

**“Series de tiempo: Una aplicación a
registros hidrométricos en una cuenca del
Estado de Oaxaca”**

Arlette Méndez Bustamante

Julio de 2013

Dedicatoria

*A quienes me han heredado el tesoro
más valioso que puede dársele a un hijo.
A quienes sin escatimar esfuerzo alguno
han sacrificado gran parte de su vida.
Por esto y más . . .
dedicado a mis padres y hermano.*

Agradecimientos

Principalmente agradezco a mis padres: Raymundo y Amalia, a mi hermano Noel, porque sin importar cuántas escuelas disponibles existan o cuan grande sea el número de profesores dispuestos a transmitir sus conocimientos, sin el apoyo emocional y económico de mi familia, no habría podido culminar mis estudios.

Quiero agradecer a todos los profesores que estuvieron presentes a lo largo de mi formación matemática durante la licenciatura, por todo el trabajo que realizaron, dedicación y sobre todo paciencia. En especial, agradezco a la M.C. Marisol López Cerino y al M.C. René Morales Luis por iniciar la idea de efectuar esta aplicación real y por el apoyo continuo en la presente tesis. También quiero reconocer el trabajo de los revisores de este proyecto: M.E. Ana Delia Olvera Cervantes, Dr. Franco Barragán Mendoza, M.C. René Morales Luis y M.C. Marisol López Cerino, porque sin ellos, no se habría podido finiquitar la presente tesis. Asimismo, reconozco la labor de los primeros revisores de este proyecto: Dra. Virginia Berrón Lara y M.M. Tirso Miguel Ángel Ramírez Solano, por el tiempo invertido en cada ponencia, sugerencias y comentarios. Un agradecimiento particular a la M.C. Verónica Borja Macías, por haber llegado en el momento indicado de mi carrera, y por su dedicación en Seminario de tesis.

A Jose Angel por haber estado a mi lado en los días de alegría, pero sobre todo, por estar en los días de tristeza y desesperación durante mi estancia en la Universidad.

A mis amigos y compañeros, quienes pasaron a convertirse en mi familia por todos los días convividos. Gracias, por compartir conmigo su conocimiento.

*Gracias Dios,
por todo cuanto
me has dado y permitido.*

Índice general

Introducción general	ix
Abreviaciones y Notación	xi
1. Introducción	1
1.1. Historia de las series de tiempo	1
1.2. Las series de tiempo en hidrología	2
1.3. Conceptos preliminares de probabilidad	3
1.3.1. Procesos estocásticos	11
2. Series de tiempo	15
2.1. Conceptos básicos del análisis de series de tiempo	15
2.1.1. Objetivos de las series de tiempo	16
2.1.2. Componentes que integran una serie de tiempo	16
2.2. Análisis de series de tiempo	20
2.2.1. Análisis de series que contienen tendencia	21
2.2.2. Análisis de series que contienen variación estacional	22
2.2.3. Modelos con tendencia y estacionalidad	24
2.3. Eliminación de las componentes tendencia y estacionalidad	25
2.3.1. Eliminación de la tendencia por diferenciación	25
2.3.2. Eliminación de la tendencia y la estacionalidad por diferenciación	27
2.4. Modelos de ajuste para series de tiempo	29
2.4.1. Medias móviles	29
2.4.2. Medianas móviles	30
2.4.3. Suavizamiento exponencial	31
2.4.4. El enfoque de Box-Jenkins	34
3. Pronósticos	57
3.1. Introducción a los pronósticos	57
3.1.1. Métodos para establecer pronósticos	58
3.1.2. Selección de una técnica de pronóstico	59
3.1.3. Errores en los pronósticos	59
3.2. Tipos de pronósticos	61
3.3. Pronóstico para procedimientos univariados	62

3.3.1.	Extrapolación de curvas de tendencia	62
3.3.2.	Suavizamiento exponencial	63
3.3.3.	Metodología de Box-Jenkins	65
3.3.4.	Sistematización de Box - Jenkins de la suavización exponencial .	66
4.	Aplicación	69
4.1.	Datos	69
4.1.1.	Gastos Mínimos Anuales Instantáneos en Metros Cúbicos por Segundo (GMAI)	72
5.	Conclusiones	83
A.	Datos	87
B.	Propiedades	89
C.	Teorema de Schur	91

Introducción general

En la teoría de probabilidad, un tema de gran relevancia por sus aplicaciones en áreas como economía, meteorología, demografía, telecomunicaciones y transportes,¹ es el *Análisis de Series de Tiempo*. Con base en la literatura, una serie de tiempo es una colección de datos obtenidos por mediciones de algún evento natural o inducido, los cuales son reunidos sobre la misma variable, bajo las mismas condiciones a lo largo del tiempo y con intervalos de la misma medida. Un propósito, del análisis de series de tiempo es modelar el mecanismo que da lugar a la serie observada, para pronosticar los valores de su comportamiento futuro, a partir de los cuales, sea posible llevar a cabo una planeación y toma de decisiones.

Considerando las características de los registros hidrométricos², es factible emplear el análisis de series de tiempo. Por lo tanto, el objetivo general en la presente tesis es: describir los modelos de ajuste de el análisis de series de tiempo, y aplicar un modelo apropiado en la determinación del comportamiento del registro histórico de datos de aforo³ mínimos anuales instantáneos, de la estación hidrométrica *Paso Ancho* en el Estado de Oaxaca, con la finalidad de pronosticar comportamientos futuros. Los resultados del análisis de series de tiempo a los datos y el pronóstico derivado de éstos, tienen un horizonte amplio de aplicación, entre ellos, la planeación para la generación de energía eléctrica o la disponibilidad de agua en los sistemas de abasto para uso y consumo humano.

Con el fin de cumplir el objetivo en la presente tesis, el contenido de este escrito se ha organizado de la siguiente manera:

El Capítulo 1, es un recorrido por la historia de las series de tiempo y la teoría de la probabilidad necesaria, para luego adentrarse en el análisis de series de tiempo.

En el Capítulo 2, se estudian los conceptos básicos de la teoría de series de tiempo. Al mismo tiempo se observa, se analiza y explica su uso, mediante modelos de suavizamiento que se pueden utilizar como métodos de predicción simples, en datos de series de tiempo. Algunos modelos de suavizamiento que se estudian son: medias y medianas móviles, suavizamiento exponencial, y los modelos del enfoque de Box-Jenkins (ARMA y ARIMA).

¹Sólo por mencionar algunas.

²Registros del caudal de agua que pasa por una sección transversal de un río.

³Datos que miden el gasto o cantidad de agua que lleva una corriente en una unidad de tiempo.

El Capítulo 3, aborda el tema de pronósticos, el cual está íntimamente ligado con la construcción misma del modelo. Por tal razón, se estudia cuáles son las consideraciones para la elección de un método de pronóstico, y medidas para el grado de incertidumbre.

En el Capítulo 4, se lleva a cabo la aplicación de la teoría de series de tiempo sobre una base de datos de aforo de la estación hidrométrica “Paso Ancho”, del Estado de Oaxaca. En el Apéndice A se encuentra dicha base de datos.

Finalmente, se presentan las conclusiones de este trabajo de tesis y se hace un planteamiento de un posible trabajo futuro.

Con la finalidad de facilitar la realización de gráficos y análisis en el desarrollo de la presente tesis, se hará uso de herramientas proporcionadas por los paquetes estadísticos: **R project** (distribuido bajo la licencia *GNU General Public Licence*) e **ITSM 2000** (incluido en el libro *Introduction to Time Series and Forecasting* [Brockwell, 2002]). Debido a que estos software son amigables ante la vista del ser humano, no se considera necesario introducir los comandos utilizados en la simulación de los datos.

Abreviaciones y Notación

\mathbb{N}	Conjunto de los números naturales.
\mathbb{Z}	Conjunto de los números enteros.
\mathbb{R}	Conjunto de los números reales.
\mathbb{C}	Conjunto de los números complejos.
\mathbb{R}^n	Espacio euclídeo n -dimensional.
$ \cdot $	Módulo de un número complejo.
\log	Logaritmo en base e o logaritmo natural.
AR	Autorregresivo.
MA	Promedio móvil.
$ARMA$	Autorregresivo de promedio móvil.
$ARIMA$	Autorregresivo e integrado de promedio móvil.
FAC/ACF	Función de autocorrelación.
$FACP/Partial ACF$	Función de autocorrelación parcial.
$GMAI$	Gastos Mínimos Anuales Instantáneos en Metros Cúbicos por Segundo.
$GMAI_t$	Variable aleatoria de $GMAI$ al tiempo t .

Capítulo 1

Introducción

Los modelos de series de tiempo, son las técnicas de pronósticos que se basan únicamente en la historia de la demanda de la variable de interés que se está pronosticando. Trabajan capturando los patrones en los datos históricos extrapolándolos en el futuro. Los modelos de series de tiempo son adecuados cuando se puede asumir una cantidad razonable de datos y una continuidad en el futuro próximo de las condiciones que se presentaron en el pasado. Estos modelos se adaptan mejor a corto plazo para el pronóstico; mientras que van perdiendo validez cuando es a largo plazo. Esto se debe a la hipótesis de que los patrones pasados y las tendencias actuales se asemejan a los patrones y tendencias que se presentan en el futuro.

1.1. Historia de las series de tiempo

El análisis de series de tiempo, utilizando como dominio el tiempo y vistas como procesos estocásticos, surge en 1970 con la obra pionera de Box-Jenkins: *Time Series Analysis: Forecasting and Control* [Box and Jenkins, 1970]. A pesar del poco tiempo de estudio del análisis de las series de tiempo, se han hecho grandes contribuciones. Los personajes más importantes a lo largo del desarrollo de dicha teoría han sido:

George Edward Pelham Box¹ y Gwilym Meirion Jenkins² en colaboración, desarrollaron un esquema de clasificación para una amplia familia de modelos de series de tiempo. En esta clasificación se distinguen los modelos autorregresivos de orden p (AR(p)), los modelos de promedio móvil de orden q (MA(q)), combinaciones de los dos anteriores, llamados modelos autorregresivos de promedio móvil (ARMA(p,q)) y el modelo más general, autorregresivo e integrado de promedio móvil (ARIMA(p,d,q)); [Box and Jenkins, 1970].

Posteriormente, se tienen las contribuciones en las series de tiempo univariantes y multivariantes, estudiadas por Gregory C. Reinsel³ [Reinsel, 2003].

¹Nació en octubre de 1919 en Inglaterra.

²Agosto de 1933, Gales - julio de 1982, Inglaterra.

³1948, Pensilvania - mayo de 2004, U.S.A.

Mientras que Peter J. Brockwell⁴ trabajaba en el análisis de las series de tiempo y procesos estocásticos, Richard A. Davis⁵ estaba trabajando en la inferencia, la estimación, predicción y las propiedades generales de los modelos de series de tiempo. Los dos personajes en colaboración desarrollaron las obras: *Time Series: Theory and Methods* [Brockwell, 1987], e *Introduction to Time Series and Forecasting*, así como el desarrollo y análisis de las series de tiempo mediante el software ITSM 2000 [Brockwell, 2002].

1.2. Las series de tiempo en hidrología

Los registros sintéticos de afluentes empezaron a ser usados en 1914 por Hazen⁶, en estudios de confiabilidad en sistemas de abastecimiento de agua, pero su construcción no estaba basada en la teoría de procesos estocásticos [Koutsoyiannis, 2005].

Las primeras aplicaciones relacionadas con la hidrología estocástica fueron alrededor de el año 1940, en medio de la revolución de la física y las matemáticas. Entre otros avances, se establece el método de Montecarlo en manos de Stanislaw Ulam en 1946, quien concibe esta metodología mientras jugaba buscaminas durante la convalecencia de una enfermedad, cuando intentaba estimar la probabilidad de tener éxito en dicho juego; [Koutsoyiannis, 2005].

En el campo de los recursos hídricos, los aportes más importantes fueron los trabajos desarrollados por Barnes⁷ en 1954, sobre la generación de flujos anuales no correlacionados asumiendo una distribución normal, y los trabajos de Tomas y Fiering⁸ para la generación de flujos temporalmente correlacionados; [Koutsoyiannis, 2005].

La obra *Time Series Analysis: Forecasting and Control* [Box and Jenkins, 1970], se convirtió en uno de los principales aportes en la materia. Este libro, pese a ser escrito para aplicaciones en una amplia variedad de campos de la ciencia, adquirió gran popularidad para aplicaciones en hidrología estocástica ([Quijano, 1977], [Cervantes, 1982], [Pankratz, 1983], [Koutsoyiannis, 2005]).

⁴<http://www.stat.colostate.edu/~pjbrock/>

⁵<http://www.stat.columbia.edu/~rdavis/biography.html>

⁶Véase [Hazen, 1914].

⁷Véase [Barnes, 1954].

⁸Véase [Thomas and Fiering, 1962].

1.3. Conceptos preliminares de probabilidad

Antes de introducirse en la teoría de series de tiempo, se dedica esta sección a definir los conceptos probabilísticos preliminares necesarios.

Definición 1.3.1 Experimento aleatorio.

Un experimento aleatorio, denotado por la letra E , es aquel experimento que presenta las tres condiciones siguientes [Ross, 1972]:

1. tiene al menos dos posibles resultados,
2. el conjunto de posibles resultados se conoce antes de que se realice el experimento,
3. puede repetirse indefinidamente bajo las mismas condiciones.

Definición 1.3.2 Espacio muestral.

El espacio muestral, denotado por el símbolo Ω , es el conjunto de todos los posibles resultados o estados de la naturaleza de un experimento aleatorio E . Los subconjuntos de Ω son llamados eventos o sucesos y se denotan por las letras mayúsculas E_1, E_2, \dots . Por ejemplo, si E es el experimento aleatorio de lanzar una moneda al aire, entonces el espacio muestral es $\Omega = \{Sol, Aguila\}$, que es el conjunto de todos los posibles valores que se pueden obtener como resultado de lanzar una moneda. El subconjunto $E_1 = \{Sol\}$ de Ω es el evento de lanzar la moneda y observar Sol; [Ross, 1972].

Definición 1.3.3 Probabilidad.

A cada evento E_i para $i \in \mathbb{N}$, en el espacio muestral Ω de un experimento aleatorio E , es necesario asignarle un número que indique qué tan factible es que E_i suceda, a este número se le llama la probabilidad de E_i , se denota por $P(E_i)$, y satisface los siguientes tres axiomas:

Axioma (1) $0 \leq P(E_i) \leq 1$.

Axioma (2) $P(\Omega) = 1$.

Axioma (3) Para cualquier sucesión infinita de eventos disjuntos E_1, E_2, \dots ,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

La probabilidad ó “distribución de probabilidad” sobre un espacio muestral Ω es una especificación de números, $P(E_i)$, que satisfacen los Axiomas 1, 2, y 3; [Kalbfleisch, 1985a].

Definición 1.3.4 Variable aleatoria.

Una variable aleatoria X es una función real definida sobre un espacio muestral Ω asociado a un experimento, es decir, $X: \Omega \rightarrow \mathbb{R}$; [Ross, 1972].

De aquí en adelante, se escribe v.a en lugar de variable aleatoria. Existen al menos dos tipos de variables aleatorias: discretas y continuas.

Definición 1.3.5 Variable aleatoria discreta.

Se dice que una variable aleatoria X es discreta si únicamente puede tomar un número finito o infinito numerable de valores distintos [Kalbfleisch, 1985a].

Ejemplo 1.3.1 Se define el experimento aleatorio E , como el lanzamiento al aire de dos monedas balanceadas. Entonces, el espacio muestral Ω consta de cuatro posibles resultados,

$$\Omega = \{HH, HT, TH, TT\},$$

y todos son igualmente probables. La variable aleatoria de interés es X igual al número de caras⁹ que se observan. Así, la v.a X asigna un número real a cada punto de Ω :

$$X(TT) = 0, \quad X(HT) = X(TH) = 1, \quad X(HH) = 2.$$

Entonces, el conjunto de posibles valores de X es

$$\{0, 1, 2\}.$$

La distribución de probabilidad de X es el conjunto de probabilidades asociadas con sus posibles valores. La probabilidad de que X asuma un valor particular x , es obtenida por la suma de las probabilidades de todos los puntos en Ω , para los cuales el número x es asignado. Por lo tanto, la distribución de probabilidad de X es:

x	$P(X = x)$
0	1/4
1	1/2
2	1/4

Definición 1.3.6 Función de probabilidad.

Sea X una variable aleatoria discreta. Se llamará a $p(x) \equiv P(X = x)$ función de probabilidad¹⁰ de la variable aleatoria X , $p(x): \mathbb{R} \rightarrow [0, 1]$, si satisface las siguientes propiedades [Kalbfleisch, 1985a]:

1. $p(x) \geq 0$, para todos los posibles valores x de X ;

2. $\sum_x p(x) = 1$.

⁹H denota a “cara”, y T denota a “cruz”.

¹⁰El nombre completo de esta función es el de *función másica de probabilidad* de una variable aleatoria discreta.

La función de probabilidad para la v.a X en el Ejemplo 1.3.1, está dada en la siguiente tabla:

x	0	1	2	Total
$p(x)$	1/4	1/2	1/4	1

Hacer referencia a la distribución de probabilidad de X no sólo implica la existencia de la función de probabilidad, sino también de la existencia de la *función de distribución acumulada* de X .

Definición 1.3.7 Función de distribución acumulada.

La función de distribución acumulada de la variable aleatoria X , denotada por $F(x)$,¹¹ es la probabilidad de que X sea menor o igual a un valor específico de x , y está dada por:

$$F(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$

En particular, si X es una v.a discreta, la función de distribución acumulada se define por la siguiente expresión [Kalbfleisch, 1985a]:

$$F(x) = \sum_{x_i \leq x} p(x_i), \quad -\infty \leq x \leq +\infty.$$

Con base en la función de probabilidad para la v.a X , en el Ejemplo 1.3.1, se tiene que la función de distribución acumulada para X está dada por:

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ p(0) = \frac{1}{4}, & \text{si } 0 \leq x < 1; \\ p(0) + p(1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}, & \text{si } 1 \leq x < 2; \\ p(0) + p(1) + p(2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1, & \text{si } x \geq 2. \end{cases}$$

Definición 1.3.8 Variable aleatoria continua.

Una variable aleatoria X , se dice que es continua si existe una función $f(x)$, llamada *función de densidad de probabilidad* de X , tal que

$$P(X \in B) = \int_B f(x) dx$$

para cualquier conjunto $B \subseteq \mathbb{R}$. De la función de distribución acumulada

$$F(x) = \int_{-\infty}^x f(x) dx,$$

se sigue que

$$f(x) = \frac{d}{dx} F(x), \quad -\infty \leq x \leq +\infty.$$

¹¹Suele denotarse por $F_X(x)$ cuando se analiza más de una v.a.

La función de densidad de probabilidad de una variable aleatoria continua X , cumple con las siguientes propiedades [Ross, 1996]:

1. $f(x) \geq 0, \quad \forall x \in \mathbb{R};$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1.$

Definición 1.3.9 Valor esperado.

El valor esperado o media de una v.a X , denotado por $E(X)$ o μ ,¹² se define por:

$$E(X) = \begin{cases} \int_{-\infty}^{+\infty} xf(x)dx, & \text{si } X \text{ es continua,} \\ \sum_x xp(x), & \text{si } X \text{ es discreta;} \end{cases} \quad (1.1)$$

siempre y cuando se pruebe que para la v.a continua, la integral existe ([Degroot, 1988], pp. 171).

El valor esperado, es un número que representa el centro de la distribución de los posibles valores que X puede tomar, el cual no está necesariamente en la mitad. Además, la Ecuación (1.1) también define el valor esperado de cualquier función de X , es decir, para $h(X)$ se sigue que:

$$E[h(X)] = \begin{cases} \int_{-\infty}^{+\infty} h(x)f(x)dx, & \text{si } X \text{ es continua;} \\ \sum_x h(x)p(x), & \text{si } X \text{ es discreta.} \end{cases}$$

Para cualquier variable aleatoria X y cualquier entero positivo k , si $h(X) = X^k$, entonces el valor esperado $E(X^k)$ se denomina el *k-ésimo momento de X* . En particular, de acuerdo con esta terminología, la media de X es el primer momento de la v.a X .

Continuando con el Ejemplo 1.3.1, se obtiene el valor esperado de la v.a discreta X , mediante su función de probabilidad:

$$\begin{aligned} E(X) &= \sum_{x=0}^2 xp(x); \\ &= 0 \left(\frac{1}{4}\right) + 1 \left(\frac{1}{2}\right) + 2 \left(\frac{1}{4}\right); \\ &= 1. \end{aligned}$$

¹²Suele denotarse por μ_X cuando se analiza más de una v.a.

El número $E(X) = 1$, es el valor que tiene mayor probabilidad de ocurrir en la distribución especificada por la función de probabilidad de X , es decir, al lanzar dos monedas al aire, se espera observar una cara.

Propiedad 1.3.1 Sean X e Y cualesquiera dos v.a.'s definidas sobre el mismo espacio muestral, con valor esperado finito, y sean a y b cualesquiera dos constantes. Entonces,

a) $E(aX + b) = aE(X) + b$;

b) $E(b) = b$;

c) $E(aX) = aE(X)$;

d) $E(aX + bY) = aE(X) + bE(Y)$.

e) Si X e Y son v.a.'s independientes,¹³ entonces $E(XY) = E(X)E(Y)$.

Demostración.

Para a), b), c) y d), véase [Kalbfleisch, 1985a], pp. 160-161; mientras que para e), véase [Degroot, 1988], pp. 182. ■

Definición 1.3.10 Varianza.

Sea X una variable aleatoria con media $\mu = E(X)$. La varianza de X , que se denotará por $Var(X)$, se define por:

$$Var(X) = E[(X - \mu)^2];$$

([Degroot, 1988], pp. 185).

El valor de $Var(X)$ se llama también varianza de la distribución de probabilidad de X . La varianza de una variable aleatoria proporciona una medida de la variación o dispersión de una distribución alrededor de su media μ .

La *desviación típica* de una variable aleatoria o de una distribución se define como la raíz cuadrada no negativa de la varianza. La desviación típica, también llamada desviación estándar de una variable aleatoria, se denota usualmente por el símbolo σ , y la varianza se denota por σ^2 .

La varianza de la v.a X en el Ejemplo 1.3.1, con media $\mu = 1$ es:

$$\begin{aligned} Var(X) &= E[(X - \mu)^2]; \\ &= \sum_{x=0}^2 (x - \mu)^2 p(x); \\ &= (0 - 1)^2 \left(\frac{1}{4}\right) + (1 - 1)^2 \left(\frac{1}{2}\right) + (2 - 1)^2 \left(\frac{1}{4}\right); \\ &= 0.5. \end{aligned}$$

¹³Véase Definición 1.3.14.

La varianza $Var(X) = 0.5$, indica el rango¹⁴ donde hay más probabilidad de que ocurra el evento.

Propiedad 1.3.2 *Para cualquier v.a X , con varianza finita, y para constantes a , b y c cualesquiera, se sigue que,*

a) $Var(X) \geq 0$;

b) $Var(X) = 0$ si, y sólo si, $X = c$;

c) $Var(aX + b) = a^2Var(X)$.

Demostración.

Para a), b) y c), véase [Degroot, 1988], pp. 185-186. ■

En muchos experimentos, es necesario considerar las propiedades de dos o más variables aleatorias simultáneamente. La distribución de probabilidad conjunta de dos variables aleatorias se denomina *distribución bivalente*. Al par (X, Y) , se le denomina vector aleatorio bivariado ([Degroot, 1988], pp. 110-118).

Ejemplo 1.3.2 *Supóngase un anuncio luminoso en el que hay tres bombillas en la primera fila y cuatro bombillas en la segunda. Sea X el número de bombillas de la primera fila que se funden en un instante t , y sea Y el número de bombillas de la segunda fila que se funden en el mismo instante de tiempo t . Entonces, existe un número finito de valores posibles distintos (x, y) para el par (X, Y) . En este caso, se dice que X e Y tienen una distribución bivalente discreta.*

Definición 1.3.11 Función de probabilidad conjunta.

Sean X, Y variables aleatorias discretas. Entonces, la función de probabilidad conjunta de (X, Y) se define como la función $p(\cdot, \cdot) : \mathbb{R}^2 \rightarrow [0, 1]$, tal que para cualquier punto (x, y) del plano XY ,

$$p(x, y) = P(X = x, Y = y),$$

donde $P(X = x, Y = y)$ indica la probabilidad de que X tome el valor x , y la v.a Y tome el valor y .

Si (x, y) no es uno de los posibles valores del par de variables aleatorias (X, Y) , entonces $p(x, y) = 0$. La función de probabilidad conjunta de (X, Y) satisface las siguientes propiedades:

1. $p(x, y) \geq 0$, para todos los posibles valores (x, y) de (X, Y) ;

2. $\sum_{(x,y)} p(x, y) = 1$.

¹⁴Amplitud de la variación de un fenómeno entre un límite menor y uno mayor.

La función de probabilidad conjunta de (X, Y) , en el Ejemplo 1.3.2, está dada en la siguiente tabla:

		Y				
		0	1	2	3	4
X	0	0.08	0.07	0.06	0.01	0.01
	1	0.06	0.10	0.12	0.05	0.02
	2	0.05	0.06	0.09	0.04	0.03
	3	0.02	0.03	0.03	0.03	0.04

Dicha función cumple con las dos propiedades. En efecto, para cada valor posible distinto (x, y) se satisface que $p(x, y) > 0$, y $\sum_{(x,y)} p(x, y) = 0.08 + 0.07 + 0.06 + 0.01 + 0.01 + 0.06 + 0.10 + 0.12 + 0.05 + 0.02 + 0.05 + 0.06 + 0.09 + 0.04 + 0.03 + 0.02 + 0.03 + 0.03 + 0.03 + 0.04 = 1$.

Definición 1.3.12 Función de densidad de probabilidad conjunta.

Sean X, Y variables aleatorias continuas. Se dice que las variables aleatorias X e Y tienen una distribución bivalente continua si existe una función no negativa f definida sobre todo el plano XY , $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que para cualquier subconjunto A del plano,

$$P[(X, Y) \in A] = \int_A \int f(x, y) dx dy.$$

La función f se denomina función de densidad de probabilidad conjunta de (X, Y) .

Tal función de densidad de probabilidad conjunta satisface las dos condiciones siguientes:

1. $f(x, y) \geq 0$, $\forall x, y \in \mathbb{R}$;
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.

Si se conoce la función de probabilidad conjunta o la función de densidad de probabilidad conjunta de (X, Y) , entonces, se puede obtener a partir de estas, la función de probabilidad marginal o la función de densidad de probabilidad marginal de cada variable aleatoria, respectivamente ([Degroot, 1988], pp. 120).

Definición 1.3.13

Sea (X, Y) un vector aleatorio bivariado.

1. Si X, Y son v.a.'s discretas con función de probabilidad conjunta $p(x, y)$, entonces las **funciones de probabilidad marginal** de X e Y , denotadas por $p_X(x)$ y $p_Y(y)$,

están dadas por:

$$p_X(x) = \sum_y p(x, y), \quad \text{para cualquier valor } x \text{ de } X;$$

$$p_Y(y) = \sum_x p(x, y), \quad \text{para cualquier valor } y \text{ de } Y.$$

2. Si X, Y son v.a.'s continuas con función de densidad de probabilidad conjunta $f(x, y)$, entonces las **funciones de densidad de probabilidad marginal** de X e Y , denotadas por $f_X(x)$ y $f_Y(y)$, están dadas por:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad \forall x \in \mathbb{R};$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx, \quad \forall y \in \mathbb{R}.$$

Las funciones de probabilidad marginal de X e Y , en el Ejemplo 1.3.2, denotadas por $p_X(x)$ y $p_Y(y)$, se pueden observar en la siguiente tabla:

		Y					$p_X(x)$
		0	1	2	3	4	
X	0	0.08	0.07	0.06	0.01	0.01	0.23
	1	0.06	0.10	0.12	0.05	0.02	0.35
	2	0.05	0.06	0.09	0.04	0.03	0.27
	3	0.02	0.03	0.03	0.03	0.04	0.15
$p_Y(y)$		0.21	0.26	0.3	0.13	0.1	

Definición 1.3.14 Variables aleatorias independientes.

Se dice que dos variables aleatorias X e Y son independientes si y solo si

$$\begin{cases} p(x, y) = p_X(x)p_Y(y), & \text{si } X \text{ e } Y \text{ son v.a.'s discretas;} \\ f(x, y) = f_X(x)f_Y(y), & \text{si } X \text{ e } Y \text{ son v.a.'s continuas;} \end{cases}$$

([Degroot, 1988], pp. 123).

En general, sean X_1, X_2, \dots, X_n v.a.'s discretas, definidas sobre el mismo espacio muestral. La función de probabilidad conjunta de X_1, X_2, \dots, X_n es una función definida sobre \mathbb{R}^n , tal que:

$$p(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

Las variables X_1, X_2, \dots, X_n son independientes si

$$p(x_1, x_2, \dots, x_n) = P_{X_1}(x_1) \cdot P_{X_2}(x_2) \cdot \dots \cdot P_{X_n}(x_n).$$

En el caso en que X_1, X_2, \dots, X_n son v.a.'s continuas, también es posible definir su función de densidad probabilidad conjunta ([Degroot, 1988], pp. 136-139).

Definición 1.3.15 Variables IID.

Si X_1, X_2, \dots, X_n son variables aleatorias independientes y tienen la misma distribución de probabilidad marginal, entonces estas son llamadas v.a's **independientes e idénticamente distribuidas**, y usualmente se denotan por IID.

Hay cantidades que permiten medir la asociación entre dos variables aleatorias, es decir, estas proporcionan información acerca de la relación entre las dos variables o acerca de su tendencia a variar juntas más que independientemente, tal es el caso de la covarianza y correlación del vector aleatorio bivariado (X, Y) ([Degroot, 1988], pp. 202-206).

Definición 1.3.16 Covarianza.

Sean X e Y variables aleatorias que tienen una distribución conjunta. La covarianza entre X e Y , que se denota por $Cov(X, Y)$, se define como:

$$Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right].$$

Propiedad 1.3.3 Sean X e Y v.a's definidas sobre el mismo espacio muestral, con varianza finita, y sean a, b, c y d constantes. Entonces,

- a) $Cov(aX + b, cY + d) = acCov(X, Y)$;
- b) $Cov(X, X) = Var(X)$;
- c) $Cov(X, Y) = Cov(Y, X)$;
- d) Si X e Y son v.a's independientes, entonces $Cov(X, Y) = 0$.

Demostración.

Para a), b) y c), véase [Kalbfleisch, 1985a], pp. 176-178; mientras que para d), véase [Degroot, 1988], pp. 204. ■

Definición 1.3.17 Correlación.

Sean X e Y v.a's que tienen una distribución conjunta. Si $0 < Var(X) < +\infty$ y $0 < Var(Y) < +\infty$, entonces la correlación entre X e Y , se denota y se define como:

$$\rho(X, Y) = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

1.3.1. Procesos estocásticos

Las series de tiempo son un caso particular de los procesos estocásticos. Por tal razón, será necesario definir a dichos procesos. Así, como conceptos que van de la mano con estos, para el análisis de las series de tiempo.

Definición 1.3.18 Proceso estocástico.

Un proceso estocástico $\{X_t: t \in T\}$ es una colección de variables aleatorias, parametrizadas por un conjunto $T \subset \mathbb{R}$ de índices, llamado espacio paramétrico. El conjunto T usualmente se interpreta como un conjunto de valores en el tiempo, y X_t es el estado del proceso al tiempo t ([Chatfield, 2003], pp. 33).

Se dice que un proceso estocástico $\{X_t: t \in T\}$ es discreto, si el conjunto de índices T es un conjunto contable (por ejemplo, $T = \{0, \pm 1, \pm 2, \dots\}$); y si los valores de T , consisten en uno o más intervalos de la recta de los reales (por ejemplo, $T = [0, \infty)$), entonces se dice que es un proceso estocástico continuo.

Se denota a la variable aleatoria al tiempo t por X_t si T es contable, y por $X(t)$ si $T \subseteq \mathbb{R}$.

El espacio de estados o muestral de un proceso estocástico contiene todos los posibles valores que las v.a's (X_t o $X(t)$) pueden tomar, y se denota por S ; [Ross, 1972].

Se mencionan algunos ejemplos de procesos estocásticos:

- Sea $X(t)$: El número de autos que pasan por un cruce durante un instante t , donde t es el tiempo medido en minutos de un intervalo $[a, b]$, el espacio de estados es $S = \{1, 2, 3, \dots\}$.
- Sea X_t : La evolución de la población de un municipio año tras año.

De aquí en adelante, cuando se haga mención de un proceso estocástico, se estará refiriendo a un proceso discreto, salvo que se diga lo contrario.

Clasificación de los procesos estocásticos.

Tipos de procesos estocásticos en relación a la v.a X_t :

- **Procesos con incrementos independientes**

Se dice que un proceso $\{X_t: t \geq 0\}$, tiene incrementos independientes, si para cada tiempo $0 \leq t_1 < t_2 < \dots < t_n$, las variables aleatorias $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes.

- **Procesos de renovación**

Un proceso de renovación es una sucesión $\{T_n\}$ de v.a's independientes e igualmente distribuidas, que representan los tiempos de vida de las unidades o los tiempos en los que se reporta una incidencia (por ejemplo, un fallo o un reclamo) en un sistema. La primera unidad empieza su ciclo de vida en el cero, y cuando falla al tiempo T_1 es reemplazada inmediatamente por una nueva unidad, la cual posteriormente falla al tiempo $T_1 + T_2$ y es reemplazada por una nueva unidad, y así sucesivamente, es decir, se van dando o renovando unidades.

- **Procesos estacionarios**

Un proceso se dice que es estacionario si X_t no depende del tiempo. En general, se dice que un proceso $\{X_t: t \geq 0\}$ es un proceso estocástico estacionario si para cualesquiera tiempos t_1, t_2, \dots, t_n , la distribución del vector $(X_{t_1}, \dots, X_{t_n})$ es la misma que la del vector $(X_{t_1+h}, \dots, X_{t_n+h})$ para cualquier $h > 0$. En particular, la distribución de X_t es igual a la distribución de la variable aleatoria X_{t+h} , $h \geq 0$, así que, si los primeros dos momentos de X_t son finitos, se tiene que

$$E(X_t) = E(X_{t+h}) = \mu, \text{ y}$$

$$Var(X_t) = Var(X_{t+h}) = \sigma^2,$$

ambos son constantes.

- **Procesos con incrementos estacionarios**

Un proceso $\{X_t: t \geq 0\}$ tiene incrementos estacionarios si para cualesquiera tiempos $s < t$, y para cualquier $h > 0$, las variables $X_{t+h} - X_{s+h}$ y $X_t - X_s$ tienen la misma distribución de probabilidad.

Definición 1.3.19 Función de autocovarianza.

Para un proceso $\{X_t: t \in \mathbb{T}\}$ tal que $Var(X_t) < \infty$, para cada $t \in \mathbb{T}$, la función de autocovarianza $\gamma_X(\cdot, \cdot) : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$ de X_t se define por:

$$\gamma_X(r, s) = E\left[(X_r - E(X_r))(X_s - E(X_s))\right] = Cov(X_r, X_s), \quad r, s \in \mathbb{T}$$

([Brockwell, 1987], pp. 11).

Definición 1.3.20 Ruido blanco

Un proceso $\{X_t: t \in \mathbb{T}\}$ se dice que es ruido blanco o un proceso puramente aleatorio si la secuencia de v.a's, $\{X_t\}$, son independientes e idénticamente distribuidas. De la definición se sigue que el proceso tiene media y varianza constante. Además, la independencia asume que $Cov(X_t, X_{t+h}) = 0$, para todo $h = \pm 1, \pm 2, \dots$ ([Chatfield, 2003], pp. 37-38).

Capítulo 2

Series de tiempo

Los métodos de análisis de series de tiempo consideran el hecho que los datos tomados en diversos periodos pueden tener algunas características de autocorrelación, tendencia o estacionalidad que se debe tomar en cuenta, para modelar el comportamiento de la serie observada. La aplicación de estos métodos tiene dos propósitos: comprender las fuerzas de influencia en los datos y descubrir la estructura que los produjo.

Las aplicaciones de la teoría de series de tiempo suelen ser: pronósticos económicos, análisis de presupuesto, análisis del mercado y análisis hidrológico, entre otros.

2.1. Conceptos básicos del análisis de series de tiempo

Muchos autores han formulado definiciones de las series de tiempo. Algunas definiciones informales que comúnmente se encuentran en la literatura son las siguientes:

- Una serie de tiempo es una colección de datos reunidos sobre la misma variable a lo largo del tiempo; [Hildebrand, 1998].
- Se llama serie de tiempo a un conjunto de mediciones de cierto fenómeno o experimento registrado secuencialmente en el tiempo; [Chatfield, 2003].
- Una serie de tiempo es un registro metódico de la medición u observación numérica, efectuada a intervalos de tiempo fijos, de las características o variables del área de interés; [Guerrero, 2003].
- Una serie de tiempo es una sucesión cronológica de observaciones de una variable particular; [Bowerman, 2009].

Una definición formal de *serie de tiempo* desde el enfoque de procesos estocásticos está dada de la forma siguiente:

Definición 2.1.1 Serie de tiempo

Una serie de tiempo es un proceso estocástico en el cual el conjunto de observaciones son tomadas secuencialmente en el tiempo, usualmente a “intervalos iguales”. Así, un modelo de serie de tiempo para datos observados $\{x_t\}$ es una especificación de la distribución conjunta (posiblemente sólo de las medias y covarianzas) de una sucesión de variables aleatorias X_t para los cuales $\{x_t\}$ se postula como una realización; [Brockwell, 2002].

2.1.1. Objetivos de las series de tiempo

Algunos de los objetivos principales del análisis de una serie de tiempo, son clasificados en: descripción, explicación, predicción y control, esto según [Chatfield, 2003]. Se analizará cada uno de los objetivos como sigue:

- **Descripción**

El primer paso en el análisis de una serie de tiempo es graficar las observaciones, para obtener medidas simples descriptivas de las características principales de la serie.

- **Explicación**

Cuando las observaciones son tomadas en dos o más variables, es posible usar la variación en una serie de tiempo para explicar la variación en otra.

- **Predicción**

Dada una serie de tiempo, es posible que se desee predecir los valores futuros que la serie puede tomar, es decir, la estimación de valores futuros de la variable de interés en función del comportamiento pasado de la serie.

- **Control**

Una serie de tiempo se analiza con el fin de mejorar el control sobre algún sistema físico o económico. En el control de calidad estadístico, las observaciones son representadas en gráficos de control, y el controlador toma acciones como resultado del estudio de los gráficos.

2.1.2. Componentes que integran una serie de tiempo

Una manera de examinar una serie de tiempo es dividiéndola en componentes. Un enfoque estándar consiste en encontrar componentes que correspondan a una tendencia a largo plazo, algún comportamiento cíclico, un componente estacional y una parte irregular, residual. Algunos de estos aspectos esenciales de los procesos no son directamente observables, es por ello que es importante disponer de procedimientos que permitan extraer esos aspectos de las observaciones.

A continuación se muestran y se definen los componentes que integran una serie de tiempo de acuerdo a [Chatfield, 2003]. Para ilustrar mejor cada uno de los componentes se analizarán algunos ejemplos, que fueron tomados de la base de datos de la obra *Introduction to Time Series and Forecasting* usando el software ITSM; [Brockwell, 2002].

- **Componente de tendencia**
- **Componente estacional**
- **Componente irregular**
- **Componente cíclico**

Componente de tendencia.

Es el componente que recoge la parte de la variable vinculada principalmente con factores de largo plazo. En la práctica resulta difícil distinguir la tendencia del componente cíclico, por tanto se combina en un solo componente llamado tendencia-ciclo.

Ejemplo 2.1.1 Las ventas de vino de una red australiana

La Figura 2.1 muestra las ventas mensuales desde enero de 1980 hasta octubre de 1991. Se observa de la gráfica que las observaciones siguen una tendencia creciente o positiva.

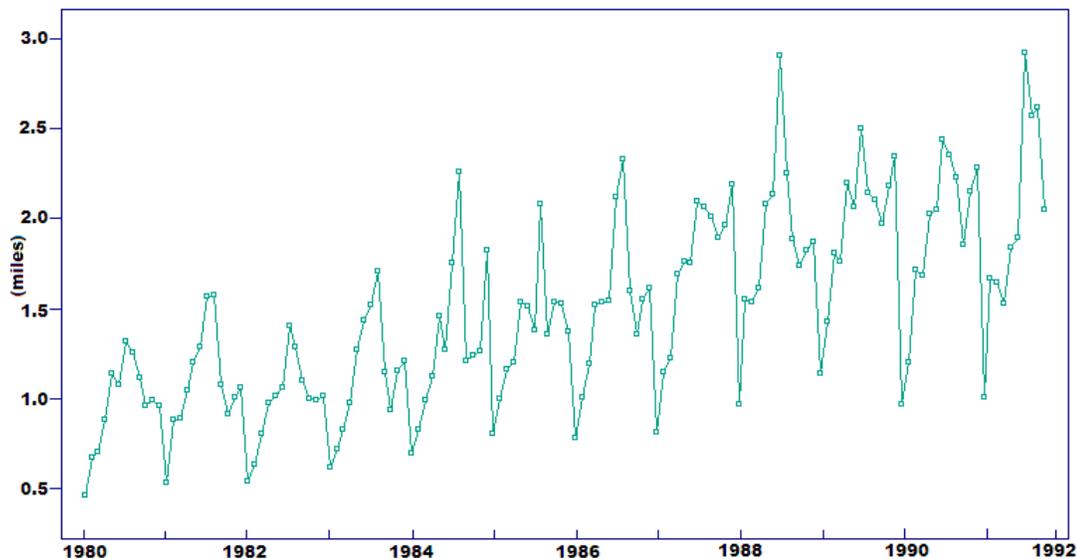


Figura 2.1: Ventas de vino de una red australiana, Ene. 1980 - Oct. 1991.

Ejemplo 2.1.2 Población de Estados Unidos

La Figura 2.2 muestra la población de Estados Unidos en intervalos de diez años, desde 1790 hasta 1990. La gráfica muestra la posibilidad de ajustar los datos a una tendencia cuadrática o exponencial.

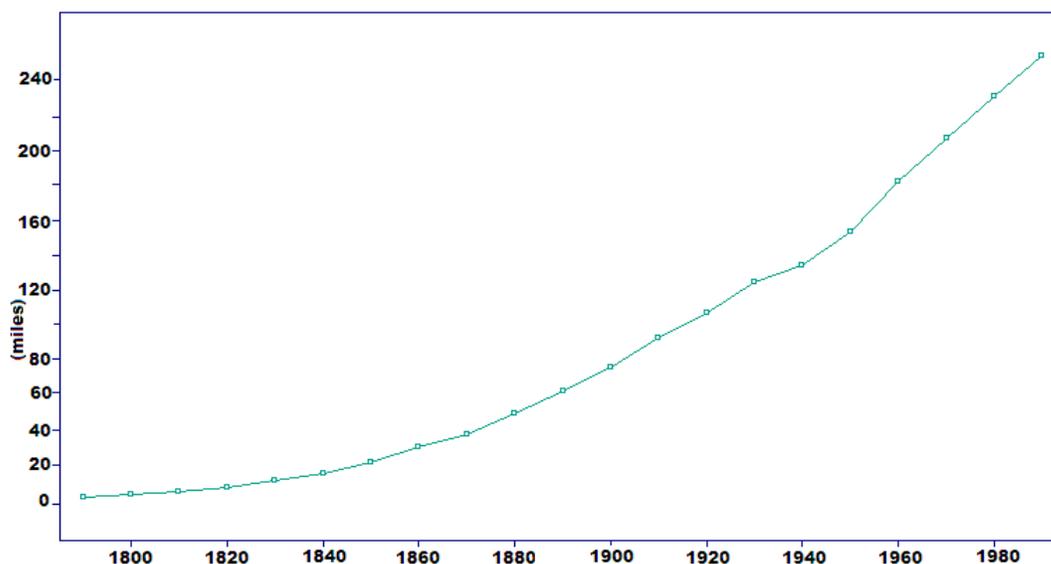


Figura 2.2: Población de U.S.A. en intervalos de diez años, 1790-1990.

Componente estacional.

Son oscilaciones cuasi-cíclicas de media cero, las cuales tienen periodicidad anual o de un submúltiplo del año (trimestrales, mensuales, etc.) y se conocen como oscilaciones estacionales.

Ejemplo 2.1.3 Muertes accidentales

La Figura 2.3 muestra los datos mensuales por muertes accidentales, desde 1973 hasta 1978. Se observa en la gráfica que hay un fuerte patrón de estacionalidad, con un máximo en julio de cada año y un mínimo en febrero de cada año. La presencia de tendencia es menos observable que en la Figura 2.1.

Componente irregular.

Son oscilaciones no sistemáticas que en general afectan a la serie en el momento en que ocurren y normalmente tienen una estructura puramente aleatoria¹.

Ejemplo 2.1.4 Huelgas en U.S.A.

La Figura 2.4 muestra los números anuales de las huelgas en U.S.A., aproximadamente de 1951 a 1980. Se aprecia que los registros fluctúan erráticamente sobre un nivel de variación lenta.

¹también llamada estructura de ruido blanco.

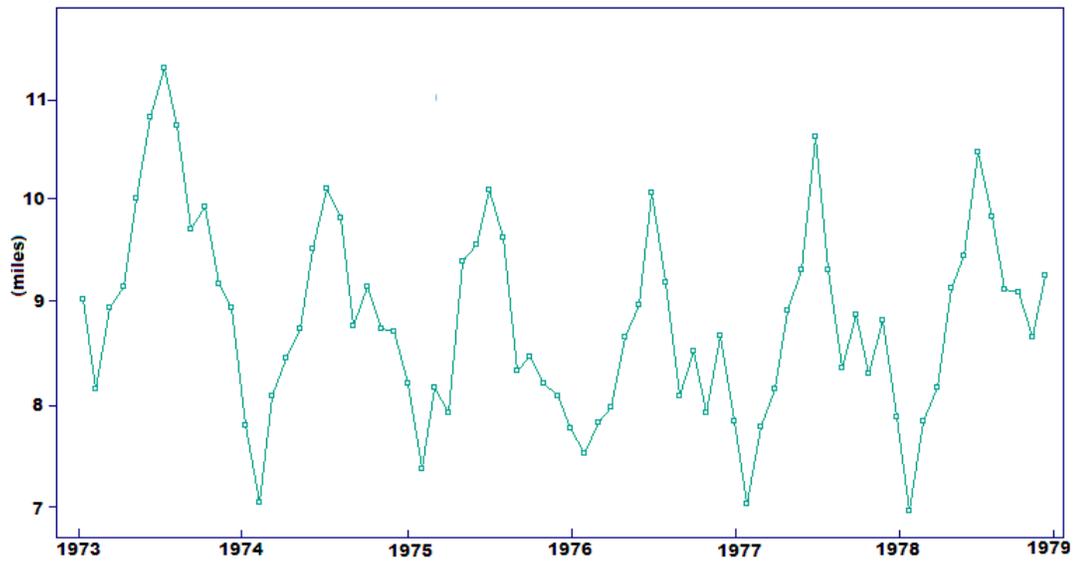


Figura 2.3: Datos mensuales por muertes accidentales, 1973-1978.

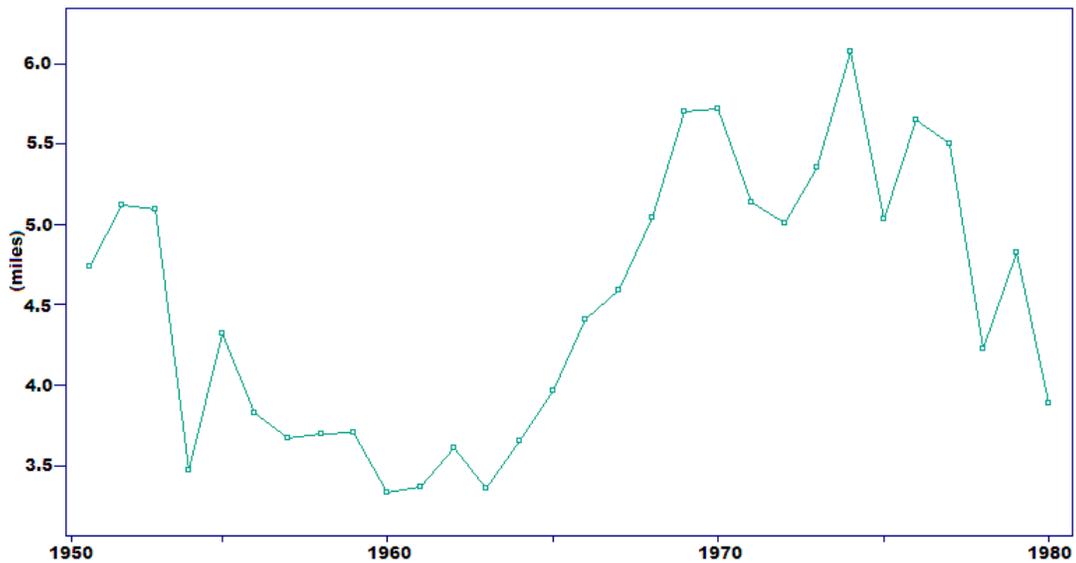


Figura 2.4: Huelgas en U.S.A., 1951-1980.

Componente cíclico.

Son oscilaciones con periodicidad de entre año y medio y hasta diez años, dependiendo de la definición del ciclo que se utilice. Suelen ser menos frecuentes y menos sistemáticas que las estacionales.

Ejemplo 2.1.5 Línea de transporte PATCO

La Figura 2.5 muestra los datos sobre el número de pasajeros en la línea pública de transporte PATCO, a lo largo de 86 semanas de 1983 a 1989, dividiéndose cada año en trece periodos de cuatro semanas. Se observa que puede haber un efecto cíclico del periodo nueve de 1983 hasta el periodo nueve de 1986, observándose tres ciclos; [Hildebrand, 1998].

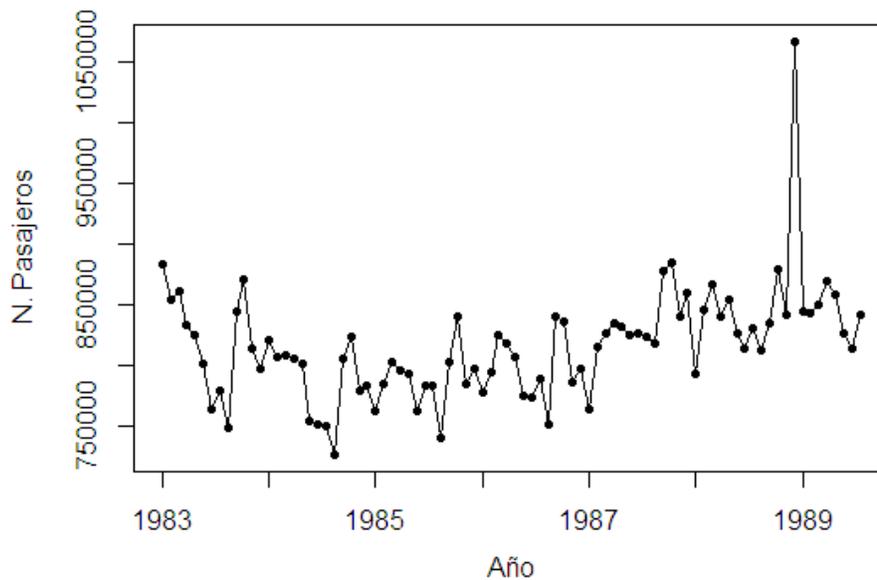


Figura 2.5: Número de pasajeros de la línea de transporte PATCO, 1983-1989.

2.2. Análisis de series de tiempo

La forma más usual de representar una serie de tiempo es en función de los componentes tendencia y estacionalidad, es decir, mediante la ecuación:

$$X_t = m_t + s_t + \varepsilon_t,$$

donde

- X_t = el valor de la serie temporal en el periodo t ;
- m_t = componente de tendencia en el periodo t ;
- s_t = es una función que representa la componente estacional;
- ε_t = el término de error en el periodo t .

El término de error ε_t representa las fluctuaciones aleatorias que ocasionan que los valores X_t se desvíen del nivel promedio μ_t , dichas fluctuaciones son un tanto más complejas de explicar dentro del análisis de series de tiempo en la mayoría de los fenómenos reales. Por este motivo, resulta conveniente introducir un componente aleatorio que permita mayor flexibilidad dentro del análisis; [Guerrero, 2003].

2.2.1. Análisis de series que contienen tendencia

Un modelo que adopta únicamente la componente de tendencia es aquel en el que la observación al tiempo t , está dada por

$$X_t = m_t + \varepsilon_t,$$

donde

m_t : denota la componente de tendencia en el periodo t ;

ε_t : denota un término de error aleatorio con media cero;

[Chatfield, 2003].

Hay varias formas matemáticas muy utilizadas para las tendencias. A continuación se presentan algunas:

- Lineal: $X_t = \alpha + \beta t + \varepsilon_t$
- Exponencial: $X_t = \alpha \exp\{\beta t\} \varepsilon_t$
- Logística: $X_t = \frac{\alpha}{1 + \beta \exp\{\lambda t\}} \varepsilon_t$
- Gompertz: $X_t = \alpha \exp\{\beta \exp\{\lambda t\}\} \varepsilon_t$

La tendencia más sencilla es la **tendencia lineal**, $m_t = \alpha + \beta t$, donde α, β son constantes, este último también es llamado componente de tendencia. El coeficiente β representa el incremento medio en X por unidad de tiempo. Otros autores prefieren denominarlo como la tendencia del parámetro β , el cual representa la pendiente del modelo lineal; con base en esta última interpretación, la tendencia es el cambio en el nivel medio por unidad de tiempo. La tendencia en la ecuación $m_t = \alpha + \beta t$ es una función determinística del tiempo y se conoce como tendencia lineal global. Recuérdese que, si β es mayor que cero, indica que hay un crecimiento lineal, y si β es menor que cero nos indica que hay una declinación durante un largo periodo. Para el caso en que $\beta = 0$, quiere decir que no hay crecimiento o declinación durante un largo periodo en la serie, y se dice que es una *serie sin tendencia*. Una forma conveniente de estimar α y β es por mínimos cuadrados; [Bowerman, 2009].

Se utiliza la ecuación de **tendencia exponencial** cuando la tendencia refleja un porcentaje de crecimiento casi constante; $\exp\{\beta\}$ es la tasa de crecimiento porcentual por unidad de tiempo, mientras que α es el valor de la tendencia en el instante t . La ecuación de tendencia exponencial se debería utilizar cuando la gráfica de logaritmo natural de X_t contra t parece casi lineal. Si

$$X_t = \alpha \exp\{\beta t\} \varepsilon_t,$$

entonces

$$\log X_t = \log \alpha + \beta t + \log \varepsilon_t;$$

podemos estimar los coeficientes: $\log \alpha$ y β a través de un modelo de regresión lineal² de $\log X_t$ contra t ; [Hildebrand, 1998].

Es un riesgo suponer que el crecimiento exponencial continuará indefinidamente. Las ecuaciones de **tendencia logística** y de **tendencia Gompertz** proporcionan tendencias similares con forma de **S**, véase la Figura 2.6. Ajustar, ya sea una ecuación logística o una ecuación de Gompertz, es un poco más difícil que ajustar ecuaciones de tendencia lineal o exponencial. No es posible transformar ninguna de estas ecuaciones con forma de **S** en un modelo de regresión, para esto hay métodos numéricos más complicados (por ejemplo, el método de Newton-Raphson, o las variantes de este: el método de Newton-Raphson modificado y el método de la secante); [Hildebrand, 1998].

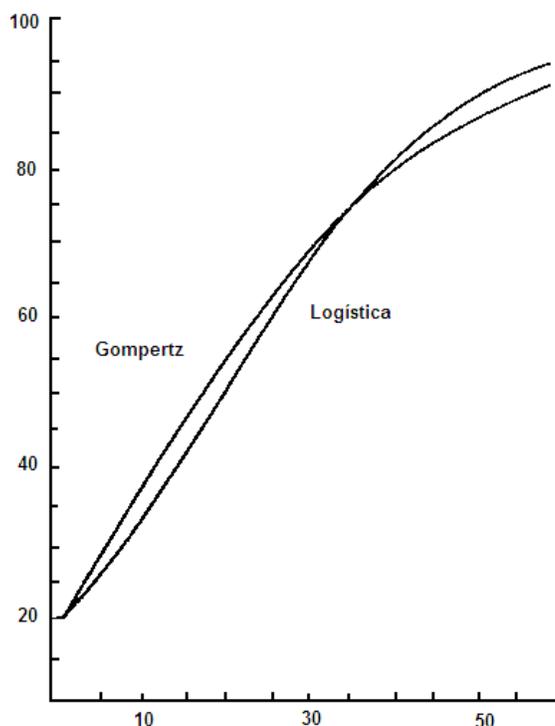


Figura 2.6: Curva típica de logística y de Gompertz.

2.2.2. Análisis de series que contienen variación estacional

En cuanto a las series de tiempo que manifiestan variación estacional, se definen dos tipos de variación estacional: si la magnitud del cambio estacional no depende del nivel de la

²El modelo de regresión lineal simple de Y contra X está dado por $Y = \beta_0 + \beta_1 X + \varepsilon$, donde Y es la variable respuesta, X la variable predictora, β_0 y β_1 son los coeficientes de regresión, y ε es un término de error (Véase [Montgomery, 2008], pp. 73-75).

serie temporal, se dice que esta manifiesta una *variación estacional constante*, observe la Figura 2.7, esta serie de tiempo tiene una tendencia lineal creciente, pero la dimensión de los cambios estacionales es la misma cuando se incrementa el nivel de la serie temporal; si la magnitud del cambio estacional depende del nivel de la serie temporal, se dice que la serie temporal muestra una *variación estacional creciente*, véase la Figura 2.8; [Bowerman, 2009].

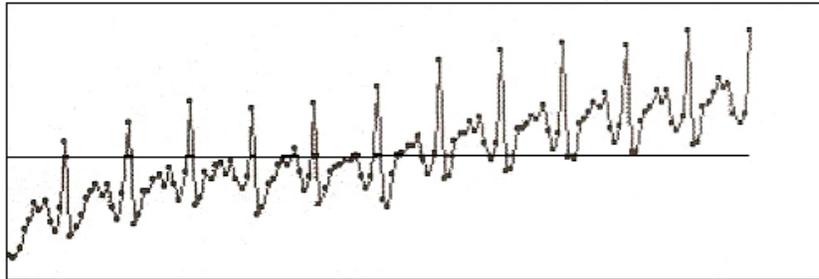


Figura 2.7: Serie temporal con variación estacional constante.

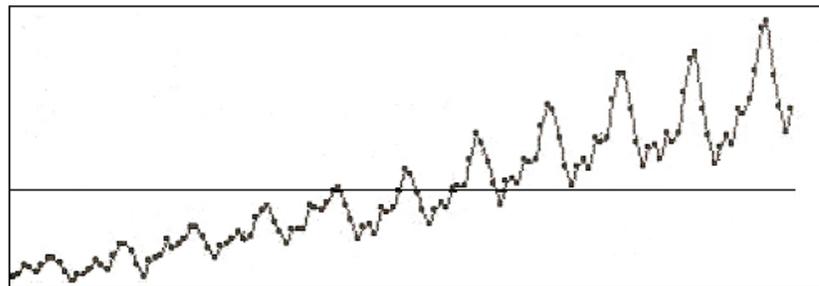


Figura 2.8: Serie temporal con variación estacional creciente.

Un modelo que adopta sólo estacionalidad, está dado por

$$X_t = s_t + \varepsilon_t, \quad (2.1)$$

donde

- s_t : denota la componente de estacionalidad;
- ε_t : denota un término de error en el periodo t .

La estacionalidad en la Ecuación (2.1) es una función del tiempo y se conoce como componente de estacionalidad.

2.2.3. Modelos con tendencia y estacionalidad

Aquí se distingue entre estacionalidad aditiva y la estacionalidad multiplicativa, las cuales son constantes de año en año; [Chatfield, 2003].

Modelos estacionales y con tendencia:

1. Aditivo:

$$X_t = m_t + s_t + \varepsilon_t.$$

2. Multiplicativo:

$$X_t = m_t * s_t * \varepsilon_t.$$

3. Mixto:

$$X_t = m_t * s_t + \varepsilon_t.$$

Donde:

X_t = el valor de la serie temporal en el periodo t ;

m_t = componente de tendencia en el periodo t ;

s_t = es una función que representa la componente estacional;

ε_t = el término de error aleatorio en el periodo t .

Una suposición usual es que ε_t sea una componente aleatoria o ruido blanco con media cero y varianza constante.

Un modelo aditivo es adecuado, por ejemplo, cuando s_t no depende de otras componentes, como m_t , es decir, se puede utilizar para modelar series de tiempo que muestran una variación estacional constante. Si por el contrario, la estacionalidad varía con la tendencia, es decir, la serie temporal manifiesta una variación estacional creciente o decreciente, el modelo más adecuado es el multiplicativo; sin embargo, dicho modelo puede ser transformado en aditivo tomando logaritmos.

Los índices estacionales $\{s_t\}$ normalmente se eligen de tal forma que cambien lentamente con el tiempo. Es decir, se debe tener $s_t \simeq s_{t-s}$, donde s es el número de observaciones por año. Regularmente los índices se normalizan y se hace la suma igual a cero en el caso aditivo, o el promedio igual a 1 en el caso multiplicativo.

Análogo al caso del análisis en que las series muestran tendencia, en las series que muestran variación estacional, el análisis depende de si se desea medir y/o eliminar esta variación.

Es usual en series que muestran una tendencia pequeña estimar el efecto estacional para un periodo de tiempo determinado. Esto se logra obteniendo el promedio de cada observación del periodo en cuestión y restándole el promedio anual; esto en el caso aditivo. En el caso multiplicativo, se debe dividir cada observación de periodo por el promedio

anual; [Hildebrand, 1998].

Ahora, el problema que se presenta, es modelar adecuadamente las componentes de la serie. Una clase importante de series de tiempo, son los procesos estacionarios, procesos que están libres de las componentes de tendencia y estacionalidad, y que juegan un papel preponderante en el análisis de serie de tiempo.

2.3. Eliminación de las componentes tendencia y estacionalidad

En esta sección se describe, cómo puede eliminarse la tendencia y estacionalidad de una serie de tiempo. La decisión más importante que se debe tomar al eliminar una tendencia es la especificación de la forma matemática de la ecuación de tendencia (Sección 2.2.1.).

Una aproximación natural al análisis de una serie de tiempo consiste en eliminar primero la tendencia y después los efectos estacionales [Hildebrand, 1998]. La esperanza es que este procedimiento ponga de relieve, con mayor claridad, cualesquiera ciclos a largo plazo.

En general, se recomienda de acuerdo a [Brockwell, 2002] graficar los datos como un primer paso en el análisis de cualquier serie de tiempo; si existen aparentes discontinuidades en la serie, tal como un cambio de nivel, se puede analizar la serie por partes o por segmentos. La inspección de la gráfica, puede también sugerir representar los datos como una realización de un proceso³ que sea del tipo

$$X_t = m_t + s_t + \varepsilon_t.$$

2.3.1. Eliminación de la tendencia por diferenciación

El método de diferenciación consiste en crear una nueva serie de tiempo Y_2, Y_3, \dots, Y_N a partir de la serie original, X_1, X_2, \dots, X_n mediante:

$$Y_t = X_t - X_{t-1} = \nabla X_t, \quad \text{para } t = 2, 3, \dots, N.$$

Definición 2.3.1 Retraso-1

Se define el retraso-1 por el operador diferencial ∇ dado por,

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

donde B es el operador de cambio de retraso,

$$BX_t = X_{t-1}.$$

³También llamado modelo de descomposición clásica.

Esto se generaliza, para obtener las potencias de B y ∇ definidas por

$$B^j X_t = X_{t-j}$$

y

$$\nabla^j(X_t) = \nabla\left(\nabla^{j-1}(X_t)\right), \quad j \geq 1.$$

En particular, $\nabla^0(X_t) = X_t$. Los polinomios en B y ∇ son trabajados de la misma forma como funciones polinomiales de variables reales. Por ejemplo,

$$\begin{aligned} \nabla^2 X_t &= \nabla\left(\nabla(X_t)\right) = (1-B)(1-B)X_t \\ &= (1-2B+B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned}$$

Si el operador ∇ es aplicado a una función de tendencia lineal, $m_t = \alpha + \beta t$, entonces se obtiene la función constante,

$$\nabla m_t = m_t - m_{t-1} = (\alpha + \beta t) - (\alpha + \beta(t-1)) = \beta.$$

En la misma forma, cualquier tendencia polinomial de grado k puede reducirse a una constante aplicando el operador ∇^k . Por ejemplo, si $X_t = m_t + \varepsilon_t$, donde

$$m_t = \sum_{j=0}^k c_j t^j$$

y ε_t es estacionario con media cero. Aplicando ∇^k se tiene

$$\nabla^k X_t = k!c_k + \nabla^k \varepsilon_t,$$

el cual es un proceso estacionario con media $k!c_k$.

Estas consideraciones sugieren posiblemente, dada cualquier sucesión $\{x_t\}$ de datos, aplicar el operador ∇ repetidamente hasta encontrar una sucesión $\{\nabla^k X_t\}$ que pueda ser modelada como una realización de un proceso estacionario [Brockwell, 2002]. Por lo tanto, la serie que originalmente constaba de N observaciones, se reducirá a una serie de solamente $N - k$ observaciones, aplicando el operador de retraso.

Ejemplo 2.3.1 Aplicando el operador ∇^k

Aplicando el operador ∇ con retraso-1, a los datos de la Figura 2.2, que muestra la población de Estados Unidos en intervalos de diez años, desde 1790 hasta 1990, se obtiene una nueva serie que reemplaza a la serie original $\{X_t\}$ por una serie una vez diferenciada $\{X_t - X_{t-1}\}$. Repitiendo este proceso se obtiene una serie dos veces diferenciada $\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$, ver Figura 2.9. Por lo tanto, se observa que es suficiente aplicar el operador de diferenciación dos veces para obtener una serie sin tendencia.



Figura 2.9: Serie dos veces diferenciada, $\{\nabla^2 X_t\}$, derivada de los datos de la población en U.S.A.

2.3.2. Eliminación de la tendencia y la estacionalidad por diferenciación

El efecto estacional en una serie de tiempo también puede eliminarse mediante un filtro simple lineal, el cual es llamado *diferenciación estacional*.

La técnica de diferenciación que se aplica a datos no estacionales se puede repartir en d periodos de estacionalidad, introduciendo el retraso- d , es decir, diferenciando el operador ∇_d , definido por

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t.$$

Nota: Este operador no debe confundirse con el operador $\nabla^d = (1 - B)^d$.

Aplicando el operador ∇_d , al modelo

$$X_t = m_t + s_t + \varepsilon_t,$$

donde s_t tiene periodo d , se obtiene

$$\nabla_d X_t = m_t - m_{t-d} + \varepsilon_t - \varepsilon_{t-d},$$

el cual da una descomposición de la diferencia $\nabla_d X_t$, entre una componente de tendencia ($m_t - m_{t-d}$) y un término de error ($\varepsilon_t - \varepsilon_{t-d}$). De esta forma, la tendencia ($m_t - m_{t-d}$) puede ser eliminada, en particular, aplicando el operador ∇^d .

Ejemplo 2.3.2 Aplicando el operador ∇_d y ∇^k

Aplicando el operador ∇_d con periodo $d = 12$ a los datos de la Figura 2.3, que muestra

los registros mensuales por muertes accidentales de 1973 a 1978, se obtiene la nueva serie $\{\nabla_{12}X_t\}$ que se observa en la Figura 2.10, la cual presenta una tendencia creciente. Si ahora se aplica el operador ∇^k con retraso $k = 1$ a la serie $\{\nabla_{12}X_t\}$, entonces se obtiene la gráfica de la serie $\{\nabla^1\nabla_{12}X_t\}$ que se muestra en la Figura 2.11, la cual no presenta las componentes de tendencia y estacionalidad.

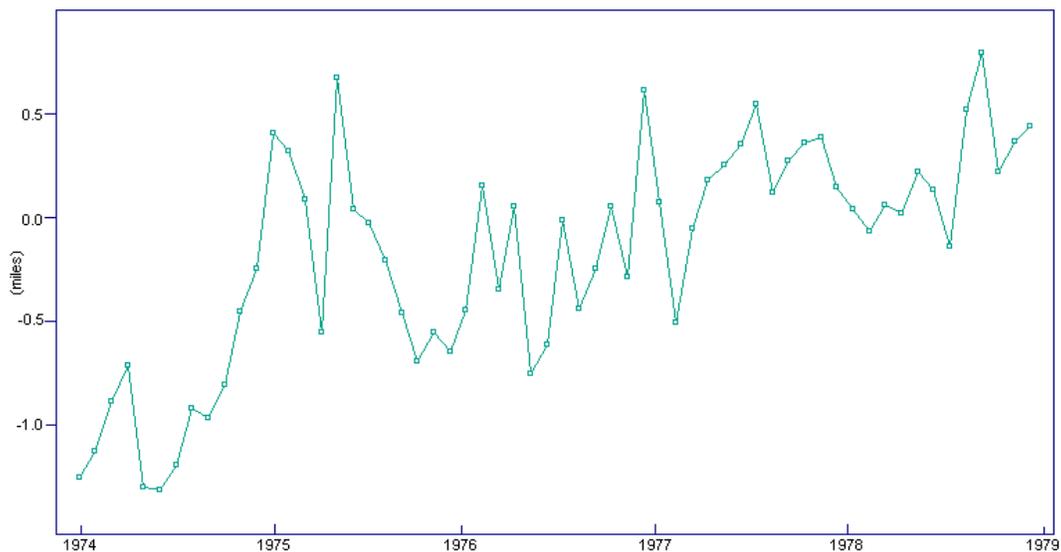


Figura 2.10: Serie diferenciada $\{\nabla_{12}X_t\}$ de las muertes accidentales.

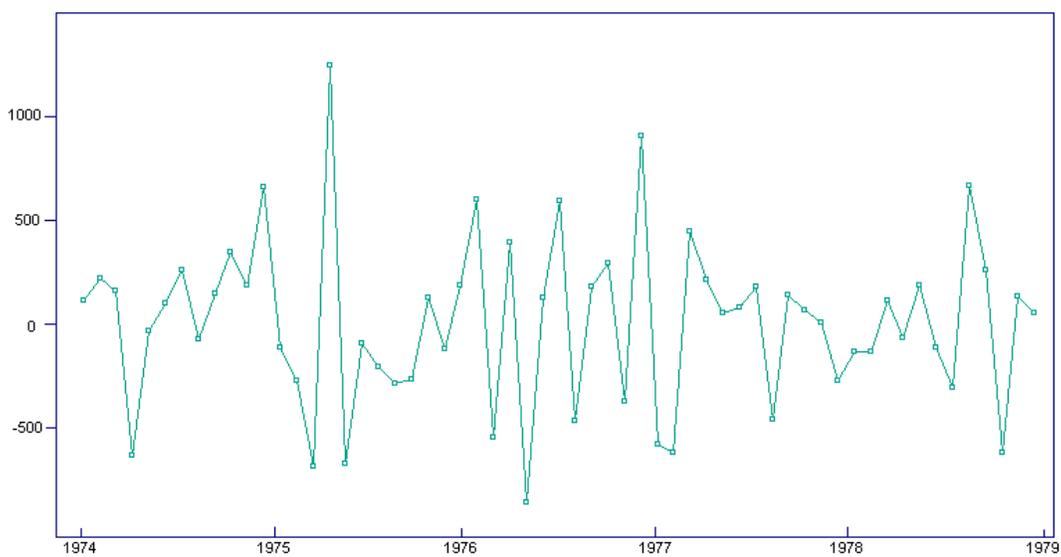


Figura 2.11: Serie diferenciada $\{\nabla^1\nabla_{12}X_t\}$ de las muertes accidentales.

2.4. Modelos de ajuste para series de tiempo

En esta sección se introducen algunos tipos de modelos probabilísticos que pueden ser usados para describir una serie de tiempo. Se discute el problema de ajustar un modelo adecuado a una serie de tiempo observada a tiempo discreto. Cada uno de los modelos, se aborda de acuerdo a su complejidad de estudio, comenzando con el modelo más sencillo hasta el modelo más sofisticado, en su aplicación.

En cada uno de los modelos se construye una nueva serie de tiempo Y_2, Y_3, \dots, Y_N a partir de la serie original, X_1, X_2, \dots, X_n .

2.4.1. Medias móviles

Medias móviles es uno de los métodos de suavizamiento o ajuste más utilizados. Para obtener la serie $\{Y_t: t \in T'\}$, se promedia los M valores (periodos) más recientes de la serie $\{X_t: t \in T\}$; [Hildebrand, 1998].

Ejemplo 2.4.1 Medias móviles

En los últimos 36 meses, la venta mensual de anuncios locales pagados por televisión (medida en minutos por día) ha sido la siguiente:

Mes	1	2	3	4	5	6	7	8	9	10	11	12
Ventas	86	80	85	93	96	102	97	89	96	87	82	81
Mes	13	14	15	16	17	18	19	20	21	22	23	24
Ventas	87	89	97	88	95	87	81	79	82	85	96	93
Mes	25	26	27	28	29	30	31	32	33	34	35	36
Ventas	99	103	96	85	78	83	90	96	85	82	89	96

Se realizarán ajustes con medias móviles de tres y cinco meses. La media móvil correspondiente al primer trimestre es $(86 + 80 + 85)/3 = 83.6667 \approx 83.7$. La media móvil correspondiente a los primeros cinco meses es $(86 + 80 + 85 + 93 + 96)/5 = 88.0$. Haciendo todos los cálculos pertinentes se obtienen los siguientes resultados:

Media móvil trimestral:

-	-	83.7	86.0	91.3	97.0	98.3	96.0	94.0	90.7	88.3	93.3
83.3	85.7	91.0	91.3	93.3	90.0	87.7	82.3	80.7	82.0	87.7	91.3
96.0	98.3	99.3	94.7	86.3	82.0	83.7	89.7	90.3	87.7	85.3	89.0

Media móvil de cinco meses:

-	-	-	-	88.0	91.2	94.6	95.4	96.0	94.2	90.2	87.0
86.6	85.2	87.2	88.4	91.2	91.2	89.6	86.0	84.8	82.8	84.6	87.0
91.0	95.2	97.4	95.2	92.2	89.0	86.4	86.4	86.4	87.2	88.4	89.6

Con estos nuevos datos, se obtiene la serie suavizada $\{Y_t: t \in T'\}$, para el caso de la media móvil trimestral ya ha perdido dos valores, es decir, esta tiene 34 observaciones (Figura 2.12); mientras que la nueva serie para la media móvil a cinco meses ha perdido cuatro observaciones (Figura 2.12).

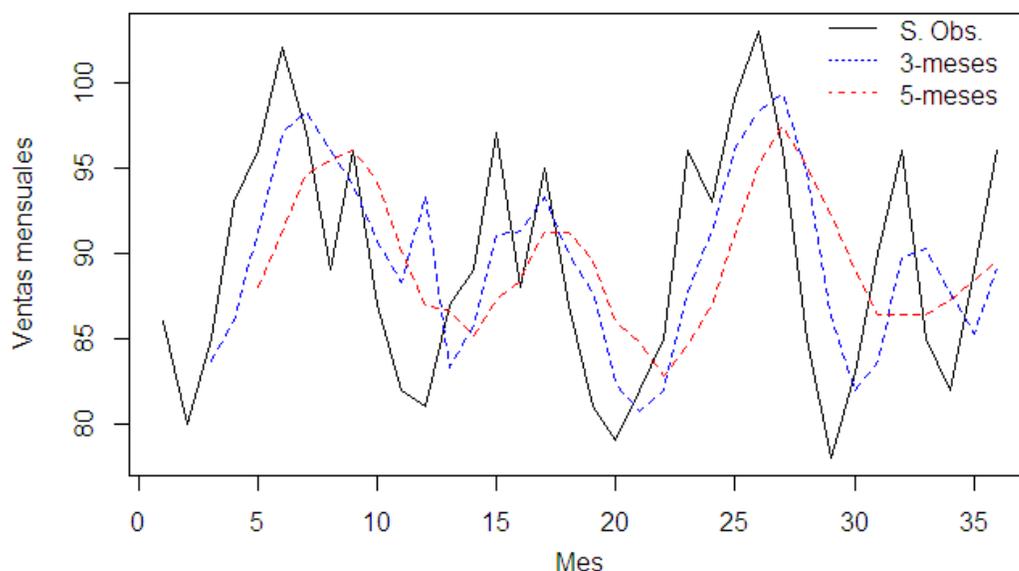


Figura 2.12: Medias móviles de tres y cinco meses.

2.4.2. Medianas móviles

Una variación de la idea de suavizamiento a través de la media móvil es el método de las medianas móviles. El método incluye el cálculo de la mediana, en vez de la media, de los M valores más recientes de la serie. En vista de que la mediana, a diferencia de la media, no se ve afectada por valores extremos, los ajustes basados en las medianas móviles no se ven afectados por un valor ocasional, sea éste muy grande o muy pequeño.

La mediana móvil para los datos del Ejemplo 2.4.1 correspondiente a los cinco primeros meses es la mediana $(86, 80, 85, 93, 96) = 86$, continuando con todos los cálculos obtenemos:

Mediana móvil de cinco meses:

-	-	-	-	86	93	96	96	96	96	89	87
87	87	87	88	89	89	88	87	82	82	82	85
93	96	96	96	96	85	85	85	85	85	89	89

La serie suavizada $\{Y_t: t \in T'\}$ ha perdido cuatro observaciones (Figura 2.13).

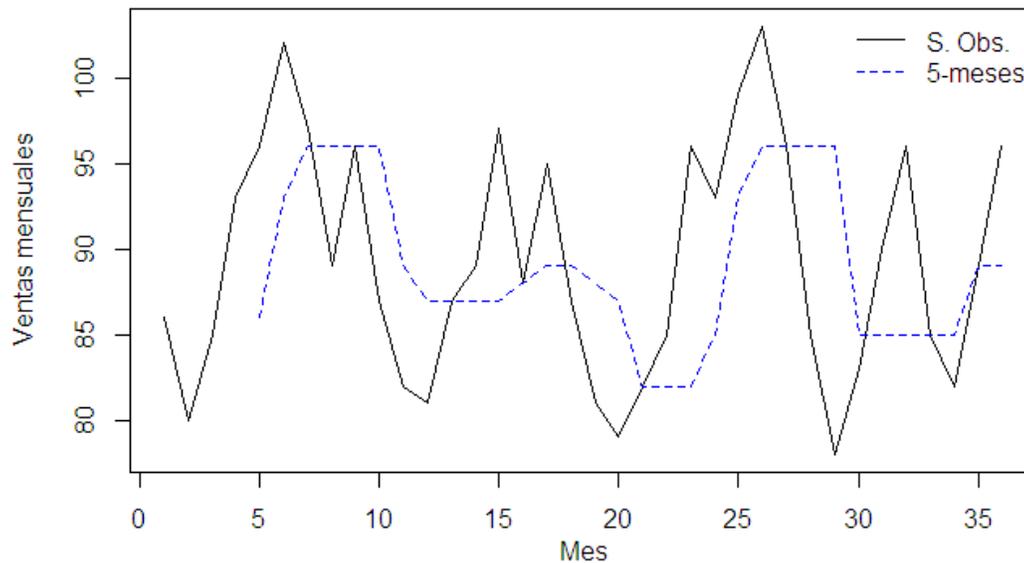


Figura 2.13: Mediana móvil a cinco meses.

► Desventajas de los modelos de medias y medianas móviles

1. La desventaja de los modelos de medias y medianas móviles radica en la elección del periodo de tiempo, puesto que un valor extremadamente grande de M ocasiona un cambio muy lento en el suavizamiento, impidiendo determinar variaciones a corto plazo; mientras que un valor muy pequeño de M ocasiona que el ajuste sea muy inestable. A menudo, para clarificar la elección de M se compara el comportamiento del ajuste con distintos valores de M .
2. Otorgan el mismo peso a los valores relativamente viejos y a los valores recientes.

2.4.3. Suavizamiento exponencial

El suavizamiento exponencial es un método alternativo a la segunda desventaja de los modelos de medias y medianas móviles, puesto que proporciona un mayor peso a los va-

lores más recientes.

La suavización exponencial es un método que muestra su mayor eficacia cuando los componentes (tendencia y variación estacional) de las series de tiempo podrían manifestar cambios en el tiempo.

En el suavizamiento exponencial, la estimación \widehat{X}_{t+1} del valor real X_{t+1} se define como

$$\widehat{X}_{t+1} = \alpha X_t + \alpha(1 - \alpha)X_{t-1} + \alpha(1 - \alpha)^2 X_{t-2} + \dots$$

El número α se encuentra entre 0 y 1 y se llama constante de suavizamiento. Obsérvese que el mayor peso se le da al valor más reciente de la serie temporal, y los pesos otorgados a los valores anteriores disminuyen rápidamente. De hecho, los pesos α , $\alpha(1 - \alpha)$, $\alpha(1 - \alpha)^2$, $\alpha(1 - \alpha)^3$, ... decrecen exponencialmente a cero; por ello el nombre de *suavizamiento exponencial* [Hildebrand, 1998].

Si se escoge $\alpha = 0.6$, en tal caso,

$$\widehat{X}_{t+1} = 0.6X_t + 0.24X_{t-1} + 0.096X_{t-2} + 0.0384X_{t-3} + \dots$$

En la práctica se hace uso de la siguiente ecuación

$$\widehat{X}_{t+1} = \alpha X_t + (1 - \alpha)\widehat{X}_t,$$

la cual rápidamente se aproxima al valor de la serie en el infinito. Para el uso de esta ecuación se debe elegir un valor inicial; la elección más común es $\widehat{X}_1 = X_1$ y α se selecciona de tal forma que minimice el promedio absoluto o el error cuadrático medio en los datos.

Hay varios métodos para suavizar una serie dentro del suavizamiento exponencial [Bowerman, 2009], se mencionarán algunos:

- **Suavización exponencial simple**

El método de suavización exponencial simple se usa para ajustar una serie temporal cuando no hay tendencia o patrón estacional, pero la media de la serie de tiempo $\{X_t\}$ cambia lentamente en el tiempo.

El procedimiento de suavización exponencial simple, inicia con el cálculo de una estimación inicial ℓ_0 de la media de la serie de tiempo en el periodo $t = 0$.

La estimación de la media en el periodo t está dada por la ecuación de suavización:

$$\ell_t = \alpha X_t + (1 - \alpha)\ell_{t-1},$$

donde α es una constante de suavización entre 0 y 1. Por lo tanto, el ajuste efectuado en el periodo t para X_{t+h} se define como

$$\widehat{X}_{t+h} = \ell_t, \quad \text{para } h = 1, 2, \dots$$

■ **Método de Hold de la suavización exponencial corregida de la tendencia**

Es un método de suavizamiento para ajustar una serie temporal que tiene localmente una tendencia, pero una tasa de crecimiento (o pendiente) que cambia con el tiempo sin ningún patrón estacional. Entonces, la estimación ℓ_t del nivel de la serie de tiempo y la estimación b_t de la tasa de crecimiento de la serie de tiempo en un periodo t están dadas por las ecuaciones de suavización:

$$\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

y

$$b_t = \gamma(\ell_t - \ell_{t-1}) + (1 - \gamma)b_{t-1},$$

donde

- α y γ son constantes de suavización entre 0 y 1.
- ℓ_{t-1} es la estimación en el tiempo $t - 1$ para la media.
- b_{t-1} es la estimación en el tiempo $t - 1$ para la tasa de crecimiento.

Así, la estimación efectuada en el periodo t para X_{t+h} se define como

$$\widehat{X}_{t+h} = \ell_t + hb_t, \quad \text{para } h = 1, 2, \dots$$

■ **Método aditivo de Hold-Winters**

Este método es apropiado para una serie que manifiesta tendencia lineal y variación estacional constante. Las ecuaciones de suavización están dadas por:

$$\ell_t = \alpha(X_t - sn_{t-d}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$

$$b_t = \gamma(\ell_t - \ell_{t-1}) + (1 - \gamma)b_{t-1},$$

y

$$sn_t = \delta(X_t - \ell_t) + (1 - \delta)sn_{t-d},$$

donde

- α , γ y δ son constantes de suavización entre 0 y 1.
- ℓ_{t-1} es la estimación en el tiempo $t - 1$ para la media.
- b_{t-1} es la estimación en el tiempo $t - 1$ para la tasa de crecimiento.
- sn_{t-d} es la estimación en el periodo $t - d$ del factor estacional.

Por lo tanto, el ajuste efectuado en el periodo t para X_{t+h} se define como

$$\widehat{X}_{t+h} = \ell_t + hb_t + sn_{t+h-d}, \quad \text{para } h = 1, 2, \dots$$

■ **Método multiplicativo de Hold-Winters**

Este método es apropiado para una serie de tiempo con tendencia lineal local con variación estacional creciente (multiplicativa). Entonces, el factor estacional de la serie en el periodo t se obtiene de las siguientes ecuaciones de suavizamiento:

$$\ell_t = \alpha(X_t/sn_{t-d}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}),$$

$$b_t = \gamma(\ell_t - \ell_{t-1}) + (1 - \gamma)b_{t-1},$$

y

$$sn_t = \delta(X_t/\ell_t) + (1 - \delta)sn_{t-d},$$

donde

- α , γ y δ son constantes de suavización entre 0 y 1.
- ℓ_{t-1} es la estimación en el tiempo $t - 1$ para la media.
- b_{t-1} es la estimación en el tiempo $t - 1$ para la tasa de crecimiento.
- sn_{t-d} es la estimación en el periodo $t - d$ del factor estacional.

Entonces, la estimación en el periodo t para X_{t+h} se define como

$$\hat{X}_{t+h} = (\ell_t + hb_t)sn_{t+h-d}, \quad \text{para } h = 1, 2, \dots,$$

donde sn_{t+h-d} es la estimación más reciente del factor estacional para la estación correspondiente al periodo $t + h$.

■ **Suavización exponencial de la tendencia amortiguada**

Método propuesto por Gardner y McKenzie, apto para ajustar una serie cuya tasa de crecimiento no se mantendrá en el futuro, los efectos de lo cual tienen que ser amortiguados, es decir, hay que reducir de tamaño la tasa de crecimiento de modo que disminuya el incremento o decremento del ajuste. Este método se desglosa en el análisis de los métodos aditivo y multiplicativo de Hold-Winters.

2.4.4. El enfoque de Box-Jenkins

Los modelos anteriores tratan de métodos primitivos para depurar datos en series de tiempo, sin considerar a fondo el proceso que los generó en primera instancia.

Definición 2.4.1 Proceso lineal general

El proceso lineal general expresa al proceso $\{X_t\}$ en función de $\{Z_t\}$ mediante la relación lineal

$$X_t = \mu + Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots \quad (2.2)$$

en donde, μ denota a la media del proceso⁴, $\{\psi_j\}$ es una sucesión de constantes, y $\{Z_t\}$ es ruido blanco con media cero y varianza σ^2 constante; [Guerrero, 2003].

⁴o, dicho de otra manera, el “nivel” del proceso.

En términos del operador de retraso B , la Ecuación (2.2) se puede reescribir como

$$X_t = \psi(B)Z_t,$$

donde $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ y $\psi_0 = 1$.

El enfoque de Box-Jenkins proporciona modelos estacionarios para el ajuste de las observaciones, dichos modelos son casos especiales del modelo (2.2).

Una serie de tiempo es estacionaria si no hay cambios sistemáticos en los datos, es decir, la media y la varianza se mantienen constantes a lo largo del tiempo. Caso contrario a las series no estacionarias, estas pueden mostrar cambios en la varianza, exhibir tendencia o presentar efectos estacionales; motivos que dificultan obtener un modelo de ajuste adecuado.

De manera formal, se define la estacionariedad y estacionariedad estricta de una serie de tiempo:

Definición 2.4.2 Estacionariedad

La serie de tiempo $\{X_t: t \in T\}$, se dice que es estacionaria si:

- $Var(X_t) < \infty$, para todo $t \in T$;
- $E(X_t) = \mu$, para todo $t \in T$;
- $\gamma_X(r,s) = \gamma_X(r+t, s+t)$, para todo $r, s, t \in T$; la función de autocovarianza toma el mismo valor para dos v.a's que estén separadas por un rezago t en el proceso, independientemente de donde se encuentren situadas estas v.a's en el tiempo.

En la literatura, el término estacionariedad es frecuentemente utilizado para referirse a la *estacionariedad de segundo orden* o *estacionariedad débil* ([Guerrero, 2003], pp. 15-17).

Definición 2.4.3 Estacionariedad estricta

La serie de tiempo $\{X_t: t \in T\}$ se dice que es estrictamente estacionaria si la distribución de probabilidad conjunta de $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ y $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$ es la misma para todo entero $h, n \geq 1$ y para todo $t_1, t_2, \dots, t_n \in T$.

La estacionariedad estricta también es conocida como *estacionariedad fuerte*.

La relación existente entre la estacionariedad y estacionariedad estricta radica en que si la serie de tiempo $\{X_t: t \in T\}$ es estrictamente estacionaria, entonces es estacionaria; el inverso de la afirmación no se cumple ([Brockwell, 1987], pp. 12-13).

Ejemplo 2.4.2 Si $\{X_t\}$ es una secuencia de variables aleatorias independientes, tal que X_t tiene una distribución exponencial⁵ con media uno, cuando t es impar; y si t es par, entonces X_t tiene una distribución normal⁶ con media uno y varianza uno. Por lo tanto, $\{X_t: t \in \mathbb{T}\}$ es estacionaria con función de autocovarianza:

$$\gamma_X(t, t+h) = \text{Cov}(X_t, X_{t+h}) = \begin{cases} \text{Var}(X_t) = 1, & \text{si } h = 0; \\ 0, & \text{si } h \neq 0. \end{cases}$$

La razón por la que es cero la covarianza si $h \neq 0$, es por suponer que X_1, X_2, \dots, X_T son v.a's independientes, de aquí se sigue que

$$\text{Cov}(X_t, X_{t+h}) = E[X_t - E(X_t)]E[X_{t+h} - E(X_{t+h})],$$

y por propiedades del valor esperado, se cumple que

$$E[X_t - E(X_t)] = E(X_t) - E(X_t).$$

Sin embargo, dado que X_1 y X_2 no tienen la misma distribución de probabilidad, entonces $\{X_t: t \in \mathbb{T}\}$ no puede ser estrictamente estacionaria.

Recuérdese que siempre se hará referencia a la Definición 2.4.2 para modelos estacionarios.

Observación: Si $\{X_t: t \in \mathbb{T}\}$ es una serie estacionaria, entonces $\gamma_X(r, s) = \gamma_X(r-s, 0)$ para todo $r, s \in \mathbb{T}$. Así, la función de autocovarianza se redefine como una función que sólo depende de la diferencia de tiempo $(r-s) = h$, también llamado *retardo* o *rezago* h ([Brockwell, 1987], pp. 11-12):

$$\gamma_h = \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t), \quad \text{para todo } t \in \mathbb{T}, \text{ y } h \in \mathbb{Z}. \quad (2.3)$$

La autocovarianza mide la dependencia lineal entre dos puntos de la misma serie, observados en diferentes tiempos. Recuérdese de la estadística clásica que si $\gamma_h = 0$, entonces X_t y X_{t+h} no están linealmente relacionadas, pero todavía puede haber algún tipo de estructura de dependencia entre ellas; sin embargo, si X_t y X_{t+h} son normal bivariada⁷, $\gamma_h = 0$ asegura su independencia ([Shumway, 2010], pp. 19).

Análogamente, la **función de autocorrelación**, $\rho_h: \mathbb{Z} \rightarrow \mathbb{R}$, se define mediante:

$$\begin{aligned} \rho_h &= \text{Corr}(X_{t+h}, X_t) \\ &= \frac{\text{Cov}(X_{t+h}, X_t)}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}} = \frac{\gamma_h}{\gamma_0}, \quad \text{para todo } t \in \mathbb{T}, \text{ y } h \in \mathbb{Z}, \end{aligned} \quad (2.4)$$

⁵Véase [Degroot, 1988], pp. 276-278.

⁶Véase [Degroot, 1988], pp. 251-258.

⁷Véase [Degroot, 1988], pp. 286-290.

y se denota por FAC (*Autocorrelation Function*), la cual mide la correlación entre X_t y X_{t+h} ([Brockwell, 1987], pp. 12-13).

Debido a que la función de autocorrelación está definida en términos de la función de autocovarianza, se dice que, ρ_h indica el grado de dependencia lineal en los datos. Si $\rho_h = 0$, no existe correlación entre X_t y X_{t+h} , es decir, X_t y X_{t+h} son independientes linealmente; si $\rho_h > 0$, entonces X_t y X_{t+h} están correlacionadas positivamente, es decir, hay una dependencia lineal positiva entre ellas, pero si $\rho_h < 0$, se dice que hay una relación lineal negativa entre X_t y X_{t+h} , o dicho en otras palabras, X_t y X_{t+h} están correlacionadas negativamente.

Dado un proceso estacionario, la función de autocovarianza y autocorrelación cumplen con las siguientes propiedades:

Propiedad 2.4.1 Sea $\{X_t: t \in T\}$ un proceso estocástico estacionario, con media μ , varianza σ^2 , función de autocovarianza γ_k y función de autocorrelación ρ_k . Entonces:

- a) $\gamma_0 = \text{Var}(X_t) \geq 0$ y $\rho_0 = 1$, para todo $t \in T$;
- b) $|\rho_k| \leq 1$ y $|\gamma_k| \leq \gamma_0$, para $k \in \mathbb{Z}$;
- c) $\gamma_k = \gamma_{-k}$ y $\rho_k = \rho_{-k}$ para todo k , es decir, γ_k y ρ_k son funciones pares, y de aquí que sean simétricas con respecto a $t = 0$; [Montgomery, 2008].

Demostración.

- a) La varianza se cumple por la Ecuación (2.3) cuando $h = 0$, y $\rho_0 = 1$ se satisface por la Ecuación (2.4) para $h = 0$.
- b) Para mostrar esta propiedad, se sabe que,

$$\text{Var}(aX_t + bX_{t+k}) \geq 0$$

para cualquier constante a, b , puesto que una varianza es siempre no-negativa.

Luego,

$$\begin{aligned} 0 &\leq \text{Var}(aX_t + bX_{t+k}) \\ &= a^2\text{Var}(X_t) + b^2\text{Var}(X_{t+k}) + 2ab\text{Cov}(X_t, X_{t+k}) \\ &= a^2\sigma^2 + b^2\sigma^2 + 2ab\gamma_k \\ &= (a^2 + b^2)\sigma^2 + 2ab\gamma_k. \end{aligned}$$

- Si, $a = b = 1$, se tiene que

$$0 \leq \sigma^2 + \gamma_k.$$

Así, $-\gamma_k \leq \sigma^2$ y como se sabe que $\gamma_0 = \sigma^2$ y $\rho_k = \frac{\gamma_k}{\gamma_0}$, se cumple que

$$-1 \leq \rho_k.$$

- Si, $a = 1$ y $b = -1$, se tiene que $0 \leq \sigma^2 - \gamma_k$. Así, $\gamma_k \leq \sigma^2$ y

$$\rho_k \leq 1.$$

Por lo tanto, $-1 \leq \rho_k \leq 1$, es decir $|\rho_k| \leq 1$. Como consecuencia se obtiene que $|\gamma_k| \leq \gamma_0$.

- c) Dado que la serie es estacionaria, por definición se cumple que $\gamma_X(r, s) = \gamma_X(r+t, s+t)$. Entonces

$$\begin{aligned} \gamma_k &= \text{Cov}(X_{t+k}, X_t), && \text{por Ecuación (2.3);} \\ &= \gamma_X(t+k, t), && \text{por Definición 1.3.19;} \\ &= \gamma_X(t+k-k, t-k) \\ &= \gamma_X(t, t-k) \\ &= \text{Cov}(X_t, X_{t-k}) \\ &= \text{Cov}(X_{t-k}, X_t) \\ &= \gamma_{-k}, && \text{por Ecuación (2.3).} \end{aligned}$$

Luego,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}. \quad \blacksquare$$

Ejemplo 2.4.3 Sea $\{X_t: t \in \mathbb{T}\}$ un proceso estocástico estacionario, tal que

$$X_t = \frac{1}{3}(Z_{t-1} + Z_t + Z_{t+1}),$$

donde, $\{Z_t\}$ es ruido blanco con media cero y varianza constante σ_Z^2 . En este caso,

$$\gamma_h = \text{Cov}(X_{t+h}, X_t) = \text{Cov}\left\{\frac{1}{3}(Z_{t+h-1} + Z_{t+h} + Z_{t+h+1}), \frac{1}{3}(Z_{t-1} + Z_t + Z_{t+1})\right\}.$$

Cuando $h = 0$,

$$\begin{aligned} \gamma_0 &= \frac{1}{9} \text{Cov}\left\{(Z_{t-1} + Z_t + Z_{t+1}), (Z_{t-1} + Z_t + Z_{t+1})\right\} \\ &= \frac{1}{9} E\left[\left((Z_{t-1} + Z_t + Z_{t+1}) - 0\right)^2\right] \\ &= \frac{1}{9} E\left(Z_{t-1}^2 + Z_{t-1}Z_t + Z_{t-1}Z_{t+1} + Z_tZ_{t-1} + Z_t^2 + Z_tZ_{t+1} + Z_{t+1}Z_{t-1}\right. \\ &\quad \left.+ Z_{t+1}Z_t + Z_{t+1}^2\right) \end{aligned}$$

Dado que $\{Z_t\}$ son v.a.'s independientes, se cumple que $E(Z_i Z_j) = 0$, para $i \neq j$.

$$\begin{aligned} &= \frac{1}{9} \left[E(Z_{t-1}^2) + E(Z_t^2) + E(Z_{t+1}^2) \right] \\ &= \frac{1}{9} \left[\sigma_Z^2 + \sigma_Z^2 + \sigma_Z^2 \right] \\ &= \frac{3}{9} \sigma_Z^2. \end{aligned}$$

Cuando $h = 1$,

$$\begin{aligned}
\gamma_1 &= \frac{1}{9} \text{Cov} \left\{ \left(Z_t + Z_{t+1} + Z_{t+2} \right), \left(Z_{t-1} + Z_t + Z_{t+1} \right) \right\} \\
&= \frac{1}{9} E \left[\left(\left(Z_t + Z_{t+1} + Z_{t+2} \right) - 0 \right) \left(\left(Z_{t-1} + Z_t + Z_{t+1} \right) - 0 \right) \right] \\
&= \frac{1}{9} E \left(Z_t Z_{t-1} + Z_t^2 + Z_t Z_{t+1} + Z_{t+1} Z_{t-1} + Z_{t+1} Z_t + Z_{t+1}^2 + Z_{t+2} Z_{t-1} + Z_{t+2} Z_t \right. \\
&\quad \left. + Z_{t+2} Z_{t+1} \right) \\
&= \frac{1}{9} \left[E(Z_t^2) + E(Z_{t+1}^2) \right] \\
&= \frac{1}{9} \left[\sigma_Z^2 + \sigma_Z^2 \right] \\
&= \frac{2}{9} \sigma_Z^2.
\end{aligned}$$

Cuando $h = -1$,

$$\begin{aligned}
\gamma_{-1} &= \frac{1}{9} \text{Cov} \left\{ \left(Z_{t-2} + Z_{t-1} + Z_t \right), \left(Z_{t-1} + Z_t + Z_{t+1} \right) \right\} \\
&= \frac{1}{9} E \left[\left(\left(Z_{t-2} + Z_{t-1} + Z_t \right) - 0 \right) \left(\left(Z_{t-1} + Z_t + Z_{t+1} \right) - 0 \right) \right] \\
&= \frac{1}{9} E \left(Z_{t-2} Z_{t-1} + Z_{t-2} Z_t + Z_{t-2} Z_{t+1} + Z_{t-1}^2 + Z_{t-1} Z_t + Z_{t-1} Z_{t+1} + Z_t Z_{t-1} + Z_t^2 \right. \\
&\quad \left. + Z_t Z_{t+1} \right) \\
&= \frac{1}{9} \left[E(Z_{t-1}^2) + E(Z_t^2) \right] \\
&= \frac{1}{9} \left[\sigma_Z^2 + \sigma_Z^2 \right] \\
&= \frac{2}{9} \sigma_Z^2.
\end{aligned}$$

De forma similar, se obtiene $\gamma_2 = \sigma_Z^2/9 = \gamma_{-2}$, y cuando $|h| > 2$, entonces $\gamma_h = 0$. Luego,

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\frac{2}{9} \sigma_Z^2}{\frac{3}{9} \sigma_Z^2} = \frac{2}{3} = \frac{\gamma_{-1}}{\gamma_0} = \rho_{-1}.$$

Análogamente, se calcula $\rho_2 = 1/3 = \rho_{-2}$, y $\rho_h = 0$, cuando $|h| > 2$. Por lo tanto, para cada $t \in \mathbb{T}$ y $h \in \mathbb{Z}$, la función de autocovarianza está dada por:

$$\gamma_h = \begin{cases} \frac{3}{9} \sigma_Z^2, & \text{si } h = 0; \\ \frac{2}{9} \sigma_Z^2, & \text{si } |h| = 1; \\ \frac{1}{9} \sigma_Z^2, & \text{si } |h| = 2; \\ 0, & \text{si } |h| > 2; \end{cases}$$

y la función de autocorrelación queda definida por:

$$\rho_h = \begin{cases} 1, & \text{si } h = 0; \\ \frac{2}{3}, & \text{si } |h| = 1; \\ \frac{1}{3}, & \text{si } |h| = 2; \\ 0, & \text{si } |h| > 2. \end{cases}$$

La función de autocorrelación, sirve para especificar, junto con la media y la varianza, a un proceso estocástico estacionario; sin embargo, desde un punto de vista práctico de series de tiempo, debe recordarse que se cuenta sólo con una realización finita del proceso, a partir de la cual se hace necesario obtener una estimación de la media, la función de autocovarianza y de autocorrelación de una serie de tiempo de longitud finita n , para poder construir un modelo apropiado para los datos ([Montgomery, 2008], pp. 28, 30).

La media μ del proceso, puede estimarse como la media muestral \bar{X} de la serie observada:

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t.$$

La función de autocovarianza muestral C_k , se usa como estimador de γ_k :

$$C_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad \text{para } k = 0, 1, 2, \dots,$$

y la función de autocorrelación ρ_k es estimada por la función de autocorrelación muestral r_k :

$$r_k = \frac{C_k}{C_0}, \quad \text{para } k = 0, 1, 2, \dots$$

Proposición 2.4.1 *El proceso lineal general es estacionario si ψ_0, ψ_1, \dots es una serie convergente, esto es, si la suma de los valores absolutos de los pesos ψ_i es finita, $\sum_{i=0}^{\infty} |\psi_i| < \infty$; [Guerrero, 2003].*

Demostración.

Se verifica que $E(X_t) = \mu$, y

$$\begin{aligned} \text{Var}(X_t) &= E\left[(X_t - \mu)^2\right] \\ &= E\left[(Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots)^2\right] \\ &= E(Z_t^2 + \psi_1^2 Z_{t-1}^2 + \psi_2^2 Z_{t-2}^2 + \dots) \\ &\quad + E(\text{productos cruzados de } \psi_i Z_{t-i} \text{ y } \psi_j Z_{t-j}) \\ &= \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty. \end{aligned}$$

La razón por la que desaparece el valor esperado de los productos cruzados, es porque en la secuencia $\{Z_t\}$ las v.a's son independientes con media cero, de aquí se sigue que

$$E(\psi_i Z_{t-i} \cdot \psi_j Z_{t-j}) = \psi_i E(Z_{t-i}) \cdot \psi_j E(Z_{t-j}) = 0, \quad \text{para toda } i \neq j.$$

$$\begin{aligned} \gamma_h &= E\left[(Z_t - \mu)(Z_{t+h} - \mu)\right] \\ &= E\left[(Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots)(Z_{t+h} + \psi_1 Z_{t+h-1} + \psi_2 Z_{t+h-2} + \dots)\right] \\ &= E(\psi_h Z_t^2 + \psi_1 \psi_{h+1} Z_{t-1}^2 + \psi_2 \psi_{h+2} Z_{t-2}^2 + \dots) \\ &\quad + E(\text{productos cruzados de } \psi_i Z_{t-i} \text{ y } \psi_j Z_{t-j}) \\ &= \sigma_Z^2 (\psi_h + \psi_1 \psi_{h+1} + \psi_2 \psi_{h+2} + \dots) \\ &= \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i \psi_{h+i} < \infty. \end{aligned}$$

En la función de autocovarianza no existe dependencia en el tiempo, pero sí de la separación h que hay entre las variables. ■

► MODELO AUTORREGRESIVO

El modelo más simple del enfoque de Box-Jenkins para el análisis de datos de series de tiempo es el *modelo autorregresivo*. La idea es utilizar los valores pasados de una serie temporal como variables independientes para estimar el siguiente valor.

Definición 2.4.4 Proceso autorregresivo

La serie de tiempo $\{X_t: t \in \mathbb{T}\}$ se dice que es un proceso autorregresivo de orden p , para algún $p \in \mathbb{N}$, y se denota por AR(p) (Autoregressive) si puede ser escrito como

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t, \quad (2.5)$$

donde $\{Z_t: t \in \mathbb{T}\}$ es un proceso de ruido blanco con media cero y varianza σ^2 constante, y $\{\alpha_i\}_{i=1}^p$ son constantes; [Chatfield, 2003].

Básicamente, (2.5) es una ecuación de regresión lineal múltiple⁸, con la característica especial de que el valor de la variable respuesta X en el periodo t , depende no de los valores en un cierto conjunto de v.a.'s predictoras, como sucede en el modelo de regresión, sino de sus propios valores observados en periodos anteriores a t y ponderados de acuerdo con los coeficientes autorregresivos $\alpha_1, \dots, \alpha_p$.

⁸El modelo de regresión lineal múltiple para k variables predictoras, está dado por $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$, donde y es la variable respuesta, $\beta_0, \beta_1, \dots, \beta_k$ son parámetros desconocidos, y ε es un término de error (Véase [Montgomery, 2008], pp. 73-75).

► MODELO DE PROMEDIO MÓVIL

Los procesos de promedios móviles fueron introducidos por Yule y Slutsky ([Yule, 1927]; [Slutzky, 1927]). La idea básica de estos procesos consiste en representar a un proceso estocástico $\{X_t: t \in T\}$ (cuyos valores pueden ser dependientes unos de otros), como una suma finita ponderada de choques aleatorios independientes $\{Z_t\}$.

Definición 2.4.5 Proceso de promedio móvil

La serie de tiempo $\{X_t: t \in T\}$ se dice que es un proceso de promedio móvil de orden q , para algún $q \in \mathbb{N}$, y se denota por MA(q) (Moving Average), si puede ser escrito como

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$

donde $\{\beta_i\}_{i=0}^q$ son constantes y $\{Z_t: t \in T\}$ es un proceso de ruido blanco con media cero y varianza σ^2 constante. Los Z_t 's se toman de tal forma que $\beta_0 = 1$; [Chatfield, 2003].

Se piensa en los valores de $\{Z_t\}$ como “impactos” aleatorios que perturban un sistema, y en $\{X_t\}$ como la salida del sistema. En el modelo de promedios móviles, el efecto de un impacto persiste durante un número específico de periodos de tiempo y después desaparece súbitamente; MA(q) tiene una memoria limitada a q periodos.

► MODELO AUTORREGRESIVO DE PROMEDIO MÓVIL

Una generalización de los modelos AR(p) y MA(q), consiste en combinar ambas clases de modelos para obtener lo que se conoce como *procesos autorregresivos de promedios móviles*, denotados por ARMA(p, q), los cuales fueron estudiados por Wold y Bartlett ([Wold, 1938]; [Bartlett, 1946]).

Definición 2.4.6 Proceso autorregresivo de promedio móvil

Una serie de tiempo $\{X_t: t \in T\}$ es un proceso ARMA(p, q) si para cada t se cumple

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad (2.6)$$

$\{Z_t\}$ es un proceso de ruido blanco con media cero y varianza constante [Chatfield, 2003].

La Ecuación (2.6) puede redefinirse simbólicamente en una forma más compacta mediante el operador de cambio de retraso B :

$$\alpha_p(B)X_t = \beta_q(B)Z_t,$$

donde α_p es el polinomio autorregresivo de orden p y β_q el polinomio de promedio móvil de orden q , es decir:

$$\alpha_p(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p, \quad (2.7)$$

y

$$\beta_q(B) = 1 + \beta_1 B + \cdots + \beta_q B^q. \quad (2.8)$$

En las Ecuaciones (2.7) y (2.8), B se considera como una variable y no como un operador.

► MODELO AUTORREGRESIVO E INTEGRADO DE PROMEDIO MÓVIL

Tal como se ha formulado, los modelos de Box-Jenkins fueron diseñados para series de tiempo estacionarias, series sin tendencia con variabilidad constante y correlaciones estables a lo largo del tiempo; pero la mayor parte de las series de tiempo con las que se trabaja son no estacionarias.

Yaglom en 1955 sugirió la posibilidad de que un cierto tipo de no estacionariedad mostrado por algunas series de tiempo, podía representarse mediante la simple toma sucesiva de diferencias de la serie original [Yaglom, 1955]. La idea es convertir ese tipo de series en estacionarias. En muchos casos prácticos, la no estacionariedad de la serie es *homogénea*, lo cual significa que es únicamente el nivel de la serie el que se ve afectado por la no estacionariedad, debido a que existe alguna tendencia polinomial adaptativa; en estos casos, es posible eliminar la tendencia mediante la diferenciación de la serie (Sección 2.3.1), y por lo tanto, volver la serie estacionaria.

Los procesos *autorregresivos e integrados de promedios móviles* denotados por ARIMA(p,d,q) se pueden ver como una generalización de los modelos ARMA(p,q), donde d denota el número de veces que se diferencia la serie de tiempo observada.

La forma general del modelo ARIMA(p,d,q) es:

$$W_t = \alpha_1 W_{t-1} + \cdots + \alpha_p W_{t-p} + Z_t + \cdots + \beta_q Z_{t-q},$$

donde

$$W_t = \nabla^d X_t = (1 - B)^d X_t;$$

[Chatfield, 2003].

Condiciones necesarias y suficientes para la existencia y unicidad de una solución estacionaria e invertible de un proceso ARMA(p,q)

La solución $\{X_t\}$ de un proceso ARMA(p,q) debe ser estacionaria e invertible, por tal razón es preciso definir los términos *causal* e *invertible* y conocer bajo qué términos se establece una solución $\{X_t\}$ con las condiciones deseadas.

Definición 2.4.7 Causal

Un proceso ARMA(p,q) definido por la ecuación $\alpha_p(B)X_t = \beta_q(B)Z_t$, se dice que es *causal* o una función causal de $\{X_t\}$ si existen constantes $\{\psi_j\}$ tales que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \dots \quad (2.9)$$

[Brockwell, 1987].

Definición 2.4.8 Invertible

Un proceso $ARMA(p, q)$ definido por la ecuación $\alpha_p(B)X_t = \beta_q(B)Z_t$, se dice que es invertible si existen constantes $\{\pi_j\}$ tales que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \dots \quad (2.10)$$

[Brockwell, 1987].

Propiedad 2.4.2 Sea $\{X_t\}$ un proceso estacionario con media cero y función de autocovarianza $\gamma(\cdot)$, y $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Entonces, para cada $t \in \mathbb{T}$, la serie

$$Y_t = \psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j},$$

es estacionaria con función de autocovarianza

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k).$$

Prueba: Véase en la presente tesis, el Apéndice B. ■

Teorema 2.4.1 Sea $\{X_t\}$ un proceso $ARMA(p, q)$ para el cual los polinomios $\alpha_p(\cdot)$ y $\beta_q(\cdot)$ no tienen ceros comunes. Entonces $\{X_t\}$ es causal si y solo si $\alpha_p(z) \neq 0$ para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\psi_j\}$ en (2.9) son determinados por la relación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \beta_q(z)/\alpha_p(z), \quad |z| \leq 1. \quad (2.11)$$

Prueba: Primero asumimos que $\alpha_p(z) \neq 0$ y $|z| \leq 1$. Esto implica que existe $\varepsilon > 0$ tal que $1/\alpha_p(z)$ tiene una expansión en series de potencias,

$$1/\alpha_p(z) = \sum_{j=0}^{\infty} \xi_j z^j = \xi(z), \quad |z| < 1 + \varepsilon.$$

Por consiguiente, $\xi_j(1 + \varepsilon/2)^j \rightarrow 0$ cuando $j \rightarrow \infty$, así que existe $k \in (0, \infty)$ para el cual

$$|\xi_j| < k(1 + \varepsilon/2)^{-j}, \quad \text{para todo } j = 0, 1, 2, \dots$$

En particular, se tiene que $\sum_{j=0}^{\infty} |\xi_j| < \infty$ y $\xi(z)\alpha_p(z) \equiv 1$ para $|z| \leq 1$. Por Propiedad 2.4.2, se puede aplicar el operador $\xi(B)$ en ambos lados de la ecuación $\alpha_p(B)X_t = \beta_q(B)Z_t$ para obtener

$$X_t = \xi(B)\beta_q(B)Z_t.$$

Así, tenemos la representación deseada,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

donde la secuencia $\{\psi_j\}$ es determinada por (2.11).

Ahora asumimos que $\{X_t\}$ es causal, es decir, $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ para alguna secuencia $\{\psi_j\}$ tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Entonces,

$$\beta_q(B)Z_t = \alpha_p(B)X_t = \alpha_p(B)\psi(B)Z_t.$$

Si hacemos $\eta(z) = \alpha_p(z)\psi(z) = \sum_{j=0}^{\infty} \eta_j z^j$, $|z| \leq 1$, entonces podemos reescribir esta ecuación como

$$\sum_{j=0}^q \beta_j Z_{t-j} = \sum_{j=0}^{\infty} \eta_j Z_{t-j},$$

y tomando los productos internos de cada lado con Z_{t-k} (recordando que $\{Z_t\}$ es ruido blanco con media cero y varianza σ^2), se obtiene $\eta_k = \beta_k$, para $k = 0, \dots, q$, y $\eta_k = 0$, para $k > q$. Por lo tanto,

$$\beta_q(z) = \eta(z) = \alpha_p(z)\psi(z), \quad |z| \leq 1.$$

Puesto que $\beta_q(z)$ y $\alpha_p(z)$ no tienen ceros comunes y ya que $|\psi(z)| < \infty$ para $|z| \leq 1$, se concluye que $\alpha_p(z)$ no puede ser cero para $|z| \leq 1$; [Brockwell, 1987]. ■

Observación 1: Si $\{X_t\}$ es un proceso ARMA(p,q) para el cual los polinomios $\alpha_p(\cdot)$ y $\beta_q(\cdot)$ tienen ceros comunes, entonces existen dos posibilidades:

1. Ninguno de los ceros comunes se encuentran sobre el círculo unitario, en cuyo caso $\{X_t\}$ es la única solución estacionaria de las ecuaciones del proceso ARMA(p,q) con ceros no comunes, obtenida mediante la cancelación de los factores comunes de $\alpha_p(\cdot)$ y $\beta_q(\cdot)$.
2. Al menos uno de los ceros comunes se encuentra sobre el círculo unitario, en tal caso las ecuaciones del proceso ARMA(p,q) pueden tener más de una solución estacionaria.

Observación 2: La primera parte de la prueba del Teorema 2.4.1 muestra que si $\{X_t\}$ es una solución estacionaria de las ecuaciones de ARMA(p,q) con $\alpha_p(z) \neq 0$ para $|z| \leq 1$, entonces se tiene que $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ donde $\{\psi_j\}$ está definida por (2.11). Inversamente si $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, entonces $\alpha_p(B)X_t = \alpha_p(B)\psi(B)Z_t = \beta_q(B)Z_t$. Así, el proceso $\{\psi(B)Z_t\}$ es la única solución estacionaria de las ecuaciones del proceso ARMA(p,q) si $\alpha_p(z) \neq 0$ para $|z| \leq 1$; [Brockwell, 1987].

Observación 3: Si $\alpha_p(\cdot)$ y $\beta_q(\cdot)$ no tienen ceros comunes y si $\alpha_p(z) = 0$ para alguna $z \in \mathbb{C}$ con $|z| = 1$, entonces no existe una solución estacionaria de $\alpha_p(B)X_t = \beta_q(B)Z_t$; [Brockwell, 1987].

Teorema 2.4.2 Sea $\{X_t\}$ un proceso ARMA(p, q) para el cual los polinomios $\alpha_p(\cdot)$ y $\beta_q(\cdot)$ no tienen ceros comunes. Entonces $\{X_t\}$ es invertible si y solo si $\beta_q(z) \neq 0$, para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\pi_j\}$ en (2.10) son determinados por la relación

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \alpha_p(z)/\beta_q(z), \quad |z| \leq 1. \quad (2.12)$$

Prueba: Primero asumimos que $\beta_q(z) \neq 0$ y $|z| \leq 1$. Por el mismo argumento como en el Teorema 2.4.1, $1/\beta_q(z)$ tiene una expansión en series de potencias

$$1/\beta_q(z) = \sum_{j=0}^{\infty} \eta_j z^j = \eta(z), \quad |z| < 1 + \varepsilon,$$

para algún $\varepsilon > 0$. Se tiene que $\sum_{j=0}^{\infty} |\eta_j| < \infty$. La Propiedad 2.4.2 permite aplicar $\eta(B)$ para ambos lados de la ecuación $\alpha_p(B)X_t = \beta_q(B)Z_t$, para obtener

$$\eta(B)\alpha_p(B)X_t = \eta(B)\beta_q(B)Z_t = Z_t.$$

Así, se obtiene la representación deseada

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

donde la secuencia $\{\pi_j\}$ es determinada por (2.12).

Inversamente, si $\{X_t\}$ es invertible, entonces $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ para alguna secuencia $\{\pi_j\}$ tal que $\sum_{j=0}^{\infty} |\pi_j| < \infty$. Entonces

$$\alpha_p(B)Z_t = \pi(B)\alpha_p(B)X_t = \pi(B)\beta_q(B)Z_t.$$

Estableciendo $\xi(z) = \pi(z)\beta_q(z) = \sum_{j=0}^{\infty} \xi_j z^j$, $|z| \leq 1$, se puede reescribir esta ecuación como

$$\sum_{j=0}^p \alpha_j Z_{t-j} = \sum_{j=0}^{\infty} \xi_j Z_{t-j},$$

y tomando productos internos de cada lado con Z_{t-k} se obtiene $\xi_k = \alpha_k$, para $k = 0, \dots, p$, y $\xi_k = 0$ para $k > p$. Por lo tanto,

$$\alpha_p(z) = \xi(z) = \pi(z)\beta_q(z), \quad |z| \leq 1.$$

Puesto que $\alpha_p(z)$ y $\beta_q(z)$ no tienen ceros comunes y ya que $|\pi(z)| < \infty$ para $|z| \leq 1$, se concluye que $\beta_q(z)$ no puede ser cero para $|z| \leq 1$; [Brockwell, 1987]. ■

Observación 4: Si $\{X_t\}$ es una solución estacionaria de $\alpha_p(B)X_t = \beta_q(B)Z_t$, $\{Z_t\}$ es ruido blanco con media cero y varianza σ^2 constante, y si $\alpha_p(z)\beta_q(z) \neq 0$ para $|z| \leq 1$, entonces

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

donde $\sum_{j=0}^{\infty} \psi_j z^j = \beta_q(z)/\alpha_p(z)$ y $\sum_{j=0}^{\infty} \pi_j z^j = \alpha_p(z)/\beta_q(z)$, $|z| \leq 1$; [Brockwell, 1987].

Observación 5: Si $\{X_t\}$ es un proceso ARMA(p, q), es decir, $\alpha_p(B)X_t = \beta_q(B)Z_t$, con $\alpha_p(z) \neq 0$, para todo z tal que $|z| = 1$, entonces es posible encontrar polinomios $\tilde{\alpha}_p(\cdot)$, $\tilde{\beta}_q(\cdot)$ y un proceso de ruido blanco $\{\tilde{Z}_t\}$ tal que $\tilde{\alpha}_p(B)X_t = \tilde{\beta}_q(B)\tilde{Z}_t$ y tal que $\{X_t\}$ es una función causal de $\{\tilde{Z}_t\}$. Si además, $\beta_q(z) \neq 0$ cuando $|z| = 1$, entonces $\tilde{\beta}_q(\cdot)$ puede ser elegido de tal manera que $\{X_t\}$ es también una función invertible de $\{\tilde{Z}_t\}$, es decir, tal que $\tilde{\beta}_q(z) \neq 0$ para $|z| \leq 1$; [Brockwell, 1987].

Si $\beta_q(B) = 1$ en un proceso ARMA(p, q) se tiene que

$$\alpha_p(B)X_t = Z_t,$$

es un proceso AR(p), el cual es estacionario si el módulo de las raíces de $\alpha_p(B) = 0$ están fuera del círculo unitario.

El caso más simple es el **modelo autorregresivo de orden uno**, es decir, AR(1), que se representa como

$$X_t = \alpha X_{t-1} + Z_t,$$

y que genera a la serie de tiempo conocida como serie de Markov, debido a que el valor de X_t está completamente determinado a partir de X_{t-1} . Para que dicha serie sea estacionaria se requiere que el módulo de la raíz de la ecuación

$$1 - \alpha B = 0,$$

se encuentre fuera del círculo unitario, es decir, $|\alpha| < 1$; [Chatfield, 2003].

El siguiente esquema autorregresivo a considerar es el **autorregresivo de segundo orden** (AR(2)), definido mediante

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t,$$

o bien

$$(1 - \alpha_1 B - \alpha_2 B^2)X_t = Z_t,$$

que fue introducido en [Yule, 1927].

El módulo de las raíces del polinomio $\alpha_2(B) = 1 - \alpha_1 B - \alpha_2 B^2$ están fuera del círculo unitario siempre que

$$|\alpha_2| < 1, \alpha_2 + \alpha_1 < 1 \text{ y } \alpha_2 - \alpha_1 < 1. \quad (2.13)$$

En efecto, al multiplicar la ecuación característica del proceso, $(1 - \alpha_1 x - \alpha_2 x^2) = 0$ por x^{-2} , y aplicando el *Teorema de Schur*⁹, se sigue que los determinantes a considerar son

$$D_1 = \begin{vmatrix} -1 & \alpha_2 \\ \alpha_2 & -1 \end{vmatrix} = 1 - \alpha_2^2$$

y

$$D_2 = \begin{vmatrix} -1 & 0 & \alpha_2 & \alpha_1 \\ \alpha_1 & -1 & 0 & \alpha_2 \\ \alpha_2 & 0 & -1 & \alpha_1 \\ \alpha_1 & \alpha_2 & 0 & -1 \end{vmatrix} = (1 + \alpha_2)^2 \left[(1 - \alpha_2)^2 - \alpha_1^2 \right],$$

ahora bien, la condición $D_1 > 0$ es equivalente a requerir que $|\alpha_2| < 1$, mientras que $D_2 > 0$ si y sólo si $(1 - \alpha_2)^2 > \alpha_1^2$; por consiguiente, las condiciones para que el proceso sea estable en términos de los coeficientes autorregresivos α_1 y α_2 vienen a ser (2.13); ([Guerrero, 2003], pp. 72).

Ahora, si consideramos $\alpha_p(B) = 1$ en un proceso ARMA(p,q), se tiene que

$$X_t = \beta_q(B)Z_t,$$

el cual es un proceso MA(q) y será invertible si el módulo de las raíces de la ecuación característica $\beta_q(B) = 0$, se encuentran fuera del círculo unitario; dichas condiciones de invertibilidad de un proceso MA se obtienen de manera similar a las condiciones de estacionariedad para procesos AR.

La importancia del concepto de invertibilidad radica en que todo proceso invertible está determinado de manera única por su FAC, lo cual no ocurre con procesos que no son invertibles. Además, todo proceso MA(q) es estacionario, esto debido a que la suma $\sum_{i=1}^q |\beta_i|$, al considerar a un número finito de sumandos, es una constante finita (a menos que algún β_i fuese infinito, pero esto implicaría que X_t es infinito); mientras que el proceso AR(p) siempre es invertible, pues $\sum_{j=1}^{\infty} |\pi_j| = \sum_{i=1}^p |\alpha_i| < \infty$; [Guerrero, 2003].

El Cuadro 2.1 resume las condiciones de estacionariedad e invertibilidad para algunos procesos de la metodología de Box-Jenkins. Una parte importante de esta metodología está pensada para liberar al investigador de la tarea de especificación de los modelos dejando que los propios datos temporales de la variable a estudiar indiquen las características de la estructura probabilística subyacente.

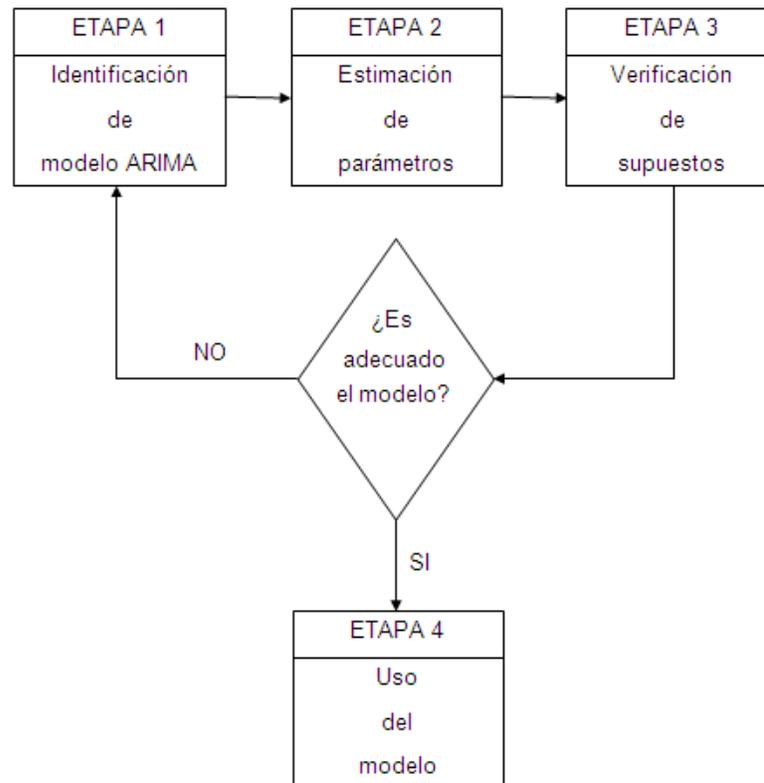
⁹Véase en la presente tesis, Apéndice C: Teorema de Schur.

Modelo	Estacionariedad	Invertibilidad
AR(1)	$ \alpha < 1$	Ninguna
AR(2)	$\alpha_1 + \alpha_2 < 1$ $\alpha_2 - \alpha_1 < 1$	Ninguna
MA(1)	Ninguna	$ \beta < 1$
MA(2)	Ninguna	$\beta_1 + \beta_2 < 1$ $\beta_2 - \beta_1 < 1$
ARMA(p, q)	$ \alpha_i < 1, i = 1, \dots, p$	$ \beta_i < 1, i = 1, \dots, q$

Cuadro 2.1: Condiciones de estacionariedad e invertibilidad para modelos de Box-Jenkins.

Construcción de modelos ARIMA(p, d, q)

La Figura 2.14 presenta gráficamente la estrategia de construcción de modelos para series de tiempo de acuerdo a la metodología de Box-Jenkins. Dicha estrategia, consta de cuatro etapas fundamentales, que se puede ver como un proceso iterativo.

Figura 2.14: Proceso iterativo de Box-Jenkins para construir modelos ARIMA(p, d, q).

***** ETAPA 1: Identificación de un proceso ARIMA(p,d,q) *****

Esta etapa tiene como objetivo determinar los órdenes de los polinomios autorregresivo y de promedio móvil (p y q respectivamente), así como el número (d) de veces que deberá aplicarse el operador diferencia para cancelar la no estacionariedad homogénea, es decir, consiste en determinar primero una serie estacionaria en función de la serie original, para la cual se pueda tener una representación ARIMA(p,d,q).

a) Estabilización de la varianza

Con el fin de volver estacionaria a la serie, lo primero que se podría hacer sería utilizar el método para seleccionar una *transformación estabilizadora de varianza*, la cual indica que si existe tendencia en la serie y la varianza incrementa con la media, entonces es recomendable hacer una transformación a los datos. Si la desviación estándar es directamente proporcional a la media se debe aplicar una transformación logarítmica, pero si la varianza cambia a través del tiempo sin que una tendencia esté presente, entonces una transformación no ayudará; en su lugar, debe considerarse un modelo que permita cambios en la varianza; [Chatfield, 2003].

La transformación logarítmica es un caso especial de la transformación general de Box-Cox [Box and Cox, 1964]: dada una serie de tiempo $\{X_t\}$ y un parámetro de transformación λ , la serie transformada está dada por:

$$Y_t = \begin{cases} (X_t^\lambda - 1)/\lambda, & \lambda \neq 0; \\ \log X_t, & \lambda = 0. \end{cases}$$

b) Estabilización del nivel

Una vez determinada la transformación apropiada para estabilizar la varianza de una serie, la cual se denotará por $T(X_t)$, se procede a estabilizar el nivel de la serie mediante la aplicación del operador diferencia B un número apropiado de veces; B es particularmente usado para remover la tendencia en el enfoque de Box-Jenkins (Sección 2.3.1.).

Para datos no estacionales, la primera diferencia es usualmente suficiente para eliminar la tendencia. Se debe evitar la *sobre-diferenciación* de la serie, porque esto podría causar problemas al tratar de identificar un posible proceso generador de la serie observada. Para evitar dicho problema, se debe calcular la desviación estándar muestral de las series: $\{T(X_t)\}$, $\{\nabla T(X_t)\}$ y $\{\nabla^2 T(X_t)\}$, y ver quién proporciona el mínimo de estos valores que estarán denotados por $S(0)$, $S(1)$ y $S(2)$ respectivamente. Una vez que se ha encontrado el valor $S(i)$ para $i = 0, 1, 2$ mínimo, se estará sugiriendo que se diferencie i -veces la serie, para convertirla en estacionaria; sin embargo, muchas veces la diferencia entre estos valores es menor a 3%, entonces se tendrá que considerar a todos los procesos resultantes ([Guerrero, 2003], pp. 113-115).

Otro criterio para conocer el número de diferencias requeridas en una serie es mediante el empleo del correlograma.

Definición 2.4.9 Correlograma

Un correlograma es un gráfico en el que los coeficientes de autocorrelación muestral r_k se trazan contra el rezago k para $k = 0, 1, \dots, M$, donde M es un número menor que el tamaño de la serie observada, M suele ser mayor o igual a 20. Al correlograma también se le conoce como función de autocorrelación muestral; [Chatfield, 2003].

Para conocer cuántas veces se debe diferenciar una serie, para convertirla en estacionaria, se debe graficar el correlograma para las series: $\{T(X_t)\}$, $\{\nabla T(X_t)\}$ y $\{\nabla^2 T(X_t)\}$; se sugiere graficar los primeros 40 coeficientes del correlograma. Si el número de coeficientes positivos en el correlograma, es mayor a 20, entonces se está indicando una ligera tendencia creciente; por el contrario, si de los 40 coeficientes encontrados, resulta que más de 20 son negativos, entonces se está indicando una tendencia decreciente; mientras tanto, el correlograma que muestre aproximadamente el mismo número de coeficientes positivos como negativos indica que la serie es estacionaria;¹⁰ ([Chatfield, 2003]; [Guerrero, 2003]).

c) Orden de p y q

Una vez que se tiene una serie estacionaria, interesa el orden del proceso. Así, para poder identificar el orden de p y q en un proceso ARIMA(p,d,q) es necesario observar nuevamente el correlograma de la serie estacionaria, puesto que es de gran ayuda para identificar un modelo adecuado dada una serie de tiempo. Primero se identificará el orden del proceso MA, es decir, el valor de q ; posteriormente el orden del proceso AR, es decir, el valor de p , para sustituir ambos dentro del modelo ARIMA(p,d,q).

- *Determinación del orden de un proceso MA*

La determinación del orden de un proceso MA es usualmente evidente de la función de autocorrelación muestral, es decir, el número de autocorrelaciones muestrales significativamente distintas de cero indicará el orden del proceso MA; r_k es significativamente distinta de cero, si

$$|r_k| > 2\sqrt{\frac{1}{n-d} \left(1 + 2\sum_{j=1}^q r_j^2\right)}, \quad \text{para } k > q.$$

La idea antes expuesta es con base en que la función de autocorrelación de un proceso MA(q) es fácil de reconocer, pues se corta en el periodo (o, también llamado retardo) q ; [Chatfield, 2003]. En efecto, dado un proceso MA(q),

$$X_t = Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q},$$

¹⁰Las consideraciones hechas en este párrafo, están sujetas a la elección del investigador.

donde $\{Z_t\}$ es un proceso de ruido blanco con media cero y varianza σ_Z^2 constante. Puesto que $\{X_t\}$ es una combinación lineal finita de términos de ruido blanco, el proceso es estacionario con media y varianza

$$E(X_t) = \sum_{i=0}^q \beta_i E(Z_{t-i}) = 0,$$

$$Var(X_t) = \sigma_Z^2 \sum_{i=0}^q \beta_i^2.$$

Como los Z_t 's son independientes,

$$Cov(Z_s, Z_t) = \begin{cases} \sigma_Z^2, & \text{si } s = t; \\ 0, & \text{si } s \neq t. \end{cases}$$

Entonces,

$$\begin{aligned} \gamma_h &= Cov(X_{t+h}, X_t) \\ &= Cov\left(\sum_{i=0}^q \beta_i Z_{t+h-i}, \sum_{j=0}^q \beta_j Z_{t-j}\right) \\ &= \begin{cases} \sigma_Z^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h}, & 0 \leq h \leq q; \\ 0, & h > q. \end{cases} \end{aligned}$$

Dividiendo γ_h por γ_0 , obtenemos la función de autocorrelación de un proceso MA(q):

$$\rho_h = \begin{cases} \frac{\sum_{i=0}^{q-h} \beta_i \beta_{i+h}}{1 + \beta_1^2 + \dots + \beta_q^2}, & 1 \leq h \leq q; \\ 0, & h > q. \end{cases}$$

La cual se corta después del retardo q , esta característica es propia de un proceso MA(q); [Shumway, 2010].

- *Determinación del orden de un proceso AR*

Un instrumento que permite efectuar la identificación de procesos AR de manera clara y simple, lo constituye la *función de autocorrelación parcial (FACP)*, la cual adquiere determinadas características que dependen del orden del proceso y del tipo de parámetros involucrados. Con el fin de introducir el concepto de autocorrelación parcial es preciso conocer la función de autocorrelación de un proceso AR(p):

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

luego, multiplicando en ambos lados por X_{t-h} y tomando el valor esperado se obtiene la función de autocovarianza, esto es,

$$\begin{aligned} E(X_{t-h} X_t) &= E(\alpha_1 X_{t-h} X_{t-1} + \dots + \alpha_p X_{t-h} X_{t-p} + X_{t-h} Z_t) \\ &= \alpha_1 E(X_{t-h} X_{t-1}) + \dots + \alpha_p E(X_{t-h} X_{t-p}) + E(X_{t-h} Z_t) \\ &= \alpha_1 \gamma_{h-1} + \dots + \alpha_p \gamma_{h-p}, \end{aligned}$$

para $h > 0$, y por lo tanto,

$$\gamma_h = \alpha_1 \gamma_{h-1} + \dots + \alpha_p \gamma_{h-p}.$$

Dividiendo γ_h por γ_0 , obtenemos la función de autocorrelación de un proceso AR(p) [Chatfield, 2003]:

$$\rho_h = \alpha_1 \rho_{h-1} + \dots + \alpha_p \rho_{h-p}, \quad h > 0. \quad (2.14)$$

Si todas las raíces para un proceso AR son reales, entonces ρ_h decae exponencialmente a cero cuando $h \rightarrow \infty$. Si algunas de las raíces son complejas, entonces estarán en pares conjugados y ρ_h decaerá de una manera sinusoidal con velocidad exponencial a cero cuando $h \rightarrow \infty$; en el caso de raíces complejas, la serie de tiempo parecerá ser de naturaleza cíclica. Esto, por supuesto, también es cierto para los modelos ARMA en el que la parte AR tiene raíces complejas; [Shumway, 2010].

Definición 2.4.10 Función de autocorrelación parcial (FACP)

La función de autocorrelación parcial de un proceso estacionario X_t , la cual se denota por ϕ_{hh} , para $h = 1, 2, \dots$, se define por

$$\phi_{11} = \text{Corr}(X_{t+1}, X_t) = \rho_1$$

y

$$\phi_{hh} = \text{Corr}(X_{t+h} - \widehat{X}_{t+h}, X_t - \widehat{X}_t), \quad h \geq 2.$$

Ambos $(X_{t+h} - \widehat{X}_{t+h})$ y $(X_t - \widehat{X}_t)$ son no correlacionados con $\{X_{t+1}, \dots, X_{t+h-1}\}$. La FACP es la correlación entre X_{t+h} y X_t con la dependencia lineal¹¹ de $\{X_{t+1}, \dots, X_{t+h-1}\}$ en cada uno, removido; [Shumway, 2010].

Los parámetros $\phi_{11}, \phi_{22}, \dots, \phi_{pp}$ pueden obtenerse a partir de las ecuaciones de Yule-Walker correspondientes al proceso AR(p) ([Montgomery, 2008], pp. 249), esto es, para cualquier valor fijo de k , $k = 1, 2, \dots$, las ecuaciones de Yule-Walker para la FAC de un proceso AR(p) están dadas por:

$$\begin{aligned} \rho_1 &= \phi_{1k} + \phi_{2k} \rho_1 + \dots + \phi_{kk} \rho_{k-1} \\ \rho_2 &= \phi_{1k} \rho_1 + \phi_{2k} + \dots + \phi_{kk} \rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{1k} \rho_{k-1} + \phi_{2k} \rho_{k-2} + \dots + \phi_{kk} \end{aligned}$$

¹¹Si la distribución conjunta de X e Y en el plano XY está relativamente concentrada alrededor de una recta, entonces se dice que X e Y son linealmente dependientes.

y aplicando la regla de Cramer se obtiene

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots,$$

$$\phi_{pp} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-4} & \rho_3 \\ & & & \dots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-4} & \rho_{p-3} \\ & & & \dots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{vmatrix}}, \quad (2.15)$$

y mediante la Ecuación (2.14), se observa que en ϕ_{pp} la última columna del numerador es la combinación lineal de las columnas anteriores, por lo tanto para un modelo AR(p), la función de autocorrelación parcial en el rezago i , $\phi_{ii} = 0$ para $i > p$. De esta manera, *el número de autocorrelaciones parciales distintas de cero indica el orden del proceso AR* ([Guerrero, 2003], pp. 118-119).

Para estimar las autocorrelaciones parciales de los datos observados, $\hat{\phi}_{ii}$, se sustituye en la Ecuación (2.15) a las autocorrelaciones $\{\rho_h\}$ por sus correspondientes estimaciones, $\{r_h\}$. Obteniendo así un método recursivo:

$$\hat{\phi}_{11} = r_1, \quad \hat{\phi}_{k+1,k+1} = \frac{r_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} r_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} r_j}$$

y

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}, \quad j = 1, \dots, k.$$

Una aproximación sugerida por Quenouille¹², indica que si el proceso es AR(p), las autocorrelaciones parciales muestrales se distribuyen de manera independiente, con media $E(\hat{\phi}_{i,i}) = \phi_{i,i}$ y con

$$\text{Var}(\hat{\phi}_{i,i}) = 1/(n-d), \quad \text{para } i > p.$$

¹²Véase [Quenouille, 1949], citado por [Guerrero, 2003] pp. 119.

Por lo tanto, $\phi_{i,i}$ es distinto de cero (al nivel de significación de aproximadamente el 5%) si el valor de $\hat{\phi}_{i,i}$ se encuentra fuera del intervalo definido por:

$$\pm 2\sqrt{\text{Var}(\hat{\phi}_{i,i})} = \pm 2/\sqrt{n-d}, \quad \text{para } i > p.$$

Aunque la FAC y FACP son auxiliares en la elección de los órdenes del modelo ARMA, en su mayoría, el comportamiento de los datos no es de forma regular, lo cual hace difícil la identificación de dichos valores, así otra manera de elegir el orden de p y q es con base en el error de pronóstico, que se abordará en la Sección 3.1.3.

*** ETAPA 2: Estimación de parámetros de un modelo ARIMA(p,d,q) ***

Una vez que son conocidos los órdenes de los polinomios autorregresivo y de promedios móviles p y q , así como el grado de diferenciación d , se postula que el modelo

$$\alpha_p(B)\nabla^d T(X_t) = \beta_q(B)Z_t,$$

resulta adecuado para representar a la serie $\{T(X_t)\}$ y se requiere entonces asignar valores a $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$, lo cual podría hacerse de manera arbitraria, pero ya existen métodos objetivos y estadísticamente apropiados, tales como los métodos de momentos, máxima verosimilitud, y mínimos cuadrados que se pueden emplear para estimar los parámetros en el modelo tentativamente identificado ([Brockwell, 1987], pp. 249-251). En paquetes de software como MINITAB, JMP, R y SAS, el usuario puede tener la elección del método de estimación y en consecuencia puede elegir el método más adecuado según las especificaciones del problema ([Montgomery, 2008], pp. 266).

*** ETAPA 3: Verificación ***

Esta etapa de la metodología de Box-Jenkins tiene su origen en la idea que *todo modelo es erróneo*, puesto que los modelos son representaciones simplificadas de la realidad, entonces habrá que elegir aquél que presente menos fallas, o poco importantes; una de las formas más claras y simples para detectar violaciones a los supuestos que fundamentan al modelo es a través del *análisis residual*, esto es, aplicando la función de autocorrelación a los residuos, considerando como *residuo* aquella parte de las observaciones que no la explica el modelo ([Chatfield, 2003], pp. 67), la cual es generalmente definida por

$$\text{residual} = \text{observación} - \text{valor ajustado}.$$

De aquí, que los residuales para un proceso ARMA(p,q) pueden ser obtenidos de

$$Z_t = X_t - \left(\sum_{i=1}^p \hat{\alpha}_i X_{t-i} + \sum_{i=1}^q \hat{\beta}_i Z_{t-i} \right) \quad (2.16)$$

donde los $\hat{\alpha}_i$'s y $\hat{\beta}_i$'s son los estimadores de los α_i 's y β_i 's respectivamente.

Si el modelo especificado es adecuado y por lo tanto los órdenes correspondientes a p y q son identificados, esto debe transformar las observaciones en un proceso de ruido blanco. Así, los residuos de la Ecuación (2.16) deben comportarse como ruido blanco ([Montgomery, 2008], pp. 266).

Para verificar que los residuales son ruido blanco, la función de autocorrelación muestral residual no debe diferir significativamente de cero para todos los rezagos mayor que uno, usando el límite aproximado del 95 % de confianza, $\pm 2/\sqrt{n}$, donde n es el número total de observaciones; sin embargo, si sólo uno (o dos) valores de las primeras 20 autocorrelaciones caen fuera de la banda de confianza, entonces esto no se considera como evidencia convincente para rechazar el modelo [Chatfield, 2003]. El número de autocorrelaciones a considerar en el análisis residual está sujeto a las diferentes literaturas.

En caso de que el modelo se rechace, entonces se modifica el modelo en una forma adecuada mediante el establecimiento de condiciones adicionales para tener en cuenta la autocorrelación significativa.

***** ETAPA 4: Uso del modelo *****

El uso del modelo consiste en los fines que el investigador haya tenido en mente al construirlo; dichos fines son por lo general de *pronóstico*, control, simulación o explicación del fenómeno en estudio. Para llevar acabo la aplicación en el desarrollo de la presente tesis se tiene como finalidad el pronóstico.

Capítulo 3

Pronósticos

Uno de los fines más frecuentes al construir un modelo para una serie de tiempo dada, es el pronóstico y como, además la construcción misma del modelo está íntimamente ligada con el pronóstico, puesto que para calcular los residuos se obtienen primero los valores estimados por el modelo, que no son otra cosa más que los valores pronosticados de la variable en el tiempo t , con base en las observaciones hasta el tiempo $t - 1$, es decir, los pronósticos un periodo hacia adelante.

Los pronósticos son una de las herramientas fundamentales para la toma de decisiones dentro de las organizaciones, tanto productivas como sin fines de lucro. Algunas de las áreas en la industria donde se utilizan los pronósticos son: la planeación y control de inventarios, producción, finanzas, ventas, comercialización, entre muchas otras.

Los pronósticos tienen como objetivo reducir la incertidumbre acerca de lo que puede acontecer en el futuro, proporcionando información cercana a la realidad que permita tomar decisiones sobre los cursos de acción a tomar tanto en el presente como en el futuro.

3.1. Introducción a los pronósticos

Las predicciones de los hechos y condiciones futuras se llama **pronósticos** y al acto de hacer tales predicciones se denomina **pronosticar**; [Bowerman, 2009].

Supóngase que se tiene una serie de tiempo con observaciones X_1, X_2, \dots, X_n . Entonces, el problema consiste en estimar los valores futuros tales como X_{n+h} , donde el entero h es llamado **horizonte de pronóstico**. El pronóstico de X_{n+h} hecho al tiempo n para h pasos es típicamente denotado por $\hat{X}(n+h)$ o $\hat{X}_n(h)$; [Chatfield, 2003].

3.1.1. Métodos para establecer pronósticos

Hay un gran número de métodos para realizar pronósticos que se pueden usar para predecir eventos futuros, los cuales se dividen para su estudio en dos tipos básicos: métodos **cualitativos** y **cuantitativos**; [Bowerman, 2009].

Los **métodos cualitativos** usan la opinión de expertos para establecer pronósticos para predecir en forma subjetiva hechos futuros. Dichos métodos se requieren a menudo cuando los datos históricos no están disponibles o son pocos. Por ejemplo, cuando se introduce un nuevo producto en el mercado, la empresa debe confiar en la opinión de expertos para la construcción de un modelo de predicción, puesto que en este caso no hay datos de ventas anteriores.

Los métodos cualitativos también son utilizados para describir el comportamiento de las variaciones del patrón que fue utilizado para el desarrollo del modelo. Existen varias técnicas cualitativas para realizar un pronóstico:

- **Ajuste de curva subjetiva.** Esta técnica se utiliza cuando una empresa está interesada en conocer el pronóstico de las ventas de un nuevo producto. Al pronosticar, la empresa recurre a la opinión de expertos para construir de forma subjetiva una curva **S** con base a la experiencia con otros productos y todo el conocimiento relacionado con el nuevo producto, con objeto de predecir cuánto tomará para que inicie el incremento rápido de las ventas, cuánto tiempo durará este crecimiento rápido y cuándo empezarán a estabilizarse las ventas del producto.

Uno de los problemas más grandes al aplicar esta técnica es decidir la forma de la curva que se usará. Se pueden utilizar muchas otras formas funcionales. Por ejemplo, una curva exponencial o curvas logarítmicas. La construcción subjetiva de dichas curvas es difícil y requiere gran experiencia y criterio.

- **Método Delphi.** Un equipo de expertos plantea predicciones relacionadas con una cuestión específica, suponiendo que el conocimiento combinado de los miembros hará predicciones más factibles. Delphi para evitar problemas entre los miembros, es decir, desacuerdo en las ideas, mantiene a los miembros del equipo separados físicamente. El pronóstico ejercido por cada uno de los miembros es entregado a un coordinador del equipo que reorganizará los pronósticos. Esto se lleva a cabo por varias rondas hasta obtener un pronóstico adecuado. El método Delphi permite diferencias justificadas de opinión y no intenta lograr una unanimidad.
- **Comparaciones técnicas independientes del tiempo.** Este método se utiliza para predecir cambios técnicos, los cambios en un área se predicen mediante el monitoreo de cambios que tienen lugar en un área distinta.

Las técnicas objetivas mejor conocidas como **métodos cuantitativos** de pronóstico se basan en el manejo de datos numéricos históricos para obtener un pronóstico preciso

y se sustentan en la suposición de que el comportamiento o patrón de los datos históricos permanece durante un periodo de extensión significativa en el futuro. Los modelos cuantitativos para establecer pronóstico se pueden agrupar en dos clases:

- **Modelo univariable** para pronóstico, los cuales predicen valores futuros de la variable de interés sólo con base en el patrón anterior de esa variable, pero se supone que el patrón anterior seguirá siendo el mismo.
- **Modelos causales.** Estos modelos predicen valores futuros de la variable de interés con base en la relación entre dicha variable y otras variables. La variable de interés recibe el nombre de variable dependiente y las otras variables serían las variables independientes. Por ejemplo, las ventas de un producto podrían estar relacionadas con el precio del producto, gastos de publicidad y precios de los competidores de productos similares.

3.1.2. Selección de una técnica de pronóstico

La elección de un método para pronóstico depende de una variedad de consideraciones [Chatfield, 2003]:

1. Tipo de serie de tiempo (por ejemplo, series macroeconómicas o de ventas) y sus propiedades (por ejemplo, tendencia o estacionalidad). Algunas series son muy regulares y por tanto fáciles de predecir, pero otras no lo son.
2. Cuántas observaciones pasadas se tienen disponibles.
3. La longitud o periodo del pronóstico. La duración del periodo se clasifica básicamente en cuatro clases ([Bowerman, 2009], pp. 19):
 - *Inmediato*, el cual contempla un plazo de menos de un mes.
 - *Corto plazo*, el cual va desde un mes hasta tres meses.
 - *Medio plazo*, son periodos de más de tres meses y menos de dos años.
 - *Largo plazo*, tienen una duración de dos años en adelante.
4. Costo del pronóstico. La complejidad, y por lo tanto el costo, de desarrollar un modelo varía de una técnica a otra. Así mismo, el costo de la operación real de la técnica de pronóstico es importante. Algunos de los métodos de pronóstico son sencillos de operar, pero otros son muy complejos.

3.1.3. Errores en los pronósticos

Todas las situaciones en las que hay un pronóstico contienen algún grado de incertidumbre. Recuérdese que la componente irregular de una serie de tiempo representa las fluctuaciones inexplicables o impredecibles en los datos, esto quiere decir que es de esperarse algún

error en el pronóstico. Si el efecto de la componente irregular es notable, entonces se tendrán pronósticos inexactos. Por el contrario, si el efecto de la componente irregular es bajo, entonces las demás componentes permitirán predecir un pronóstico confiable.

Aunque la componente irregular no es la única fuente de error en los pronósticos, la exactitud con la cual se pueda predecir cada uno de los otros componentes de la serie de tiempo también influye en la magnitud de error de los pronósticos, este es el motivo por el cual no se puede realizar con exactitud un pronóstico en situaciones prácticas. De aquí que errores de predicción grandes indicarían que la componente irregular es tan grande que ninguna técnica de predicción podría generar pronósticos exactos, o bien, podría indicar que la técnica de predicción no es capaz de predecir con exactitud la tendencia, la estacionalidad o el componente cíclico; [Bowerman, 2009]).

Medición del error de pronóstico

Se ha visto la importancia de que concuerde una técnica de predicción y el patrón de datos que caracteriza una serie de tiempo. Con frecuencia, un examen de los errores de pronóstico en el tiempo indica si la técnica de predicción va de acuerdo o no con el patrón.

Si los errores de pronóstico en el tiempo indican que la metodología usada para el pronóstico es la adecuada, entonces lo importante ahora es medir la magnitud de los errores de modo que permita determinar si puede realizarse un pronóstico exacto.

Primero denotemos el valor real de la variable de interés en el periodo de tiempo t como X_t y el valor predicho como \hat{X}_t . Así, el error del pronóstico e_t se define y se denota como sigue:

Definición 3.1.1 Error de pronóstico.

El error de pronóstico para un pronóstico particular \hat{X}_t es

$$e_t = X_t - \hat{X}_t;$$

[Bowerman, 2009].

Para medir la magnitud de los errores existen diversas maneras de realizarlo, las más comunes son: la **desviación absoluta media** (DAM), el **error cuadrático medio** (ECM) y el **criterio de información de Akaike** (AIC). Las cuales se definen de la siguiente forma:

- $\text{DAM} = \frac{\sum_{t=1}^n |e_t|}{n}$
- $\text{ECM} = \frac{\sum_{t=1}^n (e_t)^2}{n}$

- $AIC = -2\log L(\theta, \sigma^2) + 2r$,

donde,

$L(\theta, \sigma^2)$ es la *función de verosimilitud*¹ para el modelo ajustado, evaluada en los estimadores de máxima verosimilitud de los parámetros desconocidos $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, para algún entero positivo k , y varianza σ^2 ; r es el número de parámetros independientes que se ajustan al modelo que se está evaluando.

La diferencia básica entre estas medidas es que el ECM, a diferencia de DAM, penaliza más a la técnica para pronosticar en los errores grandes que en los errores pequeños. Por ejemplo, un error de 2 genera un error cuadrático de 4, y un error de 4 genera un error cuadrático de 16 (cuatro veces más grande). Dado que la suma de los cuadrados de los residuales generalmente decrecerá cuando el número de parámetros sea incrementado a pesar de si la complejidad adicional es realmente valiosa, no es bueno elegir un modelo para dar un mejor ajuste minimizando la suma de los cuadrados de los residuales; [Bowerman, 2009].

El criterio de información Akaike es una medida que de alguna forma penaliza el número de parámetros del modelo a evaluar. El AIC es usado como un criterio de selección en el análisis de series de tiempo cuando existen varios modelos, esencialmente escoge el modelo con el mejor ajuste, moderado por la función de verosimilitud, sometido a un término de penalización, para prevenir el sobre-ajuste, que se incrementa con el número de parámetros en el modelo. En principio el criterio de selección será escoger modelos con valores más bajos de AIC. Para un modelo ARMA(p, q), se tiene que $r = p + q + 1$, σ^2 es la varianza de los residuos y $\theta = (\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q)$; [Chatfield, 2003].

Estas medidas se pueden usar de dos maneras distintas. Primero, pueden ayudar en el proceso de elegir un modelo de pronóstico. Una estrategia común usada para efectuar dicha selección requiere simular datos anteriores. Entonces se usa cada uno de los modelos para generar “predicciones” de los datos anteriores. Luego se comparan dichas “predicciones” con los valores reales de los datos histórico y finalmente se mide la exactitud. Segundo, estas medidas se pueden utilizar para monitorear un sistema de pronóstico con el objeto de detectar cuándo algo va “mal” con el sistema. Por ejemplo, una situación en la cual el patrón de información que ha persistido durante un largo periodo de tiempo cambia en forma repentina. En una situación como esta interesa descubrir el cambio en el patrón tan rápido como sea posible antes de que el pronóstico se vuelva inexacto.

3.2. Tipos de pronósticos

El hecho de que las técnicas para elaborar pronósticos generen con frecuencia predicciones que contienen un cierto grado de error se relaciona con la forma de los pronósticos que se

¹Véase [Kalbfleisch, 1985b], Capítulos: 9 y 10.

requieren. Se estudiarán dos tipos de pronóstico: el **pronóstico puntual** y el **pronóstico de intervalo de predicción**.

El pronóstico puntual es donde el pronóstico para un periodo de tiempo futuro en particular consiste de un único número. Mientras que un intervalo de predicción frecuentemente ayuda a dar una mejor indicación de la incertidumbre en el futuro. En lugar de un sólo valor, un intervalo de predicción consiste de un límite superior e inferior entre el cual se espera que un valor futuro se encuentre con una probabilidad determinada; [Chatfield, 2003].

3.3. Pronóstico para procedimientos univariados

Entre las técnicas cuantitativas de los pronósticos que se estudian con frecuencia se encuentran la extrapolación de curvas de tendencia, suavizamiento exponencial, metodología de Box-Jenkins y una mezcla de suavizamiento exponencial con la metodología de Box-Jenkins.

3.3.1. Extrapolación de curvas de tendencia

Para un pronóstico con un periodo de predicción largo, de datos no estacionales, con frecuencia se usa una curva de tendencia para ajustar los datos y así extrapolar o ampliar la curva de ajuste. Esta aproximación se usa con mayor frecuencia cuando los datos son anuales, y por lo tanto, claramente no hay estacionalidad. Hay una variedad de curvas de ajuste, por ejemplo, la curva polinomial, exponencial, logística o de Gompertz. Estas curvas se requieren cuando los datos históricos anuales son de 7 a 10 años. Considérese un periodo de predicción largo igual o menor a la mitad del total de datos, puesto que un número mayor a este haría que el pronóstico fuera más inexacto; [Chatfield, 2003].

El caso más simple para pronosticar es cuando los datos se ajustan a una curva de tendencia lineal.

Ejemplo 3.3.1 Curva de tendencia lineal

Las ventas de una pequeña cadena de tiendas de comestibles, desde 1997 son:

Año	Ventas(MDD)
1997	7
1998	10
1999	9
2000	11
2001	13

La curva de ajuste para estos datos es:

$$\widehat{X} = a + bt = 6.1 + 1.3t,$$

donde a y b se calculan por mínimos cuadrados. Las ventas se expresan en millones de dólares y el origen o año cero es 1996, t aumenta una unidad por cada año. El valor 1.3 indica que las ventas aumentan a razón 1.3 millones (de dólares) por año. El valor 6.1 es el de ventas estimadas cuando $t = 0$. Esto es, el monto de las ventas estimadas para 1996 (el año base) es igual a 6.1 millones.

Si se desea el pronóstico estimado de ventas para el año 2004, debe tomarse en cuenta que está indizado con $t = 8$. Entonces,

$$\widehat{X} = 6.1 + 1.3t = 6.1 + 1.3(8) = 16.5,$$

por lo tanto, con base en las ventas anteriores, la estimación de ventas para el 2004 es de 16.5 millones de dólares; [Lind, 2004].

3.3.2. Suavizamiento exponencial

El pronóstico puntual efectuado en el periodo t para X_{t+h} coincide con el ajuste efectuado en el periodo t para X_{t+h} en cada uno de los métodos de suavizamiento exponencial vistos en la Sección 2.4.3.; mientras que para cualquier h , $h = 1, 2, \dots$, un intervalo de predicción del 95 % calculado en el periodo t para X_{t+h} en cada uno de los métodos de suavizamiento exponencial [Bowerman, 2009] están dados por:

- **Suavización exponencial simple**

$$\left[\ell_t - z_{[0.025]} s \sqrt{1 + (h-1)\alpha^2}, \ell_t + z_{[0.025]} s \sqrt{1 + (h-1)\alpha^2} \right]$$

donde el error estándar s está determinado por:

$$s = \sqrt{\frac{SSE}{t-1}} = \sqrt{\frac{\sum_{k=1}^t [X_k - \ell_{k-1}]^2}{t-1}}.$$

Nótese que si n indica el tamaño de la muestra, entonces para $t = n$ y para todo $h \geq 1$ se tiene un pronóstico constante.

- **Doble suavización exponencial**

El método de Doble suavizamiento exponencial consiste en realizar dos suavizaciones exponenciales, a partir de las cuales se obtendrá el valor estimado, o pronóstico que buscamos realizar, mediante un cálculo realizado con una expresión sencilla. La primera se aplica a los valores observados en la serie de tiempo y la segunda a la serie atenuada obtenida mediante la primera atenuación. Entonces la primera suavización es

$$\widehat{X}_t = \alpha X_t + (1 - \alpha)\widehat{X}_{t-1},$$

mientras que la segunda suavización es

$$\widehat{X}'_t = \alpha \widehat{X}_t + (1 - \alpha) \widehat{X}'_{t-1}.$$

Luego

$$a = 2\widehat{X}_t - \widehat{X}'_t, \text{ y } b = \alpha \frac{\widehat{X}_t - \widehat{X}'_t}{1 - \alpha}.$$

Así, el pronóstico es

$$P_{t+m} = a_t + b_t m,$$

donde m representa el número de periodos hacia el futuro del que se pretende hacer el pronóstico.

■ **Método de Hold de la suavización exponencial corregida de la tendencia**

Para $h = 1$,

$$[(\ell_t + b_t) \pm z_{[0.025]} s].$$

Para $h \geq 2$,

$$\left[(\ell_t + hb_t) \pm z_{[0.025]} s \sqrt{1 + \sum_{i=1}^{h-1} \alpha^2 (1 + i\gamma)^2} \right]$$

donde el error estándar s está determinado por:

$$s = \sqrt{\frac{SSE}{t-2}} = \sqrt{\frac{\sum_{k=1}^t [X_k - (\ell_{k-1} + b_{k-1})]^2}{t-2}}.$$

■ **Método aditivo de Hold-Winters**

$$[\widehat{X}_{t+h} \pm z_{[0.025]} s \sqrt{c_h}]$$

Si $h = 1$, entonces $c_1 = 1$.

Si $2 \leq h \leq d$, entonces $c_h = \left[1 + \sum_{i=1}^{h-1} \alpha^2 (1 + i\gamma)^2 \right]$.

Si $h \geq d$, entonces $c_h = 1 + \sum_{i=1}^{h-1} [\alpha(1 + i\gamma) + U_{i,d}(1 - \alpha)\gamma]^2$, donde $U_{i,d} = 1$ si $i = d$, y cero en otro caso.

El error estándar s está determinado por:

$$s = \sqrt{\frac{SSE}{t-3}} = \sqrt{\frac{\sum_{k=1}^t [X_k - (\ell_{k-1} + b_{k-1} + sn_{k-d})]^2}{t-3}}.$$

■ **Método multiplicativo de Hold-Winters**

$$[\widehat{X}_{t+h} \pm z_{[0.025]} s_r (\sqrt{c_h})(sn_{t+h-d})]$$

Si $h = 1$, entonces $c_1 = (\ell_t + b_t)^2$.

Si $2 \leq h \leq d$, entonces $c_h = \sum_{i=1}^{h-1} \alpha^2 [1 + (h-i)\gamma]^2 (\ell_t + ib_t)^2 + (\ell_t + hb_t)^2$.

El error estándar relativo s_r está determinado por:

$$s_r = \sqrt{\frac{\sum_{k=1}^t \left[\frac{X_k - \widehat{X}_{k-1}}{\widehat{X}_{k-1}} \right]^2}{t-3}} = \sqrt{\frac{\sum_{k=1}^t \left[\frac{X_k - (\ell_{k-1} + b_{k-1})s_{n_{k-d}}}{(\ell_{k-1} + b_{k-1})s_{n_{k-d}}} \right]^2}{t-3}}.$$

3.3.3. Metodología de Box-Jenkins

Una vez que se ha encontrado un modelo de ajuste satisfactorio, interesa calcular los pronósticos. Dada una serie de tiempo con N datos, los pronósticos involucran las observaciones y los residuos o errores del pronóstico hasta el tiempo N ; [Chatfield, 2003]. El pronóstico con error cuadrado mínimo de X_{N+h} al tiempo N es la esperanza condicional de X_{N+h} al tiempo N , es decir,

$$\widehat{X}_N(h) = E(X_{N+h} | X_N, X_{N-1}, \dots).$$

Con lo cual el pronóstico puntual de un proceso ARMA, el cual sabemos que es de la forma $\phi(B)X_t = \theta(B)Z_t$, donde X_t es una serie estacionaria y el proceso $\{Z_t\}$ es ruido blanco, queda definido de la forma:

$$\widehat{X}_t(h) = \phi_1 E(X_{t+h-1}) + \dots + \phi_p E(X_{t+h-p}) + E(Z_{t+h}) - \theta_1 E(Z_{t+h-1}) - \dots - \theta_q E(Z_{t+h-q}),$$

donde,

$$E(X_{t+h-j}) = \begin{cases} X_{t+h-j}, & \text{si } j \geq h; \\ \widehat{X}_t(h-j), & \text{si } j < h, \end{cases}$$

y

$$E(Z_{t+h-j}) = \begin{cases} X_{t+h-j} - \widehat{X}_{t+h-j-1}(1), & \text{si } j \geq h; \\ 0, & \text{si } j < h. \end{cases}$$

Por ejemplo, supóngase que el comportamiento de X_t lo indica el proceso

$$(1 - 0.6B)X_t = (1 + 0.2B)Z_t,$$

entonces

$$X_t = 0.6X_{t-1} + Z_t + 0.2Z_{t-1},$$

de tal forma que los pronósticos $\widehat{X}_t(h)$ se obtienen como

$$\begin{aligned} \widehat{X}_t(1) &= 0.6E(X_t) + E(Z_{t+1}) + 0.2E(Z_t) \\ &= 0.6X_t + 0.2[X_t - \widehat{X}_{t-1}(1)]. \\ \widehat{X}_t(2) &= 0.6E(X_{t+1}) + E(Z_{t+2}) + 0.2E(Z_{t+1}) \\ &= 0.6\widehat{X}_t(1). \end{aligned}$$

de forma general,

$$\widehat{X}_t(h) = 0.6\widehat{X}_t(h-1), \quad \text{para } h \geq 2,$$

es decir, la generación de pronósticos se vuelve un proceso recursivo. El pronóstico $\widehat{X}_t(1)$ involucra al pronóstico $\widehat{X}_{t-1}(1)$, el cual a su vez utiliza a $\widehat{X}_{t-2}(1)$. En efecto,

$$\widehat{X}_{t-1}(1) = 0.6X_{t-1} + 0.2[X_{t-1} - \widehat{X}_{t-2}(1)],$$

y se continua así sucesivamente, hasta llegar a

$$\widehat{X}_1(1) = 0.6X_1 + 0.2[X_1 - \widehat{X}_0(1)],$$

en donde, ya no existe información para calcular $\widehat{X}_0(1)$, se supone que $\widehat{X}_0(1) = X_1$, es decir, que se hace $Z_1 = 0$. Así, para obtener los pronósticos $\widehat{X}_t(h)$, con $h \geq 2$, hay necesidad de calcular los valores

$$\begin{aligned} \widehat{X}_1(1) &= 0.6X_1. \\ \widehat{X}_2(1) &= 0.8X_2 - 0.2\widehat{X}_1(1). \\ &\vdots \\ \widehat{X}_t(1) &= 0.8X_t - 0.2\widehat{X}_{t-1}(1). \end{aligned}$$

Nótese que el uso del método recursivo anterior para generar pronósticos es poco práctico para aplicarlo manualmente, pero no para emplearlo con un software; [Guerrero, 2003].

Los límites del $100(1 - \alpha)\%$ de probabilidad para X_{N+h} están dados por:

$$\widehat{X}_N(h) \pm z_{\alpha/2} \sqrt{Var[e_N(h)]}, \quad \text{para } h \geq 1,$$

donde, $e_N(h) = X_{N+h} - \widehat{X}_N(h)$ denota el error de pronóstico al tiempo N con h periodos de predicción, y $z_{\alpha/2}$ es el punto porcentual de una distribución normal estándar, de modo que $P(Z > z_{\alpha/2}) = \alpha/2$. Con lo cual, el valor observado de X_{N+h} será cubierto por el intervalo con una probabilidad de $100(1 - \alpha)\%$ ([Chatfield, 2003], pp. 85-86).

3.3.4. Sistematización de Box - Jenkins de la suavización exponencial

Se muestra cómo sistematizar el método de la suavización exponencial simple y un proceso $MA(q)$ de la metodología de Box-Jenkins.

Recuérdese que el método de suavización exponencial simple se usa para pronosticar una serie temporal cuando no hay tendencia o patrón estacional, pero la media de la serie de tiempo $\{X_t\}$ cambia lentamente en el tiempo, y que la estimación de la media en el periodo t está dada por la ecuación de suavización:

$$\widehat{X}_t = \alpha X_t + (1 - \alpha)\widehat{X}_{t-1},$$

donde α es una constante de suavización entre 0 y 1. Esta ecuación es algunas veces reescrita en términos del error de la forma:

$$\widehat{X}_t = \alpha e_t + \widehat{X}_{t-1},$$

donde $e_t = X_t - \widehat{X}_{t-1}$, el error del pronóstico al tiempo t .

Haciendo uso de la constante de suavizamiento α dentro de los modelos MA de Box-Jenkins, se tiene que el suavizamiento exponencial es óptimo, es decir, si el modelo de la serie de tiempo está dado por

$$X_t = Z_t + \alpha \sum_{j < t} Z_j,$$

donde $\{Z_t\}$ denota los procesos puramente aleatorios; [Chatfield, 2003].

Para el caso de un MA(1), el modelo de Box-Jenkins es de la forma:

$$X_t = Z_t - \theta_1 Z_{t-1},$$

donde el parámetro de media móvil θ_1 y la constante de suavizamiento quedan relacionados mediante:

$$\theta_1 = 1 - \alpha;$$

[Bowerman, 2009].

Por lo tanto, el pronóstico para el modelo MA(1) puede ser calculado de la misma forma en que se calcula el pronóstico para el suavizamiento exponencial, dado que nuestro modelo MA(1) queda escrito de forma similar al del suavizamiento exponencial simple aunque en términos del proceso de ruido blanco, es decir:

$$X_t = Z_t + (1 - \alpha)Z_{t-1}.$$

Capítulo 4

Aplicación

La utilidad de realizar un análisis en registros hidrométricos¹, se fundamenta en la posibilidad del uso de la información en la planeación para la generación de energía eléctrica, en la prevención de desastres por inundación, o en la disponibilidad de agua en los sistemas de abasto.

Con el fin de cumplir en su totalidad el objetivo general de la presente tesis, se analizan los datos históricos de aforo² mínimos anuales instantáneos de la estación hidrométrica “Paso Ancho”, perteneciente a la cuenca del río Verde en el Estado de Oaxaca, buscando con ello, establecer su posible comportamiento futuro.

Los gráficos presentados en este capítulo se realizaron con la ayuda del software estadístico R project, mediante los paquetes para series de tiempo [Cowpertwait, 2009].

4.1. Datos

Los datos utilizados en el análisis con la teoría de series de tiempo, son obtenidos de una base de datos emitida por el Boletín Hidrométrico de la Comisión Federal de Electricidad (CFE, 2005). Dicha base de datos contiene registros del gasto de agua (entiéndase por *gasto* como el volumen de escurrimiento por unidad de tiempo en una sección dada de un canal, río o conducto, medido en m^3/s) recolectados por las *estaciones hidrométricas* en el Estado de Oaxaca, México (Figura 4.1).

La estación hidrométrica seleccionada fue la de “Paso Ancho”, con clave 20036 (Figura 4.1), por ser una de las más completas en el registro del gasto de agua que fluye por el río Atoyac.

Paso Ancho se localiza en la corriente del río Verde o Atoyac, los cuales pertenecen a

¹Registros del caudal de agua que pasa por una sección transversal de un río.

²Datos que miden el gasto o cantidad de agua que lleva una corriente en una unidad de tiempo.

la cuenca del río Verde, en la Región Hidrológica No. 20: Costa Chica - Río Verde (Figura 4.2).

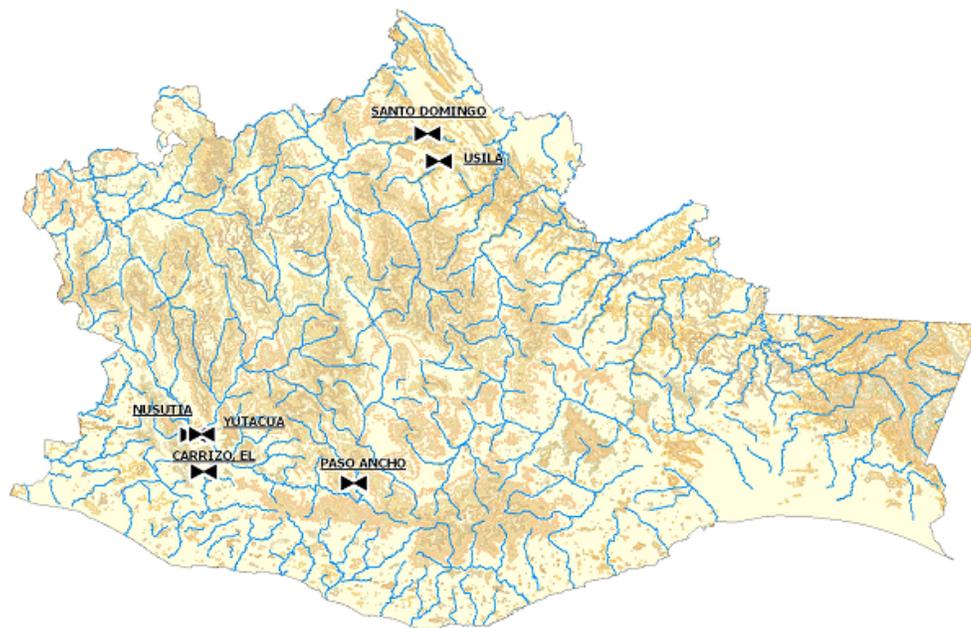


Figura 4.1: Estaciones hidrométricas en Oaxaca, México.

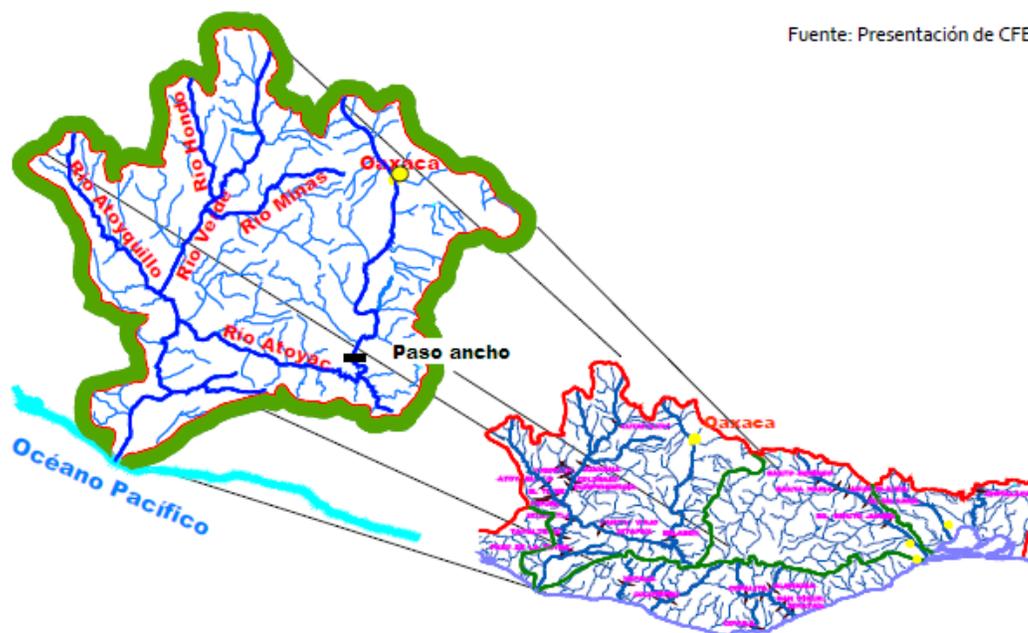


Figura 4.2: Cuenca del río Verde, Oaxaca.

Esta cuenca colinda al norte con el Sistema Hidrológico Papaloapan de la Región Golfo, al sur con el océano Pacífico, al este con la cuenca del río Tehuantepec, al sureste con las cuencas de los ríos Colotepec y Copalita, al suroeste con la cuenca del río Arena, al oeste con la cuenca del río Ometepec del Sistema Hidrológico Costa de Guerrero y al noroeste con la cuenca alta del río Atoyac del Sistema Hidrológico Balsas. El área de la cuenca es de $18\,277\text{ km}^2$, lo que equivale a 19.8% de la superficie estatal.

Paso Ancho se encuentra ubicada próxima a la localidad San Vicente Coatlán, perteneciente al municipio de Ejutla, $16^\circ 22' N$, $96^\circ 53' 27'' O$, y a una altitud de $1\,247\text{ m}$. El principal escurrimiento colector es el río Verde o Atoyac.

Se ha seleccionado esta estación por ser la más cercana a la zona de los valles centrales, específicamente a la Ciudad de Oaxaca, la cual es un sitio que sufre de problemas de inundación en cada temporada de lluvias. La aplicación del resultado obtenido puede ayudar en la planeación para la generación de energía eléctrica o en la disponibilidad de agua en los sistemas de abasto para uso y consumo humano en la Ciudad.

La *variable de estudio* a considerar para la estación hidrométrica “Paso Ancho” es *Gastos Mínimos Anuales Instantáneos en Metros Cúbicos por Segundo*. De aquí en adelante, dicha variable será denotada por *GMAI*, para la cual deseamos obtener el modelo que mejor explique el comportamiento de los valores observados, y posteriormente, pronosticar el comportamiento futuro del gasto de agua en la zona de estudio.

De acuerdo a los Capítulos 2 y 3, los pasos para llevar a cabo el análisis de las series de tiempo son los siguientes:

1. Ajuste del modelo.
2. Validación del modelo ajustado.
3. Pronóstico.

4.1.1. Gastos Mínimos Anuales Instantáneos en Metros Cúbicos por Segundo (GMAI)

La base de datos del gasto mínimo anual instantáneo (véase Apéndice A) consta de valores observados desde el año de 1957 hasta el año 2005. El ajuste del modelo se realizará a partir de el año 1957 hasta el 2002, para así poder simular el pronóstico de los periodos: 2003, 2004 y 2005; finalmente, se va a verificar el pronóstico obtenido con los datos reales observados.

Como primer paso se grafica la serie simple de la variable en estudio (Figura 4.3), para así poder apreciar las características más relevantes de la variable en consideración, tales como tendencia, variación, estacionalidad y ciclos en los datos. Se descarta la estacionalidad debido a que en esta variable se tratan datos anuales.

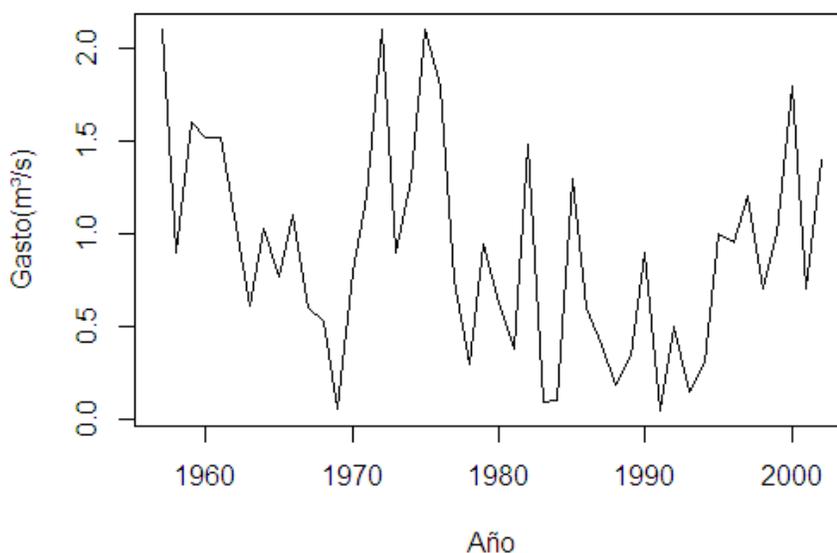


Figura 4.3: Serie simple de GMAI Paso ancho, clave 20036.

Se advierte de la Figura 4.3 que, hasta 1969, el gasto de agua presenta una declinación notoria a simple vista. Así, una medida que indique el nivel de la serie podría usarse para describir a la parte de la serie del periodo entre 1957 y 1969. Esto no ocurre así durante el periodo 1970-2002, ya que a partir de 1970 el gasto de agua comienza a crecer y llega a alcanzar un valor máximo de $2.1 \text{ m}^3/\text{s}$ durante el año de 1972, valor que se repite en el año de 1975 después de sufrir una caída en 1973, posteriormente, el gasto de agua decrece y presenta un nivel relativamente constante, alrededor de $0.5 \text{ m}^3/\text{s}$, desde 1978 hasta 1994, aunque las variaciones al rededor de dicho nivel son, para todo fin práctico muy irregulares. Para el año 1995 el gasto de agua muestra un nuevo nivel, cercano a $1 \text{ m}^3/\text{s}$, hasta el año 2002, con variación similar a la anterior.

A simple vista se a mostrado que la serie observada presenta una tendencia decreciente, lo cual se prueba rigurosamente mediante la función de autocorrelación de la serie $\{GMAI_t\}$ (Figura 4.4). Se observa en el correlograma que, las primeras autocorrelaciones caen lentamente a cero en forma de onda, lo cual significa que la serie no es estacionaria, al menos en lo que se refiere al nivel.

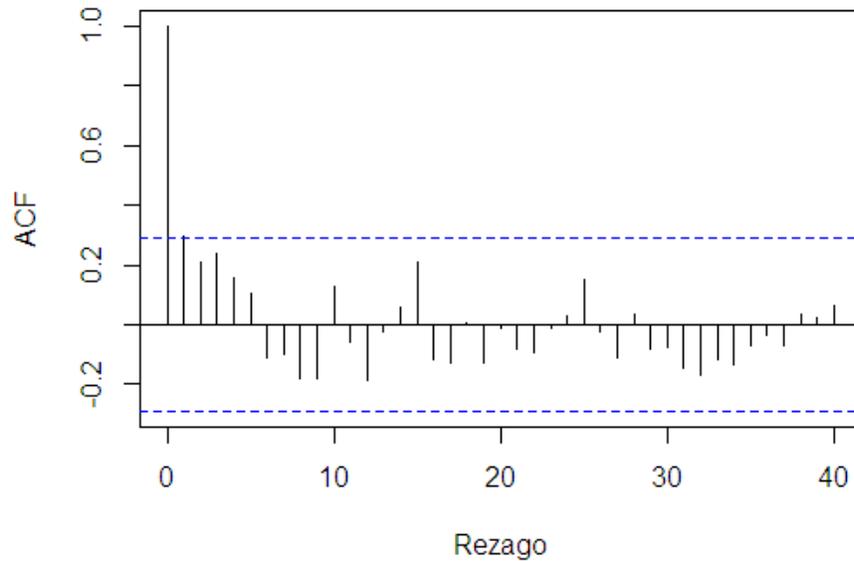


Figura 4.4: Función de autocorrelación muestral de $\{GMAI_t\}$
Paso ancho, clave 20036.

Con el fin de cancelar cualquier tendencia en los datos del gasto de agua se prosigue a aplicar el operador ∇ a la serie $\{GMAI_t\}$. La gráfica de la serie $\{\nabla GMAI_t\}$, así como su correspondiente función de autocorrelación muestral, aparecen en la Figura 4.5. Aunque sí parece existir media constante para $\{\nabla GMAI_t\}$, su varianza presenta un patrón de crecimiento, el cual se podría intentar cancelar por medio de una transformación logarítmica (Figura 4.5 a)). Asimismo, en la Figura 4.5 b) se observa que las autocorrelaciones tienden a desaparecer rápidamente, sin presentar un patrón específico, lo que se toma como indicador de estacionariedad en la serie, en lo que se refiere al nivel.

La posible transformación ya antes mencionada para estabilizar aproximadamente la varianza de $\{GMAI_t\}$ es

$$T(GMAI_t) = \log(GMAI_t),$$

así, la nueva serie a analizar es $\{T(GMAI_t)\}$, con la cual se espera que el supuesto de estacionariedad se vea satisfecho más adecuadamente (se omite la gráfica de la serie $\{T(GMAI_t)\}$ dado que carece de interpretación práctica).

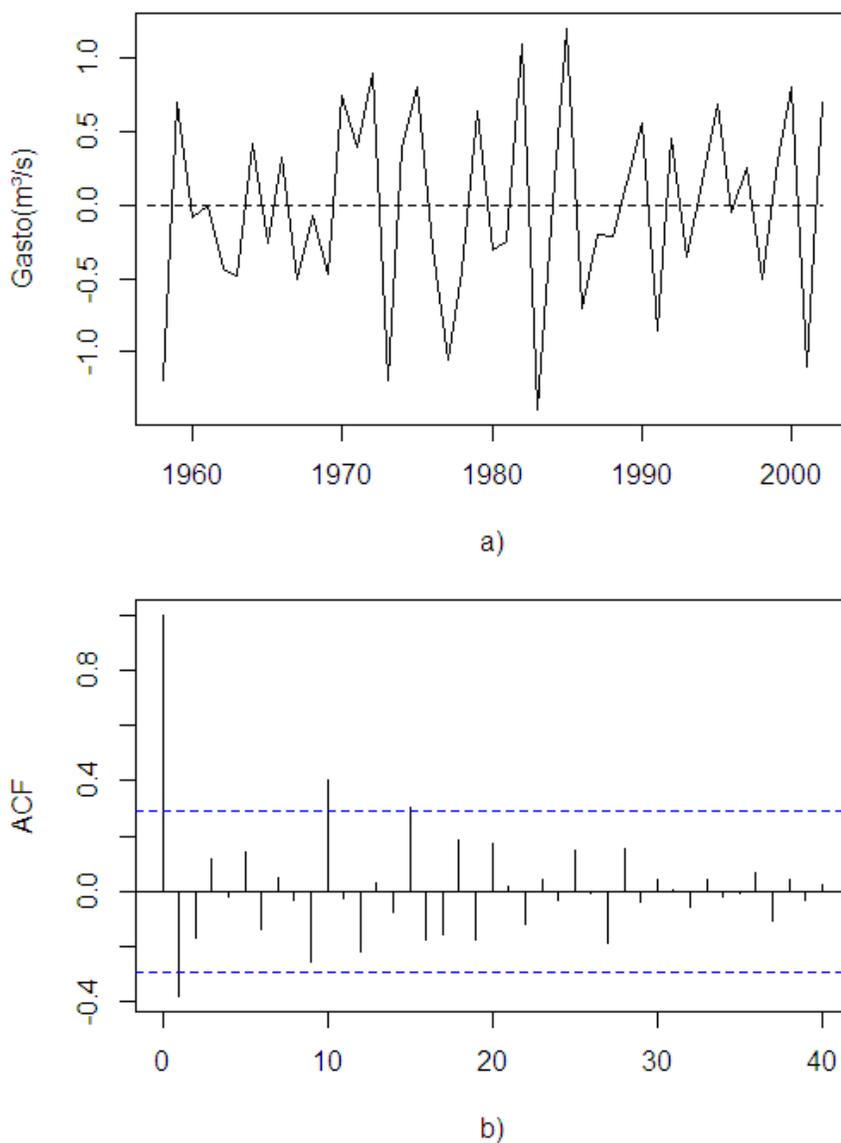


Figura 4.5: a) Gráfica de la serie $\{\nabla GMAI_t\}$ y
b) Función de autocorrelación muestral.

Una vez determinada la transformación, se procede nuevamente a estabilizar el nivel de la serie, por tal razón se grafican los correlogramas para las series $\{T(GMAI_t)\}$, $\{\nabla T(GMAI_t)\}$ y $\{\nabla^2 T(GMAI_t)\}$ (Figura 4.6). Se observa del correlograma a) que las primeras autocorrelaciones caen lentamente a cero en forma de onda, lo cual está indicando que persiste la tendencia en los datos. Observando los correlogramas b) y c), se puede ver que las primeras autocorrelaciones tienen un decaimiento muy rápido a cero, indicativo de que la serie puede ser estacionaria en la primera o en la segunda diferencia, en cuanto a nivel se refiere; sin embargo, si nuevamente se observa el correlograma c), se puede apre-

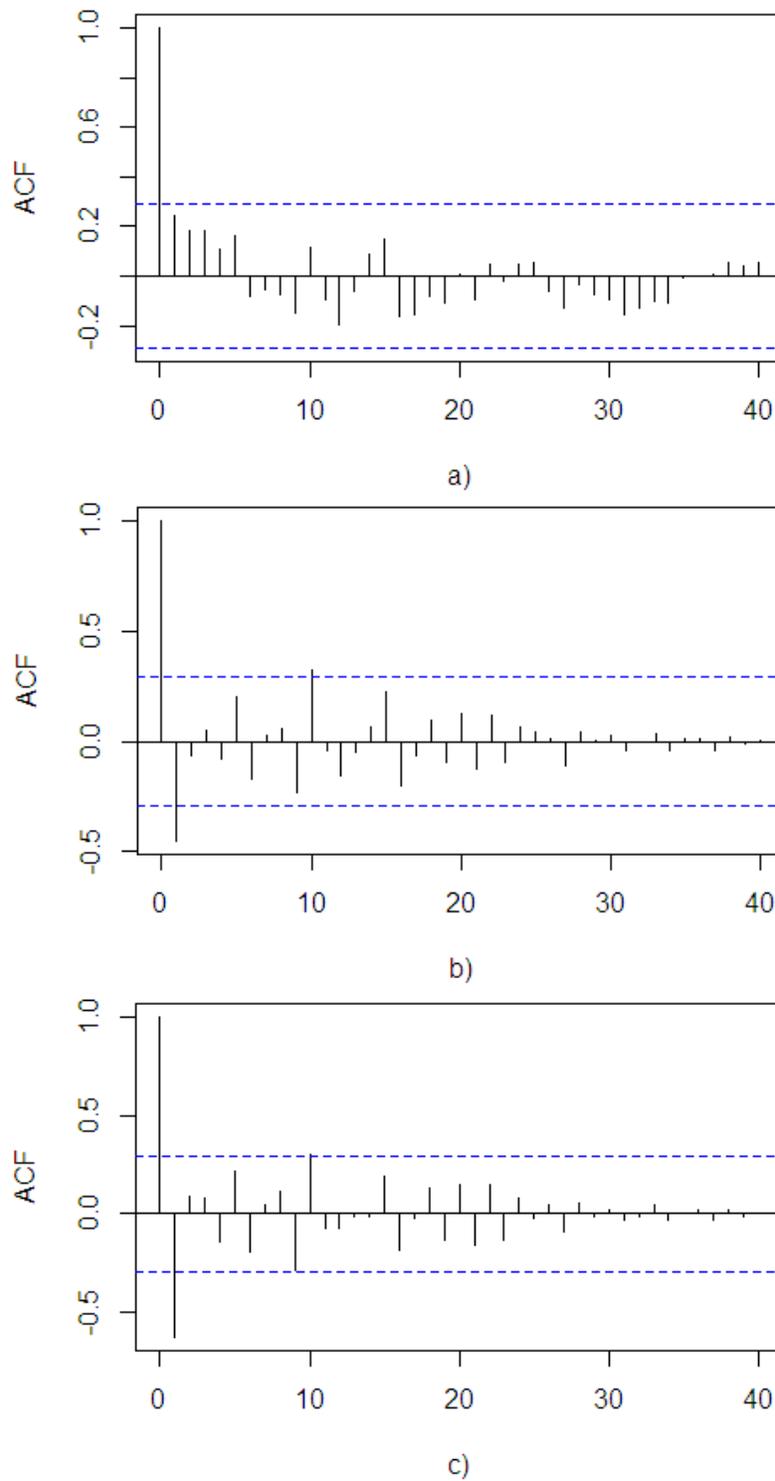


Figura 4.6: FAC muestral de: a) $\{T(GMAI_t)\}$, b) $\{\nabla T(GMAI_t)\}$ y c) $\{\nabla^2 T(GMAI_t)\}$.

ciar que tres de las 40 autocorrelaciones no son significativamente cero, además, de las 40 autocorrelaciones se tiene que 18 son positivas; mientras que el correlograma b) muestra que únicamente dos autocorrelaciones se encuentran fuera de la banda de confianza y que el número de autocorrelaciones positivas son 21 de 40 autocorrelaciones. Por lo tanto, basta con diferenciar una sola vez la serie $\{T(GMAI_t)\}$ para obtener una serie estacionaria, con ello evitando la sobrediferenciación de la serie.

En la Figura 4.7, se ilustra la serie resultante después de aplicar la primera diferencia a la serie $\{T(GMAI_t)\}$.

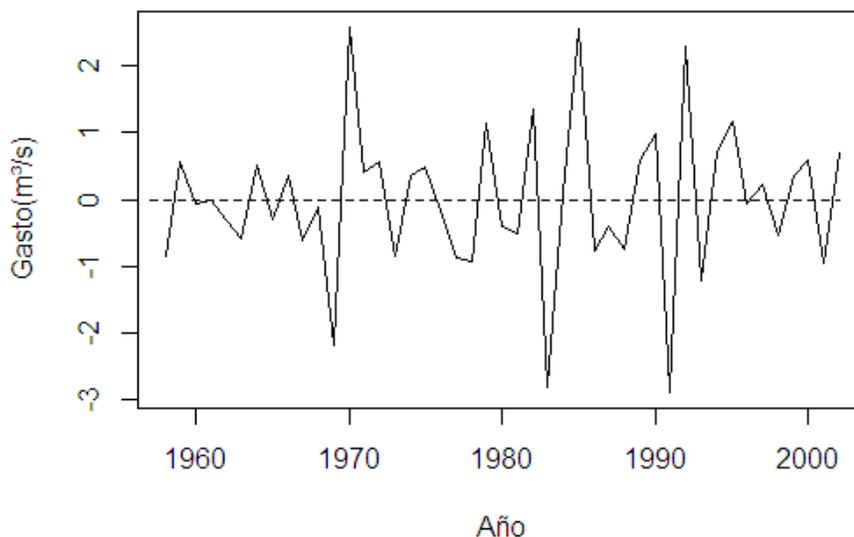


Figura 4.7: Gráfica de la serie $\{\nabla T(GMAI_t)\}$ con media constante y varianza aproximadamente estable.

Ajuste del modelo

Un vez que se obtuvo la serie estacionaria, se procede a la identificación de modelos para la serie $\{T(GMAI_t)\}$, mediante la función de autocorrelación muestral y la función de autocorrelación parcial muestral de $\{\nabla T(GMAI_t)\}$, las cuales se presentan en el Cuadro 4.1; posteriormente, se elige el modelo que tenga el mínimo valor del estadístico AIC.

Al inspeccionar la FAC muestral en el Cuadro 4.1, se aprecia que la primera autocorrelación es sensiblemente grande en comparación con las otras autocorrelaciones, entonces es razonable postular un valor de $q = 1$ para un modelo ARIMA; mientras que en la FACP muestral se observa que no hay un corte notorio en las autocorrelaciones parciales para un cierto retraso p , sino que la caída de la FACP es un poco suave, así pues, es razonable proponer al 1 y 2 como posibles valores para representar al parámetro p ; para ver de

manera gráfica el porqué de esta sugerencia, se grafican los valores del Cuadro 4.1 en la Figura 4.8.

FAC muestral										
k	1	2	3	4	5	6	7	8	9	10
r_k	-0.450	-0.060	0.051	-0.075	0.198	-0.171	0.024	0.055	-0.229	0.321
k	11	12	13	14	15	16	17	18	19	20
r_k	-0.042	-0.157	-0.048	0.064	0.223	-0.199	-0.064	0.094	-0.094	0.125
k	21	22	23	24	25	26	27	28	29	30
r_k	-0.128	0.116	-0.093	0.067	0.045	0.011	-0.112	0.045	0.005	0.027
k	31	32	33	34	35	36	37	38	39	40
r_k	-0.039	-0.004	0.037	-0.043	0.015	0.015	-0.038	0.016	-0.011	0.007
FACP muestral										
i	1	2	3	4	5	6	7	8	9	10
$\hat{\phi}_{ii}$	-0.450	-0.330	-0.187	-0.222	0.082	-0.049	-0.032	0.014	-0.282	0.059
i	11	12	13	14	15	16	17	18	19	20
$\hat{\phi}_{ii}$	0.181	0.000	-0.213	-0.105	0.152	0.053	-0.050	-0.056	-0.116	-0.048
i	21	22	23	24	25	26	27	28	29	30
$\hat{\phi}_{ii}$	-0.197	0.049	0.057	0.173	-0.094	0.065	0.068	0.103	-0.045	-0.097
i	31	32	33	34	35	36	37	38	39	40
$\hat{\phi}_{ii}$	0.031	0.007	0.019	-0.060	0.022	0.006	-0.107	-0.086	-0.009	0.000

Cuadro 4.1: FAC y FACP muestrales para $\{\nabla T(GMAI_t)\}$.

Por consiguiente, de la etapa de identificación de modelos para la serie $\{T(GMAI_t)\}$, surgen como posibles los siguientes modelos (Cuadro 4.2):

Modelo	Varianza estimada	Log- Verosimilitud	Criterio AIC
ARIMA(2,1,1)	0.829	-60.01	128.02
ARIMA(1,1,1)	0.8286	-60.01	126.02

Cuadro 4.2: Modelos propuestos para el ajuste de $\{T(GMAI_t)\}$.

Como se mostró en la Sección 3.1.3, uno de los criterios para elegir el modelo de ajuste es el criterio AIC, el cual consiste en elegir aquel modelo que tenga el mínimo valor de AIC. Así, de acuerdo al Cuadro 4.2, el modelo de ajuste sugerido para los datos de gastos mínimos anuales instantáneos en metros cúbicos por segundo, de la estación hidrométrica Paso ancho con clave 20036 es un modelo $ARIMA(1, 1, 1)$.

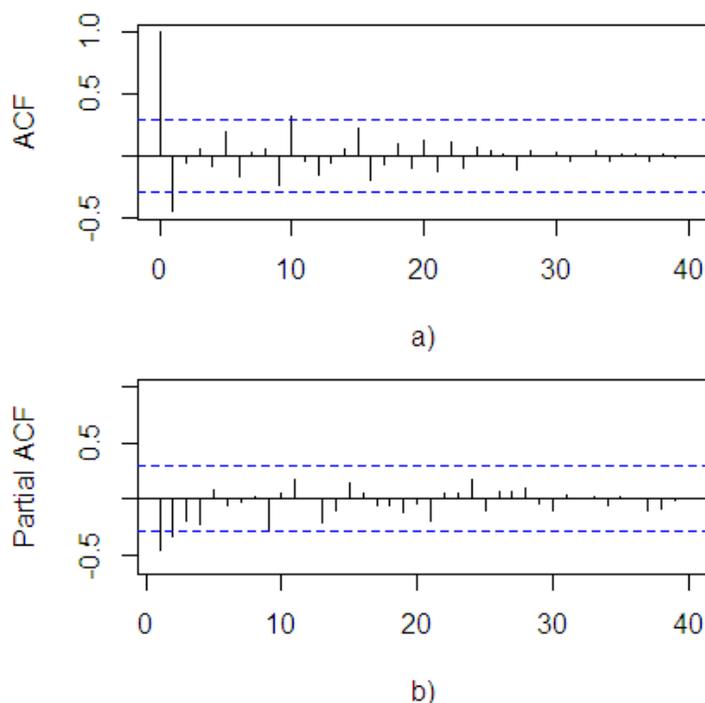


Figura 4.8: a) FAC y b) FACP muestrales para $\{\nabla T(GMAI_t)\}$.

Resta verificar las condiciones de estacionariedad e invertibilidad proporcionadas en el Cuadro 2.1 (Sección 2.4.4.). Los valores estimados de las constantes de peso para el modelo ARIMA(1,1,1) son: $\alpha_1 = 0.0413$ y $\beta_1 = -0.7553$, donde

$$|0.0413| < 1,$$

$$|-0.7553| < 1,$$

de aquí que, ambas condiciones se satisfacen. Por lo tanto, el modelo ARIMA(1,1,1) queda expresado de la siguiente forma:

$$X_t = 0.0413X_{t-1} + Z_t - 0.7553 Z_{t-1},$$

donde Z_t es ruido blanco con media cero y varianza 0.8286.

Validación del modelo

Con el fin de detectar violaciones a los supuestos del modelo ARIMA(1,1,1), se lleva a cabo la validación del modelo mediante el *análisis residual*, esto es, aplicando la función de autocorrelación a los residuos.

Nótese de la Figura 4.9 a) que, los residuos presentan media significativamente cero, y la varianza no parece seguir algún patrón de crecimiento o de decrecimiento, de hecho, la

gráfica de los residuos presenta varianza no monótona, indicativo de que la transformación aplicada para estabilizar la varianza de la serie $\{GMAI_t\}$ sea fructífero. Por otro lado, la Figura 4.9 b) indica que las autocorrelaciones son las correspondientes a un proceso de ruido blanco, esto debido a que cada una de las autocorrelaciones se encuentran dentro de la banda de confianza, es decir, no difieren significativamente de cero.

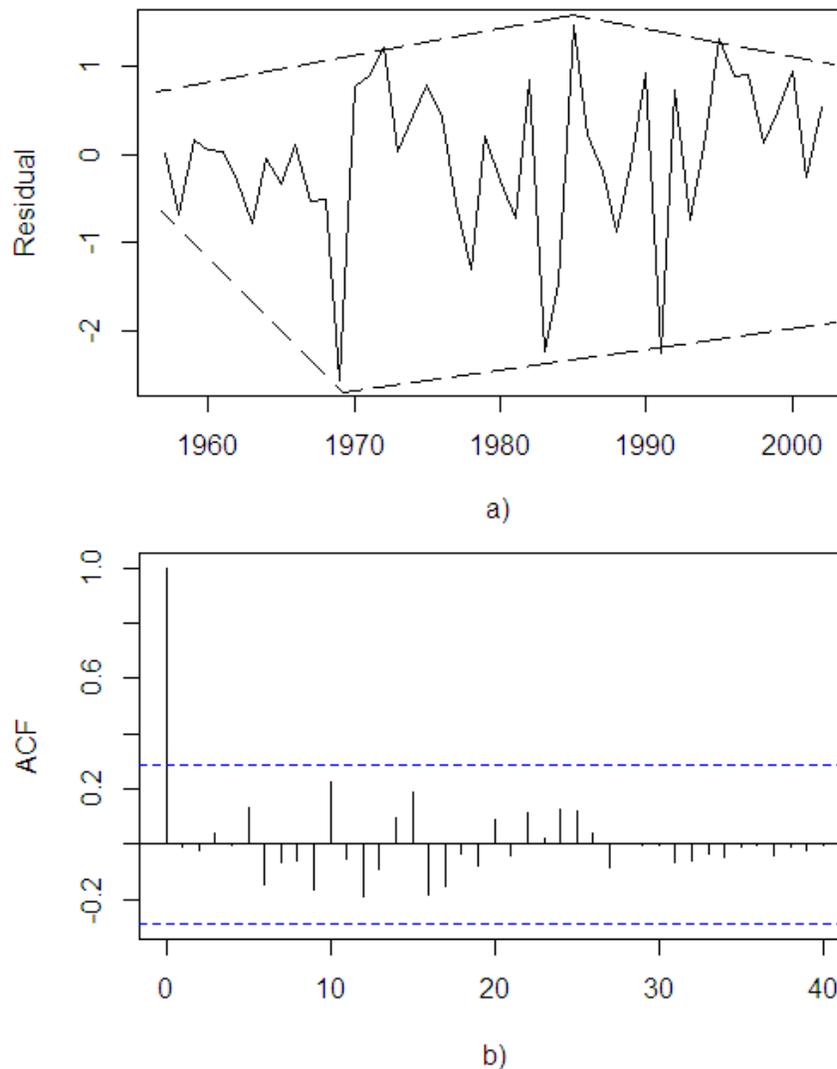


Figura 4.9: a) Gráfica de los residuos con varianza no monótona y
b) Función de autocorrelación muestral de los residuos.

Por consiguiente, de la etapa de validación del modelo ARIMA(1,1,1) se concluye que, no se encuentra evidencia convincente para rechazar dicho modelo. En la Figura 4.10 se ilustra el modelo ajustado a la serie $\{GMAI_t\}$.

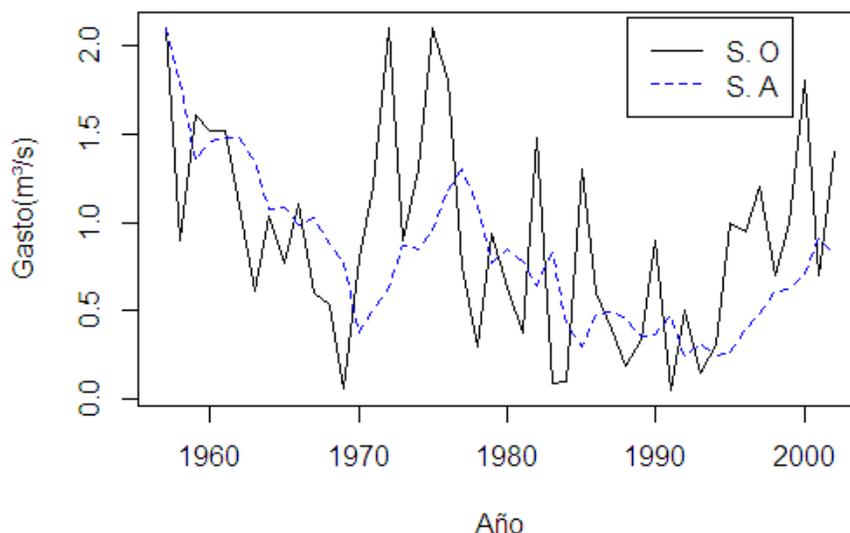


Figura 4.10: S.O: serie observada $\{GMAI_t\}$ y
S. A: serie ajustada mediante ARIMA(1,1,1).

Pronóstico

Después de la validación del modelo, se lleva a cabo el pronóstico para el periodo del año 2003 hasta el año 2005.

Periodo	Valor real	ARIMA(1,1,1)				ARIMA(3,2,1)	
		Pronóstico puntual	Error absoluto	Horizonte negativo	Horizonte positivo	Pronóstico puntual	Error absoluto
2003	0.40	0.9627	0.5627	0.1616	5.7320	1.35	0.95
2004	0.95	0.9479	0.0021	0.1482	6.0626	1.21	0.26
2005	0.60	0.9473	0.3473	0.1401	6.4052	1.20	0.60

Cuadro 4.3: Tabla comparativa de modelos ARIMA.

En el Cuadro 4.3 se muestra el valor real observado, el pronóstico puntual, el error absoluto, horizonte negativo y horizonte positivo para cada periodo predicho mediante el modelo ARIMA(1,1,1). Obsérvese, que también aparece el pronóstico puntual obtenido mediante el modelo ARIMA(3,2,1), dicho modelo, fue encontrado sin antes haber aplicado una transformación estabilizadora de varianza a los datos de $\{GMAI_t\}$. La finalidad de proporcionar ambos pronósticos puntuales consiste en observar la importancia de estabilizar la varianza, antes de identificar un modelo ARIMA, ya que de no ser así, se podría encontrar un modelo sobre-parametrizado, como es el caso del modelo ARIMA(3,2,1), que tiene el doble de parámetros en comparación con el modelo ARIMA(1,1,1), lo cual implica una pérdida de datos innecesaria y además, un pronóstico puntual más distante del valor

real observado. Ambos modelos son correctos, pero sin duda, el error absoluto del modelo ARIMA(1,1,1) es más aceptable que el error absoluto del modelo ARIMA(3,2,1) en cada periodo, por lo que se concluye que la bondad del modelo ARIMA(1,1,1) es mejor a la del modelo ARIMA(3,2,1).

Sin embargo, nótese del Cuadro 4.3 que aunque el pronóstico puntual de ARIMA(1,1,1) no coincide con el valor real observado en cada periodo, dichos valores sí se encuentran dentro del horizonte negativo y positivo, con un nivel de confianza del 95 %. Gráficamente, esto se ilustra en la Figura 4.11.

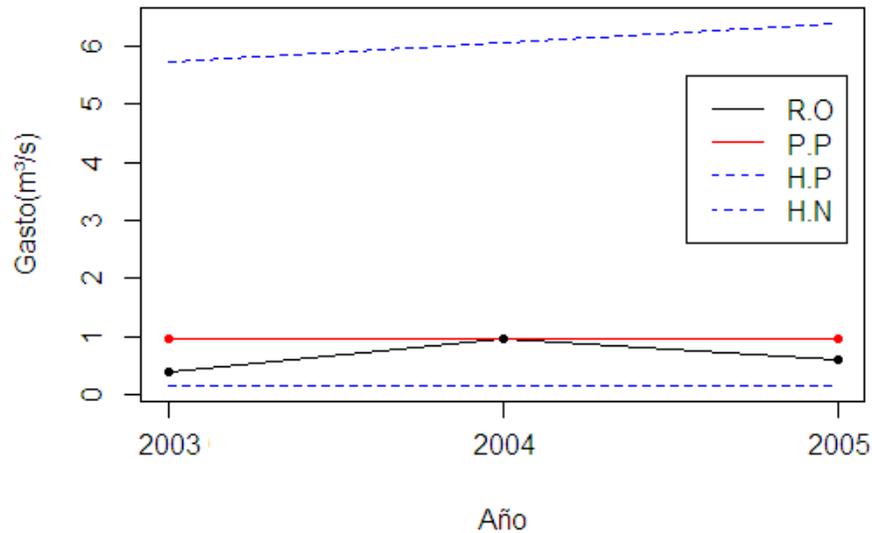


Figura 4.11: Pronóstico de la serie $\{GMAI_t\}$ Paso ancho, clave 20036, mediante el modelo ARIMA(1,1,1).

R.O: valor real observado, P.P: pronóstico puntual,
H.P: horizonte positivo y H.N: horizonte negativo.

Capítulo 5

Conclusiones

La estimación de las aportaciones que se espera ocurran en periodos futuros son de gran interés en fenómenos hidrológicos, por ejemplo, para una mejor planeación que permita la producción de energía eléctrica o la prevención de desastres por inundación. Es por ello, que en el objetivo de la presente tesis se propuso utilizar una base de datos hidrométricos, que permitiera aplicar conocimientos de estadística y probabilidad que puedan utilizarse para tratar de determinar el comportamiento de estos fenómenos.

Con el fin de satisfacer tal objetivo en la presente tesis, se empleó la base de datos “*gastos de agua mínimos anuales instantáneos en metros cúbicos por segundo*” (*GMAI*), de la estación hidrométrica *Paso Ancho* con clave 20036, dentro de la cuenca del río Verde del Estado de Oaxaca, para determinar el comportamiento de los registros históricos de dichos datos de aforo, con la finalidad de pronosticar comportamientos futuros.

Se descartó el uso de el análisis de *medias y medianas móviles* en los datos de *GMAI*, debido a que éstas únicamente son formas útiles de ajustar las aportaciones de manera laboriosa mediante comparaciones del comportamiento de los datos, sin permitir estimar comportamientos futuros. Por otro lado, aunque la *extrapolación de curvas de tendencia* es un método útil para ajuste y pronóstico, también se descartó, debido a que no funcionaría en el sentido de que las variaciones en los datos alrededor de la media son, para todo fin práctico muy irregulares (véase, Figura 4.3), lo cual no facilita el ajuste mediante una curva de tendencia.

Más aún, los *métodos cualitativos* no fueron considerados como posibles formas de analizar los datos de *GMAI*, ya que dichos métodos se usan cuando los datos históricos no se encuentran disponibles o son relativamente pocos; pero, este no es el caso para las aportaciones de *GMAI* Paso Ancho, ya que se tiene conocimiento del comportamiento de los datos históricos y además, se cuenta con una cantidad considerable de estos.

Otro de los métodos de ajuste y pronóstico es el *suavizamiento exponencial*; sin embargo, los modelos empleados en el suavizamiento exponencial imponen limitantes en lo que respecta a la representación de fenómenos reales, debido a que no consideran un componente aleatorio que les permita mayor flexibilidad. Por tal razón, para llevar a cabo la aplicación de la teoría de series de tiempo a los datos de GMAI se empleó un modelo de la clase de modelos de suavizamiento de la metodología de Box-Jenkins, conocido como $ARIMA(p, d, q)$.

El modelo resultante para ajustar los datos de los gastos de agua mínimos anuales instantáneos en metros cúbicos por segundo, de la estación hidrométrica Paso Ancho, después de la etapa de validación fue $ARIMA(1, 1, 1)$, el cual queda expresado por la siguiente ecuación:

$$X_t = 0.0413X_{t-1} + Z_t - 0.7553 Z_{t-1},$$

donde Z_t es ruido blanco con media cero y varianza 0.8286. Dicho modelo reproduce el comportamiento de las aportaciones deducidas por el gasto de agua en la estación Paso Ancho, por lo que se realizaron predicciones estadísticamente válidas para los periodos 2003, 2004 y 2005, y se observó que estas siguen un comportamiento similar a las aportaciones reales observadas, pese al comportamiento irregular de la serie.

La predicción que produjo el modelo $ARIMA(1, 1, 1)$ tiene la propiedad de ser óptima, en el sentido que ningún otro modelo da una predicción con menor error cuadrático medio. Además, la metodología de Box-Jenkins es innovadora para los datos de *GMAI* Paso Ancho.

Se concluye que, de acuerdo con los resultados matemáticos obtenidos en la presente tesis, se puede contribuir en el análisis de registros hidrométricos mediante el enfoque de Box-Jenkins, siendo así; la teoría de series de tiempo una alternativa más para la descripción, explicación, predicción y control de aportaciones hidrométricas.

Sin embargo, los resultados obtenidos en la presente tesis pueden diferir con el punto de vista dado en la ingeniería; se busca un ajuste perfecto sin que el modelo necesariamente alcance la convergencia matemática que lo justifique. Lo anterior no es más que el resultado de lo divididas que están las ciencias, que no permiten concebir a los procesos naturales como parte de todo un sistema general. Esta discrepancia de visiones ofrece la oportunidad del desarrollo de nuevas investigaciones, de hecho, si se desea concebir una idea hidrológica en los datos de *GMAI*, el análisis de series de tiempo aquí expuesto, debe ser complementado con un análisis subjetivo, sobre todo para explicar en un modelo de ajuste el comportamiento de los gastos pico.

Otro de los métodos de predicción estudiados recientemente, al igual que la metodo-

logía de Box-Jenkins, son las *redes neuronales artificiales*¹. Por lo tanto, un posible trabajo futuro de análisis consistiría en trabajar las aportaciones del gasto de agua mediante redes neuronales artificiales y comparar las nuevas estimaciones con las ya obtenidas en la presente tesis.

¹Una red neuronal artificial es una estructura que está diseñada para resolver ciertos tipos de problemas al tratar de emular la manera en que el cerebro humano podría resolver el problema. En el entorno de predicción, las redes neuronales son algunas veces usadas para resolver problemas de predicción en lugar de utilizar un enfoque de construcción de un modelo formal o el desarrollo del conocimiento subyacente del sistema que es necesario, para desarrollar un procedimiento de predicción analítica. Véase [Montgomery, 2008], Sección 7.7.

Apéndice A

Datos

En este apartado, se muestra la base de datos del gasto de agua mínimo anual instantáneo en metros cúbicos por segundo que se usó para llevar a cabo la aplicación de la teoría de series de tiempo, la cual fue extraída del Boletín Hidrométrico de la Comisión Federal de Electricidad (CFE, 2005).

La base de datos cuenta con 49 registros hidrométricos que van del periodo 1957 a 2005, también se indica la fecha en la que fueron tomados cada uno de los registros.

Año	Fecha	Gasto (m^3/s)	Año	Fecha	Gasto (m^3/s)
1957	ABR./vs	2.10	1974	ABR./vs	1.30
1958	ABR./vs	0.90	1975	MAY./vs	2.10
1959	ABR./vs	1.60	1976	JUN./08	1.80
1960	MAY./23	1.52	1977	MAR./27	0.75
1961	MAY./22	1.52	1978	ABR./vs	0.30
1962	MAY./vs	1.09	1979	MAY./vs	0.94
1963	ABR./vs	0.61	1980	ABR./03	0.63
1964	MAY./07	1.03	1981	MAY./vs	0.38
1965	ABR./28	0.77	1982	DIC./26	1.48
1966	MAY./19	1.10	1983	MAY./02	0.09
1967	MAY./vs	0.60	1984	MAY./03	0.10
1968	MAR./vs	0.53	1985	MAR./29	1.30
1969	MAY./10	0.06	1986	MAR./vs	0.60
1970	MAY./vs	0.80	1987	ABR./05	0.40
1971	MAY./vs	1.20	1988	ABR./vs	0.19
1972	ABR./vs	2.10	1989	ABR./28	0.34
1973	MAY./vs	0.90	1990	MAR./15	0.90

Año	Fecha	Gasto (m^3/s)	Año	Fecha	Gasto (m^3/s)
1991	MAR./31	0.05	1999	vs./vs	1.00
1992	MAY./vs	0.50	2000	ABR./vs	1.80
1993	JUN./07	0.15	2001	MAR./27	0.70
1994	ABR./vs	0.31	2002	SEP./03	1.40
1995	vs./vs	1.00	2003	vs./vs	0.40
1996	ABR./vs	0.95	2004	ABR./vs	0.95
1997	MAY./vs	1.20	2005	ABR./vs	0.60
1998	MAY./26	0.70			

Apéndice B

Propiedades

Propiedad 2.4.1 Si $\{X_t\}$ es un proceso estacionario con media cero y función de autocovarianza $\gamma(\cdot)$ y si $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, entonces para cada $t \in \mathbb{T}$, la serie

$$Y_t = \psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j},$$

es estacionaria con función de autocovarianza

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k).$$

Prueba: Como se tiene que $\{X_t\}$ es estacionario,

$$E(Y_t) = \sum_{j=-\infty}^{\infty} \psi_j E(X_{t-j}) = 0,$$

y

$$\begin{aligned} \gamma_Y(h) &= E \left[\left(\sum_{j=-\infty}^{\infty} \psi_j X_{t+h-j} \right) \left(\sum_{k=-\infty}^{\infty} \psi_k X_{t-k} \right) \right] \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k E(X_{t+h-j} X_{t-k}) \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k). \end{aligned}$$

Así, $E(Y_t)$ y $\gamma_Y(h)$ ambos son finitos e independientes de t ; [Brockwell, 1987].

■

Apéndice C

Teorema de Schur

Los módulos de las raíces de la ecuación

$$g^p - a_1g^{p-1} - a_2g^{p-2} - \dots - a_{p-1}g - a_p = 0$$

serán todos menores que la unidad, si y sólo si los p determinantes que se muestran a continuación son positivos:

$$D_1 = \begin{vmatrix} -1 & a_p \\ a_p & -1 \end{vmatrix},$$

$$D_2 = \begin{vmatrix} -1 & 0 & a_p & a_{p-1} \\ a_1 & -1 & 0 & a_p \\ a_p & 0 & -1 & a_1 \\ a_{p-1} & a_p & 0 & -1 \end{vmatrix},$$

$$\vdots$$

$$D_p = \begin{vmatrix} -1 & 0 & \cdots & 0 & a_p & a_{p-1} & \cdots & a_1 \\ a_1 & -1 & \cdots & 0 & 0 & a_p & \cdots & a_2 \\ \cdots & & \cdots & & \cdots & & \cdots & \\ a_{p-1} & a_{p-2} & \cdots & -1 & 0 & 0 & \cdots & a_p \\ a_p & 0 & \cdots & 0 & -1 & a_1 & \cdots & a_{p-1} \\ a_{p-1} & a_p & \cdots & 0 & 0 & -1 & \cdots & a_{p-2} \\ \cdots & & \cdots & & \cdots & & \cdots & \\ a_1 & a_2 & \cdots & a_p & 0 & 0 & \cdots & -1 \end{vmatrix}$$

([Guerrero, 2003], pp. 58).

Bibliografía

- [Barnes, 1954] Barnes, F. B. (1954). Storage required for a city water supply. In *J. Instn. Engrs. Australia*, volume 26, pages 198–203.
- [Bartlett, 1946] Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. R. Stat. Soc. Ser. B* 8, pages 27–41.
- [Bowerman, 2009] Bowerman, Bruce L., R. L. K. O. A. B. (2009). *Pronósticos, series de tiempo y regresión: un enfoque aplicado*. CENGAGE Learning(original), Progreso No. 10 Col. Centro Ixtapaluca Edo. De México, 4 edition. Impreso en Cosegraf la versión en Español.
- [Box and Cox, 1964] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. In *Journal of the Royal Statistical Society. Series B (Methodological)*, volume 26, number 2, pages 211–252. Wiley.
- [Box and Jenkins, 1970] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Series in Probability and Statistics. Holden-Day, U.S.A., first edition.
- [Brockwell, 1987] Brockwell, Peter J., R. A. D. (1987). *Time Series: Theory and Methods*. Series in Statistics. Springer, Colorado, U.S.A., first edition.
- [Brockwell, 2002] Brockwell, Peter J., R. A. D. (2002). *Introduction to Time Series and Forecasting*. Series in Statistics. Springer, United States of America, 2nd edition. Dimensions: 9.2 x 7.9 x 1.4 inches and includes ITSM 2000.
- [Cervantes, 1982] Cervantes, J. E. (1982). El uso de modelos estocásticos en el análisis de excedencias fluviales. *Boletín técnico No. 31, Chapingo (México)*, pages 75–83.
- [Chatfield, 2003] Chatfield, C. (2003). *The Analysis of Time Series, An Introduction*. Chapman and Hall/CRC, Blvd, Boca Raton, Florida 33431. United States of America, 6 edition. The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale.
- [Cowpertwait, 2009] Cowpertwait, Paul S. P., A. V. M. (2009). *Introductory Time Series with R*. Springer, New York, U.S.A.

-
- [Degroot, 1988] Degroot, M. H. (1988). *Probabilidad y Estadística*. Addison Wesley Iberoamericana, México, 2da edition.
- [Guerrero, 2003] Guerrero, V. M. (2003). *Análisis estadístico de series de tiempo económicas*. Thomson, México, D.F., 2da edition.
- [Hazen, 1914] Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. In *Transactions of the American Association of Civil Engineers*, volume 77, pages 1539–1669.
- [Hildebrand, 1998] Hildebrand, David K., R. L. O. (1998). *Estadística aplicada a la administración y a la economía*. Serie de probabilidad y estadística. Addison Wesley Longman., Boulevard de las Cataratas No. 3, Col Jardines del Pedregal, Mexico, DF., 3ra. edition. Esta versión en español es la única autorizada con 4000 ejemplares.
- [Kalbfleisch, 1985a] Kalbfleisch, J. G. (1985a). *Probability and Statistical Inference*, volume 1 : Probability. Springer-Verlag, New York, U.S.A., second edition.
- [Kalbfleisch, 1985b] Kalbfleisch, J. G. (1985b). *Probability and Statistical Inference*, volume 2: Statistical Inference. Springer-Verlag, New York, U.S.A., second edition.
- [Koutsoyiannis, 2005] Koutsoyiannis, D. (2005). Stochastic simulation of hidrosystems. In *Water Encyclopedia*, volume 1. John Wiley & Sons, Inc.
- [Lind, 2004] Lind, Douglas A., W. G. M. R. D. (2004). *Estadística para Administración y Economía*. Alfaomega, Colombia, 11^a edition.
- [Montgomery, 2008] Montgomery, Douglas C., C. L. J. M. K. (2008). *Introduction to time series analysis and forecasting*. Series in probability and statistics. Hoboken, N.J. John Wiley, United States of America.
- [Pankratz, 1983] Pankratz, A. (1983). *Forecasting with Univariate Box - Jenkins Models: Concepts and Cases*. Series in Probability and Statistics. Wiley, U.S.A., first edition.
- [Quenouille, 1949] Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *J. R. Stat. Soc. Ser. B 11*, pages 68–84.
- [Quijano, 1977] Quijano, J. (1977). Predicción de lluvias mediante el análisis de series de tiempo. Chapingo, México.
- [Reinsel, 2003] Reinsel, G. C. (2003). *Elements of Multivariate Time Series Analysis*. Series in Statistics. Springer, U.S.A., second edition.
- [Ross, 1972] Ross, S. M. (1972). *Introduction to probability models*. Academic Press, United States of America, sixth edition.
- [Ross, 1996] Ross, S. M. (1996). *Stochastic Processes*. John Wiley & Sons, Inc., New York, second edition.
-

-
- [Shumway, 2010] Shumway, Robert H., D. S. S. (2010). *Time Series Analysis and Its Applications*. Series Springer Texts in Statistics. Springer, U.S.A., 3rd edition.
- [Slutzky, 1927] Slutzky, E. (1927). The summation of random causes as the source of cyclic processes. In *Econometrica*, volume 5, number 2 (Apr., 1937), pages 105–146. The Econometric Society.
- [Thomas and Fiering, 1962] Thomas, H. A. J. and Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences in the analysis of river basins. In *Design of Water Resources Systems*, chapter 12, pages 459–493, Ed. A. Maass et al. Harvard University Press, Cambridge, Massachusetts.
- [Wold, 1938] Wold, H. O. (1938). *A Study in the Analysis of Stationary time Series*. Almqvist & Wiksell, Uppsala, Sweden, first edition.
- [Yaglom, 1955] Yaglom, A. M. (1955). Correlation theory of processes with random stationary n th increments. In *Matematicheskii Sbornik (N. S.)*, volume 37(79), number 1, pages 141–196.
- [Yule, 1927] Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to wolfer’s sunspot numbers. *Phil. Trans. R. Soc. London Ser. A 226*, pages 267–298.
-