comenzar por calcular los promedios anuales sucesivos que mostrarán una descripción de la tendencia. Una aproximación de este tipo en algunas situaciones es perfectamente adecuada, en especial cuando la tendencia es pequeña, pero algunas otras veces es necesaria una aproximación más sofisticada.

En seguida se describen algunos métodos para analizar la tendencia<sup>3</sup>:

- i) *Ajuste de una curva*: un método tradicional para tratar los datos no-estacionales que presentan tendencia, es ajustarlos a una función simple como una curva polinomial (lineal, cuadrática, etc.), una curva de Gompertz o una curva logística.

La curva de Gompertz puede ser escrita de la forma:

$$\log x_t = a + br^t$$

donde:  $a, b, r$  son parámetros tales que  $0 < r < 1$ , o en su forma alternativa:

$$\log x_t = \alpha \exp[\beta \exp(-\gamma t)]$$

con  $\gamma > 0$ . Por otro lado, la curva logística está dada por:

$$x_t = a / (1 + be^{-ct})$$

Para todas las curvas de este tipo, la función ajustada mide la tendencia y, los residuales (diferencias entre las observaciones y sus valores correspondientes en la curva ajustada) dan como resultado una estimación de fluctuaciones locales.

- ii) *Filtros*: un segundo método para tratar la tendencia es usar un filtro lineal, el cual transforma una serie de tiempo,  $\{x_t\}$ , en otra,  $\{y_t\}$ , a través de una operación lineal:

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

---

<sup>3</sup>Chris Chatfield, *op. cit.*, pp. 15-20.

donde:  $\{a_r\}$  es un conjunto de pesos tales que  $\sum a_r = 1$ . Esta operación es conocida como *promedios móviles (o medias móviles)*.

- iii) *Diferenciación*: un tipo especial de filtro, que es particularmente útil para remover la tendencia, es un operador diferencia aplicado a la serie de tiempo para convertirla en una serie estacionaria.

Para datos no-estacionales, basta con aplicar una sola vez el operador diferencia para convertirlos en estacionarios, por lo que la nueva serie,  $\{y_2, \dots, y_N\}$ , está formada por la serie original,  $\{x_1, \dots, x_N\}$ , donde se cumple que:

$$\begin{aligned} y_t &= x_t - x_{t-1} = \nabla x_{t+1} & t = 2, 3, \dots, N \\ \nabla^j x_t &= x_t - x_{t-j} = (1 - B)^j x_t \\ (1 - B)x_t &= x_t - x_{t-1} \end{aligned}$$

donde  $B$  es el *operador retardo* tal que  $Bx_t = x_{t-1}$  y  $B^j x_t = x_{t-j}$ .

## 1.7. Variación Estacional

Al igual que en una serie que presenta tendencia, el análisis de una serie de tiempo que presenta fluctuaciones estacionales, dependerá de si lo que se desea es medir el efecto estacional, o bien, eliminarlo.

Para series que muestran una tendencia pequeña, es adecuado estimar el efecto estacional para un periodo particular (por ejemplo, un mes del año: enero) para encontrar el promedio de cada observación de dicho periodo, excepto el correspondiente al promedio anual en el caso aditivo (ecuación 1.1); o la observación del periodo (enero) dividido entre el promedio anual en el caso multiplicativo (ecuación 1.2 y ecuación 1.3).

Para series que presentan una tendencia substancial, es necesario un enfoque más sofisticado. Dada una serie de tiempo con observaciones mensuales, la manera más común de eliminar el efecto estacional es calculando:

$$Sm(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + x_{t-4} + \dots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

Nótese que la suma de los coeficientes es 1. Para datos trimestrales, el efecto estacional puede ser eliminado calculando:

$$Sm(x_t) = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}$$

El efecto estacional puede ser calculado como  $x_t - Sm(x_t)$  ó  $x_t/Sm(x_t)$ , dependiendo del tipo de efecto estacional: aditivo o multiplicativo. Si la variación estacional

se mantiene aproximadamente en el mismo nivel, sin tomar en cuenta el nivel de la media, entonces se dice que es **aditiva**, pero si se incrementa de forma directamente proporcional al nivel de la media, se dirá que es **multiplicativa**.

Un efecto estacional puede ser eliminado a través de un operador diferencia, por ejemplo: para datos mensuales se utiliza el operador  $\nabla_{12}$  (operador diferencia estacional de periodo 12), donde:

$$\nabla_{12}x_t = x_t - x_{t-12}$$

## 1.8. Autocorrelación

El término de autocorrelación se puede definir como la correlación entre elementos de series de observaciones ordenadas en el tiempo. Existen diversas razones por las cuales ocurre la autocorrelación, algunas de las cuales son las siguientes:

- i) *Inercia*: una característica relevante de la mayoría de las series de tiempo económicas es la inercia (o lentitud). Como se sabe, las series de tiempo tales como el PIB, índices de precios, etc., presentan ciclos económicos. Cuando se inicia la recuperación económica, la mayoría de estas series comienzan a “moverse” hacia arriba. En este movimiento, el valor de la serie en un punto del tiempo es mayor que su valor anterior, y así continuará hasta que algo suceda para reducirlo nuevamente.
- ii) *Fenómeno de la telaraña*: la oferta de muchos productos agrícolas refleja este fenómeno, en donde la oferta reacciona al precio, con un rezago de un periodo de tiempo debido a que la implementación de las decisiones de oferta toman tiempo. Supóngase que al final del periodo  $t$ , el precio  $P_t$  resulta ser inferior a  $P_{t-1}$ . Por consiguiente, es probable que los agricultores decidan producir, en el periodo  $t + 1$ , menos de lo que produjeron en el periodo  $t$ . Si los agricultores producen excedentes en el año  $t$ , es probable que reduzcan su producción en  $t + 1$  y así sucesivamente, produciendo un patrón de telaraña.
- iii) *Manipulación de datos*: un ejemplo de esto, son las series de tiempo que contienen información trimestral, que por lo general se obtiene de información mensual, agregando simplemente las observaciones de tres meses dividida entre 3. Este procedimiento de promediar las cifras induce cierto suavizamiento en los datos al eliminar las fluctuaciones en la información mensual. Este suavizamiento puede inducir a un patrón sistemático en la parte considerada como aleatoria, introduciendo con esto autocorrelación. Otra fuente de manipulación es la interpolación o extrapolación de la información. Estas técnicas de manipulación de datos, pueden imponer el patrón sistemático ya antes mencionado, que pudiera no estar presente en la información original.

Una guía importante acerca de las propiedades de una serie de tiempo está dada por la **serie de coeficientes de autocorrelación**. Dichos coeficientes miden la correlación entre las observaciones a diferentes distancias. Éstos nos dan idea del modelo de probabilidad que genera los datos.

Dadas  $N$  parejas de observaciones de dos variables  $x$  y  $y$ ,  $\{(x_1, y_1), \dots, (x_N, y_N)\}$ , el coeficiente de correlación muestral entre ellas está dado por:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1.4)$$

donde  $r \in [-1, 1]$ . Aplicando una fórmula análoga a las series de tiempo, se puede saber si las observaciones sucesivas se correlacionan.

Dadas  $N$  observaciones  $x_1, \dots, x_N$ , en una serie de tiempo, se pueden formar  $N - 1$  parejas de observaciones:  $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$ . Sea la primera observación de cada pareja una variable, y la segunda observación de cada pareja una segunda variable, entonces, por analogía con la ecuación (1.4), se puede calcular el coeficiente de correlación entre observaciones adyacentes,  $x_t$  y  $x_{t+1}$ , usando:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2}} \quad (1.5)$$

donde:

$$\bar{x}_{(1)} = \sum_{t=1}^{N-1} x_t / (N - 1)$$

es la media de las primeras  $(N-1)$  observaciones, y

$$\bar{x}_{(2)} = \sum_{t=2}^N x_t / (N - 1)$$

es la media de las últimas  $(N-1)$  observaciones. Como el coeficiente dado por la ecuación (1.5) mide la correlación entre observaciones sucesivas, es llamado *coeficiente de autocorrelación* o *coeficiente de autocorrelación serial*.

En la ecuación (1.5),  $\bar{x}_{(1)} \simeq \bar{x}_{(2)}$  para  $N$  grande, por lo que comúnmente se aproxima por:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(N - 1) \sum_{t=1}^N (x_t - \bar{x})^2 / N}$$

donde  $\bar{x} = \sum_{t=1}^N x_t / N$  es la media general. Frecuentemente se simplifica esta última expresión suprimiendo el factor  $N/(N - 1)$ , ya que  $\lim_{N \rightarrow \infty} N/(N - 1) = 1$ , con lo que

se obtiene:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}$$

De manera análoga, se puede calcular la correlación entre observaciones que se encuentran a  $k$  “pasos” de distancia:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad k = 0, 1, 2, \dots$$

$$= \text{corr}(x_t, x_{t+k})$$

$$= \text{Coeficiente de autocorrelación a distancia } k \text{ (lag } k)$$

### 1.8.1. Autocorrelación Parcial

De manera similar a la función de autocorrelación muestral, la función de autocorrelación parcial es el conjunto de coeficientes de autocorrelación parcial desde el retardo  $k = 1$  hasta un máximo posible de retardos igual a  $N/2$ .

La función de autocorrelación parcial para el retardo  $k$  se define como una correlación condicional, es decir:

$$\phi_{kk} = \text{corr}(x_t, x_{t+k} | x_{t+1}, x_{t+2}, \dots, x_{t+k-1})$$

La expresión anterior puede interpretarse como la estimación del coeficiente de autocorrelación muestral para el retardo  $k$  con la eliminación del efecto producido por las autocorrelaciones para retardos menores que  $k$ , las cuales están presentes en la estimación del coeficiente de autocorrelación muestral. Dicho de otra forma: el coeficiente de autocorrelación parcial no considera las correlaciones acumuladas hasta el retardo  $k$  para el que se estima.

Otra manera de calcular la autocorrelación parcial entre  $x_t$  y  $x_{t+k}$  es hallando la razón entre los determinantes de dos matrices de autocorrelaciones, donde la matriz del numerador es la matriz simétrica del denominador pero sustituyendo la última columna por  $(r_1, r_2, \dots, r_k)$ . Por lo tanto, al considerar  $k = 1, 2, \dots$ , se tiene la siguiente expresión<sup>4</sup>:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}$$

donde  $P_k$  es la matriz de autocorrelación  $k \times k$ . De este modo se tiene que:

$$\phi_{11} = r_1$$

---

<sup>4</sup>Jaume Arnau, *op. cit.*, pp. 33-37.

es la función de autocorrelación de un retardo, puesto que no se elimina ninguna autocorrelación intermedia.

Para  $\phi_{22}$  se tiene que:

$$\phi_{22} = \frac{\begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix}} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

es el valor de la función de autocorrelación parcial de dos retardos.

Para  $\phi_{33}$ , el valor de la función de autocorrelación parcial es:

$$\phi_{33} = \frac{\begin{vmatrix} 1 & r_1 & r_1 \\ r_1 & 1 & r_2 \\ r_2 & r_1 & r_3 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{vmatrix}} = \frac{r_3 - r_1^3 + r_1 r_2^2 - 2r_1 r_2 - r_1^2 r_3}{1 + 2r_1^2 r_2 - r_2^2 - 2r_1^2}$$

es decir, la función de autocorrelación a tres retardos, eliminando las autocorrelaciones relativas a los retardos 1 y 2.

La expresión general para  $\phi_{kk}$  es:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{k-2} & r_1 \\ r_1 & 1 & r_1 & \dots & r_{k-1} & r_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & r_k \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 & \dots & r_{k-2} & r_{k-1} \\ r_1 & 1 & r_1 & \dots & r_{k-1} & r_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & r_1 & 1 \end{vmatrix}}$$

### 1.8.2. Correlograma

Una herramienta muy útil para interpretar un conjunto de coeficientes de autocorrelación es la gráfica llamada **correlograma**, en la cual los coeficientes de autocorrelación muestral,  $r_k$ , son graficados contra *lag*  $k$  (se usará *lag* o *retraso* indistintamente), para  $k = 0, 1, \dots, M$ , donde por lo general  $M$  es menor que  $N$ .

Interpretar un correlograma no siempre es fácil; a continuación se describe el procedimiento a seguir, para la interpretación, según el tipo de serie:

1. *Series aleatorias*: una serie de tiempo es llamada *puramente aleatoria* si consiste de una serie de observaciones independientes con la misma distribución. Entonces para una serie de tiempo puramente aleatoria con  $N$  grande se tiene que  $r_k \simeq 0$  para todos los valores de  $k$  distintos de cero. De hecho, para series de tiempo aleatorias, la distribución de  $r_k$  se aproxima a  $N(0,1/N)$ .
2. *Correlación en periodos cortos*: las series estacionarias generalmente muestran correlación en periodos cortos, que se caracterizan por un valor alto de  $r_1$ , seguido de uno o dos coeficientes que son un poco más grandes que cero, con la tendencia a disminuir paulativamente. Los valores de  $r_k$  para *retrasos* largos suelen ser aproximadamente cero (Figura 1.3 y Figura 1.4). Para las series de este tipo es conveniente usar un modelo denominado autorregresivo.

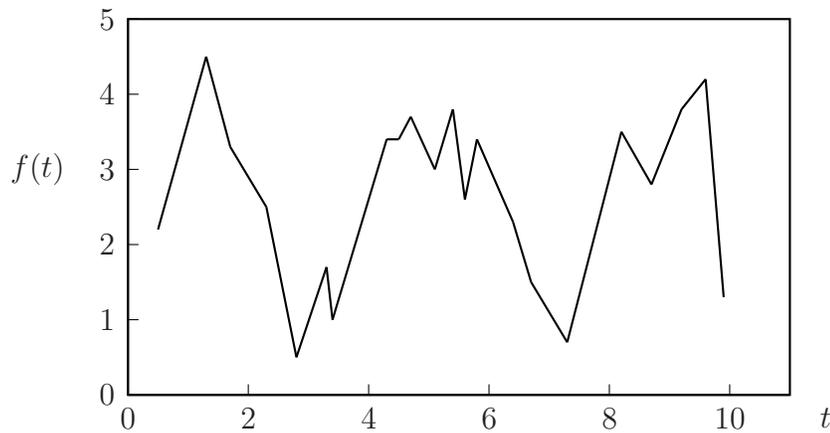


Figura 1.3: Correlación en periodos cortos.

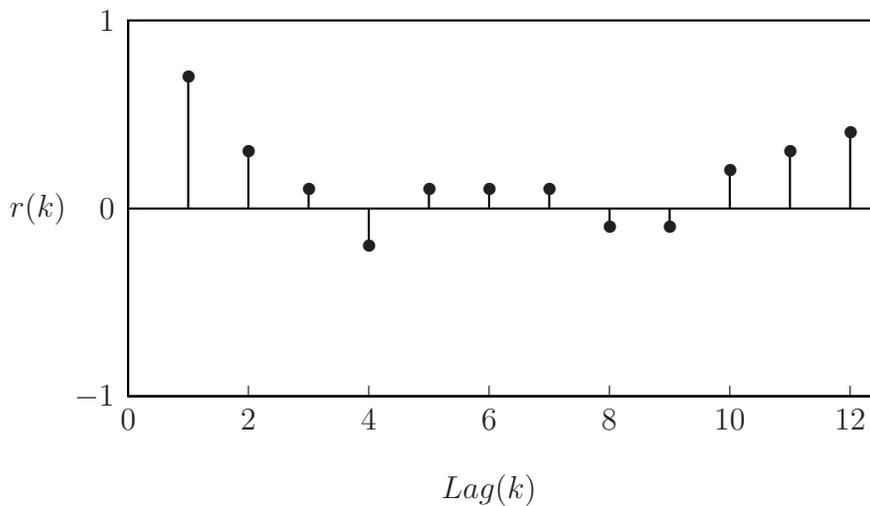


Figura 1.4: Correlograma correspondiente a la Figura 1.3.

3. *Series alternantes*: si los valores de una serie de tiempo alternan de un lado de la media al otro, el correlograma también tiende a alternar. Aquí, el valor de  $r_1$  será negativo, sin embargo, el valor de  $r_2$  será positivo siempre y cuando las observaciones para  $lag(2)$  se encuentren del mismo lado de la media (Figura 1.5 y Figura 1.6)

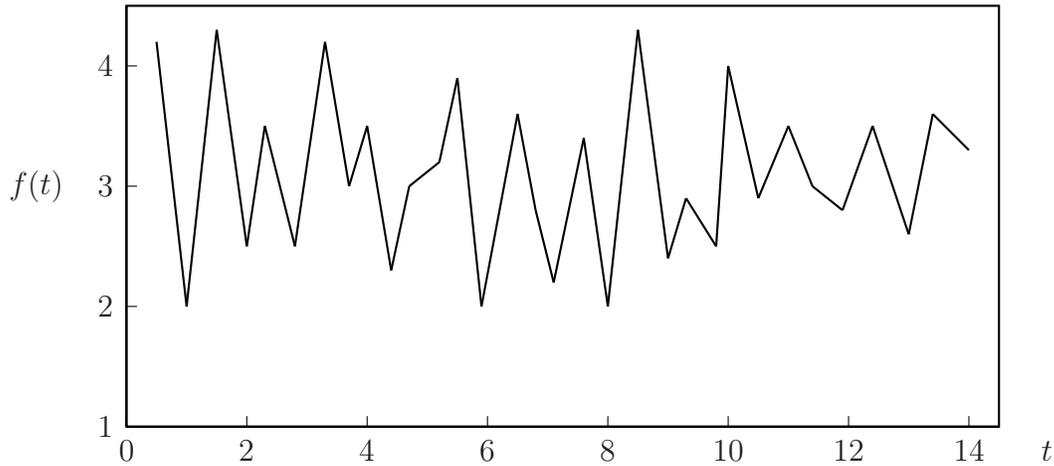


Figura 1.5: Gráfica de una serie alternante.

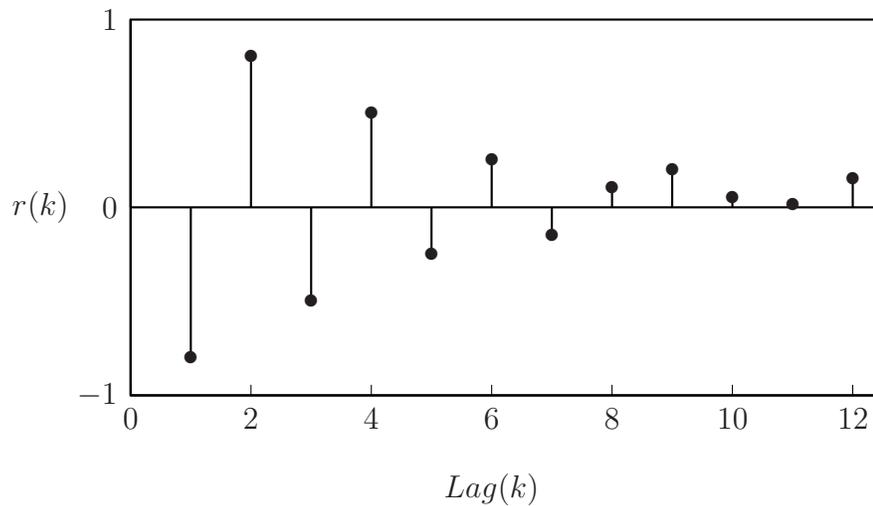


Figura 1.6: Correlograma correspondiente a la Figura 1.5.

4. *Series no estacionarias*: si una serie de tiempo tiene tendencia, entonces los valores de  $r_k$  no serán menores que cero, excepto para valores muy grandes de retrasos. Esto se debe a que una observación que se encuentra a un lado de la media tiende a ser seguida por observaciones del mismo lado de la media

gracias a la tendencia. Poco se puede inferir de un correlograma de este tipo de series, puesto que la tendencia domina todas las características. De hecho, la serie formada por la función de autocorrelación  $\{r_k\}$  sólo debe ser calculada para series de tiempo estacionarias, por lo que cualquier tendencia deber ser removida antes de calcular  $\{r_k\}$  (Figura 1.7 y Figura 1.8).

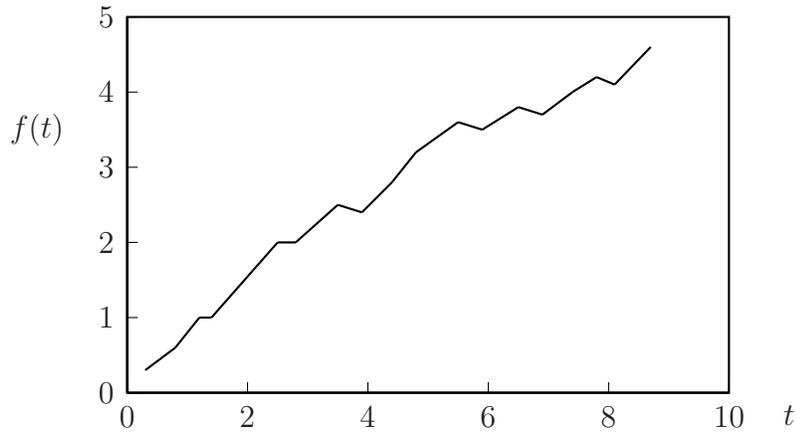


Figura 1.7: Gráfica de una serie no estacionaria.

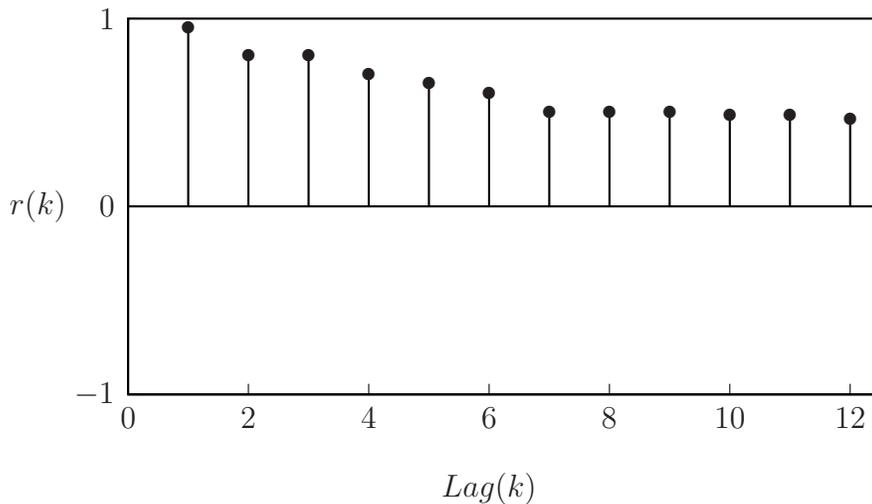


Figura 1.8: Correlograma correspondiente a la Figura 1.7.

5. *Series estacionarias*: si una serie de tiempo tiene fluctuaciones estacionales, su correlograma también mostrará oscilaciones en las mismas frecuencias. Por ejemplo, en observaciones mensuales,  $r_6$  será grande y negativo, mientras que  $r_{12}$  será grande y positivo. Si la variación estacional es removida de los datos, entonces el correlograma podrá brindar información útil; para remover la variación

estacional basta con calcular 12 promedios mensuales y quitarles el correspondiente a cada observación. El correlograma de las series resultantes, muestra que los tres primeros coeficientes son grandes y tienden a disminuir, lo que indica una correlación en un periodo corto (Figura 1.9 y Figura 1.10).

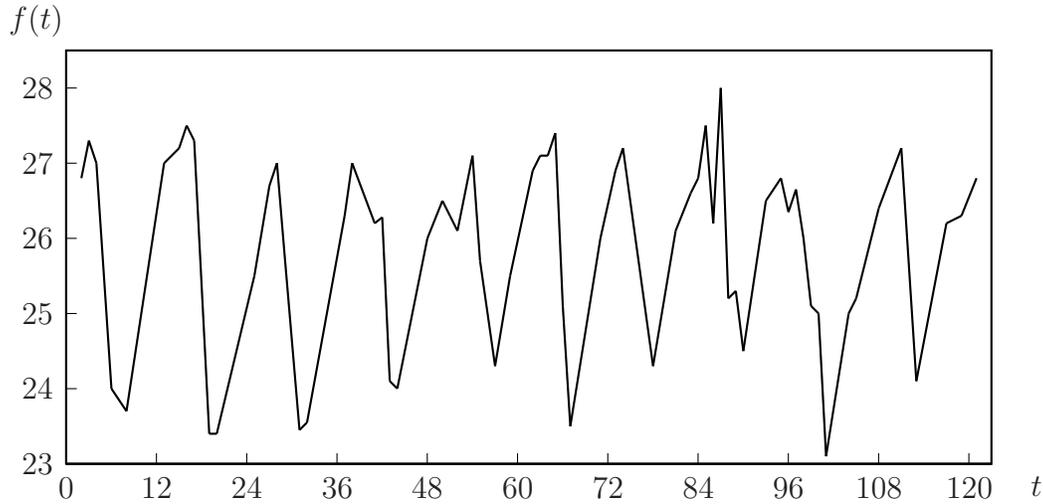


Figura 1.9: Gráfica de una serie estacionaria.

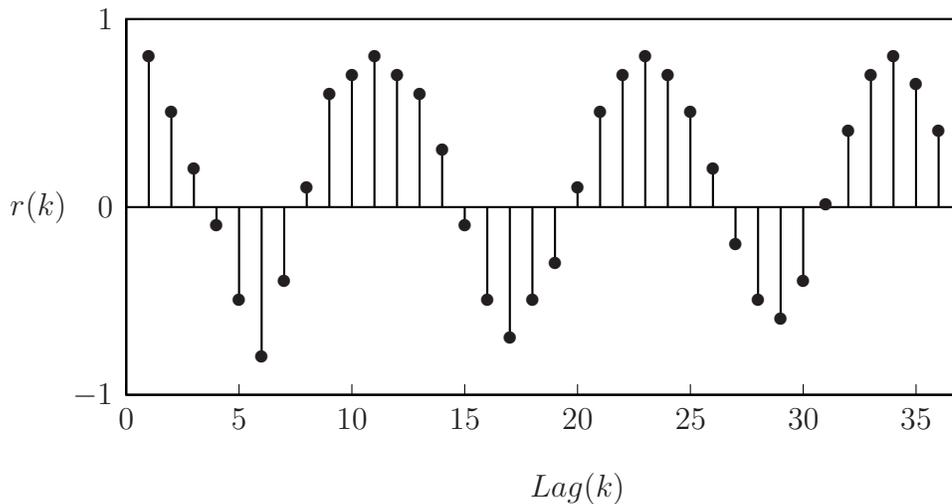


Figura 1.10: Correlograma correspondiente a la Figura 1.9.

6. *Outliers*: si una serie de tiempo contiene uno o más *Outliers*, el correlograma se verá seriamente afectado, debido a que los *Outliers* ocasionan que los coeficientes de correlación tomen el valor cero, por lo que resulta conveniente ajustarlos antes de hacer el correlograma<sup>5</sup>.

<sup>5</sup>Chris Chatfield, *op. cit.*, pp. 22-27.

## 1.9. Representación Autoregresiva de un Proceso Lineal

### 1.9.1. Procesos de Ruido Blanco

Uno de los procesos que con mayor frecuencia se utiliza en la modelización de una serie temporal es llamado *proceso de ruido blanco* o *proceso puramente aleatorio*<sup>6</sup>. Un proceso de ruido blanco esta formado por una secuencia de variables aleatorias mutuamente independientes e idénticamente distribuidas, cuya representación matemática es:

$$\begin{aligned} X_t &= \mu + a_t \\ \Rightarrow a_t &= X_t - \mu \end{aligned}$$

Así un proceso  $\{a_t\}$  (proceso de impulsos aleatorios) es considerado de ruido blanco si éste tiene las siguientes propiedades:

1. media constante  $\mathbb{E}(a_t) = \mu_a$  (usualmente 0)  $\forall t$ ,
2. varianza constante  $\text{Var}(a_t) = \sigma_a^2$   $\forall t$ ,
3.  $\gamma_k = \text{Cov}(a_t, a_{t+k}) = 0$   $\forall k \neq 0$ ,  
 $\gamma_k = \sigma_a^2$  si  $k = 0$

Puesto que son *v.a.'s* no correlacionadas, se sigue que un proceso de ruido blanco  $\{a_t\}$  es un proceso estacionario con la función de autocorrelación:

$$r_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0, \end{cases}$$

y la función de autocorrelación parcial:

$$\phi_{kk} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0. \end{cases}$$

### 1.9.2. Representaciones para los Procesos de Series de Tiempo

En el análisis de series de tiempo, existen dos representaciones útiles para expresar un proceso de serie de tiempo. Una de ellos, es escribir un proceso  $Z_t$  como una combinación lineal de una secuencia de variables aleatorias no correlacionadas<sup>8</sup>, *i.e.*,

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

---

<sup>6</sup>William Wei, *op. cit.*

donde  $\psi_0 = 1$ ,  $\{a_t\}$  es un proceso de ruido blanco con media cero, y  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ .

Otra manera útil es escribir un proceso  $Z_t$  en una representación autoregresiva ( $AR$ ), en la cual retrocedemos el valor de  $Z$  al tiempo  $t$  sobre sus propios valores pasados, *i.e.*, una observación depende de las observaciones anteriores:

$$Z_t = \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \dots + a_t \quad (1.6)$$

donde  $a_1, a_2, \dots$  es una secuencia de variables aleatorias no correlacionadas (errores independientes, proceso de ruido blanco).

### 1.9.3. Proceso Autoregresivo: Proceso $AR$

Sea la serie  $X_1, X_2, \dots$  y supongamos que en la ecuación (1.6) hay un número finito de pesos (o ponderaciones)  $\pi \neq 0$ , *i.e.*,  $\pi_1 = \alpha_1, \pi_2 = \alpha_2, \dots, \pi_p = \alpha_p$  y  $\pi_k = 0$  para  $k > p$ , entonces el proceso resultante es llamado *proceso autoregresivo de orden  $p$*  denotado como  $AR(p)$ . Éste está dado por:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \epsilon_t$$

donde  $\epsilon_1, \epsilon_2, \dots$  es una secuencia de *v.a.'s* no correlacionadas (errores independientes) con distribución normal e igual, con valor esperado 0, varianza  $\sigma_\epsilon^2$  y  $\sum_{j=1}^p |\alpha_j| < \infty$ .

Los procesos  $AR$  son útiles para describir situaciones en las que el valor presente de una serie de tiempo depende de sus valores precedentes.

#### Proceso $AR(1)$

Para el proceso autoregresivo de primer orden, escribimos:

$$X_t = \alpha_1 X_{t-1} + \epsilon_t$$

Para que el proceso  $AR(1)$  sea un proceso estacionario se debe cumplir  $|\alpha| < 1$ . Si  $|\alpha| > 1$ , el proceso se hace explosivo y rápidamente tiende a crecer desmesuradamente.

Algunas veces los procesos  $AR(p)$  son llamados **procesos de Markov**, porque el valor de  $X_t$  es completamente determinado por el conocimiento de  $X_{t-1}$ .

### 1.9.4. Proceso de Media Móvil: Proceso $MA$

En la representación de un proceso de media móvil, si un número finito de pesos (o ponderaciones)  $\psi \neq 0$ , *i.e.*,  $\psi_1 = -\beta_1, \psi_2 = -\beta_2, \dots, \psi_q = -\beta_q$  y  $\psi_k = 0$  para  $k > q$ , entonces el proceso resultante es llamado *proceso de media móvil de orden  $q$*  denotado como  $MA(q)$ . Éste está dado por:

$$X_t = \epsilon_t - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q}$$

donde  $\epsilon_1, \epsilon_2, \dots$  es una secuencia de errores independientes y  $1 + \beta_1^2 + \dots + \beta_q^2 < \infty$ . Un proceso de media móvil finito, siempre es un proceso estacionario. Los procesos de medias móviles son útiles para describir fenómenos en los cuales los eventos producen un efecto inmediato que solo dura cortos períodos de tiempo.

### Proceso $MA(1)$

Para el proceso autoregresivo de primer orden, escribimos:

$$X_t = \epsilon_t - \beta_1 \epsilon_{t-1}$$

El proceso  $MA$  tiene memoria corta, y absorbe rápidamente los impactos, esto quiere decir que ante un impacto externo (o lo que hemos llamado error  $-\epsilon_t$ ), la serie  $AR(1)$ , tarda en volver a su nivel, mientras que la serie  $MA(1)$  vuelve inmediatamente, en un par observaciones por ejemplo.

### 1.9.5. Proceso $ARMA(p, q)$

Un problema al que nos podemos enfrentar al utilizar alguna de las representaciones anteriores es que pueden contener demasiados parámetros. Esto sucede incluso para un proceso de media móvil de orden finito (o un proceso autoregresivo de orden finito), ya que a menudo se toma un modelo de orden superior para obtener una buena aproximación. En general, un gran número de parámetros reduce la eficiencia en la estimación. Por lo tanto, en la construcción de modelos, puede ser necesario incluir tanto modelos autorregresivos como de media móvil. Esto conlleva a los siguientes procesos mixtos, conocidos como procesos  $ARMA(p, q)$ : *procesos autoregresivos de media móvil*, cuya forma es<sup>7</sup>:

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q} \\ \Rightarrow X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} &= \epsilon_t - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q} \\ \Rightarrow \epsilon &= X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q} \end{aligned}$$

### Proceso $ARMA(1, 1)$

Para el proceso  $ARMA(1, 1)$  escribimos:

$$X_t = \alpha_1 X_{t-1} + \epsilon_t - \beta_1 \epsilon_{t-1}$$

Para que el proceso anterior sea estacionario se debe cumplir  $|\alpha_1| < 1$ . Cuando  $\alpha_1 = 0$  el proceso se reduce a un proceso  $MA(1)$ , y cuando  $\beta = 0$ , éste se reduce a un proceso  $AR(1)$ . Por lo tanto, podemos considerar los procesos  $AR(1)$  y  $MA(1)$  como casos especiales del proceso  $ARMA(1, 1)$ .

<sup>7</sup>Arnau Jaume, *op.cit.*, pp. 47-54

### 1.9.6. Proceso Circular $CAR(p)$

Un proceso circular  $AR(p)$  (conocido como proceso  $CAR(p)$ ) puede definirse directamente, como uno en el que la distribución de  $\Theta_t$ , dados los valores pasados  $\Theta_{t-1} = \theta_{t-1}, \Theta_{t-2} = \theta_{t-2}, \dots$ , tiene una distribución von Mises<sup>8</sup>  $VM(\mu, \kappa)$  (aunque podría ser utilizado con cualquier otra distribución circular simétrica unimodal alrededor de su media), donde la dirección  $\mu_t$  está dada por:

$$\mu_t = \mu + g[\alpha_1 g^{-1}(\theta_{t-1} - \mu) + \dots + \alpha_p g^{-1}(\theta_{t-p} - \mu)] \quad (1.7)$$

y  $\kappa$  es un parámetro de concentración constante. Una vez que los parámetros de la ecuación (1.7) se han estimado, la media del siguiente valor en la serie puede ser previsto.

Para el proceso  $AR(1)$ , existe una única distribución marginal para que el proceso sea estacionario. Esta distribución es la distribución asintótica del proceso  $AR(1)$  (Markov) y no es una von Mises.

Para ajustar modelos  $CAR(p)$  en series dispersas ( $\kappa < 2$ , *i.e.*, para valores pequeños de  $p$ ), los parámetros pueden ser estimados al realizar la máxima verosimilitud (maximizando):

$$\prod_{t=p+1}^n f\{\theta_t - \mu - g\{\alpha_1 g^{-1}[(\theta_{t-1} - \mu)/2] + \dots + \alpha_p g^{-1}[(\theta_{t-p} - \mu)/2]\}\} \prod_{t=1}^p f(\theta_t - \mu)$$

donde  $f$  denota la función de densidad de probabilidad de la distribución von Mises  $VM(0, \kappa)$ .

Para ajustar modelos  $LARMA(p, q)$  en series concentradas ( $\kappa \geq 2$ , para cualquier valor de  $\mu$ ), deben transformarse los datos circulares a datos lineales mediante una función de enlace<sup>9</sup> (del inglés *Link Function*). Una función adecuada para estos datos es la denominada *función de enlace probit*, cuya inversa está dada por:

$$x_t(\mu) = \Phi^{-1} \left( \frac{\theta_t - \mu}{2\pi} + \frac{1}{2} \right) \quad (1.8)$$

donde el valor de  $\theta - \mu$  se encuentra entre  $-\pi$  y  $\pi$ .

---

<sup>8</sup>Veáse la Sección 2.4.

<sup>9</sup>Veáse la Sección 3.1.3.

**Ejemplo.** A continuación se aplicarán los dos métodos a los datos correspondientes a la Figura 3.1.

Obsérvese detenidamente la función de autocorrelación circular (Figura 3.2). Basándose en la Figura 3.3, resulta evidente que el modelo  $AR(1)$  (o  $CAR(1)$ ) es apropiado. Las gráficas de la serie de tiempo lineal enlazada así como las correspondientes funciones de autocorrelación y autocorrelación parcial, se muestran en la Figura 1.11.

Para utilizar el método indirecto, suponemos que las distribuciones condicionales de  $\Theta_t|\theta_{t-1}$  son von Mises con parámetro de concentración  $\kappa$ . Para estimar los parámetros se puede utilizar una regresión para un valor fijo de  $\mu$ . Por lo tanto, el resultado de maximizar el valor de la probabilidad es una función de  $\mu$ , y podemos estimar  $\mu$  por el valor que maximiza dicha función. Los errores estándar pueden ser obtenidos por maximización de la log-verosimilitud.

Los valores iniciales  $\mu$  y  $\kappa$  en el algoritmo del modelo adecuado pueden obtenerse en el tratamiento de los datos como una muestra circular y un valor inicial de  $\alpha$  puede ser estimado a partir de un prototipo de la gráfica de la función de autocorrelación circular, obteniendo:

$$\hat{\mu}_0=291.2^\circ, \quad \hat{\kappa}_0=1.9, \quad \hat{\alpha}_0=0.5$$

Maximizando el valor de la probabilidad, se obtienen los parámetros estimados:

$$\hat{\mu}=289.5^\circ, \quad \hat{\kappa}=2.467, \quad \hat{\alpha}=0.678$$

con errores estándar de  $13.8^\circ$ , 0.352 y 0.138, respectivamente. Debido a que  $\hat{\kappa}$  excede a 2.0, consideramos cambiar al modelo probit enlazado, para lo cual consideramos la expresión (1.8) para transformar los datos. Al maximizar ésta sobre  $\mu$  obtenemos los parámetros estimados para este modelo:

$$\hat{\mu}_0=297.2^\circ, \quad \hat{\phi}=0.52, \quad \hat{\sigma}^2=0.146$$

Nótese el resultado para el parámetro  $\mu$  en ambos modelos; son muy distintos entre sí por lo que no son directamente comparables.

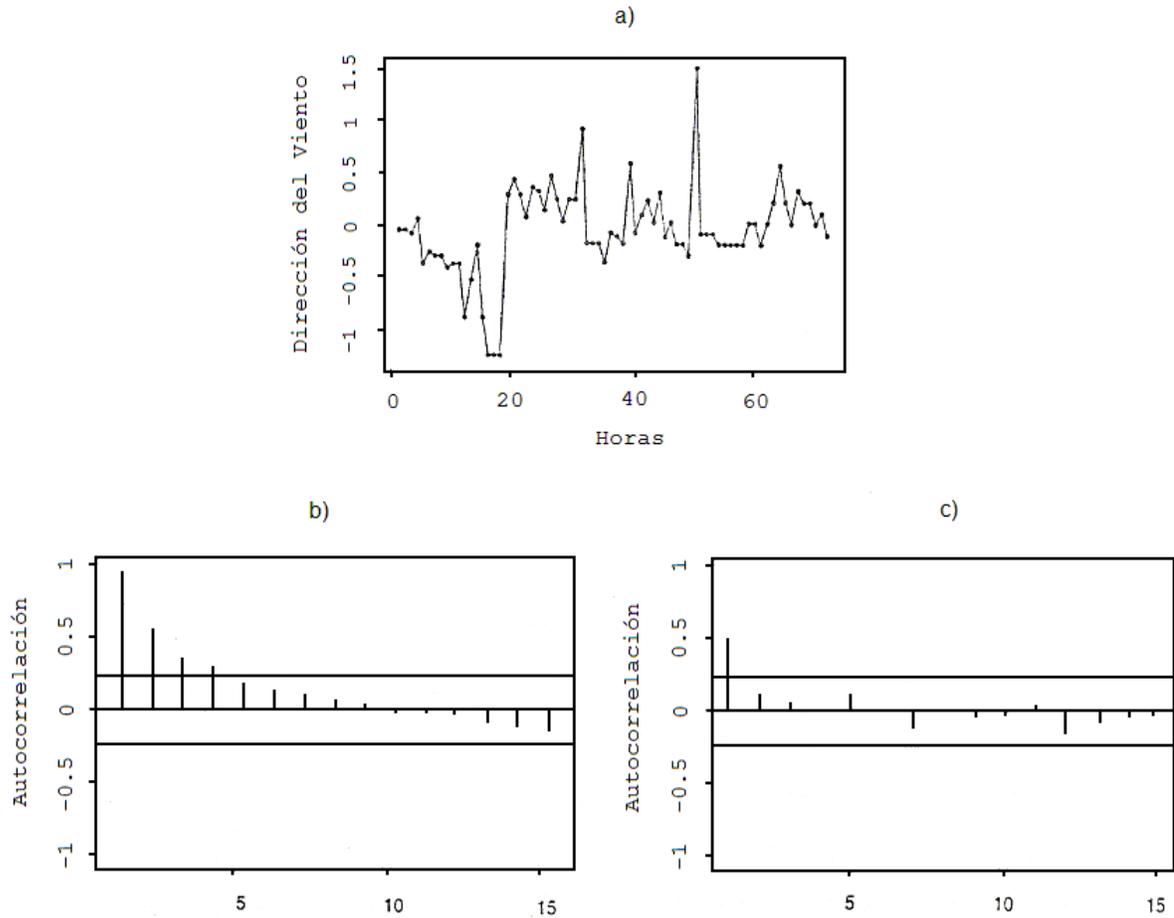


Figura 1.11: Análisis de las series de las 72 mediciones de la dirección del viento, usando un modelo probit para ajustar un modelo  $LAR(1)$ : a) Serie de tiempo lineal enlazada; b) Función de autocorrelación para series enlazadas; c) Función de autocorrelación parcial para series enlazadas (las bandas en b) y c) indican que las correlaciones son significativamente distintas de cero).

## Capítulo 2

# Introducción a la Estadística Circular

En el capítulo anterior se describieron los conceptos básicos para el correcto manejo estadístico de las series temporales, particularmente los relativos a la función de autocorrelación y el correlograma, los cuales, como se verá en los próximos capítulos, poseen una importancia decisiva en el proceso de identificación de aquellos modelos que mejor se ajustan a las series. A continuación se abordarán los fundamentos básicos de la estadística circular, así como algunas de sus distribuciones más comunes.

La estadística circular, área derivada tanto de la estadística como de la matemática, se sitúa entre la estadística tradicional, comúnmente llamada “estadística lineal”, y la estadística esférica (la cual se distribuye en tres dimensiones); es una herramienta utilizada para el análisis de *datos circulares*, es decir, de tipo angular (vectores), los cuales se distribuyen en una circunferencia. Los datos circulares aparecen en numerosos campos de la ciencia tan diversos como: la astronomía, medicina, biología, geografía, geología, meteorología, oceanografía, etc. Algunos ejemplos de las aplicaciones por área son:

» Biología:

- > Estudios de trayectorias de animales a partir de un punto referencial.
- > Ecología: Es una de las ramas de la Biología que más ha hecho uso de la estadística circular, las primeras aplicaciones se realizaron en el vuelo de las aves y orientaciones de otros animales como salamandras, hormigas e insectos.

» Física:

- > Al medir direcciones preferentes de ejes ópticos en los cristales de distintos especímenes de rocas.

- » Meteorología:
  - > Se utiliza para la representación de mapas estelares, cuya notación se realiza en grados ( $^{\circ}$ ), en el intervalo  $[0^{\circ}, 360^{\circ}]$ .
- » Geografía:
  - > Estudios de la ocurrencia de terremotos en una región. La variación de dichos terremotos se estudia por año y/o por día, es decir, se aplica el análisis de series de tiempo, obteniéndose de estos estudios datos de tipo angular.
- » Geología
  - > Estudios de direcciones magnéticas: la estadística circular se utiliza para interpretar estudios de paleomagnetismo de corriente y magnetismo polar en los tiempos geológicos.
  - > Estudios de orientaciones de fracturas y elementos presentes en rocas: interpretación de fuerzas tectónicas.
- » Medicina:
  - > La estadística circular y sus datos angulares se aplican a estudios cardiológico-cíclicos.
  - > Estudios de salto o de nado con el fin de observar la habilidad de orientación.
  - > Estudios de percepción bajo condiciones experimentales, por ejemplo, la simulación de la gravedad cero en un viaje al espacio.

La estadística circular es ideal para analizar datos de tiempo, trayectorias y orientaciones. Cabe mencionar que la estadística circular tiene un enorme potencial, sin embargo no existe mucha literatura al respecto; en los últimos años se ha visto un desarrollo vigoroso de métodos estadísticos circulares para analizar registros de datos angulares, particularmente métodos no paramétricos y métodos básicos que contribuyen al análisis de datos, así como métodos bayesianos.

## 2.1. Datos Circulares

La estadística circular se basa en una circunferencia unitaria para realizar el análisis de trayectorias con base en ángulos. Los datos de orientación reciben el nombre de *datos circulares*. La representación numérica de direcciones se realiza con referencia a un punto cero y a una rotación (en el sentido de las manecillas del reloj).

Una observación hecha en una circunferencia unitaria, donde el ángulo  $\alpha$  ( $0^\circ < \alpha < 360^\circ$ ) se encuentra entre dos puntos, P y P' medidos en grados, se le denomina vector unitario y dicha observación puede describirse como un dato circular (Figura 2.1). En el círculo unitario el eje vertical se transforma en el eje X y, el eje horizontal en el eje Y, esta transformación sucede porque el punto  $0^\circ$  se debe anclar hacia el norte en el eje positivo, además, como ya se mencionó antes, las observaciones deberán representarse con una rotación en sentido de las manecillas del reloj (Figura 2.1).

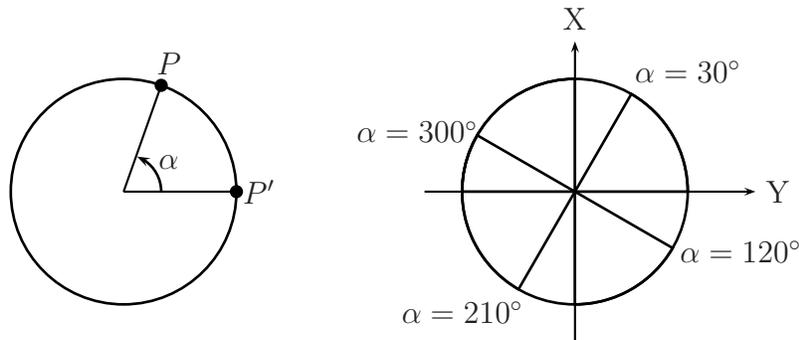


Figura 2.1: Representación de datos circulares en la circunferencia unitaria.

Los procedimientos circulares utilizan dos funciones básicas de la trigonometría: seno y coseno. La ordenada X indica el valor de coseno y la abscisa Y indica el valor de la función seno, por lo que matemáticamente, en un círculo unitario se define:

$$\begin{cases} x = \cos(\alpha) \\ y = \text{sen}(\alpha) \end{cases}$$

Así, para un círculo con radio ( $r$ ) distinto de 1, tenemos:

$$\begin{cases} \cos(\alpha) = \frac{x}{r} \\ \text{sen}(\alpha) = \frac{y}{r} \end{cases}$$

Por lo que la localización de un punto en un plano circular está determinado por un radio ( $r$ ) y un ángulo ( $\alpha$ ) localizados por las funciones trigonométricas seno y coseno (coordenadas polares).

Los datos angulares se pueden representar de dos maneras en un círculo: una de ellas es por medio de puntos alrededor de la circunferencia (Figura 2.2, a)) y, la otra es por medio de radios (Figura 2.2, b)), esta última, muestra la orientación desde el origen hasta el dato angular.

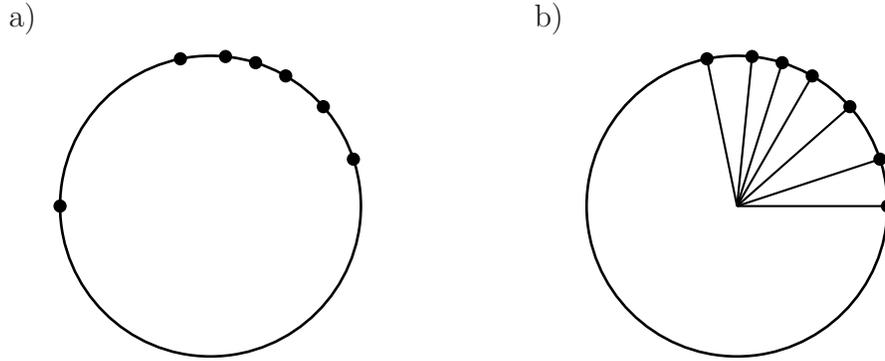


Figura 2.2: Representación de datos circulares.

## 2.2. Momentos Trigonométricos

Sea  $f(\phi)$  la función de densidad de probabilidad de una variable aleatoria  $\Phi$ . Definimos el  $p$ -ésimo momento trigonométrico de  $\Phi$  (o de  $f(\phi)$ ),  $p = 1, 2, \dots$ , por:

$$\begin{aligned}\mu'_p &= \rho_p e^{i\mu'_p} \\ &= \rho_p \cos \mu'_p + i \rho_p \sen \mu'_p \\ &= \int_0^{2\pi} \cos(p\phi) f(\phi) d\phi + i \int_0^{2\pi} \sen(p\phi) f(\phi) d\phi \\ &= \alpha'_p + i\beta'_p\end{aligned}$$

donde  $\alpha'_p$  es el  $p$ -ésimo momento coseno y  $\beta'_p$  el  $p$ -ésimo momento seno.

Cuando  $p = 1$ , se reescribe a  $\mu_1$  simplemente como  $\mu$ , y a  $\rho_1$  como  $\rho$ :

$$\mu'_p = \rho e^{i\mu}$$

donde  $\mu$  es la *dirección media* y  $\rho$  la *longitud media resultante* ( $0 \leq \rho \leq 1$ ).

Por otro lado, el  $p$ -ésimo momento trigonométrico central de  $\Phi$  es justamente el  $p$ -ésimo momento trigonométrico de  $\Phi - \mu$  (i.e.  $\Phi$  centrado con respecto a su dirección media  $\mu$ ):

$$\begin{aligned}\mu_p &= \rho_p e^{i\mu_p} \\ &= \rho_p \cos(\mu_p) + i \rho_p \sen(\mu_p) \\ &= \int_0^{2\pi} \cos(p(\phi - \mu)) f(\phi) d\phi + i \int_0^{2\pi} \sen(p(\phi - \mu)) f(\phi) d\phi \\ &= \alpha_p + i\beta_p\end{aligned}$$

Cuando  $p = 1$ ,  $\alpha_1 = \rho$  y  $\beta_1 = 0$ .

La *varianza circular* de  $\Phi$  está dada por:

$$\nu = 1 - \rho, \quad 0 \leq \nu < 2\pi$$

La *desviación estándar circular* no se calcula como de costumbre ( $\sqrt{\nu}$ ), sino como:

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

si  $\rho \rightarrow 1$  entonces:

$$\sigma \simeq (2\nu)^{1/2} \quad \text{ó bien} \quad \sigma \simeq [2(1 - \rho)]^{1/2}$$

Una medida asociada a la extensión es la llamada *dispersión circular*, dada por:

$$\delta = \frac{(1 - \rho^2)}{2\rho^2}$$

Además de los momentos trigonométricos existen otras medidas poblacionales, como lo son la *moda direccional* y los *cuartiles*, entre otras<sup>1</sup>.

## 2.3. Distribuciones Circulares

Si podemos ajustar un modelo de probabilidad a nuestros datos, entonces el conjunto de datos se puede resumir de manera eficiente utilizando dicho modelo, por esta razón los modelos de probabilidad son un aspecto de gran importancia dentro del análisis estadístico. Para datos lineales, tenemos las distribuciones univariadas que pueden tener un rango finito tal como la distribución binomial, o un rango  $+\infty$  como la distribución Poisson, o incluso pueden extenderse a lo largo de la recta al igual que la distribución normal.

En contraste, las distribuciones circulares **siempre son finitas**, su rango es  $[0^\circ, 360^\circ]$  ó bien,  $[0, 2\pi]$ . Por lo general las distribuciones circulares son continuas; existe una función de densidad de probabilidad<sup>2</sup>, la cual depende del ángulo  $\phi$ :

$$\begin{aligned} f(\phi), & \quad 0 \leq \phi < 2\pi \\ f(\phi) = f(\phi + 2\pi) & \quad (\text{f es periódica con periodo } 2\pi) \end{aligned}$$

con las siguientes características:

1. Es no negativa:

$$f(\phi) \geq 0$$

<sup>1</sup>Ver Fisher, *Statistical Analysis of Circular Data*, pp. 42,43.

<sup>2</sup>Edward Batschelet, *op. cit.*, pp. 275,276; Nicholas Fisher, *op. cit.*, pp. 39-41

2. La integral sobre el rango es igual a la probabilidad total:

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

La función de distribución  $F(\phi)$  correspondiente a  $f(\phi)$  se define sobre cualquier intervalo  $(\phi_1, \phi_2)$  por<sup>3</sup>:

$$F(\phi_2) - F(\phi_1) = \int_{\phi_1}^{\phi_2} f(\phi)d\phi$$

Sin embargo, resulta conveniente que  $\phi_1 = 0^\circ$  (ó  $-180^\circ$ ), entonces:

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

en particular:

$$F(2\pi) = 1$$

Una distribución circular puede ser considerada como una distribución bivariada sujeta a la restricción de que la probabilidad total está dispersa sobre el círculo unitario. Al igual que en la estadística bivariada ordinaria, existe un *vector medio* que apunta al *centro de masa*.

Para tener una idea más intuitiva de una distribución circular, es recomendable utilizar la llamada *gráfica circular*, donde la densidad  $f(\phi)$  es graficada en dirección radial en una circunferencia de radio unitario. Un ejemplo se muestra en la Figura 2.5. En la práctica resulta más claro utilizar una *gráfica lineal*, donde  $f(\phi)$  es representada como una coordenada rectangular. Sin embargo existe una desventaja: el punto inicial (sobre la línea recta) es arbitrario por lo que este hecho puede provocar alguna confusión (Figura 2.6).

### 2.3.1. Distribución Uniforme o Isotrópica

Con este modelo, todas las direcciones que se encuentran entre  $0^\circ$  y  $360^\circ$  tienen un valor constante en su función de probabilidad, por lo que la función de densidad es constante sobre la circunferencia del círculo unitario (Figura 2.3):

$$f(\phi) = \text{constante}, \quad (0 \leq \phi \leq 2\pi)$$

Si  $\phi$  está medido en radianes, entonces:

---

<sup>3</sup>Nicholas Fisher, *op. cit.*, 41

\* La función de densidad de probabilidad será:

$$f(\phi) = \frac{1}{2\pi}, \quad (0 \leq \phi < 2\pi)$$

\* La función de distribución será:

$$F(\phi) = \frac{\phi}{2\pi}, \quad (0 \leq \phi < 2\pi)$$

\* Momentos:

Dirección media	$\mu$ está indefinida
Longitud media resultante	$\rho = 0$
Dispersión circular	$\delta = \infty$
	$\alpha_p = 0, \quad p \geq 1$
	$\beta_p = 0, \quad p \geq 1$

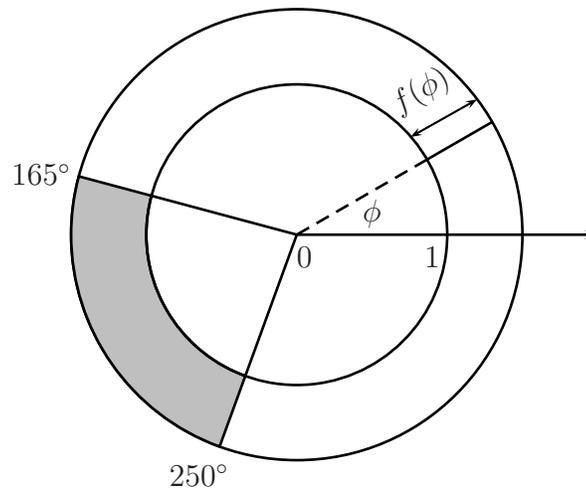


Figura 2.3: Gráfica circular de una distribución uniforme (la parte sombreada representa la probabilidad del sector que va de  $165^\circ$  a  $250^\circ$ ).

### 2.3.2. Distribución Cardioide

Es una distribución simétrica unimodal (Figura 2.4) con dos parámetros:  $\mu$  (dirección media) y  $\rho$  (longitud media resultante). En esta distribución, si  $\phi$  está medido en radianes entonces:

\* La función de densidad de probabilidad será:

$$f(\phi) = \frac{1}{2\pi} \{1 + 2\rho \cos(\phi - \mu)\}, \quad (0 \leq \phi < 2\pi, \quad 0 \leq \rho \leq 1/2)$$

\* La función de distribución será:

$$F(\phi) = \left(\frac{\rho}{\pi}\right) \sin(\phi - \mu) + \left(\frac{\phi}{2\pi}\right), \quad (0 \leq \phi \leq 2\pi)$$

\* Momentos:

Dirección media	$\mu$
Longitud media resultante	$\rho (\leq 1/2)$
Dispersión circular	$\delta = 1/(2\rho^2)$
	$\alpha_p = 0, \quad p \geq 2$
	$\beta_p = 0, \quad p \geq 1$

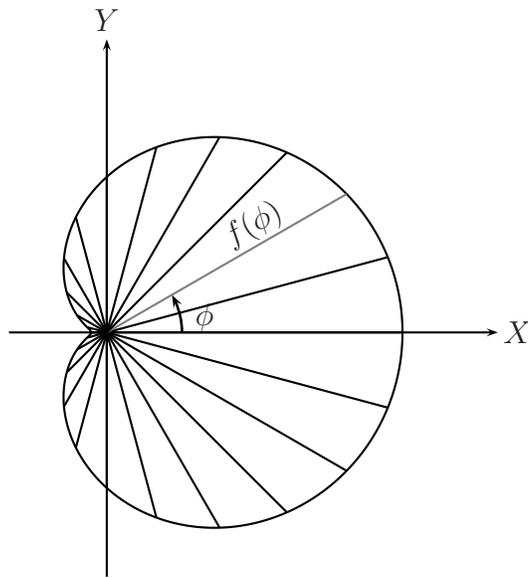


Figura 2.4: Gráfica circular de una distribución cardioide.

### 2.3.3. Distribución Envuelta

Si  $X$  es una variable aleatoria sobre la recta con función de densidad  $g(x)$  y función de distribución  $G(x)$ , entonces se puede obtener una variable aleatoria  $\Phi$  tal que:

$$\Phi \equiv X \pmod{2\pi}$$

es decir, se puede obtener una variable  $\Phi$  tal que  $\Phi - X$  resulte ser múltiplo de  $2\pi$ .

La función de densidad de probabilidad  $f(\phi)$  de  $\Phi$  se obtiene al “envolver”  $g(x)$  alrededor de la circunferencia de un círculo unitario.

\* La función de densidad de probabilidad para este modelo es:

$$f(\phi) = \sum_{k=0}^{\infty} g(\phi + 2k\pi)$$

\* La función de distribución correspondiente es:

$$F(\phi) = \sum_{k=-\infty}^{\infty} [G(\phi + 2k\pi) - G(2k\pi)]$$

Cabe mencionar que un número importante de distribuciones circulares son obtenidas con este método.

### 2.3.4. Distribución Cauchy Envuelta

Al igual que la distribución cardioide, la distribución Cauchy envuelta es simétrica unimodal con dos parámetros ( $\mu$  y  $\rho$ ). Ésta puede obtenerse al “envolver” la distribución Cauchy alrededor de la circunferencia.

\* Su función de densidad de probabilidad está dada por:

$$f(\phi) = \frac{1}{2\pi} \left( \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\phi - \mu)} \right), \quad (0 \leq \phi < 2\pi, \quad 0 \leq \rho \leq 1)$$

\* La función de distribución correspondiente es:

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

\* Momentos:

Dirección media	$\mu$
Longitud media resultante	$\rho$
Dispersión circular	$\delta = (1 - \rho^2)/(2\rho^2)$
	$\alpha_p = \rho^p$
	$\beta_p = 0, \quad p \geq 1$

### 2.3.5. Distribución Normal Envuelta

Es una distribución simétrica unimodal con dos parámetros ( $\mu$  y  $\rho$ ) que puede obtenerse al “envolver” la distribución Normal alrededor de la circunferencia.

\* Su función de densidad de probabilidad está dada por:

$$f(\phi) = \frac{1}{2\pi} \left( 1 + 2 \sum_{k=1}^{\infty} \rho^{k^2} \cos(k(\phi - \mu)) \right), \quad (0 \leq \phi < 2\pi, \quad 0 \leq \rho \leq 1)$$

\* Momentos:

Dirección media	$\mu$
Longitud media resultante	$\rho$
Dispersión circular	$\delta = (1 - \rho^4)/(2\rho^2)$
	$\alpha_p = \rho^{p^2}$
	$\beta_p = 0, \quad p \geq 1$

## 2.4. Distribución von Mises

Introducida en 1918, la distribución von Mises es una distribución simétrica unimodal; ésta juega un papel similar al de la distribución normal, por lo que también es llamada “distribución normal circular”. La distribución von Mises es el modelo más utilizado para el análisis de datos circulares (Figura 2.5).

\* Su función de densidad de probabilidad está dada por:

$$f(\phi) = \frac{1}{2\pi I_0(k)} \exp[k \cos(\phi - \mu)] \quad (0 \leq \phi < 2\pi, \quad 0 \leq k < \infty)$$

donde:

$$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} \exp[k \cos(\varphi - \mu)] d\varphi,$$

$k$  es el parámetro de concentración y

$\mu$  es el ángulo medio ó dirección media (como ya se ha visto).

\* La función de distribución correspondiente es:

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

\* Momentos:

Dirección media	$\mu$
Longitud media resultante	$\rho = A_1(k)$
Dispersión circular	$\delta = [kA_1(k)]^{-1}$
	$\alpha_p = A_p(k)$
	$\beta_p = 0, \quad p \geq 1$

donde:

$$A_p(k) = I_p(k)/I_0(k), \quad p = 0, 1, 2, 3, \dots$$

entonces:

$$A_1(k) = I_1(k)/I_0(k)$$

Para lo anterior, reescribimos la expresión para  $I_0(k)$ :

$$I_0(k) = \sum_{r=0}^{\infty} (r!)^{-2} \left(\frac{1}{2}k\right)^{2r}$$

en general sucede que:

$$I_p(k) = \sum_{r=0}^{\infty} [(r+p)!r!]^{-1} \left(\frac{1}{2}k\right)^{2r+p} \quad p = 1, 2, 3, \dots$$

Cabe mencionar que la distribución von Mises es simétrica con respecto al ángulo medio ( $\mu$ ). Ahora bien, si  $k \rightarrow 0$ , la distribución von Mises converge a la distribución uniforme; por otro lado, si  $k \rightarrow \infty$ , la distribución tiende a concentrarse alrededor del ángulo promedio  $\mu$  (Figura 2.6).

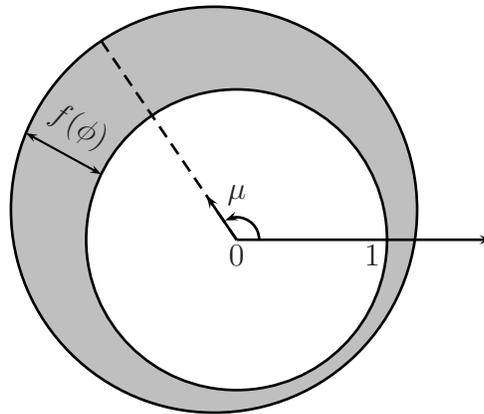


Figura 2.5: Gráfica circular de la distribución von Mises. Se muestra el ángulo medio  $\mu$ , con  $k \neq 0$ .

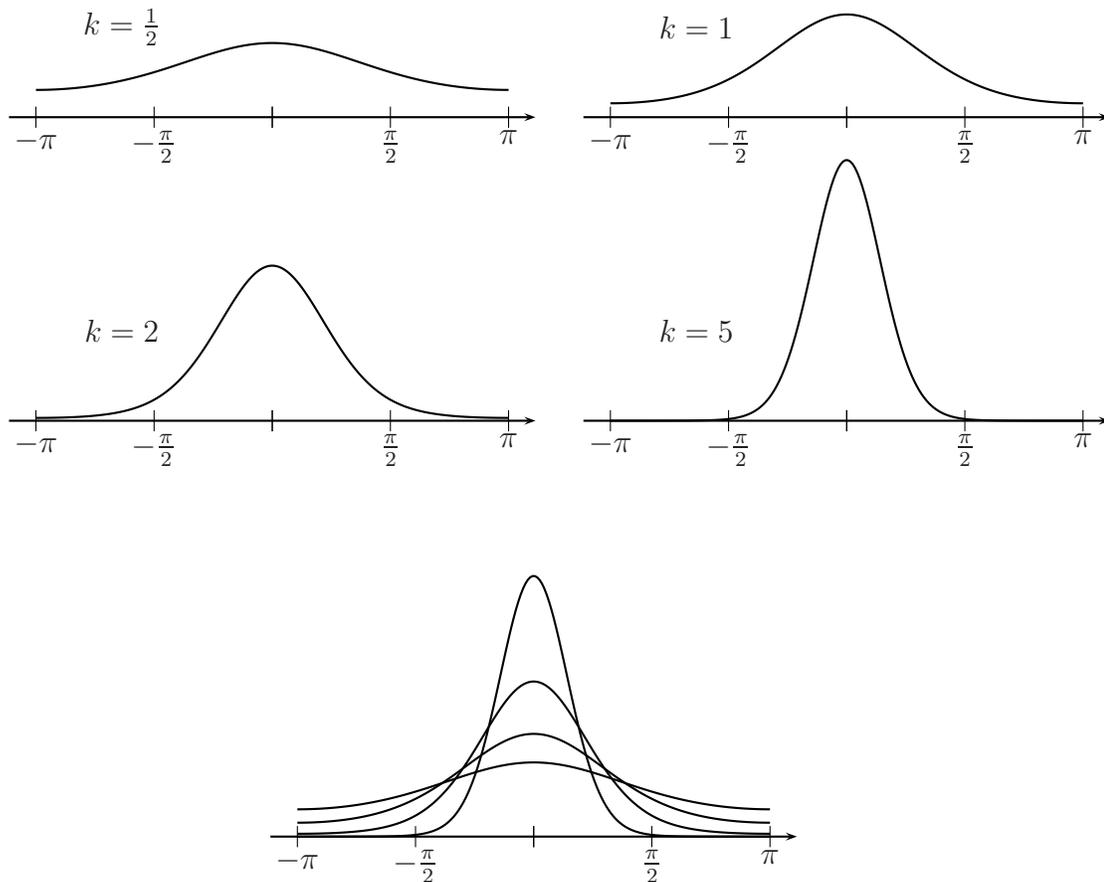


Figura 2.6: Gráficas lineales de la distribución von Mises con parámetros de concentración  $k = \frac{1}{2}, 1, 2$  y 5. Por simplicidad, el ángulo medio es igual a cero. La última gráfica muestra las cuatro anteriores, con la finalidad de poder compararlas mejor<sup>4</sup>.

---

<sup>4</sup>Fuente: Edward Batschelet, *op. cit.*, p. 280.

# Capítulo 3

## Análisis de Datos Circulares con Estructura Temporal

En el **Capítulo 1** se estudiaron las nociones básicas para analizar series de tiempo lineales, éstos conceptos serán útiles para el tratamiento de series de tiempo circulares, pues al igual que los datos lineales, los datos circulares pueden presentar una estructura temporal (Figura 3.1); a pesar de que los métodos para estudiar estos últimos no son los mismos que para los primeros, las bases del análisis son las mismas.

En este capítulo se considerará el problema de modelar y analizar series de tiempo de medidas angulares, para ello partiremos de la hipótesis de que los datos son observaciones circulares  $(\theta_1, \theta_2, \dots, \theta_n)$  en relación a una secuencia de tiempo  $(t_1, t_2, t_3, \dots)$ , donde la mayoría de las veces  $t_i$  y  $t_{i+1}$  son equidistantes.

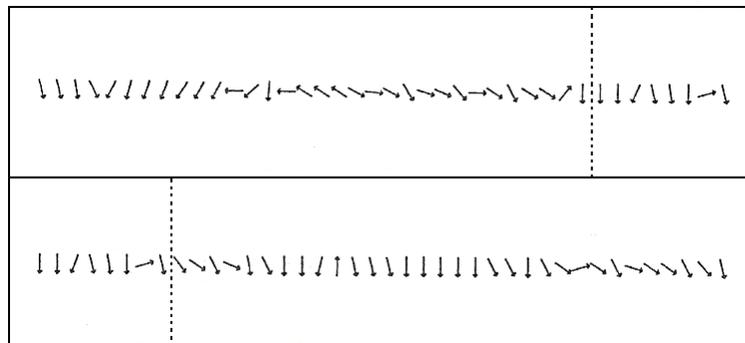


Figura 3.1: Serie de tiempo de 72 direcciones de viento<sup>1</sup>, que comprende mediciones por hora durante tres días en un sitio de Australia.

---

<sup>1</sup>Fuente: Nicholas Fisher, pp. 181,254. Véase Apéndice A, Tabla 1.

El inicio del análisis de una serie de tiempo es realizar un *análisis exploratorio*, el cual consiste en extraer y exhibir la estructura básica de los datos. Para lo anterior se utilizan dos tipos de métodos:

*Tendencias:* Sea  $x_i = \cos(\theta_i)$ ,  $y_i = \sin(\theta_i)$ ,  $i = 1, 2, \dots, n$ . El procedimiento más común para suavizar la tendencia dada una secuencia de datos, es suavizar la  $x$ - $y$ - $y$ -*secuencia* (por separado), y después combinarlas en una secuencia de medidas circulares. Existen numerosos métodos para suavizar una secuencia de medidas lineales  $x_1, x_2, \dots$ , que, aunque específicamente están diseñados para datos independientes, también son útiles para “explorar” secuencias correlacionadas. Una clase simple de métodos comprende a los llamados métodos *kernel*. El kernel estimado de  $x(t)$  (o  $y(t)$ ) en un punto  $t$  es un promedio ponderado de los valores de  $x_i$  (o de  $y_i$ ) al tiempo  $t_i$  cercano a  $t$ . Específicamente:

$$\hat{x}(t) = \frac{\sum_{i=1}^n K_h(t - t_i)x_i}{\sum_{i=1}^n K_h(t - t_i)} \quad \text{y} \quad \hat{y}(t) = \frac{\sum_{i=1}^n K_h(t - t_i)y_i}{\sum_{i=1}^n K_h(t - t_i)}$$

donde:

$$K_h(t) = h^{-1}K(t/h)$$

Generalmente,  $h$  depende del tamaño de la muestra ( $n$ ).

Por lo tanto, la suavidad estimada para  $\phi(t)$  es:

$$\hat{\theta}(t) = \tan^{-1}\{\hat{y}(t)/\hat{x}(t)\}$$

*Estructura de dependencia de series:* Si la sucesión de datos  $\{\theta_i\}$  constituye una serie de tiempo, resulta de interés examinar la estructura de dependencia después de haber removido la tendencia de los datos. Para ello se mostrará un método análogo para series de tiempo lineales, utilizando el coeficiente de correlación circular  $\rho_T$ . Para cualquier entero  $k = 1, 2, \dots$ , se define la muestra del coeficiente de correlación  $k$ -retraso (o *lag- $k$* ) circular de  $\{\theta_i\}$  por el valor  $\hat{\rho}_{T,\kappa}$ , es decir, el valor de  $\hat{\rho}_T$  calculado por parejas:

$$(\theta_1, \theta_{k+1}), \dots, (\theta_{n-k}, \theta_n)$$

Así, un correlograma circular puede obtenerse al graficar los puntos  $(k, \hat{\rho}_{T,\kappa})$ , para  $k = 1, 2, \dots$ . Este correlograma puede ser interpretado de una manera similar a un correlograma lineal (Figura 3.2).

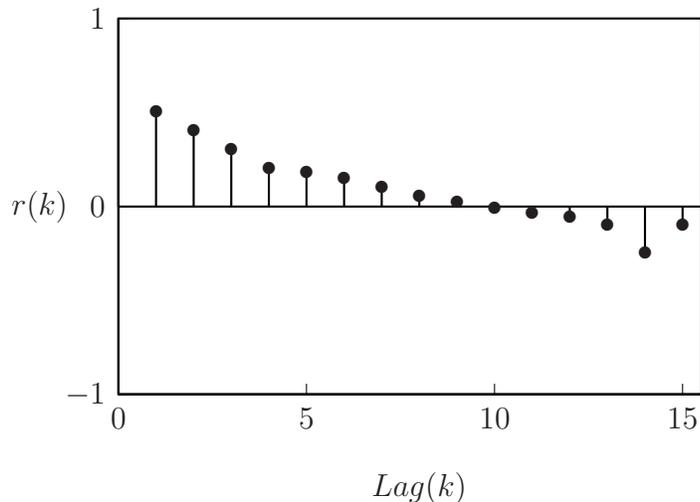


Figura 3.2: Correlograma circular para la serie de tiempo de las direcciones de viento de la Figura 3.1.

### 3.1. Algunos Modelos para Series de Tiempo de Datos Circulares

Opuesto a lo que sucede con las series de tiempo lineales, los métodos para analizar las series de tiempo de datos circulares no han recibido mucha atención. Los métodos que se han propuesto se pueden clasificar en dos grandes grupos, dependiendo de qué tipo de modelo subyacente es el caso: *procesos de Markov* o *procesos alrededor del círculo* (modelos basados en una “envoltura” lineal autoregresiva (AR)).

Haber llevado a cabo la exploración inicial de una serie temporal, a veces es de interés para buscar un modelo paramétrico que se inscribe en la serie. Si ese modelo se puede encontrar, entonces podemos predecir valores futuros de la serie. En seguida se describen cuatro modelos para series de tiempo circulares, así como la autocorrelación circular para cada uno.

#### 3.1.1. Procesos Gaussianos Projectados

Sean  $\{X_t\}$  y  $\{Y_t\}$  dos eventos independientes del mismo proceso Gaussiano, con media estacionaria igual a 0 y varianza 1. Si  $(R_t, \Theta_t)$  es la representación del punto  $(X_t, Y_t)$  en coordenadas polares, entonces  $\{\Theta_t\}$  es una serie de tiempo circular estacionaria con la propiedad de que su distribución marginal es **uniforme**. Si  $\{X_t\}$  y  $\{Y_t\}$  no tienen media igual a 0 entonces el resultado del proceso proyectado no tiene marginales uniformes sino marginales **normales**.

Para estudiar la correlación de  $\{\Phi_t\}$ , se requiere medir la correlación entre dos

variables aleatorias circulares  $\Theta$  y  $\Phi$ .

Sean  $(\Theta_1, \Phi_1)$  y  $(\Theta_2, \Phi_2)$  eventos independientes de  $(\Theta, \Phi)$ , entonces la correlación circular esta dada por<sup>2</sup>:

$$\rho_T = \frac{\mathbb{E}[\text{sen}(\Theta_1 - \Theta_2)\text{sen}(\Phi_1 - \Phi_2)]}{\{\mathbb{E}[\text{sen}(\Theta_1 - \Theta_2)^2] \cdot \mathbb{E}[\text{sen}(\Phi_1 - \Phi_2)^2]\}^{1/2}} \quad (3.1)$$

o bien:

$$\rho_T = \frac{\sum_{1 \leq i < j \leq n} \text{sen}(\Theta_1 - \Theta_2)^2 \cdot \text{sen}(\Phi_1 - \Phi_2)^2}{\{\sum \text{sen}(\Theta_1 - \Theta_2)^2 \cdot \sum \text{sen}(\Phi_1 - \Phi_2)^2\}^{1/2}} \quad (3.2)$$

La medida  $\rho_T$  tiene las siguientes propiedades:

- i)  $\rho_T \in [-1, 1]$
- ii)  $\rho_T = 0 \Leftrightarrow \Theta$  y  $\Phi$  son independientes
- iii)  $\rho_T = 1 \Leftrightarrow \Theta$  y  $\Phi$  guardan la siguiente relación (*asociación positiva*):

$$\Theta = \Phi + \alpha_0 \quad [\text{mod } 2\pi]$$

- iv)  $\rho_T = -1 \Leftrightarrow \Theta$  y  $\Phi$  guardan la siguiente relación (*asociación negativa*):

$$\Theta = -\Phi + \alpha_0 \quad [\text{mod } 2\pi]$$

- v)  $\rho_T$  es invariante bajo la elección de origen de  $\Theta$  y de  $\Phi$ , y la reflexión de  $\Theta$  sobre  $\Phi$  cambia el signo de  $\rho_T$  pero no su magnitud.
- vi) Si las distribuciones de  $\Theta$  y de  $\Phi$  son unimodales cada una y altamente concentradas, entonces:

$$\rho_T(\Theta, \Phi) \simeq \rho(\Theta, \Phi)$$

### 3.1.2. Procesos “Envueltos”

Sea  $\{X_t\}$  una serie de tiempo lineal de observaciones univariadas. La serie de tiempo “*envuelta*” circular correspondiente ( $\{\theta_t\}$ ) está dada por:

$$\theta_t = X_t \quad [\text{mod } 2\pi]$$

---

<sup>2</sup>Definición introducida por Fisher y Lee (1983).

Así,  $X_t$  puede ser descompuesta como:

$$X_t = \theta_t + 2\pi k_t$$

donde  $k_t$  es un entero.

Si  $(X, Y)$  tienen una distribución normal bivariada con varianzas  $\sigma_x^2$  y  $\sigma_y^2$ , respectivamente, y una correlación igual a  $\rho$ , entonces la correlación circular  $\rho_T$  de  $\theta_t = X_t \pmod{2\pi}$  es:

$$\rho_T = \frac{\sinh(2\rho\sigma_x\sigma_y)}{\{\sinh(2\sigma_x^2)\sinh(\sigma_y^2)\}^{1/2}}$$

Por lo tanto, si  $\{X_t\}$  es un proceso  $AR(p)$ , la función de autocorrelación circular del  $\theta_t$ -proceso  $WAR(p)$ <sup>3</sup> está dada por:

$$\rho_T(k) = \frac{\sinh\{2\rho_k\sigma^2/(1 - \phi_1\rho_1 - \dots - \phi_p\rho_p)\}}{\{\sinh\{2\sigma^2/(1 - \phi_1\rho_1 - \dots - \phi_p\rho_p)\}}$$

donde  $\rho_k$  es el  $k$ -ésimo-retraso de autocorrelación de  $\{X_t\}$  y,  $\sigma^2, \phi_1, \dots, \phi_p$  son sus  $AR(p)$  parámetros.

### 3.1.3. Procesos Derivados Utilizando Funciones de Enlace

Dentro de estos procesos encontramos dos modelos dependiendo de la variación de la serie. Para series con variación alta se utilizan modelos adecuados para datos circulares, mientras que para aquellas series con menor variación se utiliza un método que consiste en transformar los datos circulares a datos lineales, y así aprovechar la riqueza de métodos y softwares disponibles para series de tiempo lineales.

Sea  $g(x)$  una función impar monótona en el intervalo  $(-\pi, \pi)$  tal que  $g(x) = 0$ . Por ejemplo la siguiente función:

$$g_T(x) = 2 \tan^{-1}(x)$$

utilizada por Downs (1986) para formular un modelo de regresión para un ángulo de respuesta y un ángulo de covarianza.

Fisher y Lee (1992) formularon una clase general de modelos de regresión para ángulos de respuesta aleatorios y covarianzas lineales mediante un “enlace”  $g$ . Para dichas funciones de enlace  $g$ , si  $X$  es una variable aleatoria real, entonces:

$$\theta = g(x) \text{ es una variable aleatoria circular y,}$$

---

<sup>3</sup> $WAR(p)$ : es un proceso  $AR(p)$  Gaussiano envuelto.

análogamente, si  $\theta$  es una variable aleatoria circular

$$X = g^{-1}(\theta) \text{ es una variable aleatoria real.}$$

Usando las transformaciones anteriores, decimos que un proceso estacionario circular  $\{\theta_t\}$  con dirección media  $\mu$  es un *proceso con media móvil autoregresivo enlazado* (conocido como proceso *LARMA*, por sus siglas en inglés) si su proceso lineal enlazado  $g^{-1}(\theta_t - \mu)$  es un *proceso con media móvil autoregresivo*  $(p, q)$  (conocido como proceso *ARMA* $(p, q)$ , por sus siglas en inglés)<sup>4</sup>.

Supongamos que  $\theta_t$  es un proceso *LARMA* $(p, q)$  con dirección media  $\mu$ , entonces  $X_t = g^{-1}(\theta_t - \mu)$  es un proceso *ARMA* $(p, q)$ , y la autocorrelación circular de  $\theta_t$  es:

$$\rho_T(k) = \rho_T\{g(X_t), g(X_{t+k})\}$$

donde  $\rho_T$  denota el coeficiente de correlación circular de la ecuación (3.1). En la Figura 3.3 se muestran distintos autocorrelogramas para el caso de *LAR*(1) (todas las clases de modelos tienen un patrón similar en la función de autocorrelación, y estos patrones son similares a los del proceso lineal *AR*(1)).

Para un proceso *AR*, un modelo alternativo puede ser definido usando distribuciones condicionales.

Por lo tanto, para el modelo *AR*( $p$ ) se tienen dos formas circulares:

1. La *forma directa* o “*enlazada*” *LAR*( $p$ ), donde los ángulos transformados siguen un modelo lineal ordinario *AR*( $p$ ). Las ventajas de este modelo son dos:
  - a) Si el proceso  $\{\theta_t\}$  es estacionario, entonces también lo será el proceso resultante  $X_t$ .
  - b) Los parámetros son fáciles de estimar usando cualquier software estándar.
2. La *forma inversa*, donde las distribuciones condicionales tienen medias condicionales (ecuación 1.7, Capítulo 1). Al igual que el modelo anterior, éste tiene dos ventajas:
  - a) Permite la especificación directa de la estructura de los errores circulares.
  - b) Es adecuado para la dispersión de datos.

## 3.2. Selección del Modelo e Identificación

Como se estudió en el **Capítulo 1**, al ajustar un modelo a una serie de tiempo lineal se suele utilizar una sola familia de modelos de media móvil autoregresivos (es

---

<sup>4</sup>Véase Sección 1.9.

decir, aquéllos con errores Gaussianos), lo mismo sucede para el caso de series de tiempo circulares, donde por lo general se utilizan cuatro familias de modelos<sup>5</sup>.

En la sección anterior se discutieron los medios para realizar una elección adecuada de uno de estos modelos. Ahora, se considerará la forma de elegir un modelo en particular dentro de una determinada clase.

## Selección del Modelo

Debido a que las cuatro clases de modelos tienen patrones similares en la correlación (como lo muestra la Figura 3.3) y son similares en forma al patrón de descenso exponencial en los modelos lineales  $AR(1)$ , al llevar a cabo una elección de la clase de modelo, se deben inspeccionar otras características de la serie de tiempo.

*El modelo de proyección* se caracteriza por tener marginales uniformes. Si los procesos lineales que se han proyectado presentan una correlación alta, se prevé que el proceso tenderá a ocupar arcos no fijos en el círculo. Para proyecciones de baja correlación, el proceso circular se moverá sin dirección determinada alrededor del círculo rápidamente.

*Para el proceso normal envuelta*, tanto la correlación como la varianza del proceso lineal envuelto influyen en la forma del proceso circular. Para valores muy grandes de la varianza, el proceso circular tenderá a extenderse de manera uniforme alrededor del círculo, para datos de este tipo tanto el modelo de proyección como el modelo normal envuelta serán apropiados. Por otro lado, para varianzas pequeñas, el proceso circular envuelto tenderá a ocupar un arco del círculo, en lugar de extenderse de manera uniforme en la circunferencia.

*El proceso inverso* puede tener un comportamiento similar. Cuanto más pequeño es el valor del parámetro  $\kappa$ , habrá una extensión de los datos de manera más uniforme alrededor del círculo. Por otro lado, cuanto mayor sea el valor de  $\alpha$  (de la ecuación 1.7) mayor será la tendencia a permanecer, por períodos cortos de tiempo, en un pequeño arco sobre  $\mu$ .

*Finalmente, el proceso enlazado* se caracteriza por una brecha que dejan los datos al distribuirse alrededor del círculo, causada por el hecho de que valores muy grandes del proceso subyacente pueden ser proyectados en un arco alrededor de  $\mu + \pi$ .

Tomando en cuenta las observaciones anteriores, podemos seguir las siguientes recomendaciones:

1. Para datos que se distribuyen uniformemente alrededor del círculo, tanto el modelo de proyección como el modelo normal envuelta serán apropiados.

---

<sup>5</sup>Nicholas Fisher y J. Lee, *op. cit.*, pp. 327,327.

2. En contraste, si los datos tienden a agruparse en un arco, los métodos inversos y enlazados serán los más apropiados. Además, si se observa una brecha en los datos, el modelo directo es el adecuado.
3. No hay que olvidar que los cuatro modelos son muy flexibles y capaces de capturar una amplia gama de comportamientos.

## Identificación del Modelo

Al igual que en el análisis de series de tiempo lineales, la selección de un modelo apropiado para series de tiempo circulares se realiza con ayuda del correlograma.

Para los modelos inverso, normal envuelta, y proyectado, el correlograma muestral de la serie se puede calcular, de una manera eficiente, utilizando una expresión alternativa para la autocorrelación muestral:

$$\hat{\rho}_T(k) = \frac{\det \left( \sum_{t=1}^{T-k} X_t X'_{t+k} \right)}{\left\{ \det \left( \sum_{t=1}^{T-k} X_t X'_{t+k} \right) \det \left( \sum_{t=k+1}^T X_t X'_t \right) \right\}^{1/2}} \quad (3.3)$$

donde  $X_t = (\cos(\theta_t), \sin(\theta_t))'$ .

En el modelo de proyección, podemos obtener una estimación aproximada de la función de autocorrelación  $\rho(k)$  de los procesos lineales resolviendo la ecuación 3.4 para  $\rho(k)$  usando la autocorrelación muestral (ecuación 3.3) en lugar de  $\rho_T(k)$ .

$$\rho_T(k) = \frac{\pi^2}{16} \rho^2(k) \{1 - \rho^2(k)\}^2 \left\{ {}_2F_1 \left( \frac{3}{2}, \frac{3}{2}; 2; \rho^2(k) \right) \right\}^2 \quad (3.4)$$

donde  ${}_2F_1$  es la notación usual para la función hipergeométrica<sup>6</sup>.

Para estimar la autocorrelación del proceso envuelto en el modelo normal envuelta calculamos:

$$\hat{\rho}_T(k) = \frac{\sinh(2c_0 \hat{\rho}_k)}{\sinh(2c_0)}$$

donde  $c_0$  es la varianza, (*i.e.*, la autocorrelación al retraso 0) del proceso envuelto. Ésta puede ser estimada de la relación  $\bar{R} = \exp(-c_0/2)$  donde  $\bar{R}$  es la longitud media resultante de  $\{\theta_t\}$ .

Para identificar el modelo directo, una estimación preliminar  $\hat{\mu}$  de  $\mu$  se realiza, por ejemplo, calculando la dirección media de la serie circular. Así, la serie transformada  $X_t = g^{-1}(\Theta_t - \hat{\mu})$  es identificada de forma habitual.

---

<sup>6</sup>Veáse Apéndice A.

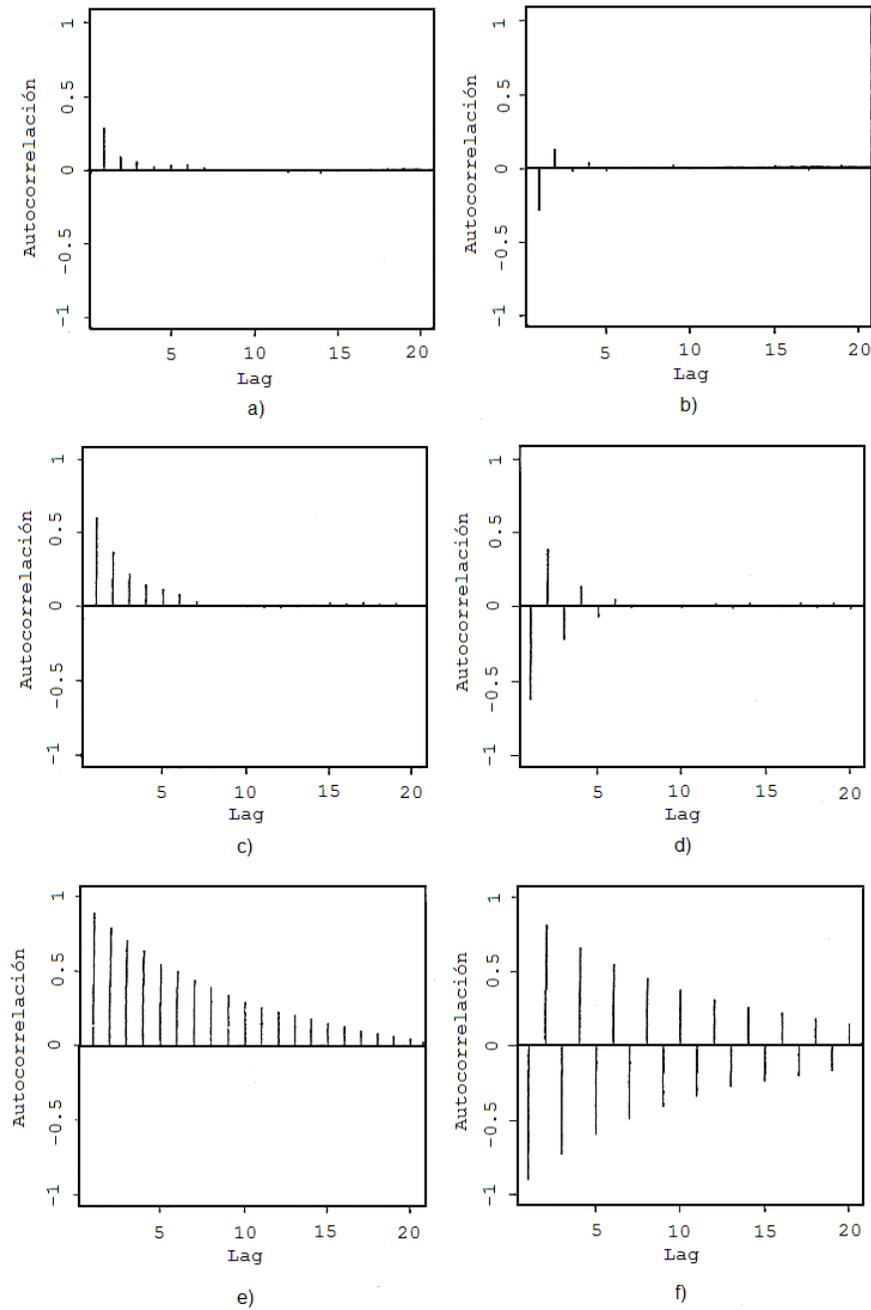


Figura 3.3: Prototipo de funciones de autocorrelación<sup>7</sup> para un modelo  $LAR(1)$  utilizando una función de enlace probit (caracterizada en la Sección 1.9.6): a)  $\phi = 0.3$ ,  $\sigma = 0.2$ ; b)  $\phi = -0.3$ ,  $\sigma = 0.2$ ; c)  $\phi = 0.6$ ,  $\sigma = 0.2$ ; d)  $\phi = -0.6$ ,  $\sigma = 0.2$ ; e)  $\phi = 0.9$ ,  $\sigma = 0.2$ ; f)  $\phi = -0.9$ ,  $\sigma = 0.2$ .

<sup>7</sup>Fuente: Nicholas Fisher y J. Lee, *op. cit.*, p. 330.

# Capítulo 4

## Aplicación

En este capítulo se presenta una aplicación de la estadística circular en datos circulares con estructura temporal. Se realizaron 70 mediciones de la dirección del viento en la Ciudad de Guadalajara, tomando las observaciones cada hora<sup>1</sup>. El primer paso a seguir en el análisis de una serie de tiempo, como ya se ha mencionado en el **Capítulo 1**, es graficar la serie para observar el comportamiento de las mediciones. La Figura 4.1 muestra la gráfica correspondiente a la serie de tiempo.

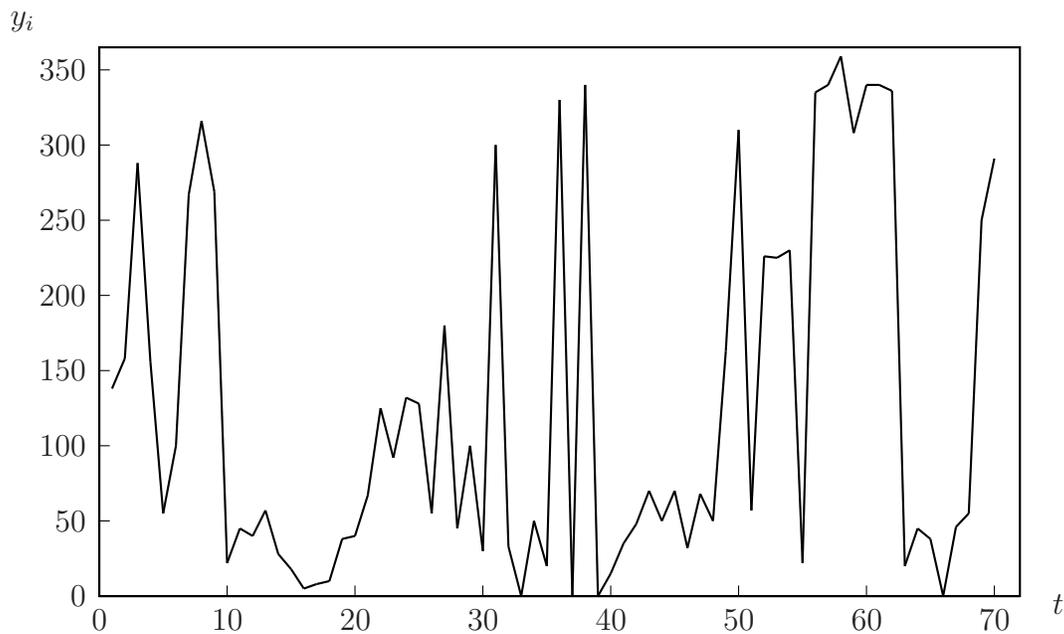


Figura 4.1: Gráfica de la serie de tiempo de 70 mediciones de la dirección del viento ( $^{\circ}\text{C}$ ), tomadas cada hora, en la Ciudad de Guadalajara.

---

<sup>1</sup>Veáse Apéndice B.

Debido a que se trata de observaciones tomadas cada una en un lapso pequeño de tiempo, es muy probable que exista cierta tendencia en los datos. Para remover la tendencia que pudiera haber en la serie se realizó una *Diferenciación de orden 1*, obteniéndose la gráfica 4.2.

Por la Figura 4.1 se concluye que la serie presenta baja tendencia, por lo que una vez removida y de acuerdo con el capítulo anterior<sup>2</sup>, se encuentra la correlación utilizando una función enlace para transformar los datos circulares en datos lineales.

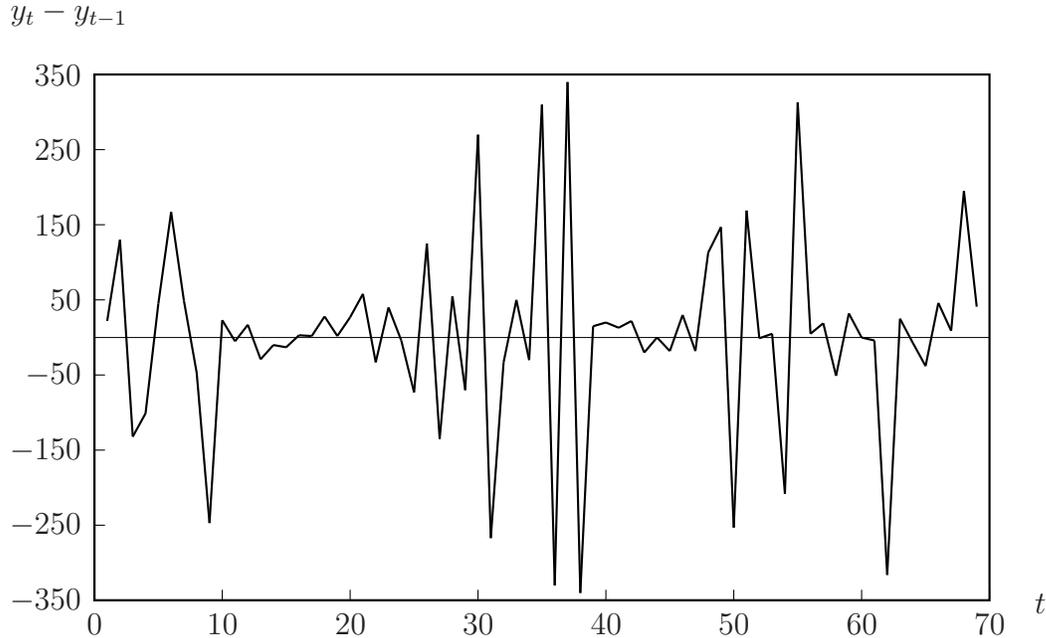


Figura 4.2: Gráfica que muestra la serie de tiempo de las 70 mediciones de la dirección del viento, después de aplicarle el método de diferenciación de orden 1.

Primero, se convierten los datos  $y_t - y_{t-1} = Y_t$  (del Apéndice B) a radianes, calculando:

$$\theta_{rad} = \frac{Y_t \cdot \pi}{180} \quad \forall i = 1, 2, \dots, 69; \quad \pi = 3.1415\dots$$

Se eligió la función:

$$g(\theta) = 4 \tan^{-1}(\theta) \quad (4.1)$$

donde  $\theta = \theta_{rad}$ .

---

<sup>2</sup>Veáse Sección 3.1.3.

Por lo tanto, para transformar los datos (circulares) a datos “lineales”, se utilizó la función:

$$g^{-1}(\theta) = \tan\left(\frac{\theta}{4}\right) = X$$

donde  $\theta = \theta_{rad}$  y  $X$  es una variable aleatoria real<sup>3</sup>.

Una vez realizados los cálculos anteriores se graficó el correlograma correspondiente; para ello, se utilizó como guía lo estudiado en la Sección 1.8, puesto que ya se realizó la transformación de los datos deseada.

Se calculó:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad \forall k = 1, 2, \dots, 11$$

En seguida se muestra el correlograma resultante:

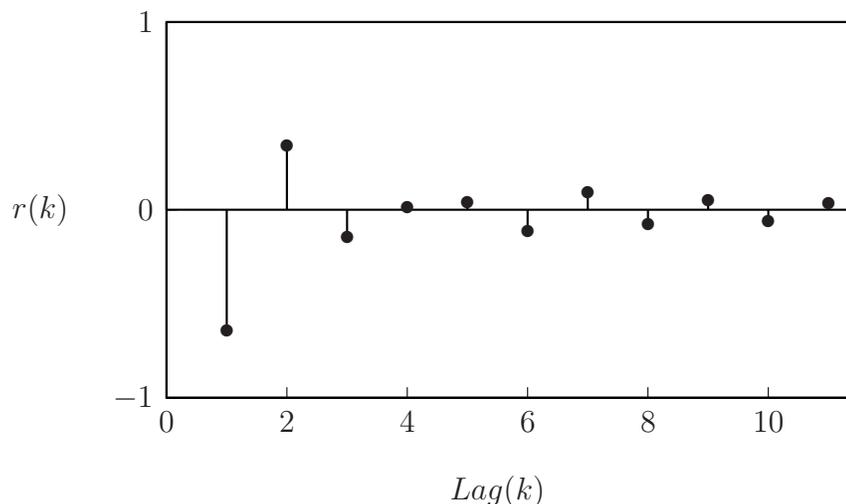


Figura 4.3: Correlograma correspondiente a la serie de tiempo de la Figura 4.1

Obsérvese que el correlograma sigue un patrón semejante al inciso d) de la Figura 3.3. Por lo tanto se ha conseguido identificar el modelo que mejor se ajusta a nuestros datos; se trata de un modelo  $LAR(1)$  con  $\phi = -0.6$  y  $\sigma = 0.2$ . En la gráfica 4.4 se muestra el correlograma estimado y el teórico.

Debido a que el modelo se estimó con base en un correlograma perteneciente a otra serie, es posible que  $\sigma$  sea mayor, esto también explica el porqué se observan algunos residuales grandes<sup>4</sup>.

<sup>3</sup>Veáse Apéndice B.

<sup>4</sup>Veáse Apéndice D.

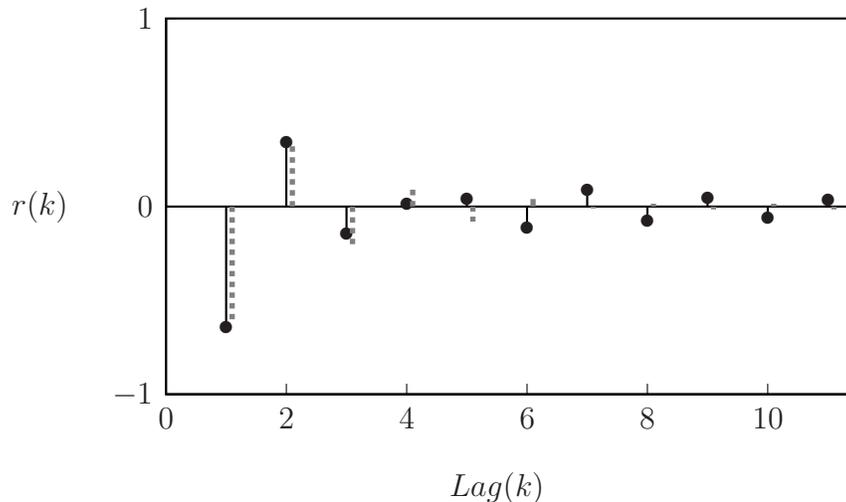


Figura 4.4: Correlogramma estimado y correlogramma teórico. Las líneas punteadas representan el correlogramma estimado.

De acuerdo a la Sección 3.1.3. se tiene que el proceso  $LAR(1)$  sigue un patrón similar al del proceso  $AR(1)$ , donde:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \epsilon_t \\ &= -0.6 X_{t-1} \quad \forall t = 2, 3, \dots, 68 \end{aligned}$$

obteniéndose los datos del Apéndice C y la gráfica 4.5.

Ahora, con el fin de convertir los datos ajustados a datos circulares, se transforman nuevamente los datos  $X_t$  del Apéndice C) mediante la función (4.1)<sup>5</sup>. Obsérvese que, una vez hecha la transformación, los datos resultantes están expresados en radianes, por lo que habrá que convertirlos a grados calculando:

$$X_t = \frac{\theta_{rad} \cdot 180}{\pi} = X'_t \quad \forall t = 2, \dots, 68; \quad \pi = 3.1415\dots$$

La última columna del Apéndice C ( $X'_t$ ) muestra los datos circulares ajustados a la serie de tiempo  $Y_t$  obtenida de la diferenciación<sup>6</sup>. Hay que recordar que la diferenciación se realizó de la siguiente manera:

$$\begin{aligned} Y_t &= y_t - y_{t-1} \\ \Rightarrow y_t &= y_{t-1} + Y_t \end{aligned} \tag{4.2}$$

<sup>5</sup>Veáse  $X_t$  transformada, Apéndice C.

<sup>6</sup>Veáse Apéndice B.

Entonces se puede obtener la serie circular ajustada mediante la ecuación (4.2).

$$\therefore x'_t = y_{t-1} + X'_t$$

donde  $y_{t-1}$  es la  $(t - 1)$ -ésima observación de la dirección del viento (en °C) y  $x'_2, x'_3, \dots, x'_{69}$  es la serie ajustada a los datos originales<sup>7</sup>, dicha serie se puede observar en la gráfica 4.6.

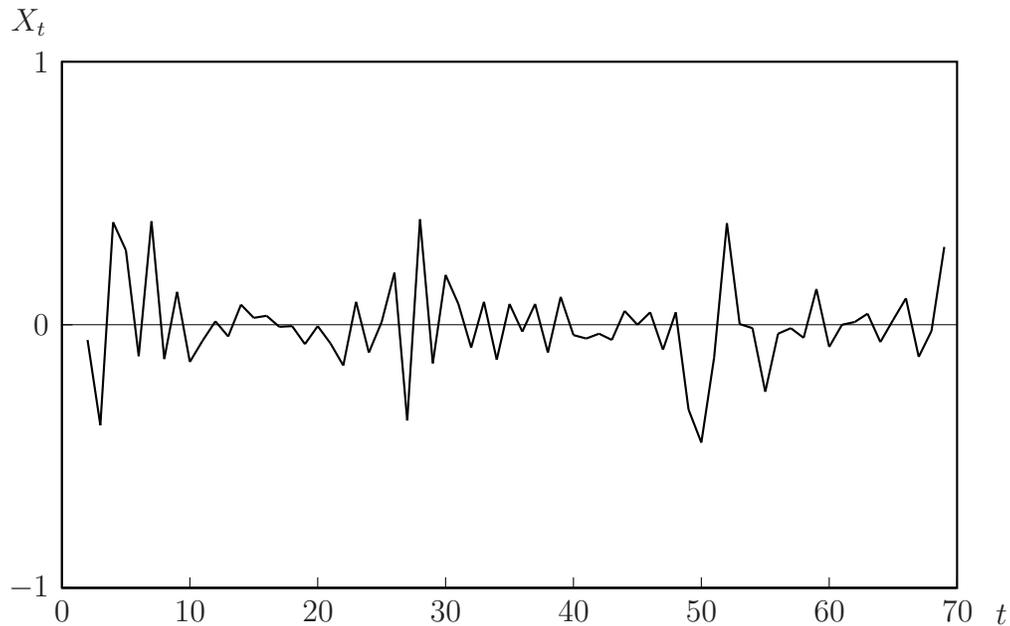


Figura 4.5: Gráfica que muestra la serie de tiempo del modelo  $LAR(1)$  con  $\phi = -0.6$  ajustada a los datos lineales del Apéndice B.

Por último, en la gráfica 4.7 se muestran las dos series de tiempo con el fin de poder compararlas mejor. La gráfica punteada muestra la serie de tiempo de los datos de la dirección del viento, mientras que la gráfica de línea continua muestra la serie ajustada.

---

<sup>7</sup>Veáse Apéndice D.

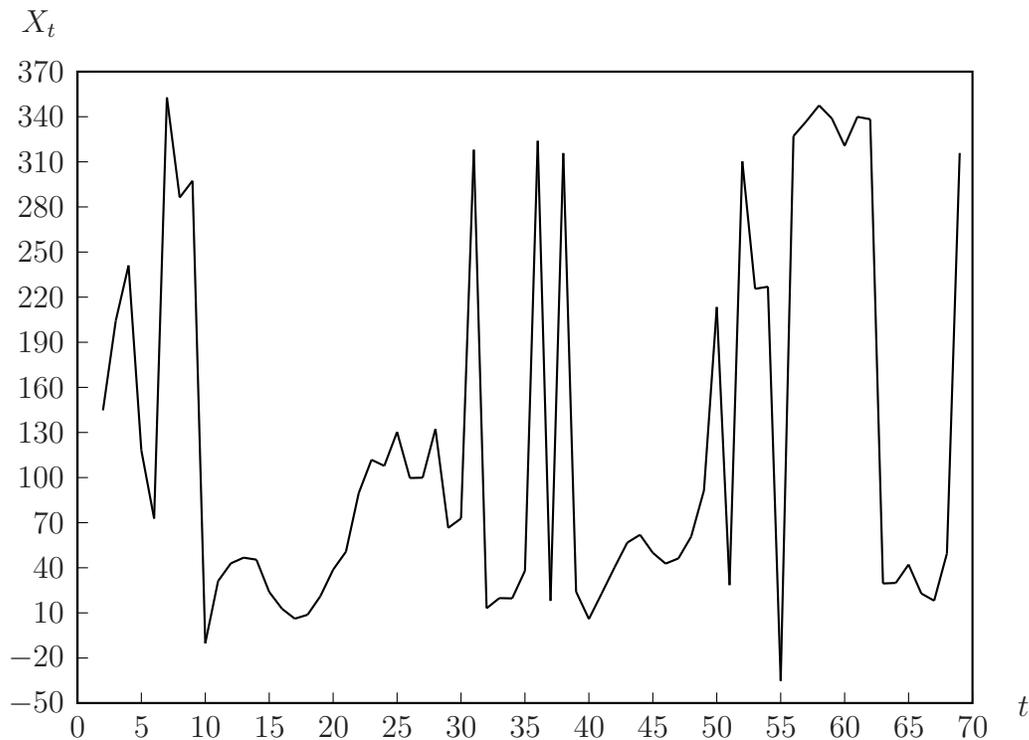


Figura 4.6: Serie de tiempo del modelo  $LAR(1)$  con  $\phi = -0.6$ , la cual ajusta los datos circulares originales.

Con el propósito de mostrar que se ha realizado un buen ajuste, se ajustará un modelo más grande ( $AR(2)$ ) con ayuda del software SAS, y en seguida se argumentará el porqué es mejor el modelo  $AR(1)$ .

Para obtener tanto  $\phi_1$  como  $\phi_2$  para el nuevo modelo, utilizaremos la última columna del Apéndice B ( $X_i$ ), puesto que queremos ajustar un modelo  $AR(2)$  a datos lineales. Para ello se escribe el siguiente código dentro de la ventana “Editor” (en SAS):

```
proc arima data=serie;
  identify var=xi nlag=12;
  estimate P=2 Q=0 noconstant method=ml
  outmodel=modelo
  run;
  quit;
```

donde  $data=serie$  indica la base donde se encuentran los datos, en este caso “*serie*” es la tabla del Apéndice B;  $var=xi$  indica la columna que contiene los datos a los cuales se les ajustará el modelo, en este caso “*xi*” corresponde a la columna  $X_i$  del Apéndice B;  $estimate P=2 Q=0 method=ml$  indica que se trata de un modelo

$AR(2)$ , por el método de máxima verosimilitud; finalmente, *outmodel=modelo* indica el nombre de la tabla donde se guardarán los resultados una vez compilado, en dicha tabla se registra el número de parámetros estimados, el valor de cada parámetros, el número de residuales, la varianza de la estimación y la varianza de cada parámetro.

Una vez compilado el código anterior, se obtuvieron los siguientes resultados:

$$\begin{array}{lll} \phi_1 = -0.728 & \text{y} & \phi_2 = -0.138 \\ \sigma_1 = 0.12 & \text{y} & \sigma_2 = 0,12 \end{array}$$

Por lo tanto, el modelo  $AR(2)$  ajustado queda expresado como:

$$x'_i = -0.728x_{i-1} + 0.138x_{i-2}$$

Después de realizar los cálculos necesarios como se hizo para el modelo  $AR(1)$ , se obtiene la serie ajustada a los datos circulares <sup>8</sup>.

Por último, se realiza la siguiente prueba de hipótesis sobre el parámetro  $\phi_2$ :

$$H_0 = \phi_2 = 0 \qquad H_a = \phi_2 \neq 0$$

para lo cual se calcula el intervalo de confianza con  $\alpha = 0.05$ :

$$\begin{aligned} & \phi_2 \pm 1,96\sigma_2 \\ & = (-0.82, 0.54) \end{aligned}$$

Por lo tanto,  $\phi_2$  no es significativo, así, el valor de  $\phi_2 = 0$  es uno de los posibles valores para  $\phi_2$ , puesto que esta contenido en el intervalo  $(-0.82, 0.54)$ .

$\therefore$  se acepta  $H_0$ , y queda como resultado el modelo  $AR(1)$  como aquél que mejor ajusta los datos.

---

<sup>8</sup>Veáse Apéndice A, Tabla 2.

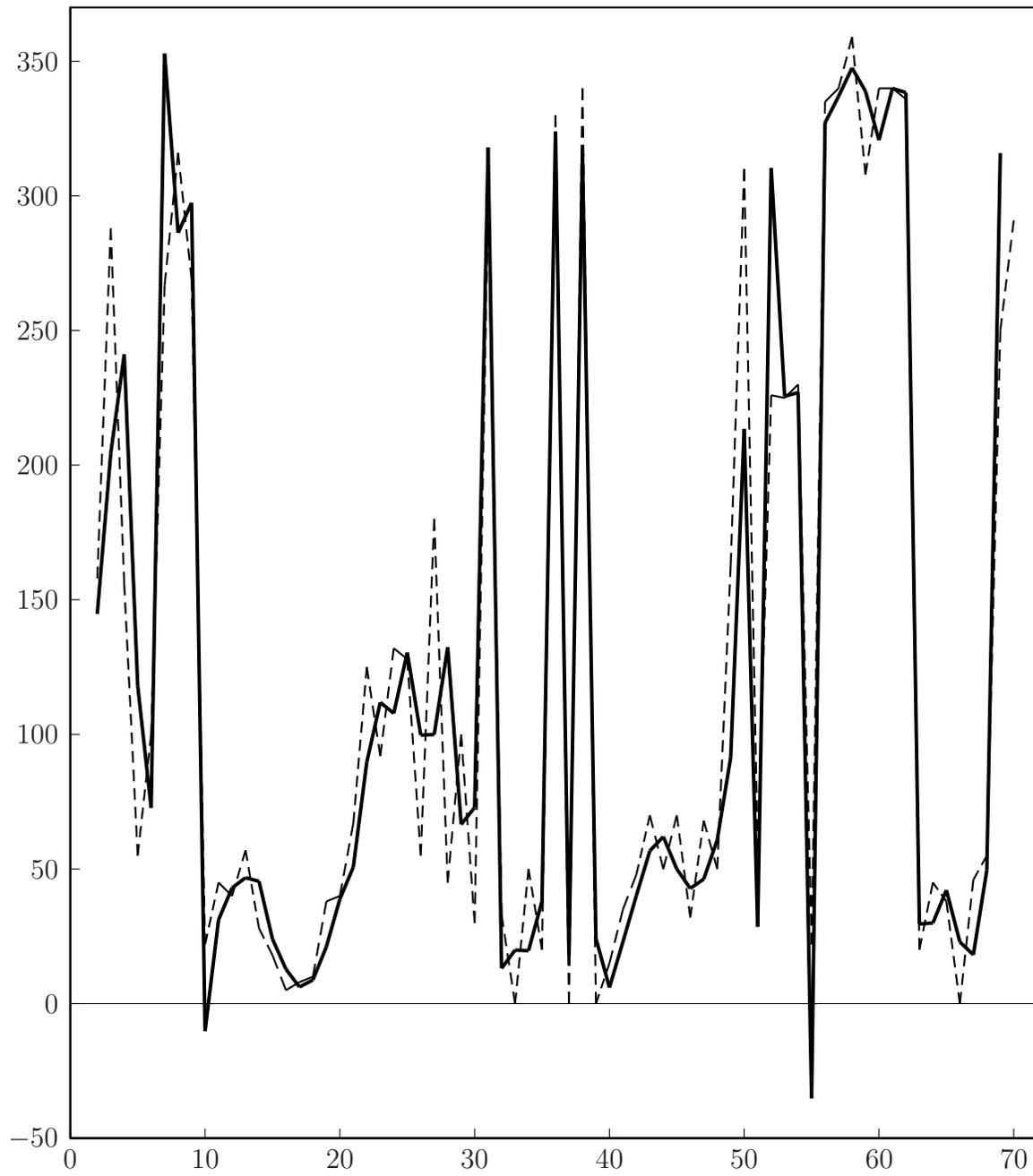


Figura 4.7: Gráfica que muestra la serie de tiempo circular de 70 mediciones de la dirección del viento (gráfica punteada), así como la serie que se genera del modelo ajustado  $LAR(1)$ .

# Conclusiones

Al igual que con los datos lineales, se pueden encontrar los mismos tipos de variación en las series de tiempo circulares: variación estacional, variación cíclica, o bien, tendencia, por lo que la exploración inicial de la serie se lleva a cabo de una manera similar a la correspondiente con datos lineales.

Una vez realizado el análisis exploratorio, es posible que al estadístico le interese encontrar un modelo con el cual se identifique la serie; si es el caso, entonces deberá elegir de entre 4 familias de modelos distintas (modelo de proyección, modelo de la normal envuelta, modelos utilizando la función enlace), lo anterior a diferencia de las series de tiempo lineales en las cuales el modelo seleccionado es uno correspondiente a la familia ARIMA; por lo tanto en este punto, el análisis se vuelve más complejo para aquél que corresponde al de datos circulares.

Una herramienta que juega un papel muy importante dentro del análisis de datos circulares, es el **correlograma**, ya que no solo se utiliza para interpretar la serie resultante, sino que también facilita la identificación del modelo al existir prototipos de funciones de autocorrelación de modelos derivados de funciones enlace.

Si el modelo seleccionado es uno que utilice una función enlace, el análisis podría resultar más fácil pero a su vez más tedioso, debido a que hay que transformar los datos circulares en lineales, una vez hecho este paso resultará más fácil estudiar la serie.

Una limitación más dentro del análisis circular, es el *pronóstico*, puesto que hasta ahora no se ha encontrado literatura al respecto; sin embargo si los datos se ajustan a un modelo derivado de una función enlace, entonces, una vez convertidos los datos circulares a lineales es posible realizar un pronóstico con ayuda de los conocimientos y paquetería que se tienen para datos lineales.

Con respecto al análisis estadístico mostrado en el **Capítulo 4**, es prudente señalar que no siempre se puede contar con observaciones que permitan la transformación a datos lineales, lo que ocasionaría un análisis mucho más complejo, pues sería necesario utilizar software especializado en datos circulares (como Oriana); aunado a lo anterior, se debe tener presente que, a pesar de tener datos que puedan ser transformados, no siempre se puede estimar el modelo mediante la comparación de correlogramas (como se hizo en la Aplicación), ya que es difícil que el correlograma de la serie transformada se asemeje a los prototipos que provee la literatura, no

obstante el análisis podría resultar no tan difícil puesto que la transformación fue posible, en estos casos se podría hacer uso de la paquetería especializada en datos lineales (como SPSS, SAS, STATISTICA).

Cabe resaltar que si se realiza el análisis  $AR(1)$  (de los datos utilizados en la Aplicación), utilizando SAS, se obtiene una  $\phi_1 = -0.638$ , resultado que además de ser aproximado al obtenido mediante la comparación de correlogramas, también confirma la teoría, pues en un modelo  $AR(1)$ ,  $\phi_1$  coincide con  $r_1$ .

## Apéndice A.

Tabla 1: Serie de tiempo de 72 direcciones del viento, tomadas cada hora, durante tres días en Australia.

285	285	280	300	240	255	250	250	235	240
240	180	220	265	180	150	150	150	335	355
335	305	345	340	315	0	330	300	330	330
50	270	270	270	245	285	280	270	15	285
310	330	300	340	280	300	270	270	255	90
285	285	285	270	270	270	270	270	300	300
270	300	330	15	330	300	345	330	330	300
315	285								

Tabla 2: Serie de tiempo estimada, resultante del modelo  $AR(2)$ , para los datos circulares.

53.981	139.197	193.644	248.767	179.574	206.169	49.701	261.568
81.598	57.456	16.302	36.833	16.299	18.847	9.612	36.130
19.294	61.662	105.027	45.746	147.986	103.263	52.336	234.517
-41.465	187.737	10.354	344.307	-205.960	214.798	142.443	-12.190
344.933	-291.897	653.928	-328.684	341.261	262.301	31.359	57.770
32.175	61.542	34.767	81.117	26.150	171.118	226.711	-69.827
439.173	140.278	202.138	18.498	506.542	46.986	222.240	293.466
374.854	323.735	331.556	22.912	345.238	166.880	1.634	74.735
26.495	237.024	131.479					

## Función Hipergeométrica

Una **función hipergeométrica**  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$  es una función que puede definirse en la forma de una serie hipergeométrica, *i.e.* una serie en la cual la relación de los términos sucesivos se pueden escribir:

$$\frac{c_{k+1}}{c_k} = \frac{P(k)}{Q(k)} = \frac{(k + a_1)(k + a_2) \dots (k + a_p)}{(k + b_1)(k + b_2) \dots (k + b_p)(k + 1)} \cdot x$$

La función  ${}_2F_1(a, b; c; x)$  correspondiente a  $p = 2$  y  $q = 1$  fue la primera función hipergeométrica en ser estudiada (surge con frecuencia en la mayoría de problemas físicos).

## Apéndice B.

Tabla 3: Tabla de datos.

$t$	$y_t$	$y_t - y_{t-1}$	$\theta$ ( <i>radianes</i> )	$X_i$
1	136	22	0.384	0.096
2	158	130	2.269	0.637
3	288	-132	-2.304	-0.649
4	156	-101	-1.763	-0.472
5	55	45	0.785	0.199
6	100	167	-2.321	-0.656
7	267	49	0.855	0.217
8	316	-47	-0.820	-0.208
9	269	-247	0.925	0.235
10	22	23	0.401	0.101
11	45	-5	-0.087	-0.022
12	40	17	0.297	0.074
13	57	-29	-0.506	-0.127
14	28	-10	-0.175	-0.044
15	18	-13	-0.227	-0.057
16	5	3	0.052	0.013
17	8	2	0.035	0.009
18	10	28	0.489	0.123
19	38	2	0.035	0.009
20	40	27	0.471	0.118
21	67	58	1.012	0.259
22	125	-33	-0.576	-0.145
23	92	40	0.698	0.176
24	132	-4	-0.070	-0.017
25	128	-73	-1.274	-0.330
26	55	125	2.182	0.607
27	180	-135	-2.356	-0.668
28	45	55	0.960	0.245
29	100	-70	-1.222	-0.315
30	30	270	-0.524	-0.132
31	300	-267	0.576	0.145
32	33	-33	-0.576	-0.145
33	0	50	0.873	0.222
34	50	-30	-0.524	-0.132
35	20	310	0.175	0.044
36	330	-330	-0.524	-0.132
37	0	340	0.698	0.176

38	340	-340	-0.698	-0.176
39	0	15	0.262	0.066
40	15	20	0.349	0.087
41	35	13	0.227	0.057
42	48	22	0.384	0.096
43	70	-20	-0.349	-0.087
44	50	0	0	0
45	50	-18	-0.314	-0.079
46	32	36	0.628	0.158
47	68	-18	-0.314	-0.079
48	50	113	1.972	0.537
49	163	147	2.566	0.747
50	310	-253	0.820	0.208
51	57	169	-2.286	-0.643
52	226	-1	-0.017	-0.004
53	225	5	0.087	0.022
54	230	-208	1.606	0.424
55	22	313	0.227	0.057
56	335	5	0.087	0.022
57	340	19	0.332	0.083
58	359	-51	-0.890	-0.226
59	308	32	0.559	0.141
60	340	0	0	0
61	340	-4	-0.070	-0.017
62	336	-316	-0.279	-0.070
63	20	25	0.436	0.110
64	45	-7	-0.122	-0.031
65	38	-38	-0.663	-0.167
66	0	46	0.803	0.203
67	46	9	0.157	0.039
68	55	195	-1.833	-0.493
69	250	41	0.716	0.181
70	291			

$y_t$  =  $t$ -ésima medición de la dirección del viento ( $^{\circ}$ C).

$y_t - y_{t-1}$  = Diferenciación de orden 1.

$\theta$  (*radianes*) =  $t$ -ésima observación en radianes.

$X_i$  =  $i$ -ésima observación convertida a dato lineal.

## Apéndice C.

Tabla 4: Serie de tiempo ajustada a los datos lineales.

$t$	$X_t$	$X_t$ transformada	$X'_t$
2	-0.058	-0.231	-13.226
3	-0.382	-1.460	-83.676
4	0.390	1.486	85.152
5	0.283	1.103	63.201
6	-0.119	-0.475	-27.224
7	0.393	1.499	85.894
8	-0.130	-0.518	-29.689
9	0.125	0.497	28.455
10	-0.141	-0.561	-32.166
11	-0.060	-0.241	-13.830
12	0.013	0.052	3
13	-0.045	-0.178	-10.212
14	0.076	0.305	17.460
15	0.026	0.105	6.002
16	0.034	0.136	7.805
17	-0.008	-0.031	-1.8
18	-0.005	-0.021	-1.2
19	-0.074	-0.294	-16.854
20	-0.005	-0.021	-1.2
21	-0.071	-0.284	-16.248
22	-0.155	-0.616	-35.281
23	0.087	0.347	19.888
24	-0.106	-0.422	-24.157
25	0.010	0.042	2.4
26	0.198	0.781	44.766
27	-0.364	-1.397	-80.024
28	0.401	1.525	87.385
29	-0.147	-0.583	-33.410
30	0.189	0.748	42.850
31	0.079	0.315	18.066
32	-0.087	-0.347	-19.888
33	0.087	0.347	19.888
34	-0.133	-0.529	-30.307
35	0.079	0.315	18.066
36	-0.026	-0.105	-6.002
37	0.079	0.315	18.066
38	-0.106	-0.422	-24.157

39	0.106	0.422	24.157
40	-0.039	-0.157	-9.008
41	-0.052	-0.210	-12.020
42	-0.034	-0.136	-7.805
43	-0.058	-0.231	-13.226
44	0.052	0.210	12.020
45	0	0	0
46	0.047	0.189	10.814
47	-0.095	-0.379	-21.714
48	0.047	0.189	10.814
49	-0.322	-1.247	-71.476
50	-0.448	-1.685	-96.537
51	-0.125	-0.497	-28.455
52	0.386	1.473	84.413
53	0.003	0.010	0.6
54	-0.013	-0.052	-3
55	-0.255	-0.998	-57.154
56	-0.034	-0.136	-7.805
57	-0.013	-0.052	-3
58	-0.050	-0.199	-11.417
59	0.136	0.540	30.926
60	-0.084	-0.337	-19.280
61	0	0	0
62	0.010	0.042	2.4
63	0.042	0.168	9.610
64	-0.066	-0.262	-15.038
65	0.018	0.073	4.201
66	0.1	0.4	22.934
67	-0.122	-0.486	-27.839
68	-0.024	-0.094	-5.402
69	0.296	1.151	65.931

$X_t$  =  $t$ -ésimo valor ajustado al  $t$ -ésimo dato lineal de la Tabla 3.

$X_t$  **transformada** =  $X_t$  en radianes.

$X'_t$  =  $X_t \rightarrow ^\circ C$ .

## Apéndice D.

Tabla 5: Datos Originales y Datos Ajustados:  $y_t$  y  $x'_t$ .

$t$	$y_t$	$x'_t$	<i>Residuales</i>
1	136		
2	158	144.774	13.226
3	288	204.324	83.676
4	156	241.152	-85.152
5	55	118.201	-63.201
6	100	72.776	27.224
7	267	352.894	-85.894
8	316	286.311	29.689
9	269	297.455	-28.455
10	22	-10.166	32.166
11	45	31.17	13.83
12	40	43	-3
13	57	46.788	10.212
14	28	45.46	-17.46
15	18	24.002	-6.002
16	5	12.805	-7.805
17	8	6.2	1.8
18	10	8.8	1.2
19	38	21.146	16.854
20	40	38.8	1.2
21	67	50.752	16.248
22	125	89.719	35.281
23	92	111.888	-19.888
24	132	107.843	24.157
25	128	130.4	-2.4
26	55	99.766	-44.766
27	180	99.976	80.024
28	45	132.385	-87.385
29	100	66.59	33.41
30	30	72.85	-42.85
31	300	318.066	-18.066
32	33	13.112	19.888
33	0	19.888	-19.888
34	50	19.693	30.307
35	20	38.066	-18.066
36	330	323.998	6.002
37	0	18.066	-18.066

38	340	315.843	24.157
39	0	24.157	-24.157
40	15	5.992	9.008
41	35	22.98	12.02
42	48	40.195	7.805
43	70	56.774	13.226
44	50	62.02	-12.02
45	50	50	0
46	32	42.814	-10.814
47	68	46.286	21.714
48	50	60.814	-10.814
49	163	91.524	71.476
50	310	213.463	96.537
51	57	28.545	28.455
52	226	310.413	-84.413
53	225	225.6	-0.6
54	230	227	3
55	22	-35.154	57.154
56	335	327.195	7.805
57	340	337	3
58	359	347.583	11.417
59	308	338.926	-30.926
60	340	320.72	19.28
61	340	340	0
62	336	338.4	-2.4
63	20	29.61	-9.61
64	45	29.962	15.038
65	38	42.201	-4.201
66	0	22.934	-22.934
67	46	18.161	27.839
68	55	49.598	5.402
69	250	315.931	-65.931
70	291		

$y_t$ =Serie original: dirección del viento en °C.

$x'_t$ =Serie ajustada a la serie original.

# Bibliografía

- [1] Aguirre, A. (1994) Introducción al Tratamiento de Series Temporales (Aplicación a las Ciencias de la Salud). *Díaz de Santos*.
- [2] Arnau, J. (2001) Diseño de Series Temporales: Técnicas de Análisis. *Edicions Universitat Barcelona*.
- [3] Chatfield, C. (2004) The Analysis of Time Series (An Introduction). *Chapman & Hall/CRC*.
- [4] Batschelet, E. (1981) Circular Statistics in Biology. *Academic Press*, 274-284.
- [5] Fisher, N.I.(1995) Statistical Analysis of Circular Data. *Cambridge University Press*.
- [6] Fisher, N.I. y Lee, A.J. (1983) A Correlation Coefficient For Circular Data. *Biometrika*, 327-332.
- [7] Fisher, N.I. y Lee, A.J. (1994) Times Series Analysis of Circular Data. *J.R. Statist. Soc. B*, 327-338.
- [8] Reyes Sánchez, Eloisa (2005) Tesis: Estadística Circular: herramienta para analizar y modelar las trayectorias de escape presa-depredador de *Litopenaeus Setiferus* y *Callinectes Sapidus*. UNAM, FES Zaragoza.
- [9] Wei, William (1990) Time Series Analysis. Univariate and Multivariate Methods. *Addison Wesley*.
- [10] [www.iam.udg.mx/Estacion/index.html](http://www.iam.udg.mx/Estacion/index.html) (agosto, 2008)
- [11] [www.mathworld.com](http://www.mathworld.com) (mayo y junio, 2008)