



UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA

**“ANÁLISIS DE LA DEMANDA DE ENERGÍA
ELÉCTRICA USANDO SERIES DE TIEMPO”**

TESIS
PARA OBTENER EL TÍTULO DE:
LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA:
MARÍA ASUNCIÓN GONZÁLEZ SIERRA

DIRECTOR DE TESIS:
M.C. JOSÉ DEL CARMEN JIMÉNEZ HERNÁNDEZ

HUAJUAPAN DE LEÓN, OAXACA
JULIO DE 2010

A MIS PADRES,
CON TODO MI AMOR.

Agradecimientos

A Dios, por permitirme vivir y por la familia maravillosa de la cual me concedió la dicha de formar parte.

A los seres a quienes más amo en la vida, mis padres, Constantino González Sierra y Reyna Sierra Rojas. Por haber puesto toda su confianza en mí, porque no dudaron ni por un instante que yo lograría cumplir este sueño. Porque siempre han procurado el bienestar y felicidad sus hijos, sin escatimar esfuerzos. Por todos los sabios consejos que me han dado y que no han pasado desapercibidos en mi vida. A ellos, que nunca terminaré de agradecer el haberme dado la vida y el permitirme realizar este sueño. Y sobre todo porque ellos son mi más grande ejemplo a seguir.

A mis hermanos, Alex, Nicolás, Reyna, Daniel y René, porque de alguna manera me motivan a superarme, y a seguir adelante día a día.

Mi más sincero agradecimiento a mi asesor de tesis, M.C. José del Carmen Jiménez H., por el tema de tesis propuesto y por dirigir mis pasos durante la realización de la misma; por su paciencia, tiempo y apoyo que me brindó. A mis sinodales M.C. Marisol López Cerino, M.C. Ana Delia Olvera Cervantes y al M.C. Vulfrano Tochihuitl Bueno, por el tiempo que dispusieron a la revisión de este trabajo y por sus valiosos comentarios y observaciones. También agradezco a la M.C. Verónica Borja Macías y al Dr. Jesús Fernando Tenorio Arvide, quienes me indujeron a concluir esta tesis a tiempo y me ayudaron con los trámites de la misma.

A David, sin quien la realización de este sueño no hubiera sido posible. Por el apoyo incondicional y comprensión que me brindó durante mi estancia en la Universidad. Porque siempre estuvo a mi lado para apoyarme, ayudarme a levantar cuando caía y seguir adelante en los momentos más difíciles de mi vida. Por haberme enseñado, que todo sueño por difícil que parezca de realizar, es posible. Porque siempre buscó la manera de motivarme para seguir adelante. Por compartir conmigo su amplio conocimiento, y sobre todo, por tener un corazón tan noble y porque nunca dejó de creer en mí, me enseñó a confiar en mí misma y me ayudó a ser una mejor persona.

A mis compañeros de grupo y demás amigos de quienes no menciono nombres para no omitir alguno, por haberme hecho más agradables los momentos que compartimos durante estos cinco años. Porque después de convivir con ellos tanto tiempo aprendí a quererlos como a una familia.

A todos los maestros que tampoco menciono nombres por no omitir alguno, que en algún momento me transmitieron sus amplios conocimientos dentro y fuera de las aulas de esta alma máter.

Y a todos aquellos que de alguna manera, influyeron para que este trabajo pudiera realizarse.

Índice general

Índice general	v
1. Preliminares	3
1.1. Demanda de energía eléctrica	4
1.2. Series de tiempo	6
1.2.1. Objetivos de las series de tiempo	7
1.2.2. Aplicaciones de las series de tiempo	8
1.2.3. Análisis de las series de tiempo	8
2. Conceptos de probabilidad y estadística	13
2.1. Definiciones	13
2.1.1. Variables aleatorias	16
2.1.2. Distribuciones de probabilidad	21
2.1.3. Variables aleatorias múltiples	23
2.2. Estadística	27
2.2.1. Estimación	29
2.2.2. Métodos de estimación puntual	31
2.3. Procesos estocásticos	33
3. Fundamentos de series de tiempo	41
3.1. Fundamentos y técnicas descriptivas de series de tiempo	41
3.1.1. Transformaciones	42
3.1.2. Análisis de series que contienen tendencia	42
3.1.3. Análisis de series que contienen variación estacional	47
3.2. Modelos que describen series de tiempo	49
3.2.1. Función autocorrelación parcial	50
3.2.2. Procesos puramente aleatorios	53
3.2.3. Funciones autocorrelación y autocovarianza muestral	54
3.2.4. Procesos lineales estacionarios	56
3.2.5. Modelos de series de tiempo estacionarios	60
3.3. Identificación de modelos	71
3.3.1. Pasos para la identificación	72

3.4. Estimación de parámetros	73
3.4.1. Estimación de los parámetros de un proceso autorregresivo	73
3.4.2. Estimación de los parámetros de un proceso de medias móviles	75
3.4.3. Estimación de los parámetros de un proceso autorregresivo de medias móviles	76
3.5. Verificación	76
3.5.1. Análisis residual	77
4. Aplicación	79
4.1. Datos	79
4.2. Análisis	80
5. Conclusiones	91
A. Datos	93
B. Correlogramas	95
Bibliografía	101

Introducción

Existen distintas herramientas estadísticas que son de utilidad cuando se requiere un modelo para ajustar a un conjunto de datos, una de estas herramientas son las series de tiempo. Muchas aplicaciones reales se han trabajado con series de tiempo y han obtenido muy buenos resultados. En general, cuando se tienen datos ordenados cronológicamente en el tiempo, resulta útil modelarlos aplicando estas herramientas. En la vida real, la mayoría de datos que por una u otra razón deben registrarse se encuentran ordenados en el tiempo, es por esto que resulta fácil modelarlos mediante una serie de tiempo.

Se pueden encontrar datos que, por su naturaleza no permiten realizar un buen ajuste, es decir, además de ser no estacionarios, su comportamiento no permite identificar claramente sus componentes. Esto complica el análisis y se tienen que probar diferentes métodos antes de obtener el que logre ajustar los datos. Sin embargo, una vez que ya se cuenta con el modelo tentativo, se debe verificar que el modelo realmente sea el adecuado, es decir, el ajuste no es tan aleatorio como se esperaría.

El objetivo de esta tesis, consiste en llevar a cabo el análisis de los datos de demanda máxima mensual de energía eléctrica correspondientes a una subestación de la Comisión Federal de Electricidad (CFE), para ver si alguna de las componentes de serie de tiempo se encuentran presentes en los datos y realizar el ajuste mediante un modelo.

En el Capítulo 1, se presenta una introducción a la demanda de energía eléctrica, así como a las series de tiempo, sus aplicaciones, objetivos y clasificación.

En el Capítulo 2, se expone parte de la teoría de probabilidad, estadística y procesos estocásticos, haciendo énfasis en estos últimos, por ser la principal base sobre la que se sustenta el estudio de las series de tiempo.

En el Capítulo 3, se muestran los modelos principales de series de tiempo, las técnicas que las describen, así como también los métodos clásicos para realizar el análisis de una serie de tiempo.

La aplicación de series de tiempo para datos de demanda de energía eléctrica se muestra

en el Capítulo 4, se lleva a cabo el análisis de los datos y el ajuste del modelo para los mismos.

Finalmente, en el Capítulo 5 se muestran las conclusiones obtenidas a partir del análisis en el Capítulo 4.

En el Apéndice A se muestran los datos de demanda máxima mensual de energía eléctrica utilizados para llevar a cabo el análisis.

En el Apéndice B se presentan los patrones más comunes para determinar los modelos que corresponden a una serie de tiempo, esto basándose en las características de la función autocorrelación y la función autocorrelación parcial.

Hoy en día, existen también herramientas computacionales que facilitan el estudio y análisis de las teorías probabilísticas y estadísticas. Una de estas herramientas nos la proporciona el software estadístico R project, bajo licencia GNU General Public Licence, para Windows. Las gráficas y análisis presentadas en este trabajo fueron realizados con la ayuda de dicho paquete.

Capítulo 1

Preliminares

Una rama de las ciencias exactas que más aplicaciones tiene en la solución de problemas reales, es la matemática. La probabilidad y la estadística son ramas de esta que a lo largo de los años han mostrado una gran variedad de aplicaciones. En cualquier análisis de un problema real que involucre probabilidad y/o estadística se tiene presencia de incertidumbre. En el proceso de modelación y solución de un problema, se usan datos incompletos o aproximaciones, por lo que no se tiene una certeza completa de la solución. Esto se debe, en gran parte, a la naturaleza misma del problema, a la aproximación de cantidades en el procedimiento de solución y a la incertidumbre de los datos, los cuales deben ser tomados en cuenta, aunque esto no quiere decir que resolver un problema usando teoría de probabilidad y estadística no resulte confiable. Actualmente existen técnicas estadísticas, que son utilizadas para describir el comportamiento de los datos recolectados en determinada área o investigación. En este trabajo se lleva a cabo un análisis de la demanda de energía eléctrica usando series de tiempo; esto con el fin de observar si alguna de sus componentes se encuentran presentes en los datos.

En nuestra vida diaria existen fenómenos que son clasificados como deterministas y aleatorios. Los fenómenos deterministas son aquellos que, repetidos un cierto número de veces, siempre se obtienen los mismos resultados. Sin embargo, es posible repetir el mismo experimento bajo las mismas condiciones y obtener diferentes resultados, a esto es a lo que se conoce como fenómenos aleatorios. La demanda de energía eléctrica de una ciudad es considerada como un fenómeno aleatorio.

Este capítulo se divide en dos secciones. En la primera se habla acerca de los aspectos generales referentes a la demanda de energía eléctrica, mientras que en la segunda se hace referencia a las de series de tiempo.

1.1. Demanda de energía eléctrica

El proceso de generación y distribución de energía eléctrica resulta ser muy complicado. Pocas personas imaginan la gran cantidad de órdenes que se mandan a distintas centrales eléctricas por el simple hecho de encender o apagar una lámpara. En este trabajo no se aborda el tema de generación más bien el de la demanda de energía eléctrica. Se realiza un análisis del comportamiento de los datos de esta demanda en una subestación de la Comisión Federal de Electricidad (CFE).

La energía eléctrica, es un concepto asociado al tiempo y a la potencia nominal de una determinada carga eléctrica. La unidad de medida de la energía eléctrica es el kilovatio-hora o kiloWatt-hora (kWh). El medidor de energía, almacena el valor acumulado de toda la energía consumida durante el ciclo de lectura.

En lo referente a la generación de energía eléctrica, en 1937 México padecía circunstancias difíciles en cuanto al abastecimiento y la cobertura del servicio, el cual estaba a cargo de instituciones privadas. Los objetivos de desarrollo tanto en áreas rurales como urbanas, no eran cubiertos por las empresas privadas que tenían menos del 38 por ciento de la cobertura de energía eléctrica, por lo que, en agosto de 1937 se creó la CFE cuyo objetivo primordial fue construir la infraestructura necesaria para dotar de energía eléctrica a toda la población.

Actualmente, la capacidad instalada que reporta la CFE es de 49,861 Megawatts, dividiéndose ésta en diferentes fuentes: las termoeléctricas, los productores independientes de energía, las hidroeléctricas, las carboeléctricas, las geotérmicas y las eoléctricas [7].

A lo largo de toda la República Mexicana, se cuenta con 177 centrales generadoras y se cuenta con una red de transmisión de 48,527 km. En los últimos años se ha presentado una preocupación por satisfacer el abastecimiento de energía, que cubra el ritmo acelerado de crecimiento urbano e industrial que el país presenta [7].

Se le denomina *demanda eléctrica* a la suma de todas y cada una de las potencias consumidas en el mismo instante. Existen varios tipos de demanda entre los cuales están: la *demanda instantánea*, que es la potencia eléctrica consumida en el sistema completo en un instante concreto; la *demanda horaria*, que es la energía utilizada en una hora; y la *demanda diaria*, que se refiere a la energía absorbida a lo largo de un día.

La demanda de energía eléctrica ha ido en aumento desde los últimos años. Actualmente, esta demanda muestra un gran auge, puesto que la electricidad es usada en todos los sectores. Por ello, la demanda de energía eléctrica, además de ser el mejor indicador de carga de trabajo a la que se está sometiendo al conjunto del sistema eléctrico, también es un claro reflejo de la actividad económica y del bienestar de un país [7].

Los valores de la demanda varían constantemente, esta variación depende de muchos factores de difícil medición, por ejemplo, condiciones ambientales, estaciones anuales, según el día de la semana, luminosidad y otros. Es decir, en tan solo un día la demanda puede cambiar según la hora, en un año según la estación en que se encuentre. En [7] se menciona que las máximas demandas se producen en invierno, en días laborables y muy fríos. Esto se debe principalmente a que en días laborables existe un gran movimiento de la población, lo cual trae como consecuencia la utilización de más energía eléctrica que la usada en días no laborables, cuando la mayor parte de la gente se encuentra en sus casas, es decir, si no es en invierno o en días muy fríos, pues de esta manera, las personas que cuentan con calefacción necesitan de este servicio. También existe incremento de gasto de energía durante las primeras horas de la tarde en los días de verano, debido principalmente al uso de aire acondicionado, realizando así un gasto considerable de energía eléctrica mayor al de otros días del año. De este modo, existen curvas típicas de invierno y verano que es cuando existe mayor consumo de energía eléctrica. Por otro lado, también se tiene una medición de las demandas bajas de energía eléctrica. Éstas se observan regularmente en las primeras horas de la mañana, en días festivos navideños o en puentes de primavera.

Para medir el consumo total de energía eléctrica en tiempo real, existen dos opciones:

- La primera consiste en realizar un conteo de la totalidad de clientes que consumen energía en todo el sistema.
- La segunda opción es llevar a cabo un conteo de los valores instantáneos de todos los grupos generadores que producen en dicho sistema.

Siempre que se habla de demanda y generación instantánea en tiempo real, se cumple que:

$$\text{Demanda instantánea real} = \text{Generación instantánea real.}$$

De esta manera se puede asegurar que la potencia total generada más/menos la potencia que circula por las líneas de interconexión con otros países es equivalente a la potencia total consumida o bien es a lo que se llama *demanda de energía eléctrica*.

Para calcular la demanda de energía eléctrica se requieren de cientos de equipos de medida que operen de manera simultánea, otorgándole a estos equipos, así como a las vías y equipos de transmisión de las medidas de potencia, una importante responsabilidad, al mismo tiempo una alta fiabilidad.

A pesar de que la mayor demanda de energía eléctrica ocurre en un número de horas muy reducido, el sistema tiene la obligación de disponer de la capacidad suficiente para atender dicha demanda.

La estructura del consumo nacional distingue dos destinos de las fuentes energéticas: por una parte el sector productor de energía, y por otra el consumo final, que a su vez se clasifica en consumo no energético y en consumo energético. Éste último absorbe la mayor parte de la demanda total y es donde se encuentra el mayor potencial para consolidar el ahorro y el uso más eficiente de los recursos. Dentro de este destino destacan cuatro sectores, que, por orden de importancia, son los siguientes:

- a) transporte;
- b) industria y minería;
- c) residencial;
- d) comercial y público;
- e) agrícola.

En resumen, se tiene que la demanda máxima representa para un instante dado, la máxima coincidencia de cargas eléctricas (motores, compresores, iluminación, equipo de refrigeración, etc.) operando al mismo tiempo. Es decir, la demanda máxima corresponde a un valor instantáneo en el tiempo. De esta manera, resulta muy importante medir y prever la demanda de energía eléctrica, esto ayuda a administrar mejor el sistema eléctrico. Además, también se dice que su evolución resulta ser un gran indicador acerca de la sociedad y su desarrollo económico.

1.2. Series de tiempo

Una *serie de tiempo* es definida como un conjunto de observaciones hechas secuencialmente en el tiempo; es decir, valores ordenados cronológicamente.

Hoy en día las series de tiempo tienen muchas aplicaciones y se usan abundantemente en distintas áreas de la ciencia, principalmente en economía, medicina, ingeniería, física, etc. Claros ejemplos de series de tiempo son, las mediciones diarias de temperatura, los precios de venta de un determinado artículo, la demanda de energía eléctrica en determinada población, el índice de natalidad y mortalidad, solo por mencionar algunos. Existen pocas ramas de la ciencia en las que no aparecen datos que puedan ser considerados como series temporales, puesto que comúnmente se hacen las mediciones siguiendo un orden cronológico.

Muchas instituciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar o prevenir, para ello suelen utilizar series de tiempo para predecir lo que ocurrirá con una variable en el futuro a partir del comportamiento de esa variable en el pasado, de aquí que las series de tiempo sean una herramienta básica para lograr este objetivo.

En esta tesis se describen algunos modelos estadísticos para analizar una serie de tiempo sin tener en cuenta otras variables que puedan influir en la misma, estos métodos son denominados univariantes. En particular, nos interesa la aproximación estocástica (frente a la determinista) es decir aquella que supone que la serie temporal tiene un carácter probabilístico, por haber sido generada por alguna variable aleatoria con una distribución de probabilidad determinada aunque comúnmente desconocida.

Existen series de tiempo *continuas* y *discretas*. Se dice que una serie de tiempo es *continua* si las observaciones son tomadas continuamente en el tiempo y es *discreta* si las observaciones son tomadas solo en tiempos específicos, por lo regular estos tiempos son igualmente espaciados.

1.2.1. Objetivos de las series de tiempo

Los objetivos del análisis de una serie de tiempo son diversos, entre los que pueden destacar, la predicción, el control de un proceso, la simulación de procesos, y la generación de nuevas teorías físicas o biológicas, entre otros. A continuación se hace mención de estos objetivos.

- *Descripción:* Al momento de graficar los datos se obtienen medidas simples descriptivas de la serie de tiempo a analizar, así como sus principales propiedades y algunas componentes de las series de tiempo.
- *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, a esto se le llama predicción. Dicho en otras palabras, es la estimación de valores futuros de la variable en función del comportamiento pasado de la serie. Este objetivo se usa ampliamente en economía e ingeniería.
- *Control de procesos:* Cuando una serie de tiempo es generada, la cual mide la calidad de un procedimiento de manufactura, es decir, se trata de seguir la evolución de una variable determinada con el fin de regular su resultado. Esta teoría se utiliza frecuentemente en medicina, en los centros de control de enfermedades y en las fábricas.
- *Simulación:* Se emplea en investigación aplicada, cuando el proceso es muy complejo para ser estudiado de forma analítica.

Además la naturaleza intrínseca de las series de tiempo es que las observaciones son dependientes o están correlacionadas y el orden de las observaciones es de mucha importancia.

1.2.2. Aplicaciones de las series de tiempo

Debido a la gran diversidad de los campos donde las series de tiempo son aplicadas, éstas se han clasificado en diferentes grupos, entre los cuales se pueden mencionar los siguientes:

- *Series económicas.* En esta categoría se consideran las series de tiempo que son usadas para analizar datos económicos como pueden ser precios, valores, deudas, la tasa de inflación, tasas de desempleo, el producto interno bruto anual, entre otras.
- *Series Físicas (aplicadas a las ciencias).* Dentro de este grupo se consideran las series que son aplicadas al análisis de datos físicos como datos meteorológicos, la temperatura máxima o mínima diaria, las precipitaciones diarias, la velocidad del viento (energía eólica), la energía solar, datos sísmológicos o vulcanológicos y geofísica.
- *Series de mercadotecnia.* Este tipo de series tienen relación con el comercio, el porcentaje de ventas, costos de productos y producción, entre otros.
- *Series demográficas.* Se usan cuando se realizan estudios poblacionales, por ejemplo el total de habitantes en una determinada población, la tasa de natalidad y mortalidad anual, tasas de dependencia y los censos que verifican el cambio poblacional.
- *Control de proceso.* Este tipo de series se usan para detectar cambios en la evolución de un proceso de manufactura, midiendo una variable, la cual muestra la calidad del proceso.
- *Procesos binarios.* Este tipo especial de series de tiempo ocurren cuando las observaciones pueden tomar solo uno de dos valores posibles usualmente estos valores son 0 y 1.
- *Procesos puntuales.* Este tipo de series tienen lugar cuando las observaciones ocurren aleatoriamente a través del tiempo, es decir, no siguen una secuencia específica.
- *Series de telecomunicación.* Son utilizadas en el análisis de señales, por ejemplo miden la intensidad o la frecuencia con que han sido transmitidas ciertas señales de comunicación, éstas no siguen un patrón, pues resulta difícil conocer el comportamiento de este tipo de series.
- *Series de transporte.* Este tipo de series analizan los datos de tráfico, es decir, analizan y miden la cantidad de automóviles en movimiento a lo largo del día, con esto se realiza el análisis de la cantidad de tráfico en determinado día.

1.2.3. Análisis de las series de tiempo

El análisis de series de tiempo es una descripción de los elementos que la constituyen. Es decir, consiste en verificar cuáles son las componentes presentes en la serie o cuál es el

comportamiento que los datos observados presentan.

Una característica en el análisis de las series de tiempo, es el hecho de que las observaciones usualmente no son independientes y además, es importante tomar en cuenta el orden en que ocurren dichas observaciones. Cuando las observaciones sucesivas son dependientes, los valores futuros pueden ser determinados (o predecidos) a partir de las observaciones pasadas.

Una serie es *determinística* si puede ser predecida con exactitud. Muchas series de tiempo son estocásticas, esto es, las predicciones son realizadas basándose en valores pasados, aunque las predicciones exactas resultan casi imposibles.

El primer paso al analizar una serie de tiempo es presentar un gráfico temporal que indique la evolución de la variable en cuestión a lo largo del tiempo, con el valor de la serie en el eje de ordenadas y en el eje de las abscisas se representan los valores del tiempo. El segundo paso es determinar si la secuencia de valores es completamente aleatoria o se puede encontrar algún patrón a lo largo del tiempo. Resulta bastante difícil realizar un análisis de las series de tiempo si no se tienen graficados los datos, pues ésto nos permite tener una visualización mucho más amplia de lo que se está estudiando y es más fácil obtener conclusiones.

Componentes de una serie de tiempo

Las técnicas convencionales del análisis de las series de tiempo consisten principalmente en descomponer la variación existente en la serie, algunos tipos de variación son tendencia, variación estacional, cambios cíclicos y fluctuaciones irregulares. Éstas se conocen también como componentes de una serie de tiempo.

- *Tendencia.* Es la dirección general de la variable en el periodo de observación, es decir, el cambio a largo plazo de la media de la serie. También se le conoce con el nombre de evolución subyacente.
- *Efecto estacional.* Existen series de tiempo tales que exhiben variación en periodos relativamente cortos de tiempo, puede ser anual, mensual, etc.; es decir, corresponde a fluctuaciones periódicas de la variable. Normalmente estos efectos suelen ocurrir debido a cambios ambientales o condiciones climáticas.
- *Cambios cíclicos.* Además de los efectos estacionales, algunas series de tiempo presentan variación en un periodo debido a otras causas físicas o ambientales. Este tipo de variación está caracterizado por oscilaciones alrededor de la tendencia con una duración aproximada de 2 o 8 años.
- *Fluctuaciones irregulares.* Después de extraer de la serie la tendencia y variaciones cíclicas, nos quedará una serie de valores residuales, que pueden ser o no totalmente aleatorios. Se vuelve al punto de partida, ahora también nos interesa determinar si esa

secuencia temporal de valores residuales puede o no ser considerada como aleatoria pura. Este tipo de variación corresponde a movimientos erráticos que no sigue un patrón específico y que obedecen a diversas causas. Esta componente es prácticamente impredecible.

La serie temporal es la suma de éstas componentes, es decir

Valor observado = Tendencia + Estacionalidad + Componente Irregular,
aunque también en ocasiones se puede ver como el producto de estas componentes.

Series de tiempo estacionarias y no estacionarias

Una serie de tiempo es *estacionaria* si no hay cambios sistemáticos en ella, es decir, no hay tendencia; dicho en otras palabras, la media y la variabilidad se mantienen constantes a lo largo del tiempo.

Análogamente, una serie es *no estacionaria* si la media y/o variabilidad cambian a lo largo del tiempo.

Las series no estacionarias pueden mostrar cambios en la varianza, exhibir tendencia (la media crece o disminuye a lo largo del tiempo), y además presentar efectos estacionales.

Una serie de tiempo estacionaria tiene ciertas ventajas sobre la no estacionaria, por ejemplo, las predicciones resultan más fáciles de obtener. Puesto que la media es constante, se puede estimar con todos los datos que se tienen y utilizar este valor para predecir una nueva observación. También se pueden obtener intervalos de predicción (confianza) para las predicciones, asumiendo que la variable considerada sigue una distribución conocida.

Los modelos cuantitativos se pueden clasificar de acuerdo a la información que se tenga en modelos econométricos o multivariados y en modelos univariados o series de tiempo. En esta tesis únicamente se tratará el modelo univariante.

Los modelos econométricos tratan de explicar el comportamiento de una o más variables en función de la evolución de otras variables que se consideran explicativas. Las variables explicadas por el modelo se denominan *endógenas*, mientras que las variables explicativas del modelo, pero no explicadas por él, se denominan *predeterminadas*. En este enfoque no se necesita conocer ninguna relación de causalidad, explicativa del comportamiento de la variable endógena, ni en su defecto, ninguna información relativa al comportamiento de otras variable explicativas, ya que en este caso no existe este tipo de variables. Es suficiente con conocer una serie temporal de la variable en estudio, para estimar el modelo que se utilizará para predecir. La predicción univariante se utiliza, en problemas económicos, principalmente con dos objetivos:

- La predicción de algunas variables explicativas de un modelo causal, cuando se espera que en el futuro conserven algunas de las características de su evolución en el pasado.

- La predicción a corto plazo (de 1 a 4 trimestres), debido a su gran capacidad para recoger la dinámica en el comportamiento de la variable estudiada. Además, en condiciones normales, cuando no existen bruscas alteraciones respecto a la experiencia reciente de la variable, estos métodos pueden proporcionar buenas predicciones.

Capítulo 2

Conceptos de probabilidad y estadística

La probabilidad es la disciplina matemática que se encarga del estudio de los fenómenos o experimentos aleatorios. Se entiende por fenómenos aleatorios, aquellos experimentos que cuando se repiten bajo las mismas condiciones, los resultados obtenidos no siempre son los mismos. Es importante aceptar que aun cuando un experimento se haya efectuado un gran número de veces, no es posible predecir los resultados del experimento, aunque se haya llevado a cabo bajo las mismas condiciones iniciales, es por esto que se considera que el experimento es aleatorio. De esta manera, la teoría de la probabilidad tiene el objetivo de desarrollar modelos matemáticos para cualquier fenómeno aleatorio.

En este capítulo se presentan conceptos de probabilidad y estadística, así como una breve introducción a los procesos estocásticos.

2.1. Definiciones

El propósito original de la teoría de probabilidad, era describir la relación estrecha que existía entre la experiencia y los juegos de azar, y el principal objetivo era calcular ciertas probabilidades. Actualmente, la probabilidad utiliza modelos matemáticos para describir fenómenos que presenten incertidumbre. Un concepto importante en el cual se basan estos modelos es el de experimento aleatorio, el cual se da a continuación.

Definición 2.1.1 *Un experimento aleatorio es aquel experimento que presenta las siguientes condiciones:*

- a) *tiene al menos dos posibles resultados;*
- b) *el conjunto de posibles resultados se conoce antes de que el experimento se realice;*

c) *puede repetirse indefinidamente bajo las mismas condiciones.*

Ejemplo 2.1.2 *Algunos ejemplos de experimentos aleatorios son:*

- *lanzar una moneda al aire;*
- *lanzar un dado al aire;*
- *abrir un libro en determinada página.*

Es importante aclarar que existen diferentes interpretaciones de la probabilidad, a continuación se mencionan éstas:

- *Interpretación frecuentista:* esta interpretación se basa en repetir un proceso un número grande de veces en condiciones similares y se estima la probabilidad de un evento usando la frecuencia con que ocurre.
- *Interpretación clásica:* se basa en la idea de los resultados “igualmente verosímiles” (cada uno de los resultados posibles tienen la misma probabilidad de ocurrir).
- *Interpretación subjetiva:* esta interpretación resulta ser personal, se basa en la experiencia del investigador, el cual asigna una probabilidad a uno de los posibles resultados de un proceso, esta asignación depende de su criterio.

El *espacio de probabilidad* es el modelo matemático creado para estudiar los fenómenos aleatorios. A continuación se tiene la definición de espacio muestral.

Definición 2.1.3 *El espacio muestral o espacio muestra asociado a un experimento aleatorio es el conjunto que consta de todos los posibles resultados de éste. El espacio muestral se denotará por el símbolo Ω .*

A cada elemento de Ω se le llama *punto muestral*. El espacio muestral más simple es el que contiene un número finito de n puntos. Si n es muy pequeño, es fácil visualizar el espacio. Si Ω es finito se denotan los puntos muestrales como E_1, E_2, \dots, E_n , y si es infinito numerable se denotará como E_1, E_2, \dots

Los eventos o sucesos en un experimento aleatorio son subconjuntos de un espacio muestral. Si el evento consta de un solo punto, entonces el evento es denominado simple. En adelante, los eventos se denotarán con letras mayúsculas.

El número $P(A)$ representa una forma de medir la posibilidad de observar la ocurrencia del evento A , al efectuar una vez el experimento aleatorio. Por ejemplo, dado un espacio muestral Ω finito con puntos muestrales E_1, E_2, \dots, E_n , existe un número asociado a cada

E_j , llamado la probabilidad de E_j y denotado por $P(E_j)$. Estos números son no negativos tales que

$$P(E_1) + P(E_2) + \cdots + P(E_2) = 1.$$

Definición 2.1.4 Dado un experimento aleatorio con un espacio muestral Ω , la función $P : \Omega \rightarrow \mathbb{R}$, tal que satisface los siguientes axiomas:

- a) $0 \leq P(A) \leq 1$ para toda $A \in \Omega$,
- b) $P(\Omega) = 1$,
- c) Para cualquier sucesión infinita de eventos disjuntos de Ω , A_1, A_2, \dots se cumple que

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

se llama función de probabilidad sobre Ω .

Cabe mencionar que la anterior es la definición axiomática de la probabilidad, existen también otros enfoques para definir la probabilidad, sin embargo, para nuestros propósitos, basta con referirse a esta.

Independencia de eventos

La independencia es utilizada frecuentemente en esta tesis, y es un rasgo distintivo de la teoría de la probabilidad. *Independencia* entre eventos quiere decir que la ocurrencia de un evento no influye en el resultado de otro.

Definición 2.1.5 Dos eventos A y B son independientes, si

$$P(A \cap B) = P(A)P(B).$$

Definición 2.1.6 Los eventos A_1, A_2, \dots, A_n son independientes si se cumplen todas y cada una de las siguientes condiciones:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) & i \neq j, \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) & i \neq j \neq k, \\ &\vdots \\ P(A_1 \cap \cdots \cap A_n) &= P(A_1) \cdots P(A_n). \end{aligned}$$

2.1.1. Variables aleatorias

Una variable aleatoria es una función real definida sobre el espacio muestral. Intuitivamente se dice que es un valor que depende del resultado de un experimento aleatorio. El concepto de variable aleatoria es fundamental en la teoría de probabilidad, es por eso que a continuación se da la siguiente definición.

Definición 2.1.7 *Una variable aleatoria X es una función definida sobre Ω , que toma valores reales, es decir, $X : \Omega \rightarrow \mathbb{R}$.*

En general se denotan las variables aleatorias con letras mayúsculas X, Y, \dots , en adelante se escribirá v.a. para hacer referencia a variable aleatoria.

Existen al menos dos tipos de variables aleatorias: las continuas y las discretas.

Definición 2.1.8 *Una v.a. X se dice que es discreta, si solamente puede tomar un número finito o infinito numerable de valores distintos, es decir, $X(\Omega)$ es a lo más numerable.*

Función de distribución

Toda v.a. tiene asociada una función de distribución de probabilidad. Si la v.a. X es discreta, se asigna a cada valor x de X (es decir, a cada $x \in X(\Omega)$) su probabilidad como

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

tal que cumple:

a) $p(x) \geq 0 \quad \forall x;$

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

A $p(x)$ se le conoce como *función de distribución de probabilidad* de la v.a. X .

Ejemplo 2.1.9 *Supóngase que se realiza el experimento que consiste en lanzar dos dados balanceados al aire y se observa el resultado obtenido. (x, y) representa un punto del espacio muestral, donde x es el número que aparece en el primer dado e y el número que aparece en el segundo dado.*

El espacio muestral consta de los posibles resultados al lanzar los dados, los cuales son:

$$\begin{aligned}\Omega = & \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ & (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ & (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ & (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ & (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), \\ & (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}.\end{aligned}$$

En total se tienen 36 posibles resultados o puntos muestrales, todos equiprobables, es decir cada uno de ellos con probabilidad $\frac{1}{36}$.

Algunos ejemplos de eventos o sucesos son $A = \{(1, 3)\}$, $B = \{(1, 2), (5, 5), (6, 1)\}$.

Se define la variable aleatoria X como la suma de los dos números obtenidos. El conjunto de posibles valores que la variable aleatoria X puede tomar es: $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Para obtener la probabilidad p_i de que la v.a. X tome el valor i , se debe obtener la probabilidad del evento A_i que consta de todos los pares en Ω tales que la suma de sus componentes es igual a i . Como cada punto muestral es equiprobable, la probabilidad de A_i es $P(A_i) = \frac{|A_i|}{36}$, donde $|A_i|$ representa el número de elementos en A_i . Por ejemplo, para calcular $P(X = 2)$, se tiene que sólo los elementos del par $(1, 1)$ suman 2, así $P(X = 2) = P(\{(1, 1)\}) = \frac{|\{(1, 1)\}|}{36} = \frac{1}{36}$.

Análogamente se calculan las probabilidades restantes: $p_3 = \frac{2}{36}$, $p_4 = \frac{3}{36}, \dots$,
 $p_{11} = \frac{2}{36}$, $p_{12} = \frac{1}{36}$.

Función de distribución acumulada

Definición 2.1.10 La función de distribución acumulada (FDA) de una v.a. X se define y denota por

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

Esta función, mide la probabilidad acumulada de la v.a. hasta cierto x . Algunas veces la FDA también suele denotarse por $F_X(x)$.

Propiedad 2.1.11 La FDA cumple las siguientes propiedades:

$$a) \quad \lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P(X \leq x) = 0;$$

$$b) \lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P(X \leq x) = 1;$$

$$c) \text{ Si } x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2).$$

Cuando X es una v.a. discreta, la FDA es una función escalonada.

Haciendo uso de la FDA se puede definir una v.a. continua de la siguiente manera:

Definición 2.1.12 Sea $F(x)$ la FDA de una v.a. X , entonces se dice que X es continua si $F(x)$ es continua.

Función de densidad de probabilidad

Definición 2.1.13 Sea X una v.a. continua, con FDA $F(x)$, entonces la función $f(x)$ dada por

$$f(x) = \frac{d}{dx}[F(x)],$$

siempre y cuando la derivada exista, se llama función de densidad de probabilidad (fdp) de X .

Si la v.a. X es continua, se asigna la probabilidad al suceso en que la v.a. toma valores en un determinado intervalo $[a, b]$ como,

$$P(X \in [a, b]) = \int_a^b f(x)dx,$$

donde $f(x)$ es la fdp de la v.a. X .

Propiedades de la fdp

Si $f(x)$ es la fdp de la v.a. X entonces se cumple que:

$$a) f(x) \geq 0;$$

$$b) \int_{-\infty}^{\infty} f(x)dx = 1.$$

Algunas otras propiedades que también se cumplen son:

$$a) P(a \leq x \leq b) = F(b) - F(a);$$

$$b) P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = \int_a^b f(x)dx.$$

Valor esperado

El valor esperado de una v.a. es un número que representa el promedio ponderado ó el valor medio de sus posibles valores.

Definición 2.1.14 Sea X una v.a. El valor esperado de X denotado por $E(X)$ se define como

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

Siempre que la suma y la integral sean convergentes para el primer y segundo caso respectivamente. El valor esperado también se conoce comúnmente como *esperanza*, *media*, *valor promedio* o *valor medio* y se denota por μ .

De manera más general, si $g(X)$ es una función de X , entonces el valor esperado de $g(X)$ está dado por

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

Siempre que la suma y la integral sean convergentes para cada uno de los casos. Por otro lado, si $g(X) = X^r$, entonces $E[X^r]$ se conoce como el *r-ésimo momento central*.

A continuación se especifican las propiedades que cumple el valor esperado.

Propiedad 2.1.15 Sea X una v.a con valor esperado finito, y sea c una constante. Entonces se cumplen:

- a) $E(c) = c$;
- b) $E(cX) = cE(X)$;
- c) si $X \geq 0$, entonces $E(X) \geq 0$;
- d) sean $g_1(X), g_2(X), \dots, g_n(X)$ funciones de X , entonces,

$$E \left[\sum_{i=1}^n g_i(X) \right] = \sum_{i=1}^n E[g_i(X)],$$

siempre que los valores esperados existan.

Demostración:

Véase [12], pág. 95 y [5], pág. 187.

Varianza

La varianza de una v.a. es una medida del grado de dispersión de los diferentes valores que toma la variable en cuestión.

Definición 2.1.16 Sea X una v.a. con $\mu < \infty$, la varianza de X , denotada por $\text{Var}(X)$, se define como

$$\text{Var}(X) = E[(X - \mu)^2].$$

Se tienen dos expresiones para la varianza, dependiendo si la v.a. X es discreta o continua, las cuales son:

$$\text{Var}(X) = \begin{cases} \sum (x - \mu)^2 p(x), & \text{si } X \text{ es discreta;} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, & \text{si } X \text{ es continua.} \end{cases}$$

Siempre que la suma y la integral sean convergentes para el primer y segundo caso respectivamente. La varianza se denota usualmente por σ^2 . A la raíz cuadrada de $\text{Var}(X)$ se le llama *desviación estándar* y es denotada por σ .

Al igual que el valor esperado, la varianza también cumple ciertas propiedades.

Propiedad 2.1.17 Sea X una v.a. con varianza finita, y sea c una constante. Entonces:

- a) $\text{Var}(X) \geq 0$;
- b) $\text{Var}(X) = 0 \Leftrightarrow X = c$;
- c) $\text{Var}(cX) = c^2 \text{Var}(X)$;
- d) $\text{Var}(X + c) = \text{Var}(X)$;
- e) $\text{Var}(X) = E(X^2) - [E(X)]^2$.

Demostración:

Véase [5], pp. 195-196.

Ejemplo 2.1.18 Para los datos del ejemplo 2.1.9, el valor esperado de la v.a. X es

$$\begin{aligned}
 E(X) &= \sum_i i * P(X = i) \\
 &= \sum_{i=2}^{12} i * p_i \\
 &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} \\
 &\quad + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\
 &= 7.
 \end{aligned}$$

La varianza de la v. a. X es

$$\begin{aligned}
 \text{Var}(X) &= \sum_i (i - E(X))^2 * P(X = i) \\
 &= \sum_{i=2}^{12} (i - 7)^2 * p_i \\
 &= (2 - 7)^2 \frac{1}{36} + (3 - 7)^2 \frac{2}{36} + (4 - 7)^2 \frac{3}{36} + (5 - 7)^2 \frac{4}{36} + (6 - 7)^2 \frac{5}{36} + (7 - 7)^2 \frac{6}{36} \\
 &\quad + (8 - 7)^2 \frac{5}{36} + (9 - 7)^2 \frac{4}{36} + (10 - 7)^2 \frac{3}{36} + (11 - 7)^2 \frac{2}{36} + (12 - 7)^2 \frac{1}{36} \\
 &= \frac{35}{6} \\
 &\approx 5.8.
 \end{aligned}$$

Para este ejemplo en particular se obtuvo que $E(X) = 7$, lo cual indica que 7 es el valor promedio de los valores de la v.a. X . Por otro lado, se calculó la varianza, es decir, $\text{Var}(X) \approx 5.8$, esto indica que la desviación promedio cuadrática alrededor de la media es de aproximadamente 5.8. Es decir, si se repite el experimento un número grande de veces y se observan los valores de la v.a. X se encontrará que los valores cercanos a 7 son los más frecuentes y que la medida de dispersión $(x - 7)^2$ tiene como valores más frecuentes a valores cercanos a 5.8.

2.1.2. Distribuciones de probabilidad

Una parte importante de las variables aleatorias es su función de distribución. Se escribirá " $X \sim F$ " para indicar que la v.a. X tiene cierta función de distribución F . Existen distribuciones discretas y continuas.

Distribuciones discretas

Definición 2.1.19 *La variable aleatoria X tiene distribución discreta si existe un conjunto $C \subseteq \mathbb{R}$ a lo más numerable, tal que $P(X \in C) = 1$.*

Algunas distribuciones de probabilidad discretas son:

- distribución de probabilidad binomial (Véase [12], pág. 103);
- distribución de probabilidad geométrica (Véase [12], pág. 115);
- distribución de probabilidad binomial negativa (Véase [12], pág. 122);
- distribución de probabilidad hipergeométrica (Véase [12], pág. 126).

Distribuciones continuas

Definición 2.1.20 *Una variable X tiene distribución absolutamente continua si existe una fdp $f(x)$ tal que*

$$P(X \in A) = \int_A f_X(x) dx, \quad A \subseteq \mathbb{R}.$$

Se tienen algunas distribuciones continuas importantes. La distribución de probabilidad normal es la que más se utiliza; muchos fenómenos observados en el mundo real tienen distribuciones de frecuencia relativas que pueden ser modeladas mediante una distribución de este tipo.

Distribución normal.

Definición 2.1.21 *Se dice que la v.a. continua X tiene una distribución normal o gaussiana si su fdp es*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\},$$

donde $\mu, x \in \mathbb{R}$ y $\sigma > 0$. En este caso se dice que X tiene una distribución normal con parámetros μ y σ^2 y se denota por

$$X \sim N(\mu, \sigma^2).$$

Es fácil verificar que si $X \sim N(\mu, \sigma^2)$, entonces se cumple que:

- $E(X) = \mu$;
- $\text{Var}(X) = \sigma^2$.

Una v.a. X tiene una distribución normal estándar si $\mu = 0$ y $\sigma^2 = 1$, en cuyo caso su función de densidad es la siguiente:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-x^2}{2}\right\},$$

y se escribe

$$X \sim N(0, 1).$$

Nótese que

$$X \sim N(\mu, \sigma^2) \iff Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Es decir, la v.a. X con distribución normal no estándar se transforma a una estándar $Z = \frac{X - \mu}{\sigma}$, este procedimiento se denomina estandarización.

Algunas otras distribuciones continuas son:

- distribución de probabilidad uniforme (Véase [12], pág. 174);
- distribución de probabilidad gamma (Véase [12], pág. 185);
- distribución de probabilidad beta (Véase [12], pág. 194).

2.1.3. Variables aleatorias múltiples

Definición 2.1.22 *Un vector aleatorio n -dimensional es una función del espacio muestral Ω en \mathbb{R}^n .*

En particular para $n = 2$, se tiene un *vector bivariado* o *bidimensional*.

Definición 2.1.23 *Si X y Y son variables aleatorias (v.a's) discretas, entonces se dice que el vector aleatorio (X, Y) es discreto.*

Función de distribución conjunta

Para el caso de dos variables se tiene:

Definición 2.1.24 *Sea (X, Y) un vector aleatorio discreto, la función $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $p(x, y) = P(X = x, Y = y)$ se llama función de distribución de probabilidad conjunta.*

En la definición, $P(X = x, Y = y)$ significa que es la probabilidad de que X tome el valor x y Y tome el valor y .

Para cualquier evento A en Ω se cumple:

$$P((X, Y) \in A) = \sum_{(x,y) \in A} P(X = x, Y = y).$$

Las propiedades que cumple la función de distribución de probabilidad conjunta son:

- a) $p(x, y) \geq 0 \quad \forall x, y;$
- b) $\sum_{(x,y)} p(x, y) = 1.$

Para los vectores bivariados también se define la FDA.

Definición 2.1.25 La FDA conjunta del vector aleatorio (X, Y) se define y denota como

$$F(x, y) = P(X \leq x, Y \leq y).$$

Para el caso en que (X, Y) es discreto, se tiene

$$F(x, y) = \sum_{s \leq x} \sum_{t \leq y} P(X = s, Y = t).$$

La *distribución de probabilidad marginal* para X está dada por

$$p_X(x) = P(X = x) = \sum_y p(x, y),$$

análogamente, la *distribución de probabilidad marginal* para Y está dada por

$$p_Y(y) = P(Y = y) = \sum_x p(x, y).$$

Definición 2.1.26 Sean X, Y v.a's discretas con distribución de probabilidad conjunta $p(x, y)$, entonces X y Y son independientes si y sólo si

$$p(x, y) = p_X(x)p_Y(y).$$

Teniendo los dos últimos conceptos, dadas dos v.a's independientes que tienen la misma distribución de probabilidad marginal se dice que son *v.a's independientes e idénticamente distribuidas* y se denotan por *v.a's iid*.

En el caso en que X, Y sean v.a's continuas, también se pueden definir los conceptos anteriores.

Covarianza

Definición 2.1.27 Sean X, Y v.a.'s. La covarianza de X y Y , se denota y define como

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

La covarianza cumple con una serie de propiedades, las cuales se enuncian en la siguiente propiedad.

Propiedad 2.1.28 (Propiedades de la covarianza.) Sean X, Y y Z variables aleatorias y sea c una constante, se cumple que:

- a) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$;
- b) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- c) $\text{Cov}(X, X) = \text{Var}(X)$;
- d) $\text{Cov}(c, X) = 0$;
- e) $\text{Cov}(cX, Y) = c\text{Cov}(X, Y)$;
- f) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$;
- g) Si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$;
- h) $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$.

Demostración:

Véase [12], pp. 266-267.

Obsérvese que de g) y h) se obtiene que $V(X + Y) = V(X) + V(Y)$ y $E(XY) = E(X)E(Y)$.

Coefficiente de correlación

El coeficiente de correlación entre dos variables indica el grado de dependencia lineal que hay entre ellas.

Definición 2.1.29 Sean X, Y variables aleatorias con varianzas distintas de cero, el coeficiente de correlación entre ellas, se define y denota por

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

Propiedad 2.1.30 (Propiedades del coeficiente de correlación.) *El coeficiente de correlación cumple las siguientes propiedades:*

- a) Si X y Y son independientes, entonces $\rho(X, Y) = 0$;
- b) $-1 \leq \rho(X, Y) \leq 1$;
- c) Si X y Y son independientes e idénticamente distribuidas, entonces $\rho(X+Y, X-Y) = 0$.

Cuando $\rho(X, Y) = 0$ entonces se dice que no existe correlación entre X y Y , si $|\rho(X, Y)| = 1$, entonces se dice que X y Y están *perfectamente correlacionadas*. Por otro lado, si X y Y son independientes, entonces $\rho(X, Y) = 0$, el recíproco no necesariamente es cierto.

Ejemplo 2.1.31 Sean X y Y v.a's discretas con la siguiente distribución de probabilidad:

$p_x(-1) = p_x(1) = \frac{5}{16} = p_y(-1) = p_y(1)$ y $p_x(0) = \frac{6}{16} = p_y(0)$,
y con distribución de probabilidad conjunta:

		x		
		-1	0	1
y	-1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
	0	$\frac{3}{16}$	0	$\frac{3}{16}$
	1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

Para este ejemplo se cumple que X y Y son dependientes pero con covarianza cero.

Solución.

Obsérvese que $p(0, 0) = 0$. Es claro que

$$p(0, 0) \neq p_x(0)p_y(0),$$

con lo cual queda demostrado que X y Y son dependientes.

De los datos del ejemplo se observa que $E(X) = E(Y) = 0$. También,

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyp(x, y) \\ &= (-1)(-1)(1/16) + (-1)(0)(3/16) + (-1)(1)(1/16) \\ &\quad + (0)(-1)(3/16) + (0)(0)(0) + (0)(1)(3/16) \\ &\quad + (1)(-1)(1/16) + (-1)(0)(3/16) + (1)(1)(1/16) \\ &= (1/16) - (1/16) - (1/16) + (1/16) = 0. \end{aligned}$$

Entonces,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 - 0(0) = 0.$$

Con lo cual queda probado que si dos v.a's tienen covarianza cero, no necesariamente son independientes. Nótese que $\rho(X, Y) = 0 \Leftrightarrow \text{Cov}(X, Y) = 0$, para cualesquiera v.a's X, Y . De aquí, se deduce que el hecho de que la correlación entre dos v.a's es cero no implica que X y Y son independientes.

2.2. Estadística

La estadística es la ciencia que trata de la recopilación, organización, presentación, análisis e interpretación de datos numéricos con el fin de tomar la mejor decisión y obtener significados precisos.

Es importante mencionar que dependiendo del contexto en el que se utilice, la estadística tiene tres significados: el término estadístico se refiere a una medida obtenida en una muestra; por otro lado, la palabra estadística es usada para referirse a la información estadística y ésta también es utilizada para referirse al conjunto de métodos y técnicas que se utilizan para analizar la información.

Los métodos y técnicas estadísticas, comúnmente se usan para realizar tareas de tipo descriptivas, por ejemplo, para organizar, analizar y resumir datos de tipo numérico.

Definición 2.2.1 *Una población es la colección de toda la posible información que caracteriza a un fenómeno.*

Definición 2.2.2 *Una muestra aleatoria de la población es una colección de v.a's iid X_1, X_2, \dots, X_n . El tamaño de la muestra aleatoria es representado por n .*

La estadística se divide en dos ramas: estadística descriptiva y estadística inferencial.

La estadística descriptiva consiste en representar los datos recolectados en forma de tablas y gráficas. En este tipo de estadística solo resume y describe los datos sin tratar de inferir nada más.

La estadística inferencial por su parte, como su nombre lo indica, hace inferencia acerca de los datos recopilados, es decir, hace generalizaciones que van más allá de la información que proporcionan los datos. Analiza una población basándose en una muestra de la población estudiada.

La inferencia estadística es definida como el conjunto de técnicas que permiten formular inferencias inductivas. Con inferencias inductivas se entiende como aquellos razonamientos que van de lo particular a lo general.

Medidas de posición central

Existen algunas medidas que son llamadas de posición central, entre las que destacan: la media, la moda, la mediana.

La media muestral es un valor representativo de todos los valores de la muestra, es el promedio ponderado, es decir,

Definición 2.2.3 Sean X_1, X_2, \dots, X_n v.a's iid. La media muestral se define y denota por

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

Proposición 2.2.4 Supóngase que X_1, X_2, \dots, X_n son v.a's iid, con $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$, entonces:

- a) $E(\bar{X}) = \mu$;
- b) $\text{Var}(\bar{X}) = \sigma^2/n$;
- c) $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$, para $i = 1, 2, \dots, n$.

La *mediana* de un conjunto de observaciones es el valor para el cual, cuando las observaciones se ordenan de forma creciente, la mitad de éstas es menor que este valor y la otra mitad es mayor.

La *moda* de un conjunto de observaciones es el valor que se repite con mayor frecuencia.

Medidas de dispersión

Las medidas de dispersión proporcionan el grado de propagación de una variable alrededor de una medida de posición central. Existen dos tipos de medidas de dispersión: medidas de dispersión absoluta y medidas de dispersión relativa.

Dentro de las medidas de dispersión absoluta se encuentran:

- El rango (Ra): es la diferencia entre el máximo y el mínimo de la variable en cuestión.
- Desviación absoluta media respecto a la media (d_e): indican las desviaciones de la media con respecto a la media en valor absoluto.
- Varianza: mide la dispersión de los valores de la variable respecto a la media. A mayor varianza existe mayor dispersión.
- Desviación típica o estándar.

Definición 2.2.5 Sean X_1, X_2, \dots, X_n , v.a's iid con $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$. La varianza muestral se denota y define por

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}.$$

Además, se cumple que $E(S_X^2) = \sigma^2$, lo cual indica, que la varianza muestral es un estimador insesgado de la varianza poblacional.

Definición 2.2.6 La desviación típica o estándar muestral se define como la raíz cuadrada positiva de la varianza,

$$S_X = +\sqrt{S_X^2}.$$

Las medidas de dispersión relativa miden el grado de dispersión de distintas distribuciones.

2.2.1. Estimación

Uno de los objetivos de la estadística es hacer inferencias acerca de determinadas características de la población. Estas características se conocen como parámetros. Así, un parámetro es una constante fija con valor desconocido. A continuación se tiene la definición formal de parámetro.

Definición 2.2.7 Aquella caracterización numérica de la distribución de la población que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés es a lo que se le conoce como parámetro.

Comúnmente un parámetro se denota por θ , es importante observar que θ también representa en algunas ocasiones un vector, es decir $\theta = (\theta_1, \theta_2, \dots, \theta_n)$, donde $\theta_1, \theta_2, \dots, \theta_n$ son parámetros.

Las inferencias antes mencionadas se logran haciendo estimación. Existen dos tipos de estimación, la estimación puntual y la estimación por intervalos. La primera proporciona un solo valor numérico, este valor debe ser aproximado al parámetro objetivo. La segunda consiste en dar dos números que formarán un intervalo que debe contener al parámetro objetivo. Este intervalo recibe el nombre de *intervalo de confianza estimado*.

Los estimadores puntuales más comunes son:

- la media de la muestra para estimar el valor promedio de la población;
- la desviación estándar de la muestra para estimar la desviación estándar de la población;
- la proporción en la muestra como estimador de la proporción en la población.

Por otro lado, se define un estadístico como una función que depende de la muestra, es decir,

Definición 2.2.8 *Un estadístico T , es función de las variables aleatorias observadas en la muestra que no contiene valores desconocidos y se denota por $T = f(X_1, X_2, \dots, X_n)$.*

Con lo anterior, un estadístico es usado para estimar los valores de los parámetros que se desconocen o funciones de éstos. Además, un estadístico es una variable aleatoria, mientras un parámetro es una constante. Para ilustrar el proceso de estimación se dan las siguientes definiciones:

Definición 2.2.9 *Sea θ un parámetro arbitrario. Un estimador de θ , denotado por $\hat{\theta}$ es aquel estadístico usado para estimar el valor del parámetro desconocido θ .*

La estimación de θ es un valor específico t obtenido de los datos de la muestra. De esta manera, un estimador es un estadístico mediante el cual se obtiene una estimación de las observaciones de la muestra. Se pueden definir varios estadísticos para estimar un parámetro desconocido θ .

Para fines prácticos se denotará mediante $f(x; \theta)$ a la función de densidad de probabilidad de la población de interés, donde θ es un parámetro arbitrario y la función depende de éste.

Definición 2.2.10 *Sea $\hat{\theta}$ un estimador puntual de θ . Entonces $\hat{\theta}$ es insesgado si $E(\hat{\theta}) = \theta$. En caso contrario, se dice que el estimador es sesgado.*

El sesgo B de un estimador puntual $\hat{\theta}$ se define como

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

El sesgo de un estimador puede ser positivo, negativo o cero.

Un concepto usado en estimación es el de error cuadrático medio, el cual se define a continuación.

Definición 2.2.11 *Sea $\hat{\theta}$ un estimador de un parámetro desconocido θ . El error cuadrático medio del estimador $\hat{\theta}$ del parámetro θ se define y denota por*

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

El error cuadrático medio cumple la siguiente propiedad,

$$ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left[\theta - E(\hat{\theta}) \right]^2.$$

Observación: En la práctica es común encontrar al menos dos estimadores para determinado parámetro. En estos casos se debe elegir aquél que sea el mejor en el sentido de cumplir

ciertas propiedades que aproximen mejor al parámetro. Para determinar que estimador es el más adecuado, existen dos métodos. El primero consiste en elegir el estimador tal que tenga una distribución de muestreo centrada alrededor de θ y la varianza sea la menor posible. Esto es, si $\hat{\theta}_1$ y $\hat{\theta}_2$ son insesgados, entonces se elige aquel que presente menor varianza, es decir, se elige θ_2 si

$$\text{Var}(\hat{\theta}_2) < \text{Var}(\hat{\theta}_1).$$

El segundo método consiste en usar el ECM. Para esto, se calcula el ECM para ambos estimadores y se realiza el cociente entre ellos, si el cociente es menor que 1, entonces se dirá que el estimador para el cual su ECM se encuentra en el numerador es el mejor, pues éste tendrá el menor ECM. Si el cociente es igual a 1, entonces el método no proporciona ninguna información respecto a los estimadores.

Definición 2.2.12 Sea θ un parámetro arbitrario y $\tilde{x} = (x_1, x_2, \dots, x_n)$ los datos de la muestra aleatoria. La función de verosimilitud para la muestra se denota y define por

$$L(\tilde{x}; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (2.1)$$

$L(\tilde{x}; \theta)$ es función de los parámetros, en algunas ocasiones se denota a la función de verosimilitud con $L(\theta)$.

2.2.2. Métodos de estimación puntual

Existen varios métodos para hallar la estimación de un parámetro. Los más comunes son: el método de los momentos y el método de máxima verosimilitud.

El método de los momentos consiste en igualar los momentos muestrales y los poblacionales.

El método de estimación por máxima verosimilitud consiste en hallar el valor del parámetro que maximiza la función de verosimilitud.

Definición 2.2.13 Sea $L(\tilde{x}; \theta)$ la función de verosimilitud de una muestra aleatoria X_1, X_2, \dots, X_n de una distribución con función de densidad $f(x; \theta)$. Si $t = u(x_1, x_2, \dots, x_n)$ es el valor de θ que maximiza $L(\tilde{x}; \theta)$, entonces $T = u(X_1, X_2, \dots, X_n)$ se llama estimador de máxima verosimilitud de θ , y $t = u(x_1, x_2, \dots, x_n)$ es la estimación de máxima verosimilitud.

Para simplificar la estructura de la función de verosimilitud, se busca maximizar el logaritmo natural de $L(\theta)$, lo cual se puede hacer pues la función \ln es creciente y continua en su dominio, esto asegura que la solución que maximiza $L(\theta)$ también maximiza $\ln[L(\theta)]$. Por motivos prácticos, comúnmente se escribe $l(\theta) = \ln[L(\tilde{x}; \theta)]$.

Para ilustrar este método, se tiene el siguiente ejemplo para estimar la media y la varianza de una población con distribución normal.

Ejemplo 2.2.14 Si X_1, X_2, \dots, X_n es una muestra aleatoria de una distribución normal cuya función de densidad de probabilidad está dada por

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}.$$

Determinar los estimadores de máxima verosimilitud para μ y σ^2 respectivamente.

Solución:

Para este ejemplo, es fácil notar que la función de verosimilitud depende tanto de μ como de σ , es decir, estos son los dos parámetros los cuales deben ser estimados (son los valores para los cuales la función de verosimilitud obtiene un valor máximo).

Por definición,

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ \sum_{i=1}^n -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}, \end{aligned}$$

de aquí,

$$L(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ \sum_{i=1}^n -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}.$$

Tomando logaritmos a esta última ecuación:

$$\begin{aligned} l(\mu, \sigma^2) &= \ln[L(x_1, x_2, \dots, x_n; \mu, \sigma^2)] \\ &= -n \ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ &= -n \ln(2\pi\sigma^2)^{1/2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ &= \frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Derivando con respecto a μ y σ^2 , e igualando a cero estas derivadas se obtiene

$$\frac{\partial[l(\mu, \sigma^2)]}{\partial\mu} = -\frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad (2.2)$$

y

$$\frac{\partial[l(\mu, \sigma^2)]}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0. \quad (2.3)$$

Resolviendo (2.2), para μ se obtiene

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

Sustituyendo este valor en (2.3) y resolviendo para σ^2 se tiene,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

Así, los estadísticos de máxima verosimilitud de μ y σ^2 son $T_\mu = \bar{X}$ y $T_{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$, respectivamente.

2.3. Procesos estocásticos

El estudio de los procesos estocásticos consiste en analizar el comportamiento de una variable aleatoria a través del tiempo, basándose en características no determinísticas, es decir, fenómenos aleatorios.

Para tener una mejor apreciación de las propiedades dadas en el análisis de las series de tiempo, se darán en esta sección algunos conceptos fundamentales de procesos estocásticos.

Definición 2.3.1 *Un proceso estocástico se define como una colección de v.a.'s $\{X_t : t \in T\}$ parametrizada por un conjunto $T \subset \mathbb{R}$ de subíndices, llamado espacio paramétrico. El conjunto T usualmente se interpreta como un conjunto de valores de tiempo.*

Se dice que el proceso estocástico es discreto si T es discreto (por ejemplo $T = 0, \pm 1, \pm 2, \dots$), y el proceso estocástico es continuo si T puede tomar cualquier valor en un intervalo de \mathbb{R} , por ejemplo $[0, \infty)$, es decir el tiempo es continuo. Un proceso es llamado de valores reales,

si las variables aleatorias solo pueden tomar valores reales.

Se denota la variable aleatoria en el tiempo t por $X(t)$ si el tiempo es continuo, y por X_t si el tiempo es discreto. Estas notaciones hacen referencia al estado o posición del proceso estocástico en el instante t .

Definición 2.3.2 *El espacio de estados de un proceso se define como el conjunto de todos los posibles valores que las v.a's $\{X_t\}$ pueden tomar, y se denota por E .*

Una *realización* de un proceso estocástico es una asignación de los valores de las v.a's en el espacio de estados.

Algunos ejemplos de procesos estocásticos son:

- $X(t)$: número de personas que esperan en una parada de autobús durante un instante t donde t es el tiempo medido en minutos del intervalo $[9, 10]$, el espacio de estados es $E = \{1, 2, 3, \dots\}$.
- X_t : precio del dólar en un día t del mes de junio de 2010 ($t = 1, 2, \dots, 30$), el espacio de estados para cada v.a. es continuo, por ejemplo $[0, 20]$.
- X_t : número de nacimientos en el estado de Oaxaca en el mes t ($t = 1, 2, \dots, 12$), $E = \{1, 2, 3, \dots\}$.

Similarmente a las definiciones dadas anteriormente para v.a's, se tiene la siguiente definición.

Definición 2.3.3 *Sea $\{X_t : t \in T\}$ un proceso estocástico con $E(X_t^2) < \infty$, entonces se define:*

a) *la función media del proceso por,*

$$\mu_t = E(X_t);$$

b) *la función varianza del proceso,*

$$\sigma_t^2 = E[(X_t - \mu_t)^2];$$

c) *la función covarianza entre X_r y X_s ,*

$$\gamma(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_r)(X_s - \mu_s)], \text{ donde } r, s \in T;$$

d) *y la función correlación entre X_r y X_s por,*

$$\rho(r, s) = \frac{\gamma(r, s)}{\sqrt{\sigma_r^2 \sigma_s^2}}.$$

Para definir los procesos estocásticos, se debe determinar e identificar la distribución de probabilidad asociada a cada v.a.

Clasificación de los procesos estocásticos

Los procesos estocásticos se pueden clasificar según: la estructura del conjunto de índices T y el conjunto de estados E , y según las características probabilísticas de las variables aleatorias.

Procesos estocásticos basados en la estructura de T y E

Existen cuatro tipos de procesos estocásticos dependiendo de si T y E son continuas o discretas, de esta forma se tiene la siguiente clasificación:

- *cadena* si E y T son discretos;
- *proceso puntual* si E es discreto y T es continuo;
- *sucesión de v.a's* si E es continuo y T discreto;
- *proceso continuo* si E y T son continuos.

Procesos estocásticos basados en las características probabilísticas de las v.a's.

Otra forma de clasificación de los procesos estocásticos toma en cuenta las características de las variables aleatorias en cuestión, de esta forma, los procesos estocásticos se pueden clasificar en:

- procesos estacionarios;
- procesos Markovianos;
- procesos de incrementos independientes.

En particular nos interesa conocer más acerca de procesos estocásticos estacionarios, es por esta razón que a continuación se analizan este tipo de procesos.

Procesos estacionarios

Un proceso se dice que es estacionario si X_t no depende del tiempo. Para lograr una mejor comprensión de los procesos estacionarios, es necesario definir ciertos conceptos, que son fundamentales para este tipo de procesos. Además serán de gran utilidad en el capítulo posterior.

La función de distribución conjunta *n-dimensional* para un proceso estocástico se define por

$$F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = P(X_{t_1} \leq x_{t_1}, \dots, X_{t_n} \leq x_{t_n}).$$

Los procesos estacionarios a su vez se clasifican en procesos estacionarios débiles y procesos estacionarios fuertes.

Un proceso se dice que es débilmente estacionario si todos los momentos de orden n existen y son independientes del tiempo t . Más aún, un proceso estacionario de segundo orden¹ tiene media y varianzas constantes. Formalmente se tiene la siguiente definición.

Definición 2.3.4 *Un proceso estacionario débil es aquel proceso $\{X_t : t \in T\}$, con $E(X_t^2) < \infty \forall t \in T$, tal que,*

- a) $E(X_t) = m, \quad \forall t \in T;$
- b) $\gamma(r, s) = \gamma(r + t, s + t), \quad \forall r, s, t \in T.$

En a), significa que el valor esperado es constante. El hecho de que $\gamma(r, s) = \gamma(r + t, s + t)$ significa que la función covarianza toma el mismo valor para dos v.a.'s que estén separadas por un retardo ó rezago t en el proceso, independientemente de dónde se encuentren situadas estas v.a.'s en el tiempo.

Por otro lado, un proceso se dice que es *estrictamente estacionario* ó *fuertemente estacionario*, si $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ tienen la misma distribución conjunta que $(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$.

Definición 2.3.5 *Un proceso estacionario fuerte es aquel proceso $\{X_t : t \in T\}$, tal que $\forall t \in T, E(X_t^2) < \infty$ si $\forall n \in \mathbf{N}, \forall k \in T$ y $\forall t_1, t_2, \dots, t_n \in T$,*

$$F(x_{t_1}, x_{t_2}, \dots, x_{t_n}) = F(x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k}).$$

Siempre se cumple que un proceso estrictamente estacionario implica un proceso débilmente estacionario, el recíproco no necesariamente se cumple.

Para un proceso estrictamente estacionario, la función media $\mu_t = \mu$ es una constante, siempre que $E(|X_t|) < \infty$. También si $E(X_t^2) < \infty$, entonces $\sigma_t^2 = \sigma^2$ para todo t , de aquí que la varianza también es constante.

¹Un proceso es estacionario de segundo orden si $F(x_{t_1}, x_{t_2}) = F(x_{t_1+k}, x_{t_2+k})$, para $t_1, t_2, k, t_1+k, t_2+k \in \mathbb{Z}$.

Funciones de autocorrelación y autocovarianza

Para un proceso estacionario $\{X_t : t \in T\}$, se tiene que la media $E(X_t) = \mu$ y la varianza $\text{Var}(X_t) = E(X_t - \mu) = \sigma^2$ son constantes, y la covarianza $\text{Cov}(X_t, X_s)$, la cual es una función que solo depende del rezago del tiempo. En este caso, en particular para cuando $s = t + k$ (k es un rezago), se define la *función de autocovarianza* como una función dependiente solamente del rezago k como:

$$\gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)] = \text{Cov}(X_t, X_{t+k}).$$

Por motivos prácticos, algunas veces se debe estandarizar la función autocovarianza para obtener la *función de autocorrelación* (FAC), dada por

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \quad (2.4)$$

la cual mide la correlación entre X_t y X_{t+k} ,

Para un proceso estacionario, la función de autocovarianza y la FAC cumplen las siguientes propiedades:

Propiedad 2.3.6 Sea $\{X_t : t \in T\}$ un proceso estocástico estacionario, con media μ , varianza σ^2 , función autocovarianza γ_k y función autocorrelación ρ_k , entonces:

- a) $\gamma_0 = \text{Var}(X_t) \geq 0$ y $\rho_0 = 1$;
- b) $|\rho_k| \leq 1$ y $|\gamma_k| \leq \gamma_0$;
- 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 $t = 0$;
- d) γ_k y ρ_k son positivas semidefinidas, es decir

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0$$

y

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0,$$

para cualquier conjunto de puntos t_1, t_2, \dots, t_n y cualquier número real $\alpha_1, \alpha_2, \dots, \alpha_n$.

Demostración.

- a) Se cumple por definición.

b) Se sabe que $\text{Var}(aX_t + bX_{t+k}) \geq 0$, y por propiedad (2.1.28,h), se tiene que

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

De aquí se desprenden dos casos:

■ Si $a = b = 1$, entonces

$$\begin{aligned} 0 &\leq 2\sigma^2 + 2\gamma_k \\ 0 &\leq \sigma^2 + \gamma_k \\ -\gamma_k &\leq \sigma^2. \end{aligned}$$

De esta última desigualdad y del hecho de que $\gamma_0 = \sigma^2$ y $\rho_k = \frac{\gamma_k}{\gamma_0}$, se sigue que

$$-1 \leq \rho_k. \quad (2.5)$$

■ Ahora, si $a = 1$ y $b = -1$, entonces

$$\begin{aligned} 0 &\leq 2\sigma^2 - 2\gamma_k \\ 0 &\leq \sigma^2 - \gamma_k \\ \gamma_k &\leq \sigma^2 \end{aligned}$$

De aquí se obtiene que

$$\rho_k \leq 1. \quad (2.6)$$

De (2.5) y (2.6) se obtiene que $-1 \leq \rho_k \leq 1$, es decir $|\rho_k| \leq 1$.

$|\gamma_k| \leq \gamma_0$ es consecuencia inmediata del hecho de que $-1 \leq \rho_k \leq 1$.

c) El resultado se obtiene por el hecho de que la correlación entre X_{t+k} y X_t es la misma que la correlación entre X_t y X_{t-k} , esto por ser X_t estacionario. Es decir,

$$\begin{aligned} \gamma_k &= \text{Cov}(X_{t+k}, X_t) \\ &= \text{Cov}(X_t, X_{t-k}) \\ &= \gamma_{-k}. \end{aligned}$$

d) Definiendo $X = \sum_{i=1}^n \alpha_i X_{t_i}$, el resultado es inmediato, del hecho de que

$$0 \leq \text{Var}(X) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \text{Cov}(X_{t_i}, X_{t_j}) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|}.$$

Análogamente se obtiene el resultado para ρ_k .

Lo anterior se muestra en el siguiente ejemplo.

Ejemplo 2.3.7 Caminata aleatoria

Supongamos que $\{Z_t\}$ es un proceso puramente aleatorio². Un proceso $\{X_t : t \in T\}$ es llamado caminata aleatoria si

$$X_t = X_{t-1} + Z_t.$$

Comúnmente este proceso comienza cuando $t = 0$, es decir $X_0 = 0$, de la relación recursiva anterior se obtiene que

$$X_1 = Z_1$$

y

$$X_t = \sum_{i=1}^t Z_i.$$

Es fácil verificar que $E(X_t) = t\mu_Z$ y $\text{Var}(X_t) = t\sigma_Z^2$. En efecto,

$$\begin{aligned} E(X_t) &= E\left[\sum_{i=1}^t Z_i\right] \\ &= \sum_{i=1}^t E[Z_i] \\ &= t\mu_Z. \end{aligned}$$

Y para la varianza se tiene,

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}\left(\sum_{i=1}^t Z_i\right) \\ &= \sum_{i=1}^t \text{Var}(Z_i) \\ &= t\sigma_Z^2. \end{aligned}$$

Como la media y la varianza dependen de t , el proceso es no estacionario.

Por otro lado, la función de autocovarianza está dada por,

²Para un proceso puramente aleatorio se cumple que $E(Z_i) = \mu_Z$ y $\text{Var}(Z_i) = \sigma_Z^2$, además las Z_i 's son v.a's iid.

$$\begin{aligned}\gamma_k &= \text{Cov}(X_t, X_{t+k}) \\ &= \text{Cov}(X_t, X_t + Z_{t+1} + \dots + Z_{t+k}) \\ &= \text{Cov}(X_t, X_t) \\ &= \text{Var}(X_t) \\ &= t\sigma_Z^2.\end{aligned}$$

Hasta aquí se mostraron los conceptos básicos correspondientes a la teoría de probabilidad y estadística, así como también la introducción a procesos estocásticos, los cuales son la base para el estudio de series de tiempo, dicho estudio se muestra en el siguiente capítulo.

Capítulo 3

Fundamentos de series de tiempo

No siempre es posible predecir con total exactitud algún valor de la serie de tiempo. Sin embargo, se pueden encontrar ciertos patrones en las series de tiempo como tendencia, estacionalidad, entre otros mencionados en el Capítulo 1, estos pueden describir el comportamiento de la serie. También es posible describirlas mediante modelos basados en distribuciones de probabilidad. En este capítulo se estudian las principales técnicas y modelos que describen a las series de tiempo.

3.1. Fundamentos y técnicas descriptivas de series de tiempo

Graficar los datos que conforman una serie de tiempo resulta ser de gran utilidad para llevar a cabo un buen análisis de la misma. Una vez que los datos han sido graficados es posible observar algunas discontinuidades o irregularidades, incluso datos atípicos, si es que existen en la serie de tiempo. En el Capítulo 1 se describen algunas componentes de las series de tiempo, entre las que destacan: tendencia, cambios cíclicos y efecto estacional. Tales discontinuidades pueden ser cambios de nivel repentinos, por lo que sería de mucha utilidad analizar los datos por intervalos o segmentos iguales. Es necesario analizar estas componentes y si es posible realizar transformaciones o aplicar alguna técnica para eliminarlas o modificarlas. Llevando a cabo este proceso, se logra obtener series de tiempo estacionarias. Ciertamente, esto permite realizar un mejor análisis además de predicciones en las series de tiempo. En adelante, sólo se tomarán en cuenta las series de tiempo discretas.

Una manera de representar una serie de tiempo es mediante la ecuación

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

donde X_t es un proceso, m_t es la *componente de tendencia*, s_t es una función que representa la *componente estacional* y ε_t es la *componente de ruido aleatorio*¹. A continuación se describen brevemente estos tipos de análisis.

¹ $E(\varepsilon_t) = 0$ y $\text{Var}(\varepsilon_t) = \sigma^2$.

3.1.1. Transformaciones

Es posible realizar transformaciones en los datos graficados con el fin de facilitar el análisis de la serie de tiempo. Existen transformaciones logarítmicas y transformaciones aplicando raíz cuadrada. Algunas razones para realizar una transformación son las siguientes:

- a) Para estabilizar la varianza. Si en la serie existe tendencia y la varianza incrementa con la media, puede hacerse una transformación. Si la desviación estándar es directamente proporcional a la media, se debe hacer una transformación logarítmica.
- b) Para hacer aditivo el efecto estacional. Si existe tendencia y el tamaño del efecto estacional incrementa con la media, se puede hacer una transformación para hacer constante el efecto estacional (cambios periódicos anuales), a esto se le llama *efecto estacional aditivo*. Si el tamaño del efecto estacional es directamente proporcional a la media, entonces se dice que es un efecto estacional *multiplicativo*. En este caso se debe hacer una transformación logarítmica.
- c) Para obtener una distribución normal en los datos. En muchas ocasiones es importante normalizar los datos para facilitar el análisis.

3.1.2. Análisis de series que contienen tendencia

El tipo más simple de tendencia lineal es de la forma *tendencia lineal + ruido*, aunque puede definirse de diferentes maneras. La observación en el tiempo t es una variable aleatoria X_t dada por

$$X_t = \alpha + \beta t + \varepsilon_t, \quad (3.1)$$

donde α y β son constantes y ε_t es un término de error aleatorio con media cero.

El nivel medio en el tiempo t está dado por

$$m_t = \alpha + \beta t, \quad (3.2)$$

el cual también se llama *componente de tendencia*. La tendencia en (3.2) es una función que depende del tiempo y en algunas ocasiones la llaman *tendencia lineal global*. Otros autores hacen referencia a la tendencia como el cambio en el nivel medio por unidad de tiempo, es decir toman a β como la tendencia.

Es posible ajustar un modelo para la tendencia mediante una función lineal (3.2); este tipo de ajuste tienen la ventaja de ser muy sencillo de analizar, sin embargo, en ocasiones existen cambios bruscos en la tendencia, es por esta razón que es más adecuado realizar el análisis por intervalos. Es decir, ajustar un modelo lineal a trozos. Si se ajusta un modelo lineal a trozos es más fácil visualizar los cambios de la tendencia; además de que se puede observar como α y β evolucionan con el tiempo. Esta evolución se puede dar de forma

determinista, aunque también se puede dar de forma estocástica, con lo cual se obtiene lo que se conoce como tendencia estocástica.

Formalmente definimos el modelo con componente estacional de la siguiente manera.

Definición 3.1.1 *El modelo cuya única componente es la tendencia está dado por*

$$X_t = m_t + \varepsilon_t, \quad \text{para } t = 1, 2, \dots, n, \quad (3.3)$$

donde $E(\varepsilon_t) = 0$.

Observación 3.1.2 *Si $E(\varepsilon_t) \neq 0$, entonces se reemplazan m_t y ε_t en (3.3) con $m_t + E(\varepsilon_t)$ y $\varepsilon_t - E(\varepsilon_t)$, respectivamente.*

A continuación se analizan algunas técnicas que ayudan a describir la tendencia en una serie de tiempo.

► Estimación de la tendencia

Uno de los principales métodos no paramétricos para la estimación de tendencia y construcción de modelos es el suavizamiento.

Los siguientes métodos fueron pensados con el fin de estimar la tendencia ajustando un polinomio y luego eliminarla de los datos. De esta manera es posible encontrar un modelo adecuado de serie de tiempo estacionario para la componente residual; o bien, eliminar la tendencia mediante diferenciación y luego encontrar el modelo apropiado para la serie diferenciada (componente residual).

Suavizamiento.

A continuación se presentan dos tipos de suavizamiento que pueden ser usados para estimar la tendencia en una serie de tiempo.

I. Suavizamiento por filtros. Usar filtros lineales ayuda a convertir una serie de tiempo $\{x_t\}$ en otra $\{y_t\}$ mediante

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

donde $\{a_r\}$ es un conjunto de pesos tales que $\sum a_r = 1$. Esta técnica hace referencia a un proceso de promedio móvil.

Un procedimiento de suavizado en las series puede ser por ejemplo,

$$\xrightarrow{x_t} \boxed{\text{Filtro 1}} \xrightarrow{y_t} \boxed{\text{Filtro 2}} \xrightarrow{z_t}$$

El filtro 1 con pesos $\{a_r\}$ actúa sobre $\{x_t\}$ para producir $\{y_t\}$. El filtro 2 con pesos $\{b_j\}$ actúa sobre $\{y_t\}$ para producir $\{z_t\}$.

Consideremos el siguiente filtro de promedio móvil,

$$W_t = \frac{1}{2q+1} \sum_{r=-q}^q X_{t-r}, \quad \text{donde } q \in \mathbf{Z}^+,$$

del proceso $\{X_t\}$ definido por la ecuación (3.3). Para $q+1 \leq t \leq n-q$ se tiene,

$$W_t = \frac{1}{2q+1} \sum_{r=-q}^q m_{t-r} + \frac{1}{2q+1} \sum_{r=-q}^q Y_{t-r} \simeq m_t,$$

suponiendo que m_t es aproximadamente lineal en $[t-q, t+q]$ y que la media de los términos de error es muy cercana a cero, entonces

$$\hat{m}_t = \frac{1}{2q+1} \sum_{r=-q}^q X_{t-r}, \quad q+1 \leq t \leq n-q, \quad (3.4)$$

proporciona un estimador para la tendencia.

La elección de los filtros para suavizar es de carácter subjetivo, el investigador debe probar con diferentes filtros para obtener una mejor estimación de la tendencia.

II. Suavizamiento exponencial. Una técnica bastante conocida en la literatura es el *suavizamiento exponencial*, esta también es una técnica muy útil para estimar la tendencia.

Por ejemplo para cualquier número fijo $a \in [0, 1]$, el proceso \hat{m}_t , $t = 1, 2, \dots, n$ definido por

$$\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1}, \quad t = 2, \dots, n$$

y

$$\hat{m}_1 = X_1,$$

puede ser estimado mediante suavizamiento exponencial.

El método de suavizamiento exponencial asume que

$$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_t, \quad (3.5)$$

donde α es una constante y $0 < \alpha < 1$, es un estimador para la tendencia. Los pesos $a_j = \alpha(1 - \alpha)^j$ decrecen con respecto a j .

El suavizamiento exponencial utiliza valores presentes y pasados de la serie, este método es común cuando se realiza predicción.

Una vez que se estima la tendencia se pueden observar las fluctuaciones locales,

$$\begin{aligned} Res(x_t) &= \text{residuales del valor suavizado} \\ &= x_t - s_m(x_t) \\ &= \sum_{r=-q}^s b_r x_{t+r}. \end{aligned}$$

El cual también puede ser considerado como un filtro lineal tomando $b_0 = 1 - a_0$ y $b_r = -a_r$ para $r \neq 0$. Si se tiene que $\sum a_r = 1$, entonces $\sum b_r = 0$ y el filtro elimina la tendencia.

Ajuste polinomial.

Para datos anuales es posible ajustar una función simple de tiempo como una curva polinomial (lineal, cuadrática, etc.), una curva *Gompertz*, una curva logística o por algún otro método como el conocido por mínimos cuadrados, véase [4].

Por ejemplo por mínimos cuadrados lo que se pretende, una vez que se tiene la tendencia de la forma $m_t = a_0 + a_1 t + a_2 t^2$, es ajustar a los datos x_1, x_2, \dots, x_n , eligiendo los parámetros a_0, a_1, a_2 , tales que minimicen $\sum_{t=1}^n (x_t - m_t)^2$.

La curva *Gompertz* está dada por

$$\log x_t = \alpha + \beta r^t,$$

donde α y β son parámetros, $0 < r < 1$, o bien por

$$x_t = \alpha \exp\{\beta \exp(-\gamma t)\}, \quad \gamma > 0.$$

La curva logística está dada por

$$x_t = \frac{\alpha'}{1 + \beta' \exp(-ct)}.$$

En las curvas antes mencionadas, $t \rightarrow \infty$, sin embargo, la *Gompertz* converge más lento que la logística. Además, en estas curvas, la función proporciona una medida de la tendencia y los residuales, una aproximación de fluctuaciones locales donde los residuales son las diferencias entre las observaciones y los correspondientes valores de la curva ajustada.

► **Eliminación de tendencia**

Lo que se intenta ahora en lugar de eliminar los términos de error, es eliminar la tendencia en una serie mediante diferenciación.

Eliminación por diferenciación. Diferenciar una serie de tiempo dada hasta lograr hacerla estacionaria, es un tipo de filtro particular el cual elimina la tendencia de la serie. Para poder aplicar esto se necesita que los datos sean no estacionales, es decir, que no presenten fluctuaciones periódicas muy marcadas. Muchas veces basta con obtener las diferencias de primer orden para lograr la estacionariedad de la serie.

A continuación se definen dos operadores que serán de gran utilidad tanto en esta sección como en las posteriores.

Definición 3.1.3 *El operador diferencial ∇ está dado por*

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

donde B es el operador de retraso dado por

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

Análogamente se definen

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

y

$$\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t)), \quad j \geq 1, \quad \text{con } \nabla^0(X_t) = X_t.$$

El método de diferenciación consiste en formar una nueva serie y_2, \dots, y_N a partir de la original x_1, \dots, x_N mediante

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

En algunos casos es necesario obtener la segunda diferencia usando el operador ∇^2 ,

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

3.1.3. Análisis de series que contienen variación estacional

Existen básicamente 3 modelos estacionales:

a) Aditivo,

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

b) Multiplicativo,

$$X_t = m_t s_t \varepsilon_t.$$

c) Mixto,

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

donde, en cada una de las ecuaciones, m_t es el nivel medio desestacionalizado en el tiempo t , s_t es el efecto estacional en el tiempo t y ε_t es el error aleatorio.

En el caso multiplicativo, una transformación logarítmica es adecuada para obtener un modelo aditivo lineal.

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 restándole el promedio anual para el caso aditivo. En el caso multiplicativo se debe dividir cada observación de periodo por el promedio anual.

► Estimación de tendencia y componentes estacionales.

En el caso en que la serie dada presenta tendencia y variación estacional se debe trabajar con ambas componentes a la vez. Para llevar a cabo este procedimiento se debe seguir una secuencia de pasos simples.

En primer lugar se estima la tendencia, luego las componentes estacionales y por último se vuelve a estimar la tendencia, ahora sin incluir los datos estacionales².

A continuación se ilustra este procedimiento.

²Método descrito en el libro de Brockwell & Davis [2].

- a) Se estima la tendencia de la serie con alguno de los métodos descritos en la sección 3.1.2, en la parte correspondiente a estimación de tendencia, suponiendo que se tienen n observaciones. Si el periodo es par ($d = 2q$), entonces el estimador es

$$\hat{m}_t = \frac{\frac{1}{2}X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + \frac{1}{2}X_{t+q}}{d}, \quad q < t \leq n - q, \quad (3.6)$$

en caso de que el periodo sea impar entonces se toma $d = 2q + 1$ y se usa el estimador (3.4).

- b) Se debe estimar la componente estacional. Se calcula el promedio w_k de las desviaciones $\{(x_{k+jd} - \hat{m}_{k+jd}), q < k + jd \leq n - q\}$, para $k = 1, \dots, d$. Aunque esto resulta ser inapropiado pues el promedio de las desviaciones no necesariamente suma cero, entonces se estiman las componentes estacionales mediante

$$s_k = w_k - \frac{\sum_{i=1}^d w_i}{d}, \quad k = 1, \dots, d,$$

y $\hat{s}_k = \hat{s}_{k-d}$, $k > d$.

Ahora los datos desestacionalizados son estimados mediante

$$d_t = x_t - \hat{s}_t, \quad t = 1, \dots, n,$$

es decir las componentes estacionales estimadas son eliminadas del modelo para construir otro.

- c) Por último se debe estimar nuevamente la tendencia pero ahora sin las componentes estacionales (d_t) usando alguno de los métodos descritos en la sección 3.1.2.

Una vez realizado todo lo anterior, la componente del ruido se podrá estimar por

$$\hat{\varepsilon}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, n.$$

El siguiente es un ejemplo muy sencillo de como estimar los datos estacionales y tendencia.

Ejemplo 3.1.4 *Para cuando se tienen datos mensuales, una forma de eliminar el efecto estacional se obtiene al calcular*

$$s_m(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}. \quad (3.7)$$

Para datos trimestrales la ecuación a calcular es:

$$s_m(x_t) = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_{t+1} + \frac{1}{2}x_{t+2}}{4}. \quad (3.8)$$

De esta manera el efecto estacional para una serie de tiempo se puede estimar haciendo $x_t - s_m(x_t)$ en el caso aditivo ó $\frac{x_t}{s_m(x_t)}$ para el caso multiplicativo.

► **Eliminación de tendencia y componentes estacionales.**

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

Por ejemplo, para datos mensuales se debe calcular

$$\nabla_{12}X_t = X_t - X_{t-12}.$$

Es decir, la técnica de diferenciación descrita anteriormente para datos no estacionales, también se puede aplicar a datos estacionales de periodo d . Ahora se define el operador ∇_d de la siguiente manera,

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t. \quad (3.9)$$

Aplicando este operador 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 la descomposición de la diferencia $\nabla_d X_t$ entre una componente de tendencia ($m_t - m_{t-d}$) y el término de error ($\varepsilon_t - \varepsilon_{t-d}$). De esta forma, la tendencia ($m_t - m_{t-d}$) puede ser eliminada usando el método descrito en la sección 3.1.2, en particular aplicando el operador dado en (3.9).

Estos procedimientos descritos son sólo algunos de los existentes para estimar y/o eliminar tanto la tendencia como el efecto estacional en una serie de tiempo. Dichos procedimientos serán de utilidad más adelante con fines prácticos.

3.2. Modelos que describen series de tiempo

Como se mencionó en el Capítulo 2, un proceso estocástico es una familia de v.a's indexadas por un conjunto que generalmente se asocia a valores de tiempo. Los procesos estocásticos miden la evolución de alguna o algunas variables a lo largo del tiempo. En esta tesis, se hará énfasis en los procesos para los cuales el conjunto de índices es discreto, en este caso, se puede ver al proceso $\{X_t : t \in T\}$ como una sucesión de v.a's.

Una *realización* de un proceso estocástico $\{X_t : t \in T\}$, es una asignación de valores para cada una de las v.a's de este proceso.

Una serie de tiempo se puede ver como una realización de algún proceso estocástico discreto. Por motivos prácticos se supone que cada observación x_t es un valor realizado de cierta v.a. X_t .

Formalmente se tiene la siguiente definición de serie de tiempo.

Definición 3.2.1 Una serie de tiempo para los datos observados x_t es una especificación de las distribuciones conjuntas de una sucesión de v.a.'s $\{X_t\}$, de las cuales $\{x_t\}$ son tomadas como realizaciones.

3.2.1. Función autocorrelación parcial

Comúnmente, además de la autocorrelación entre las variables X_t y X_{t+k} , se desea conocer la correlación entre estas variables, excluyendo la dependencia lineal entre X_t y X_{t+k} y las variables $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Esto se denota como la correlación condicional

$$\text{Corr}(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1}),$$

la cual es conocida como *función autocorrelación parcial* (FACP). El proceso para encontrar dicha autocorrelación es el siguiente.

Considérese un proceso estacionario $\{X_t\}$, del cual se asume que $E(X_t) = 0^3$. La dependencia lineal de X_{t+k} con $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ se define como el mejor estimador lineal de X_{t+k} , en el sentido de mínimos cuadrados como una combinación lineal de $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Es decir,

$$\hat{X}_{t+k} = \alpha_1 X_{t+k-1} + \alpha_2 X_{t+k-2} + \dots + \alpha_{k-1} X_{t+1},$$

donde $\alpha_i (1 \leq i \leq k-1)$ es el coeficiente de regresión lineal de mínimos cuadrados que minimiza

$$E(X_{t+k} - \hat{X}_{t+k})^2 = E(X_{t+k} - \alpha_1 X_{t+k-1} - \dots - \alpha_{k-1} X_{t+1})^2.$$

Aplicando las condiciones necesarias de minimización, se obtiene el siguiente sistema de ecuaciones lineales para las variables $\alpha_1, \alpha_2, \dots, \alpha_{k-1}$,

$$\begin{aligned} \gamma_1 &= \alpha_1 \gamma_0 + \alpha_2 \gamma_{-1} + \dots + \alpha_{k-1} \gamma_{-k} \\ \gamma_2 &= \alpha_1 \gamma_1 + \alpha_2 \gamma_0 + \dots + \alpha_{k-1} \gamma_{1-k} \\ &\vdots \\ \gamma_{k-1} &= \alpha_1 \gamma_{k-2} + \alpha_2 \gamma_{k-3} + \dots + \alpha_{k-1} \gamma_0. \end{aligned}$$

Como ρ cumple (2.4), entonces el sistema de ecuaciones anterior se transforma en:

³Si $E(X_t) \neq 0$, entonces se puede hacer una transformación de X_t a $X_t - E(X_t)$ y aplicar el procedimiento

$$\begin{aligned}
\rho_1 &= \alpha_1\rho_0 + \alpha_2\rho_{-1} + \cdots + \alpha_{k-1}\rho_{-k} \\
\rho_2 &= \alpha_1\rho_1 + \alpha_2\rho_0 + \cdots + \alpha_{k-1}\rho_{1-k} \\
&\vdots \\
\rho_{k-1} &= \alpha_1\rho_{k-2} + \alpha_2\rho_{k-3} + \cdots + \alpha_{k-1}\rho_0,
\end{aligned}$$

cuya forma matricial es la siguiente,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}. \quad (3.10)$$

De forma similar se define la dependencia lineal de X_t con las variables $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$, como el mejor estimador de mínimos cuadrados de X_t como combinación lineal de $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$, es decir,

$$\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \cdots + \beta_{k-1} X_{t+k-1},$$

donde los escalares $\beta_i (1 \leq i \leq k-1)$ son los coeficientes de regresión lineal por mínimos cuadrados que minimizan

$$E(X_t - \hat{X}_t)^2 = E(X_t - \beta_1 X_{t+1} - \cdots - \beta_{k-1} X_{t+k-1})^2.$$

De aquí se obtiene el sistema de ecuaciones siguiente

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}. \quad (3.11)$$

Nótese que (3.10) y (3.11) son equivalentes, es decir tienen la misma solución. Esto implica que $\alpha_i = \beta_i$ para cada $i = 1, 2, \dots, k-1$.

Ahora se tiene que $(X_t - \hat{X}_t)$ es la variable que resulta al eliminar la dependencia lineal de X_t respecto de $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Análogamente $(X_{t+k} - \hat{X}_{t+k})$ es la variable que resulta de eliminar la dependencia lineal de X_{t+k} respecto de $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Así, la autocorrelación parcial entre X_t y X_{t+k} , que por definición es la correlación después de eliminar la independencia lineal antes mencionada, es igual a la correlación entre las variables $(X_t - \hat{X}_t)$ y $(X_{t+k} - \hat{X}_{t+k})$. Es decir

$$P_k = \frac{\text{Cov}[(X_t - \hat{X}_t)(X_{t+k} - \hat{X}_{t+k})]}{\sqrt{\text{Var}(X_t - \hat{X}_t)\text{Var}(X_{t+k} - \hat{X}_{t+k})}},$$

a partir de esta ecuación se obtiene que

$$\begin{aligned}
 P_1 &= \rho_1, \\
 P_2 &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_1 \end{vmatrix}}, \\
 P_3 &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \\
 &\vdots \\
 P_k &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_1 \end{vmatrix}}.
 \end{aligned}$$

Los cuales son valores de la función autocorrelación parcial entre X_t y X_{t+l} , para $l = 1, \dots, k$.

Existe otra forma de obtener la función de autocorrelación parcial entre dos variables. Considérese una variable X_{t+k} de un proceso estacionario con media cero que depende de las variables de regresión $X_{t+k-1}, X_{t+k-2}, \dots, X_t$, es decir,

$$X_{t+k} = \phi_{k1}X_{t+k-1} + \phi_{k2}X_{t+k-2} + \cdots + \phi_{kk}X_t + e_{t+k} \quad (3.12)$$

donde ϕ_{ki} denota el i -ésimo coeficiente de regresión y e_{t+k} es un término de error normal no correlacionado con X_{t+k-j} para $j \geq 1$. Multiplicando por X_{t+k-j} en ambos lados de (3.12) y tomando el valor esperado de la expresión resultante se obtiene:

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k},$$

y dividiendo por γ_0 se obtiene,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k},$$

para $j = 1, 2, \dots, k$ se obtienen las siguientes ecuaciones,

$$\begin{aligned}\rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \dots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \dots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \dots + \phi_{kk}\rho_0.\end{aligned}$$

Con la regla de Cramer se obtiene

$$\begin{aligned}\phi_{11} &= \rho_1, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_1 \end{vmatrix}}, \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \\ &\vdots \\ \phi_{kk} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_1 \end{vmatrix}}.\end{aligned}\tag{3.13}$$

Nótese que los coeficientes ϕ_{kk} de regresión lineal cumplen $\phi_{kk} = P_k$. Usualmente en la literatura se denota la función autocorrelación parcial como ϕ_{kk} .

3.2.2. Procesos puramente aleatorios

Definición 3.2.2 *Se dice que un proceso $\{Z_t : t \in T\}$ es un proceso puramente aleatorio si es una sucesión de v.a's iid no correlacionadas, donde además $E(Z_t) = \mu_Z$, la varianza es constante, $\text{Var}(Z_t) = \sigma_Z^2$ y $\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = 0$, para todo $k \neq 0$.*

Por otro lado, se dice que el proceso puramente aleatorio Z_t es estacionario y se tiene lo siguiente.

- La función autocovarianza ($\gamma_k = E(Z_t Z_{t+k})$) está dada por

$$\gamma_k = \begin{cases} \sigma_Z^2, & k = 0; \\ 0, & \text{otro caso.} \end{cases}$$

- La función autocorrelación está dada por

$$\rho_k = \begin{cases} 1, & k = 0; \\ 0, & \text{otro caso.} \end{cases}$$

- Y la función autocorrelación parcial está dada por

$$\phi_{kk} = \begin{cases} 1, & k = 0; \\ 0, & \text{otro caso.} \end{cases}$$

En particular si se toma $\mu_Z = 0$, entonces se hace referencia a un *proceso ruido blanco*.

3.2.3. Funciones autocorrelación y autocovarianza muestral

Media muestral

Del ejemplo 2.2.14, se observa que la media muestral resulta ser un buen estimador para la media poblacional. Para el caso cuando se tiene solo una realización (una asignación de valores para las v.a's.), también la media muestral proporciona un estimador para la media $\mu = E(X_t)$ de un proceso estacionario, esto es

$$\hat{X} = \frac{1}{n} \sum_{t=1}^n X_t,$$

claramente, \hat{X} es un estimador insesgado.

Función autocovarianza muestral

Se puede proponer un estimador para la función autocovarianza γ_k . Para fines prácticos algunas ocasiones resulta más útil calcular el estimador para γ_k y después usar el hecho que $\rho_k = \frac{\gamma_k}{\gamma_0}$ para obtener el estimador de ρ_k . Para una realización, el estimador $\hat{\gamma}_k$ de γ_k , se define análogamente al caso ordinario como

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad (3.14)$$

ó

$$\hat{\gamma}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad (3.15)$$

aunque ambos estimadores resultan ser sesgados, muchos autores prefieren usar (3.14) puesto que resulta mejor estimador que (3.15). En efecto,

$$\begin{aligned} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) &= \sum_{t=1}^{n-k} [(X_t - \mu) - (\bar{X} - \mu)][(X_{t+k} - \mu) - (\bar{X} - \mu)] \\ &= \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (\bar{X} - \mu) \sum_{t=1}^{n-k} (X_t - \mu) \\ &\quad - (\bar{X} - \mu) \sum_{t=1}^{n-k} (X_{t+k} - \mu) + (n-k)(\bar{X} - \mu)^2 \\ &\approx \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (n-k)(\bar{X} - \mu)^2. \end{aligned}$$

Por otro lado

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_k) &= \mathbb{E} \left[\frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \right] \\ &\simeq \frac{1}{n} \mathbb{E} \left[\sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (n-k)(\bar{X} - \mu)^2 \right] \\ &= \frac{1}{n} \left[\sum_{t=1}^{n-k} \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)] - (n-k)\text{Var}(\bar{X}) \right] \\ &= \frac{1}{n} \left[\sum_{t=1}^{n-k} \text{Cov}(X_t, X_{t+k}) - (n-k)\text{Var}(\bar{X}) \right] \\ &= \gamma_k - \frac{k}{n} \gamma_k - \left(\frac{n-k}{n} \right) \text{Var}(\bar{X}). \end{aligned}$$

Análogamente

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_k) &= \mathbb{E} \left[\frac{1}{(n-k)} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \right] \\ &\simeq \frac{1}{(n-k)} \mathbb{E} \left[\sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (n-k)(\bar{X} - \mu)^2 \right] \\ &= \gamma_k - \text{Var}(\bar{X}). \end{aligned}$$

Al ignorar el término de $\text{Var}(\bar{X})$ en las últimas dos ecuaciones, es fácil ver que $\hat{\gamma}_k$ se hace insesgado, mientras $\hat{\gamma}_k$ sigue siendo sesgado. El sesgo de $\hat{\gamma}_k$ es más grande que el sesgo de $\hat{\gamma}_k$, sobre todo cuando k es grande con respecto a n . Usando el criterio de comparar el ECM de ambos estimadores se encuentra que $\hat{\gamma}_k$ tiene menor ECM que $\hat{\gamma}_k$. Además, $\hat{\gamma}_k$ es positiva semidefinida como lo es γ_k , mientras que $\hat{\gamma}_k$ no lo es. Por estas razones, es preferible usar (3.14) como *función autocovarianza muestral* para estimar la función de autocovarianza γ_k .

Función autocorrelación muestral

Asociada a la función de autocovarianza muestral, también es posible definir la *función de autocorrelación muestral* (FAC muestral) para una serie de tiempo dada X_1, X_2, \dots, X_n , mediante,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad k = 0, 1, 2, \dots$$

Cuando se grafica $\hat{\rho}_k$ contra k , se obtiene a lo que comúnmente se le conoce como *correlograma muestral*.

Función autocorrelación parcial muestral

Además de las funciones antes mencionadas, es posible definir la *función autocorrelación parcial muestral* (FACP muestral); esto se logra, sustituyendo en (3.13), ρ_i por su respectivo estimador $\hat{\rho}_i$. Ahora se tiene un método recursivo iniciando con $\hat{\phi}_{11} = \hat{\rho}_1$ y

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j}$$

y

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

3.2.4. Procesos lineales estacionarios

Es posible representar una serie de tiempo de dos maneras. La primera consiste en escribir un proceso X_t como una combinación lineal de una sucesión de v.a.'s no correlacionadas y la segunda forma utiliza el operador de retraso B .

Definición 3.2.3 Una serie de tiempo $\{X_t\}$ es un proceso lineal si se puede representar como

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad \forall t, \quad (3.16)$$

donde $\{Z_t\}$ es un proceso ruido blanco, y $\{\psi_j\}$ es una sucesión de constantes tales que $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$.

Utilizando el operador de retraso B , se puede escribir (3.16) de la siguiente forma:

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

donde $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$.

Si $\psi_j = 0$ para todo $j < 0$, entonces se tiene

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

Esta forma de escribir un proceso se le denomina representación de *medias móviles*.

Nótese que X_t está bien definida, pues la condición $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ asegura que la suma en (3.16) converge (con probabilidad 1) puesto que $E(|Z_t|^2) = E(Z_t^2) = \text{Var}(Z_t) + (E(Z_t))^2 = \sigma^2$, de aquí que $E(|Z_t|) \leq \sigma$ y

$$E|X_t| \leq \sum_{j=-\infty}^{\infty} (|\psi_{t-j}|E|Z_{t-j}|) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_{t-j}| \right) \sigma < \infty.$$

De aquí en adelante, una suma infinita de variables aleatorias es definida como el límite en la media cuadrática de la suma finita de sumas parciales. Esto es, (3.16) converge en la media cuadrada, es decir, X_t en (3.16) es definida tal que

$$E \left[\left(X_t - \sum_{j=-n}^n \psi_j Z_{t-j} \right)^2 \right] \rightarrow 0 \text{ cuando } n \rightarrow \infty.$$

Por otro lado, para la representación de medias móviles se cumple,

$$E(X_t) = E \left(\sum_{j=0}^{\infty} \psi_j Z_{t-j} \right) = \sum_{j=0}^{\infty} \psi_j E(Z_{t-j}) = 0,$$

$$\text{Var}(X_t) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2,$$

y

$$\text{E}(Z_t X_{t-j}) = \begin{cases} \sigma_Z^2, & j = 0; \\ 0, & \text{otro caso.} \end{cases}$$

De aquí que,

$$\gamma_k = \text{E}(X_t X_{t+k}) \quad (3.17)$$

$$= \text{E} \left(\sum_{i=0}^{\infty} \psi_i Z_{t-i} \sum_{j=0}^{\infty} \psi_j Z_{t+k-j} \right) \quad (3.18)$$

$$= \text{E} \left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j Z_{t-i} Z_{t+k-j} \right) \quad (3.19)$$

$$= \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \quad (3.20)$$

es decir,

$$\gamma_k = \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \quad (3.21)$$

y

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

Es fácil ver que las funciones autocovarianza y autocorrelación dependen solo de la diferencia del tiempo k . Más aún, puesto que involucran sumas infinitas, para ser estacionarias, se debe probar que γ_k es finita para cada k . Esto es,

$$|\gamma_k| = |\text{E}(X_t X_{t+k})| \leq [\text{Var}(X_t) \text{Var}(X_{t+k})]^{1/2} = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2.$$

De aquí que, también $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ es una condición requerida para que el proceso en (3.16) sea estacionario.

Para una sucesión de autocovarianzas γ_k , $k = 0, \pm 1, \pm 2, \dots$, la *función generadora de covarianzas* se define como

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k,$$

donde la varianza del proceso γ_0 , es el coeficiente de B^0 , y la autocovarianza del rezago k , γ_k , es el coeficiente tanto de B^k como de B^{-k} . Por el hecho de que el proceso es estacionario y por (3.21) se sigue que

$$\begin{aligned}\gamma_k &= \sigma_Z^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \\ &= \sigma_Z^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} \\ &= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} \\ &= \sigma_Z^2 \psi(B) \psi(B^{-1}),\end{aligned}$$

donde $\psi_j = 0$ para $j < 0$. Análogamente, se define la *función generadora de autocorrelación*, la cual está dada por

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}.$$

Otra forma usual de escribir un proceso $\{X_t\}$ es mediante una representación autorregresiva, en la cual el valor X_t es la suma de sus propios valores regresados en el tiempo t más un proceso aleatorio, es decir,

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + Z_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + Z_t,$$

ó haciendo uso del operador B ,

$$\pi(B)X_t = Z_t, \tag{3.22}$$

donde $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, y $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$. Cuando un proceso puede ser escrito en esta forma, se dice que es invertible⁴.

Para que un proceso lineal $X_t = \psi(B)Z_t$ sea invertible, debe poder escribirse en términos de un proceso autorregresivo, las raíces de $\psi(B) = 0$ como función de B deben estar fuera del círculo unitario. Esto es, si β es una raíz de $\psi(B)$, entonces $|\beta| > 1$, donde $|\cdot|$ es la métrica Euclídeana⁵.

⁴Box-Jenkins (1976) determinan esta propiedad en los procesos lineales. Además argumentan que en predicción, un proceso que no es invertible no tiene sentido.

⁵Cuando β es un número real, $|\beta|$ es igual al valor absoluto de β , y cuando es un número complejo, $\beta = c + id$, entonces $|\beta| = \sqrt{c^2 + d^2}$.

Para que el proceso (3.22) sea estacionario, se debe poder reescribir en una representación de medias móviles, esto es,

$$X_t = \frac{1}{\pi(B)} Z_t = \psi(B) Z_t,$$

tal que $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ se satisface. Para que esto sea posible, se requiere que las raíces de $\pi(B) = 0$ se encuentren fuera del círculo unitario.

En la práctica, la construcción de estos modelos no es muy útil, puesto que contienen un número infinito de parámetros que son imposibles de estimar a partir de un número finito de observaciones disponibles. Una forma alternativa, es construir modelos con solo un número finito de parámetros.

3.2.5. Modelos de series de tiempo estacionarios

Anteriormente se mencionó que un proceso estacionario puede escribirse como una combinación lineal de v.a's no correlacionadas. Aquí se presenta la misma idea solo que para un número finito de observaciones. En este apartado se da a conocer el modelo autorregresivo de medias móviles, el cual es una combinación del modelo autorregresivo y el de medias móviles.

► El proceso autorregresivo

Definición 3.2.4 Sea $\{Z_t\}$ un proceso puramente aleatorio con media cero y varianza σ_Z^2 . Se dice que un proceso es autorregresivo de orden p , y se denota por $AR(p)$ si puede ser escrito como

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \quad (3.23)$$

o

$$\alpha_p(B) X_t = Z_t, \quad (3.24)$$

donde $\alpha_p(B) = (1 - \alpha_1 B - \cdots - \alpha_p B^p)$.

El proceso X_t es regresado en los mismos valores pasados de X_t en ambos casos, de aquí que al proceso se le denomine autorregresivo.

Puesto que $\sum_{j=1}^p |\pi_j| = \sum_{j=1}^p |\alpha_j| < \infty$, el proceso siempre es invertible. Para que sea estacionario, las raíces de $\alpha_p(B) = 0$, deben estar fuera del círculo unitario.

Los procesos AR son útiles para describir situaciones en las cuales la serie de tiempo presentada depende de sus valores pasados más un proceso aleatorio.

Para una mejor comprensión de este tipo de modelos, se consideran los siguientes casos.

■ **El proceso autorregresivo de primer orden $AR(1)$**

El proceso autorregresivo de primer orden puede escribirse como

$$X_t = \alpha_1 X_{t-1} + Z_t \quad (3.25)$$

ó usando el operador de retroceso B como,

$$(1 - \alpha_1 B)X_t = Z_t. \quad (3.26)$$

Este proceso siempre es invertible, para que sea estacionario, se debe cumplir que la raíz de $(1 - \alpha_1 B) = 0$, debe estar fuera del círculo unitario. Es decir, para un proceso estacionario se tiene que $|\alpha_1| < 1$. El proceso $AR(1)$ algunas veces es llamado proceso de Markov, pues el valor de X_t es completamente determinado a partir de X_{t-1} .

El proceso (3.25) puede escribirse como un proceso de medias móviles (3.16) de orden infinito, en efecto, si se hacen sustituciones se obtiene

$$X_t = \alpha X_{t-1} + Z_t = \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t,$$

si se siguen haciendo sustituciones se llega a que (3.25) se puede reescribir como

$$\begin{aligned} X_t &= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \alpha^3 Z_{t-3} + \dots \\ &= \sum_{i=0}^{\infty} \alpha^i Z_{t-i}. \end{aligned}$$

Análogamente, si se usa el operador B , entonces (3.26) puede ser reescrito como,

$$\begin{aligned} X_t &= \frac{Z_t}{1 - \alpha B} \\ &= (1 + \alpha B + \alpha^2 B^2 + \dots) Z_t \\ &= \left(\sum_{i=0}^{\infty} \alpha^i B^i \right) Z_t \\ &= \sum_{i=0}^{\infty} \alpha^i Z_{t-i}. \end{aligned}$$

A partir de la última ecuación se obtiene que

$$E(X_t) = 0$$

y

$$\text{Var}(X_t) = \sigma_Z^2(1 + \alpha^2 + \alpha^4 + \dots) = \frac{\sigma_Z^2}{(1 - \alpha^2)},$$

esto, pues las Z 's son v.a's iid.

La FAC del proceso AR(1)

La función autocovarianza se puede obtener como:

$$\begin{aligned} \gamma_k &= \text{Cov}(X_t, X_{t+k}) \\ &= \text{Cov}(X_t, \alpha_1 X_{t+k-1} + Z_{t+k}) \\ &= \alpha_1 \text{Cov}(X_t, X_{t+k-1}) + \text{Cov}(X_t, Z_{t+k}) \\ &= \alpha_1 \gamma_{k-1} + 0 \\ &= \alpha_1(\alpha_1 \gamma_{k-2}) \\ &\vdots \\ &= \alpha_1^k \gamma_0 \\ &= \frac{\alpha_1^k \sigma_X^2}{(1 - \alpha^2)}. \end{aligned}$$

De donde se puede inferir que,

$$\rho_k = \alpha_1^k, \quad k \geq 1,$$

se asume que $\rho_0 = 1$. De aquí que, cuando el proceso es estacionario y $|\alpha_1| < 1$, la FAC decae exponencialmente, dependiendo del signo de α_1 . Si $0 < \alpha_1 < 1$, entonces las autocorrelaciones son positivas y si $-1 < \alpha_1 < 0$, entonces las autocorrelaciones se muestran de forma alternada, es decir, una positiva y otra negativa. En ambos casos, las magnitudes de las correlaciones decaen exponencialmente como se muestra en el apéndice B. El caso a) ilustra la FAC cuando $\alpha_1 > 0$ mientras que el caso c) hace referencia a la FAC cuando $\alpha_1 < 0$.

La FACP del proceso AR(1)

A partir de (3.13), se obtiene

$$\phi_{kk} = \begin{cases} \rho_1 = \alpha_1, & k = 1; \\ 0, & k > 1. \end{cases}$$

Análogamente, en el apéndice B, en la Figura b) se puede observar la FACP para cuando $\alpha_1 > 0$ y en d) el caso cuando $\alpha_1 < 0$ además se ve, como se corta después

del primer rezago en ambos casos.

■ *El proceso autorregresivo de segundo orden AR(2)*

En el caso en que $p = 2$ se tiene el proceso autorregresivo de segundo orden dado por

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

ó

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

Este proceso siempre es invertible; para que sea estacionario, las raíces de $\alpha(B) = (1 - \alpha_1 B - \alpha_2 B^2)$ deben estar fuera del círculo unitario.

Las siguientes condiciones son necesarias para asegurar la estacionariedad del proceso, sin tomar en cuenta si las raíces son reales o complejas:

- $-2 < \alpha_1 < 2$;
- $-1 < \alpha_1 < 1$.

En efecto, supóngase que B_1 y B_2 son las raíces de $(1 - \alpha_1 B - \alpha_2 B^2) = 0$, ó equivalentemente de $\alpha_2 B^2 + \alpha_1 B - 1 = 0$, donde B_1 y B_2 están dadas por,

$$B_1 = \frac{-\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2\alpha_2},$$

y

$$B_2 = \frac{-\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2\alpha_2}.$$

De aquí,

$$\frac{1}{B_1} = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2},$$

y

$$\frac{1}{B_2} = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}.$$

La condición requerida $|B_i| > 1$ implica que $|\frac{1}{B_i}| < 1$ para $i = 1, 2$. De aquí se tiene,

$$\left| \frac{1}{B_1} \cdot \frac{1}{B_2} \right| = |\alpha_2| < 1$$

y

$$|\alpha_1| = \left| \frac{1}{B_1} + \frac{1}{B_2} \right| < 2$$

Entonces, para un modelo $AR(2)$ las condiciones de estacionariedad están dadas por:

- $\alpha_1 + \alpha_2 < 1$;
- $\alpha_2 - \alpha_1 < 1$;
- $-1 < \alpha_2 < 1$.

Con lo cual se asegura que las raíces serán mayores que 1, y con esto la estacionariedad del proceso.

Ejemplo 3.2.5 *Supóngase que se tiene el modelo $AR(2)$ dado por,*

$$X_t = X_{t-1} - 0.6X_{t-2} + Z_t$$

equivalente a $(1 - B + 0.6B^2)X_t = Z_t$, de donde las raíces de $(1 - B + 0.6B^2) = 0$ son: $B = 0.5617978 \pm 0.8988764i$, pero con módulo 1.06, lo cual asegura que el proceso es estacionario.

La FAC del proceso $AR(2)$

La autocovarianza del proceso $AR(2)$ se obtiene de la siguiente manera:

$$\begin{aligned} \gamma_k &= \text{Cov}(X_t, X_{t+k}) = \text{Cov}(X_t, \alpha_1 X_{t-k-1} + \alpha_2 X_{t-k-2} + Z_{t-k}) \\ &= \alpha_1 \text{Cov}(X_t, X_{t-k-1}) + \alpha_2 \text{Cov}(X_t, X_{t-k-2}) + \text{Cov}(X_t, Z_{t-k}) \\ &= \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2}, \quad k \geq 1. \end{aligned}$$

De la ecuación anterior se tiene que

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2}, \quad k \geq 1.$$

En particular para $k = 1$ se tiene,

$$\rho_1 = \alpha_1 + \alpha_2 \rho_1,$$

de donde

$$\rho_1 = \frac{\alpha_1}{1 - \alpha_2}.$$

La FAC caerá exponencialmente si las raíces de $(1 - \alpha_1 B - \alpha_2 B^2) = 0$ son reales y presentará ondas senoidales en caso de ser complejas.

La FACP del proceso AR(2)

A partir de (3.13) se puede calcular la función de autocorrelación parcial para $k \geq 1$, de la siguiente manera:

$$\begin{aligned}\phi_{11} &= \rho_1 = \frac{\alpha_1}{1 - \rho_2}, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ &= \frac{\left(\frac{\alpha_1^2 + \alpha_2 - \alpha_2^2}{1 - \alpha_2}\right) - \left(\frac{\alpha_1}{1 - \alpha_2}\right)^2}{1 - \left(\frac{\alpha_1}{1 - \alpha_2}\right)^2} \\ &= \frac{\alpha_2[(1 - \alpha_2)^2 - \alpha_1^2]}{(1 - \alpha_2)^2 - \alpha_1^2} = \alpha_2. \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \\ &= \frac{\begin{vmatrix} 1 & \rho_1 & \alpha_1 + \alpha_2\rho_1 \\ \rho_1 & 1 & \alpha_1\rho_1 + \alpha_2 \\ \rho_2 & \rho_1 & \alpha_1\rho_2 + \alpha_2\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0.\end{aligned}$$

Análogamente, $\phi_{kk} = 0$ para $k \geq 3$. Es por esta razón que la FACP de un proceso AR(2) se corta después del segundo rezago.

- **El proceso general autorregresivo de orden p , AR(p)**

El proceso autorregresivo de orden p está dado por las ecuaciones (3.23) ó (3.24). Para este proceso se pueden generalizar los casos particulares anteriores, obteniendo

sus FAC y FACP respectivamente.

La FAC del proceso $AR(p)$

Se multiplica por X_{t-k} en ambos lados de (3.23), para hallar la función de autocovarianza, es decir,

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

ahora tomando el valor esperado se obtiene,

$$\begin{aligned} E[X_{t-k}X_t] &= E[\alpha_1 X_{t-k}X_{t-1} + \cdots + \alpha_p X_{t-k}X_{t-p} + X_{t-k}Z_t] \\ &= \alpha_1 E[X_{t-k}X_{t-1}] + \cdots + \alpha_p E[X_{t-k}X_{t-p}] + E[X_{t-k}Z_t] \\ &= \alpha_1 \gamma_{k-1} + \cdots + \alpha_p \gamma_{k-p}, \quad k > 0, \end{aligned}$$

de aquí

$$\gamma_k = \alpha_1 \gamma_{k-1} + \cdots + \alpha_p \gamma_{k-p}.$$

A partir de la última ecuación se puede obtener la función de autocorrelación de la siguiente manera,

$$\rho_k = \alpha_1 \rho_{k-1} + \cdots + \alpha_p \rho_{k-p}, \quad k > 0.$$

La FACP del proceso $AR(p)$

De la ecuación (3.13) y usando el hecho que $\rho_k = \alpha_1 \rho_{k-1} + \cdots + \alpha_p \rho_{k-p}$, $k > 0$, entonces para la última columna del numerador de (3.13) se puede ver fácilmente que ésta es la combinación lineal de las columnas anteriores de la misma matriz. De aquí que para $k \geq p$, ϕ_{kk} será igual a cero.

► **El proceso de promedio móvil**

En la representación media móvil de un proceso, si y sólo si un número finito de pesos $\psi_i = \beta_i$ son distintos de cero, entonces el proceso resultante se dice que es un proceso o modelo de medias (promedio) móviles de orden q y se denota por $MA(q)$.

Definición 3.2.6 Sea $\{Z_t\}$ un proceso puramente aleatorio con media 0 y varianza σ_Z^2 , entonces el proceso $\{X_t\}$ se dice que es un proceso $MA(q)$ si

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad (3.27)$$

donde $\{\beta_i\}$ son constantes. Los Z 's se toman de tal forma que $\beta_0 = 1$. O bien, usando el operador de retroceso B se obtiene

$$X_t = \beta(B)Z_t, \quad (3.28)$$

donde $\beta(B) = (1 + \beta_1 B + \dots + \beta_q B^q)$.

Anteriormente se introdujo la idea de que un proceso $MA(q)$ es estacionario. Para que sea invertible, se deben imponer ciertas condiciones. Para explicar estas condiciones de invertibilidad se usa el operador de retroceso B . Es decir, por definición (3.27) puede ser reescrito en la forma (3.28). Entonces, un proceso $MA(q)$ es invertible, si las raíces de la ecuación

$$\beta(B) = 1 + \beta_1 B + \dots + \beta_q B^q = 0,$$

se encuentran fuera del círculo unitario, B en esta ecuación es considerada como una variable compleja y no como un operador.

■ **El proceso de promedio móvil de primer orden $MA(1)$**

Cuando $Z_{t-q} = 0$ para $q > 1$, se tiene el proceso de promedio móvil de primer orden

$$\begin{aligned} X_t &= Z_t + \beta_1 Z_{t-1} \\ &= (1 + \beta_1 B)Z_t. \end{aligned}$$

La FAC del proceso $MA(1)$

Haciendo uso del operador de retroceso B , se tiene que las autocovarianzas del proceso son

$$\gamma_k = \begin{cases} (1 + \beta_1^2)\sigma_Z^2, & k = 0; \\ \beta_1\sigma_Z^2, & k = 1; \\ 0, & k > 1; \end{cases}$$

y la función de autocorrelación está dada por

$$\rho_k = \begin{cases} 1, & k = 0; \\ \frac{\beta_1}{1 + \beta_1^2}, & k = 1; \\ 0, & k \geq 2; \end{cases}$$

$1 + \beta_1^2$ siempre es acotado, entonces el proceso $MA(1)$ es estacionario. El proceso será invertible, si las raíces de $(1 - \beta_1 B) = 0$ se encuentran fuera del círculo unitario, porque $B = \frac{1}{\beta_1}$, se requiere que $|\beta_1| < 1$ para que el proceso $MA(1)$ sea invertible.

La FACP del proceso MA(1)

A partir de (3.13) se obtiene que

$$\phi_{11} = \rho_1 = \frac{\beta_1}{1 + \beta_1^2} = \frac{\beta_1(1 - \beta_1^2)}{1 - \beta_1^4},$$

en general

$$\phi_{kk} = \frac{\beta_1^k(1 - \beta_1^2)}{1 - \beta_1^{2(k+1)}}, \quad k \geq 1.$$

Contrario a su FAC, la cual se corta después del primer rezago, la FACP de un modelo MA(1) decae exponencialmente dependiendo del signo de β_1 (Véase apéndice B).

■ **El proceso de promedio móvil de segundo orden MA(2)**

El proceso de promedio móvil de orden 2 está dado por

$$\begin{aligned} X_t &= Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} \\ &= (1 + \beta_1 B + \beta_2 B^2) Z_t. \end{aligned}$$

Puesto que un proceso de medias móviles de segundo orden es siempre estacionario, las raíces de $(1 + \beta_1 B + \beta_2 B^2) = 0$ deben caer fuera del círculo unitario. De aquí se desprenden las siguientes condiciones que deben cumplir los parámetros β_1 y β_2 .

- $\beta_1 + \beta_2 < 1$;
- $\beta_2 - \beta_1 < 1$;
- $-1 < \beta_2 < 1$.

■ **El proceso general de promedio móvil de orden q**

El valor esperado y la varianza del proceso de medias móviles de orden q están dadas por,

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

Como los Z 's son independientes,

$$\begin{aligned}
\gamma_k &= \text{Cov}(X_t, X_{t+k}) \\
&= \text{Cov}(\beta_0 Z_t + \cdots + \beta_q Z_{t-q}, \beta_0 Z_{t+k} + \cdots + \beta_q Z_{t+k-q}) \\
&= \begin{cases} 0, & \text{si } k > q; \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k}, & \text{si } k = 0, 1, \dots, q; \\ \gamma_{-k}, & \text{si } k < 0. \end{cases}
\end{aligned}$$

Puesto que

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

La función autocorrelación de un proceso $MA(q)$ está dada por

$$\rho_k = \begin{cases} 1, & \text{si } k = 0; \\ \sum_{i=0}^{q-k} \beta_i \beta_{i+k} / \sum_{i=0}^q \beta_i^2, & \text{si } k = 1, \dots, q; \\ 0, & \text{si } k > q; \\ \rho_{-k}, & \text{si } k < 0. \end{cases}$$

► El proceso autorregresivo de promedio móvil

Aunque la representación autorregresiva y la representación de promedio móvil resultan ser muy útiles para representar procesos invertibles y estacionarios, pueden contener muchos parámetros, lo cual es un problema. En general, un número muy grande de parámetros reduce la eficiencia en la estimación. Es por esta razón, que al construir el modelo, es necesario incluir ambas representaciones. Es decir, se obtiene una mezcla de ambos, la cual es conocida como representación autorregresiva de promedios móviles.

En adelante se denotará por $ARMA(p, q)$ un proceso autorregresivo de medias móviles de orden p y q respectivamente.

Definición 3.2.7 *Se dice que $\{X_t\}$ es un proceso $ARMA(p, q)$ si es estacionario y si para cada t , se cumple*

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad (3.29)$$

donde $\{Z_t\}$ es un proceso ruido blanco.

Usando el operador de retroceso, se puede escribir (3.29) como,

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

donde

$$\alpha_p(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p,$$

y

$$\beta_q(B) = 1 - \beta_1 B - \dots - \beta_q B^q.$$

Donde p y q representan los grados de los polinomios asociados autorregresivo y de promedio móvil, respectivamente.

Para que el proceso sea invertible, se requiere que las raíces de $\beta_q(B) = 0$ se encuentren fuera del círculo unitario. Análogamente, para que el proceso sea estacionario, las raíces de $\alpha_p(B) = 0$ deben estar fuera del círculo unitario. Además, se asume que los polinomios $\beta_q(B)$ y $\alpha_p(B)$ no tienen raíces en común.

El proceso estacionario e invertible *ARMA* puede ser escrito como una representación pura autorregresiva, es decir,

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

donde

$$\pi(B) = \frac{\alpha_p(B)}{\beta_q(B)} = (1 - \pi_1 B - \pi_2 B^2 - \dots).$$

Análogamente, el proceso también puede ser reescrito en términos de un proceso de medias móviles,

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

donde

$$\psi(B) = \frac{\beta_q(B)}{\alpha_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \dots).$$

La función autocorrelación del proceso ARMA(p,q)

Para hallar la FAC del proceso *ARMA*(p, q), se multiplica (3.29) por X_{t-k} , así se obtiene,

$$X_{t-k}X_t = \alpha_1 X_{t-k}X_{t-1} + \dots + \alpha_p X_{t-k}X_{t-p} + X_{t-k}Z_t + \beta_1 X_{t-k}Z_{t-1} + \dots + \beta_q X_{t-k}Z_{t-q}.$$

y tomando el valor esperado,

$$\begin{aligned} E[X_{t-k}X_t] &= E[\alpha_1 X_{t-k}X_{t-1} + \dots + \alpha_p X_{t-k}X_{t-p} + X_{t-k}Z_t + \beta_1 X_{t-k}Z_{t-1} + \\ &\quad \dots + \beta_q X_{t-k}Z_{t-q}]. \end{aligned}$$

De donde se obtiene,

$$\begin{aligned} \gamma_k &= \alpha_1 \gamma_{k-1} + \dots + \alpha_p \gamma_{k-p} + E(X_{t-k}Z_t) + \beta_1 E(X_{t-k}Z_{t-1}) + \\ &\quad \dots + \beta_q E(X_{t-k}Z_{t-q}). \end{aligned}$$

Como $E(X_{t-k}Z_{t-i}) = 0$ para $k > i$, entonces la última igualdad se reduce a

$$\gamma_k = \alpha_1\gamma_{k-1} + \cdots + \alpha_p\gamma_{k-p}, \quad k \geq (q+1),$$

a apartir de esto último y por definición, se obtiene la función autocorrelación parcial dada por,

$$\rho_k = \alpha_1\rho_{k-1} + \cdots + \alpha_p\rho_{k-p}, \quad k \geq (q+1).$$

Obsérvese que la función autocorrelación parcial solamente depende de los parámetros del término autorregresivo, es por esta razón, que esta función cumple con la propiedad de dicho proceso de que no se corta después de p rezagos, sino que decae exponencialmente.

La función autocorrelación parcial del proceso ARMA(p,q)

Puesto que este proceso es una mezcla de los procesos $AR(p)$ y $MA(q)$, y el proceso $ARMA$ contiene el proceso MA como un caso especial, la FAC parcial también será una mezcla de decaimiento exponencial positivo, negativo o alternando valores, dependiendo de las raíces de $\alpha(B) = 0$ y de $\beta(B) = 0$ (Véase Apéndice B).

Cabe mencionar que cuando $p = 0$ entonces el proceso se reduce a un proceso $MA(q)$ y cuando $q = 0$, entonces el proceso se reduce a un $AR(p)$.

▪ ***El proceso ARMA(1,1)***

El proceso $ARMA(1,1)$ está dado por

$$X_t = \alpha_1 X_{t-1} + Z_t - \beta_1 Z_{t-1}$$

ó usando el operador de retroceso,

$$(1 - \alpha_1 B)X_t = (1 - \beta_1 B)Z_t.$$

Para que este proceso sea estacionario e invertible, se asume que $|\alpha_1| < 1$ y $|\beta_1| < 1$ respectivamente.

3.3. Identificación de modelos

Existen diversas técnicas para identificar modelos de series de tiempo, en esta tesis, únicamente se describe la más común que consiste en observar el correlograma y la función autocorrelación parcial.

3.3.1. Pasos para la identificación

Para ilustrar la identificación, considérese el modelo general $ARMA(p, q)$

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

la identificación se refiere a la metodología usada para identificar las transformaciones requeridas en caso de ser necesarias para estabilizar la varianza, al mismo tiempo que se determinan los órdenes p y q respectivamente. De esta manera, para una serie dada, se proponen los siguientes pasos útiles para identificar el modelo adecuado.

- *Paso 1.* Graficar los datos de la serie de tiempo. En cualquier análisis estadístico, el primer paso es graficar los datos. Examinando cuidadosamente la gráfica se puede dar una idea de si la serie contiene tendencia, estacionalidad, varianza no constante, no estacionariedad, etc. Esto proporciona bases para postular una posible transformación a los datos.

Las transformaciones más comunes son estabilizar la varianza mediante diferencias o estimando y eliminando sus componentes mediante alguno de los métodos mencionados en la sección 1 de este Capítulo.

- *Paso 2.* Calcular y examinar la FAC muestral y la FACP muestral de la serie original para ver si es necesario realizar diferenciación, las reglas para determinar esto son:
 - si la FAC muestral decae muy lentamente y la FACP se corta después del primer rezago, esto indica que es necesario obtener la primer diferencia para la serie o aplicar alguna otra técnica para hacerla estacionaria;
 - de manera más general, para eliminar la no estacionariedad en la serie se puede considerar una diferencia de orden mayor. Comúnmente solo se obtienen las diferencias para d igual a 0, 1 ó 2. Obtener una diferencia de orden mayor a tres ocasiona una sobrediferenciación, es decir, en lugar de obtener un mejor ajuste para el modelo se pierde eficiencia, por lo que estimar y eliminar las componentes de la serie sería lo recomendado.

- *Paso 3.* Calcular y examinar la FAC muestral y FAC de la serie transformada o diferenciada para identificar los órdenes de p y q respectivamente (p es el orden más grande en el polinomio autorregresivo $(1 - \alpha_1 B - \cdots - \alpha_p B^p)$ y q es el orden más grande del polinomio de medias móviles $(1 - \beta_1 B - \cdots - \beta_q B^q)$). Usualmente, los órdenes obtenidos para p y q son menores o iguales a 3. Para construir un buen modelo son necesarias al menos $n = 50$ observaciones y el número de la FAC muestral y FACP a ser calculadas es de aproximadamente $n/4$.

Brevemente, para un proceso $AR(p)$ se tiene que la función autocorrelación decae exponencialmente, mientras que su función autocorrelación parcial se corta después del rezago p . Por otro lado, para el proceso $MA(q)$, la función autocorrelación parcial se corta después del q -ésimo rezago, mientras su función autocorrelación parcial decae

exponencialmente. El Cuadro 3.1 resume estas características para los procesos $AR(p)$, $MA(q)$ y $ARMA(p, q)$.

Proceso	FAC	FACP
$AR(p)$	Decae exponencialmente	Se corta después del rezago p
$MA(q)$	Se corta después del rezago q	Decae exponencialmente
$ARMA(p, q)$	Decae después del rezago $(q - p)$	Decae después del rezago $(p - q)$

Cuadro 3.1: Características de los procesos $AR(p)$, $MA(q)$ y $ARMA(p, q)$.

3.4. Estimación de parámetros

Una vez que se tiene el modelo tentativo, el siguiente paso es estimar los parámetros del modelo. Es decir, si ya se ha detectado cual es el modelo más adecuado para los datos y por consiguiente el orden del modelo, entonces se deben estimar los parámetros para que el ajuste del modelo quede completo. Considérese el modelo general $ARMA(p, q)$,

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots + \beta_q Z_{t-q}.$$

El objetivo de estimación de parámetros consiste en determinar los parámetros $\alpha_1, \alpha_2, \dots, \alpha_p$ y $\beta_1, \beta_2, \dots, \beta_q$.

3.4.1. Estimación de los parámetros de un proceso autorregresivo

Supóngase que se tiene un proceso AR de orden p , con media μ , dado por

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + Z_t.$$

Supóngase también que se tienen N observaciones x_1, x_2, \dots, x_N , entonces los parámetros $\mu, \alpha_1, \dots, \alpha_p$ pueden estimarse por mínimos cuadrados, es decir, minimizar

$$S = \sum_{t=p+1}^N [x_t - \mu - \alpha_1(x_{t-1} - \mu) - \cdots - \alpha_p(x_{t-p} - \mu)]^2$$

con respecto a $\mu, \alpha_1, \dots, \alpha_p$. Si $\{Z_t\}$ es normal, entonces los estimadores son también estimadores de máxima verosimilitud.

Otro método para estimar $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ consiste en usar el hecho de que la función correlación para un proceso $AR(p)$ está dada por $\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \dots + \alpha_p\rho_{k-p}$ para $k \geq 1$, a partir de la cual se puede obtener el sistema de ecuaciones siguiente:

$$\begin{aligned}\rho_1 &= \alpha_1 + \alpha_2\rho_1 + \alpha_3\rho_2 + \dots + \alpha_p\rho_{p-1} \\ \rho_2 &= \alpha_1\rho_1 + \alpha_2 + \alpha_3\rho_1 + \dots + \alpha_p\rho_{p-2} \\ &\vdots \\ \rho_p &= \alpha_1\rho_{p-1} + \alpha_2\rho_{p-2} + \alpha_3\rho_{p-3} + \dots + \alpha_p.\end{aligned}$$

las cuales son conocidas como ecuaciones *Yule-Walker*. Reemplazando ρ_k por $\hat{\rho}_k$ para cada $k = 1, 2, \dots, p$ se obtiene el sistema:

$$\begin{aligned}\hat{\rho}_1 &= \hat{\alpha}_1 + \hat{\alpha}_2\hat{\rho}_1 + \hat{\alpha}_3\hat{\rho}_2 + \dots + \hat{\alpha}_p\hat{\rho}_{p-1} \\ \hat{\rho}_2 &= \hat{\alpha}_1\hat{\rho}_1 + \hat{\alpha}_2 + \hat{\alpha}_3\hat{\rho}_1 + \dots + \hat{\alpha}_p\hat{\rho}_{p-2} \\ &\vdots \\ \hat{\rho}_p &= \hat{\alpha}_1\hat{\rho}_{p-1} + \hat{\alpha}_2\hat{\rho}_{p-2} + \hat{\alpha}_3\hat{\rho}_{p-3} + \dots + \hat{\alpha}_p.\end{aligned}$$

Y en forma matricial,

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}. \quad (3.30)$$

Resolviendo este sistema se obtienen los estimadores para los parámetros del proceso $AR(p)$.

En particular para $p = 1$ se tiene:

$$\hat{\alpha}_1 = \hat{\rho}_1.$$

Si $p = 2$, entonces

$$\hat{\alpha}_1 \simeq \frac{\hat{\rho}_1(1 - \hat{\rho}_2)}{(1 - \hat{\rho}_1^2)},$$

y

$$\hat{\alpha}_2 \simeq \frac{(\hat{\rho}_2 - \hat{\rho}_1^2)}{(1 - \hat{\rho}_1^2)}.$$

3.4.2. Estimación de los parámetros de un proceso de medias móviles

Estimar los parámetros de un proceso de medias móviles es un poco más complicado que para un proceso autorregresivo. Comúnmente se usan métodos numéricos para realizar esta estimación.

Considérese el proceso de medias móviles,

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

recuérdese que la FAC para este proceso está dada por

$$\rho_k = \frac{\beta_k + \beta_1 \beta_{k+1} + \cdots + \beta_{q-k} \beta_q}{1 + \beta_1^2 + \cdots + \beta_{k-1}^2 + \cdots + \beta_q^2}, \quad k = 1, 2, \dots, p.$$

En particular, para $q = 1$ se tiene que la función autocorrelación está dada por

$$\rho_1 = \frac{\beta_1}{1 + \beta_1^2},$$

de donde se obtiene la ecuación cuadrática,

$$\rho_1 \beta_1^2 - \beta_1 + \rho_1 = 0. \quad (3.31)$$

resolviendo esta ecuación para β_1 se obtienen las soluciones:

$$\beta_1 = \frac{1}{2\rho_1} + \left\{ \frac{1}{(2\rho_1)^2} - 1 \right\}^{\frac{1}{2}},$$

y

$$\beta_1 = \frac{1}{2\rho_1} - \left\{ \frac{1}{(2\rho_1)^2} - 1 \right\}^{\frac{1}{2}}.$$

Si se sustituyen en estas soluciones β por $\hat{\beta}$ y ρ por $\hat{\rho}$ se obtienen dos estimadores para β_1 . Se debe elegir $\hat{\beta}_1$ tal que $|\hat{\beta}_1| < 1$.

Alternativamente, para el proceso $MA(1)$ dado por

$$X_t = \mu + Z_t + \beta_1 Z_{t-1},$$

donde μ y β_1 son constantes y Z_t representa un proceso púramente aleatorio, se puede estimar μ y β_1 , usando un método numérico de la siguiente manera.

- Seleccionar los valores iniciales adecuados de μ y β_1 , tales que $\mu = \bar{x}$ y β_1 como el valor de la solución a la ecuación (3.31).

- Calcular la suma de residuales cuadrados usando la fórmula recursiva siguiente,

$$Z_t = X_t - \mu - \beta_1 Z_{t-1}. \quad (3.32)$$

Haciendo $z_0 = 0$, calcular $z_1 = x_1 - \mu$, $z_2 = x_2 - \mu - \beta_1 z_1, \dots, z_N = x_N - \mu - \beta_1 z_{N-1}$. Luego se calcula $\sum_{t=1}^N z_t^2$.

- Se repite este procedimiento para otros valores cercanos a μ y β_1 .
- Los valores buscados de μ y β_1 son aquellos tales que, minimicen $\sum_t z_t^2$.

Este procedimiento puede desarrollarse para procesos de orden mayor que uno. El problema con este proceso es que por ser un proceso iterativo se deben realizar muchas iteraciones para alcanzar las soluciones óptimas.

3.4.3. Estimación de los parámetros de un proceso autorregresivo de medias móviles

El procedimiento iterativo descrito para estimar los parámetros de un proceso de medias móviles, puede ser utilizado también para estimar los parámetros de un procedimiento autorregresivo de medias móviles.

En particular para un proceso $ARMA(1,1)$, se tiene el siguiente modelo,

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + Z_t + \beta_1 Z_{t-1}.$$

Para N observaciones dadas, x_1, x_2, \dots, x_N , se proponen valores para μ, α_1, β_1 y se hace $z_0 = 0, x_0 = \mu$. Calcular recursivamente los residuales por

$$\begin{aligned} z_1 &= x_1 - \mu \\ z_2 &= x_2 - \mu - \alpha_1(x_1 - \mu) - \beta_1 z_1 \\ &\vdots \\ z_N &= x_N - \mu - \alpha_1(x_{N-1} - \mu) - \beta_1 z_{N-1}. \end{aligned}$$

Luego se calcula $\sum_{i=1}^N z_i^2$. Proponer otros valores para μ, α_1, β_1 y repetir el mismo procedimiento seguido para el proceso de medias móviles.

3.5. Verificación

Una vez que se han estimado los parámetros del modelo propuesto, se debe verificar que este modelo sea el más adecuado para describir los datos, es decir, que ajuste bien los datos. Para ver si el modelo es el adecuado, se debe verificar que los supuestos del modelo se satisfagan. El supuesto básico es que $\{Z_t\}$ es de ruido blanco, es decir, que las v.a's Z_t 's iid son no correlacionadas con media cero y varianza constante. La forma más común de llevar a cabo este paso, es analizando los residuales de la serie.

3.5.1. Análisis residual

Los residuales están dados por

$$\text{residuales} = \text{observación} - \text{valor ajustado},$$

es decir, supóngase que el modelo ajustado para un proceso $ARMA(p, q)$ está dado por

$$X_t = \hat{\alpha}_1 X_{t-1} + \cdots + \hat{\alpha}_p X_{t-p} + Z_t + \hat{\beta}_1 Z_{t-1} + \cdots + \hat{\beta}_q Z_{t-q},$$

entonces los residuales estarán dados por

$$\hat{Z}_t = X_t - \hat{\alpha}_1 X_{t-1} - \cdots - \hat{\alpha}_p X_{t-p} - \hat{\beta}_1 Z_{t-1} - \cdots - \hat{\beta}_q Z_{t-q}.$$

En particular para el proceso $AR(1)$ dado por $X_t = \alpha_1 X_{t-1} + Z_t$. Si $\hat{\alpha}_1$ es el estimador para α_1 , entonces $\hat{\alpha}_1 x_{t-1}$ es el valor ajustado en el tiempo t para α_1 , es decir, el residual correspondiente para el valor observado x_t es

$$\hat{z}_t = x_t - \hat{\alpha}_1 x_{t-1}.$$

El simbolo “ $\hat{\cdot}$ ” en z_t indica que el residual es el término de error estimado. Si α se conoce con exactitud, entonces se puede calcular el error exacto $z_t = x_t - \alpha_1 x_{t-1}$, lo cual no sucede con datos reales, sólo en ejercicios simulados.

Si se tiene un buen modelo, entonces lo que se espera es que los residuales sean aleatorios y cercanos a cero, y la validación del modelo se puede llevar a cabo graficando los residuales para comprobar si se cumple esta propiedad. Los residuales también son considerados como una serie de tiempo, esto porque se encuentran ordenados en el tiempo.

Otra forma para llevar a cabo la verificación del modelo, consiste en graficar los residuales y calcular la FAC. Para probar que los residuales sean de ruido blanco iid, se examina su FAC muestral y se rechaza esta hipótesis si más de tres valores de los primeros cuarenta caen fuera de la banda de confianza $\pm 1.96/\sqrt{n}$ ó $\pm 2.58/\sqrt{n}$, donde n es el número total de observaciones, ó si un valor cae muy lejos de la banda de confianza. De lo contrario, el modelo propuesto es el adecuado.

Capítulo 4

Aplicación

En este capítulo se lleva a cabo una aplicación de series de tiempo a datos de demanda máxima mensual de energía eléctrica, realizando un ajuste para los datos.

En el capítulo anterior se sentaron las bases necesarias para llevar a cabo el análisis de los datos que forman una serie de tiempo, para realizarlo se deben tomar en cuenta algunas consideraciones para ajustar el modelo si se quieren obtener buenos resultados. Dichas consideraciones consisten en seguir los siguientes pasos: identificación del modelo, estimación de sus parámetros y por último verificación del modelo.

4.1. Datos

Los registros de demanda máxima mensual con los que se cuenta, consisten de una muestra de seis subestaciones de un sistema de distribución de la CFE. Cada una de estas subestaciones corresponden a un periodo de enero de 1995 a diciembre de 2006. Las subestaciones 1 y 2 están compuestas por dos transformadores de potencia cada uno. Para llevar a cabo el análisis se eligió la segunda subestación y el transformador 1. Debido a que las demás subestaciones presentan datos faltantes y otras comenzaron operaciones en el año 2004, desde luego que esto no es obstáculo para no llevar a cabo el análisis. Sin embargo, el objetivo es analizar datos completos y así poder identificar las componentes de la serie.

Los datos se pueden observar en la Figura 4.1, donde en el eje de las abscisas se representa el tiempo (enero de 1995 a diciembre de 2006) y en el eje de las ordenadas se muestra la demanda máxima mensual medida en MW. De primera inspección se logra apreciar, que la serie presenta fluctuaciones irregulares, lo cual indica que la varianza no es constante, además, es posible observar que existen altas de demanda de energía eléctrica en los meses de diciembre, aunque no de manera constante. Se observa también que existe la componente de tendencia la cual no es constante y la componente estacional, lo que indica que la serie no es estacionaria.

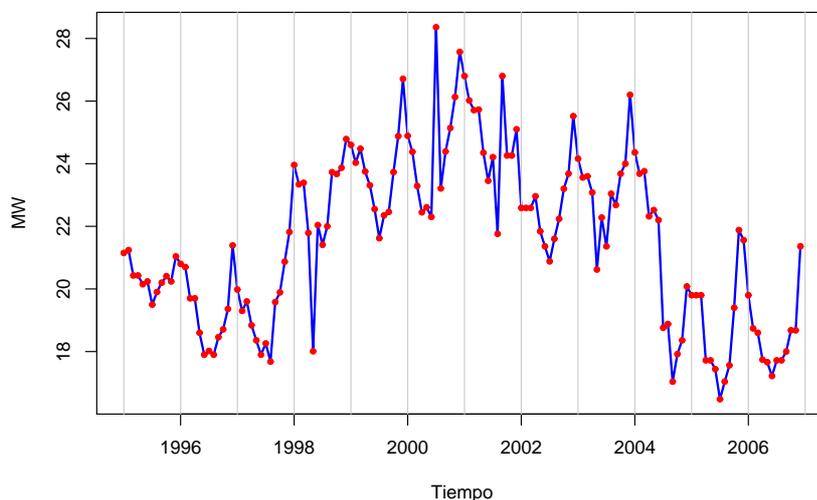


Figura 4.1: Serie de tiempo de demanda máxima de energía eléctrica.

4.2. Análisis

Para iniciar el análisis, se graficaron la FAC y la FACP de la serie original con el fin de identificar si era necesario diferenciar los datos ó aplicar un filtro de medias móviles, esto para eliminar las componentes de tendencia y estacionalidad, de esta forma se puede hacer la serie estacionaria. Las Figuras 4.2 y 4.3, muestran dichas gráficas para la serie en cuestión.

Se puede observar que la FAC decae en forma de ondas, por lo que no es necesario diferenciar la serie, por su parte la FACP no presenta un patrón específico, es por esta razón que se procedió a estimar las componentes de la serie mediante el procedimiento descrito en la sección 3.1.3., para después eliminar las componentes de tendencia y estacional, quedando únicamente la componente residual con la cual se realizó el análisis.

Se usó un filtro de medias móviles dado por (3.4) para estimar la componente de tendencia de la serie, es decir, se usó un periodo $d = 10$ y

$$\hat{m}_t = \frac{0.5x_{t-5} + x_{t-4} + \cdots + x_{t+4} + 0.5x_{t+5}}{10},$$

donde $5 < t < 139$. Dicha componente se puede observar en la Figura 4.4, ésta gráfica confirma el hecho de que la serie es no estacionaria, puesto que la tendencia no es constante.

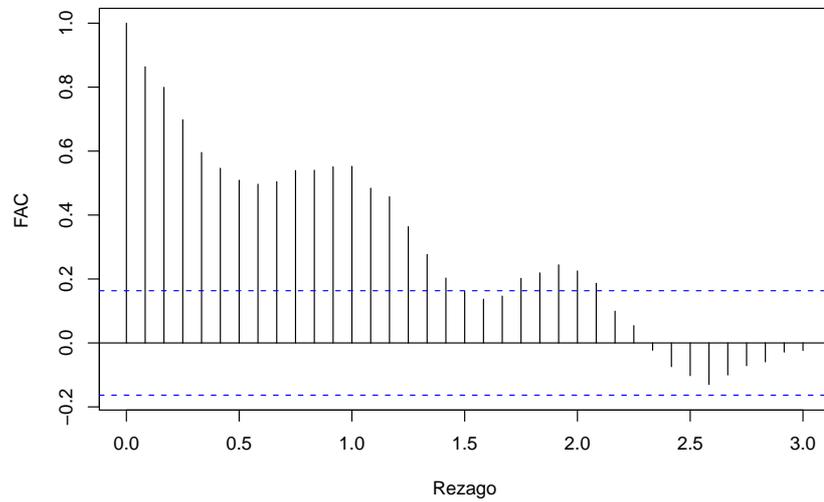


Figura 4.2: FAC de datos originales.

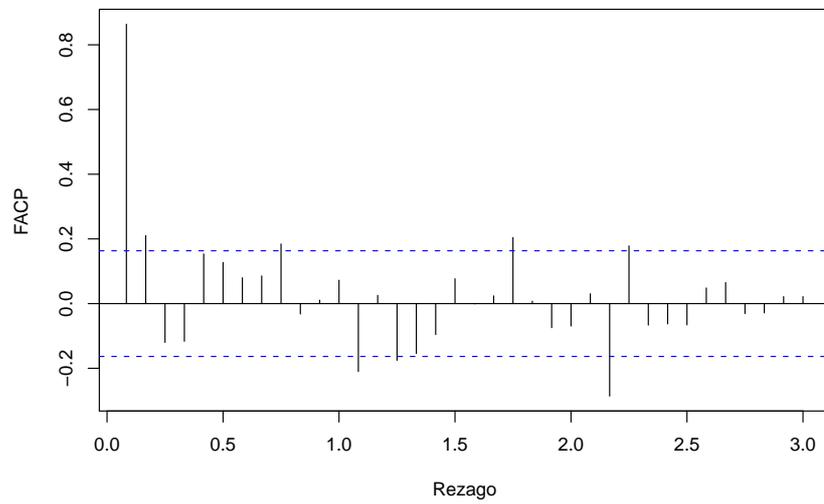


Figura 4.3: FACP de datos originales.

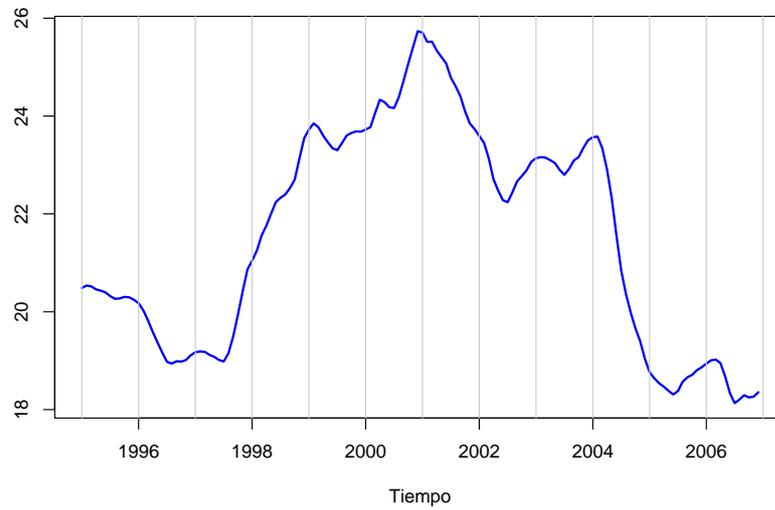


Figura 4.4: Tendencia de demanda de energía eléctrica.

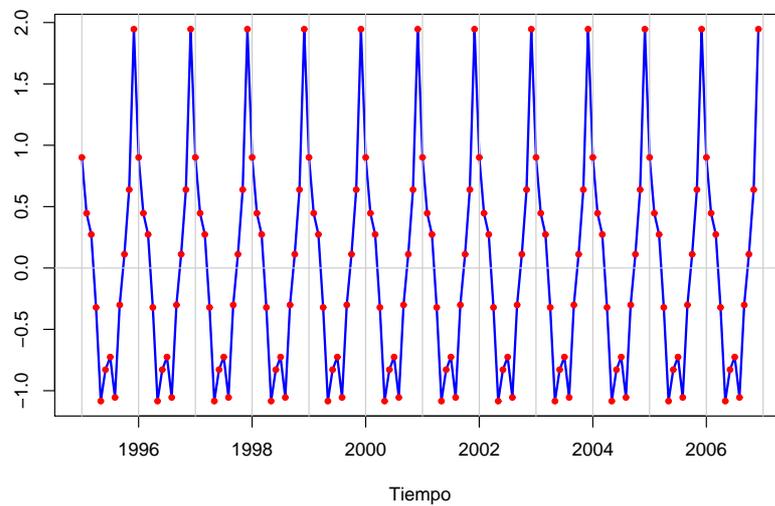


Figura 4.5: Componente estacional.

Una vez estimada la tendencia de la serie original, se procedió a eliminar esta componente m_t de la serie, es decir, se obtiene

$$y_t = x_t - m_t,$$

donde x_t representa las observaciones de la serie original.

Utilizando la serie sin tendencia y_t se estimó su componente estacional s_t , esta componente puede apreciarse en la Figura 4.5. Aquí puede observarse, que las altas de demanda de energía eléctrica tienden a presentarse en el mes de diciembre, lo cual confirma lo que se observó en la serie original, donde se logra apreciar una leve estacionalidad de la serie.

El siguiente paso consiste en desestacionalizar la serie, es decir, eliminar la componente estacional s_t , de esta forma se obtiene,

$$u_t = y_t - s_t.$$

La serie resultante correspondiente a este paso puede verse en la Figura 4.6.

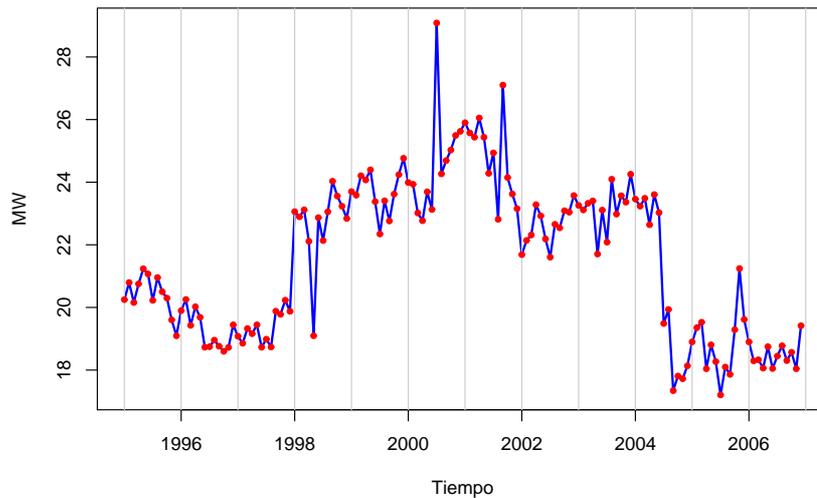


Figura 4.6: Serie desestacionalizada.

Una vez realizado esto, lo siguiente es estimar nuevamente la tendencia a esta última serie obtenida, dicha serie desestacionalizada con tendencia reestimada se muestra en la Figura 4.7. En esta serie puede apreciarse que la tendencia sigue siendo no constante, sin embargo ahora presenta una forma más definida, la cual resulta más fácil de hacerla constante y con esto obtener una serie estacionaria.

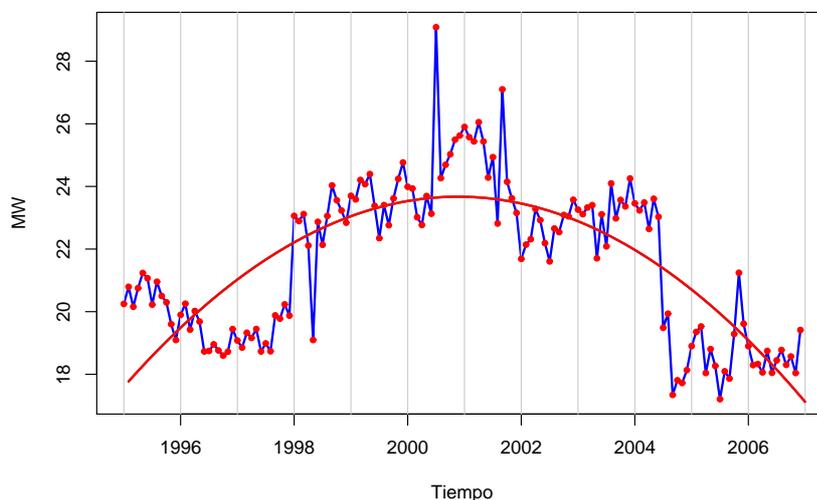


Figura 4.7: Serie desestacionalizada (con tendencia reestimada).

Por último, se eliminó la tendencia de la serie anterior, obteniendo el ruido residual,

$$r_t = x_t - m_t^1 - s_t,$$

el cual se puede observar en la Figura 4.8, donde se muestra la estacionariedad de la serie.

Para determinar cual es el modelo que mejor se ajusta a los datos y determinar el orden, se calculó y graficó la FAC y la FACP de la última serie obtenida, es decir, de la componente residual, dichas gráficas pueden apreciarse en las Figuras 4.9 y 4.10.

Como se puede observar, la FAC decae en forma de ondas, aunque el decrecimiento no es lento. La FACP es la que muestra un comportamiento interesante, pues se ve que solamente los dos primeros valores salen de la banda de confianza y los demás se encuentran dentro de esta. Lo anterior llevó a concluir que el modelo que se ajusta a los datos, corresponde a un proceso autorregresivo de medias móviles de orden (2,0) ó bien a un proceso autorregresivo de orden 2 (ARMA(2,0) ó AR(2)). Es decir, el modelo propuesto está dado por

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad (4.1)$$

donde α_1 y α_2 son los parámetros a estimar.

Para hallar $\hat{\alpha}_1$ y $\hat{\alpha}_2$ se utilizan las ecuaciones *Yule-Walker* dadas por (3.30), en donde se obtuvo el siguiente sistema para el modelo propuesto,

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \end{bmatrix}. \quad (4.2)$$

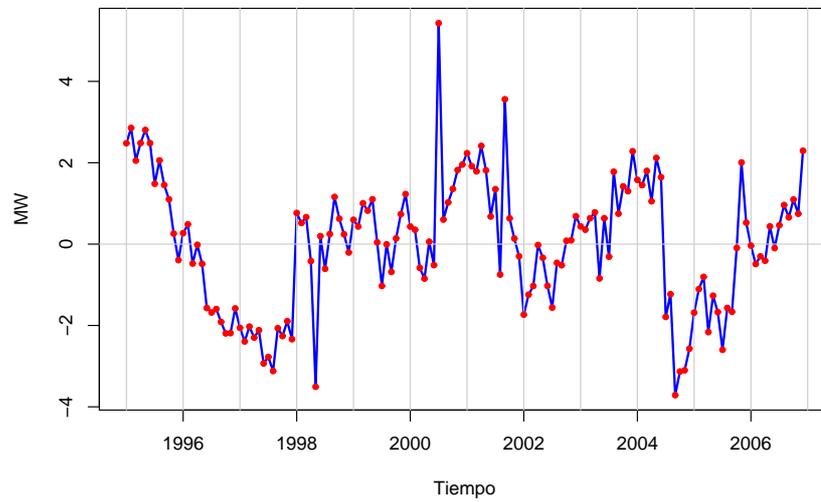


Figura 4.8: Componente residual (sin tendencia ni estacionalidad).

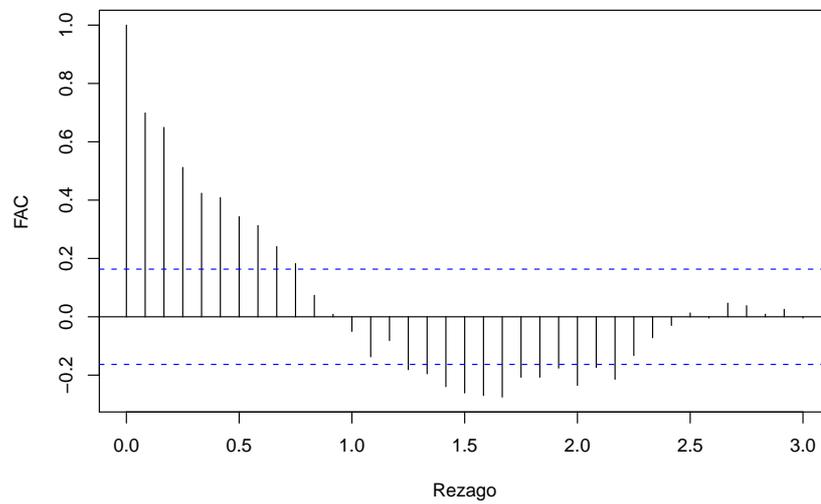


Figura 4.9: FAC de componente residual.

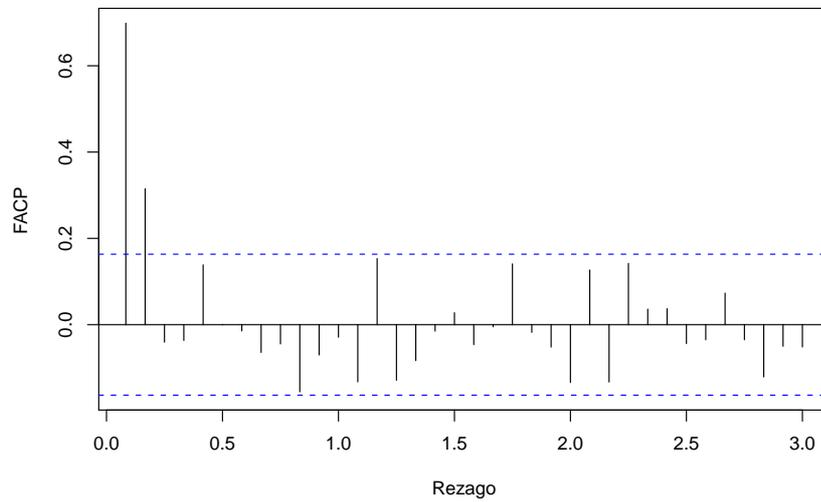


Figura 4.10: FACP de componente residual.

Ahora bien, resolviendo el sistema (4.2), usando regla de Cramer para α_1 se tiene,

$$\begin{aligned}
 \hat{\alpha}_1 &= \frac{\begin{vmatrix} \hat{\rho}_1 & \hat{\rho}_1 \\ \hat{\rho}_2 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{vmatrix}} \\
 &= \frac{\hat{\rho}_1 - \hat{\rho}_1\hat{\rho}_2}{1 - \hat{\rho}_1^2} \\
 &= \frac{\hat{\rho}_1(1 - \hat{\rho}_2)}{1 - \hat{\rho}_1^2}
 \end{aligned}$$

y para α_2 ,

$$\begin{aligned}
 \hat{\alpha}_2 &= \frac{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & \hat{\rho}_2 \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{vmatrix}} \\
 &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}.
 \end{aligned}$$

Los primeros 36 valores de la FAC que se obtuvieron se muestran en el Cuadro 4.1, usando los dos primeros valores y reemplazandolos en $\hat{\alpha}_1$ y $\hat{\alpha}_2$ se obtienen

$$\hat{\alpha}_1 = \frac{0.864(1 - 0.8)}{1 - (0.864)^2} = 0.681$$

y

$$\hat{\alpha}_2 = \frac{0.8 - (0.864)^2}{1 - (0.864)^2} = 0.211,$$

cuyos errores estándar son: 0.0817 y 0.0820.

k	1	2	3	4	5	6	7	8	9
$\hat{\rho}_k$	0.864	0.800	0.698	0.596	0.546	0.509	0.497	0.504	0.539
k	10	11	12	13	14	15	16	17	18
$\hat{\rho}_k$	0.540	0.551	0.552	0.484	0.457	0.364	0.277	0.203	0.161
k	19	20	21	22	23	24	25	26	27
$\hat{\rho}_k$	0.137	0.147	0.202	0.219	0.244	0.225	0.186	0.099	0.054
k	28	29	30	31	32	33	34	35	36
$\hat{\rho}_k$	-0.023	-0.074	-0.102	-0.130	-0.100	-0.071	-0.059	-0.028	-0.024

Cuadro 4.1: FAC de la componente residual.

Así, el modelo resultante más aproximado a los datos es:

$$X_t = 0.681X_{t-1} + 0.211X_{t-2} + Z_t. \quad (4.3)$$

La última fase en el análisis de la serie consiste en verificar que el modelo propuesto ajuste bien a los datos. Esto se llevó a cabo mediante el análisis residual. Los residuales obtenidos se muestran en la Figura 4.11.

De esta figura se puede observar que los residuales tienden a concentrarse alrededor de la media, es decir, los residuales son no correlacionados, con media cero y varianza constante.

Por otro lado se calculó la FAC para los residuales cuyos valores se muestran en el Cuadro 4.2.

Gráficamente, los residuales se muestran en la Figura 4.12. La gráfica muestra, que de los primeros 40 valores, solamente tres valores salen de la banda de confianza, y los demás se encuentran dentro de ella, es decir, son relativamente cercanos a cero. Con este análisis se llega a la conclusión de que el modelo (4.3) ajusta bien a los datos.

Finalmente, la Figura 4.13 muestra el modelo ajustado (4.3) a la serie.

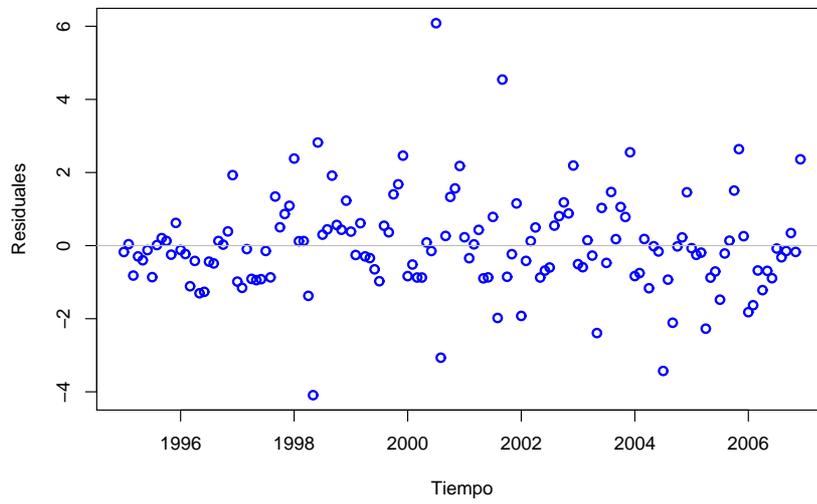


Figura 4.11: Residuales.

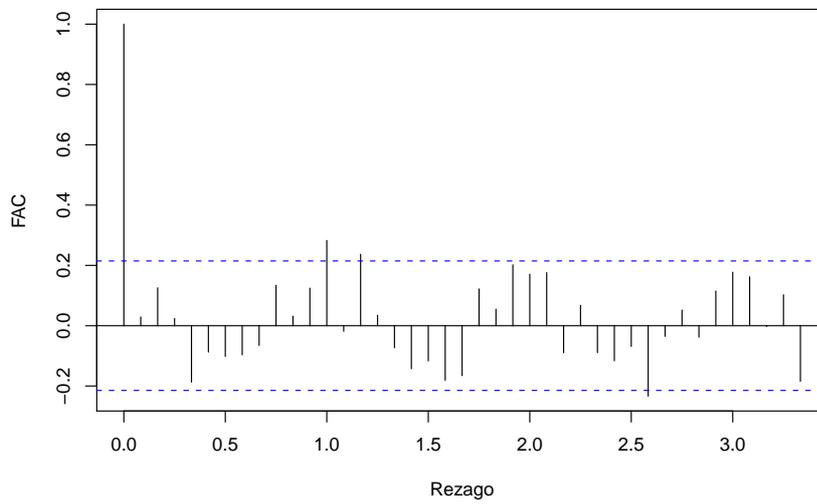


Figura 4.12: FAC de los residuales.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	0.029	0.126	0.024	-0.187	-0.087	-0.102	-0.097	-0.065	0.134	0.031
k	11	12	13	14	15	16	17	18	19	20
$\hat{\rho}_k$	0.125	0.283	-0.019	0.237	0.034	-0.073	-0.143	-0.117	-0.181	-0.166
k	21	22	23	24	25	26	27	28	29	30
$\hat{\rho}_k$	0.122	0.055	0.202	0.171	0.176	-0.090	0.068	-0.089	-0.116	-0.069
k	31	32	33	34	35	36	37	38	39	40
$\hat{\rho}_k$	-0.234	-0.035	0.052	-0.038	0.115	0.177	0.162	-0.002	0.103	-0.184

Cuadro 4.2: FAC de los residuales.

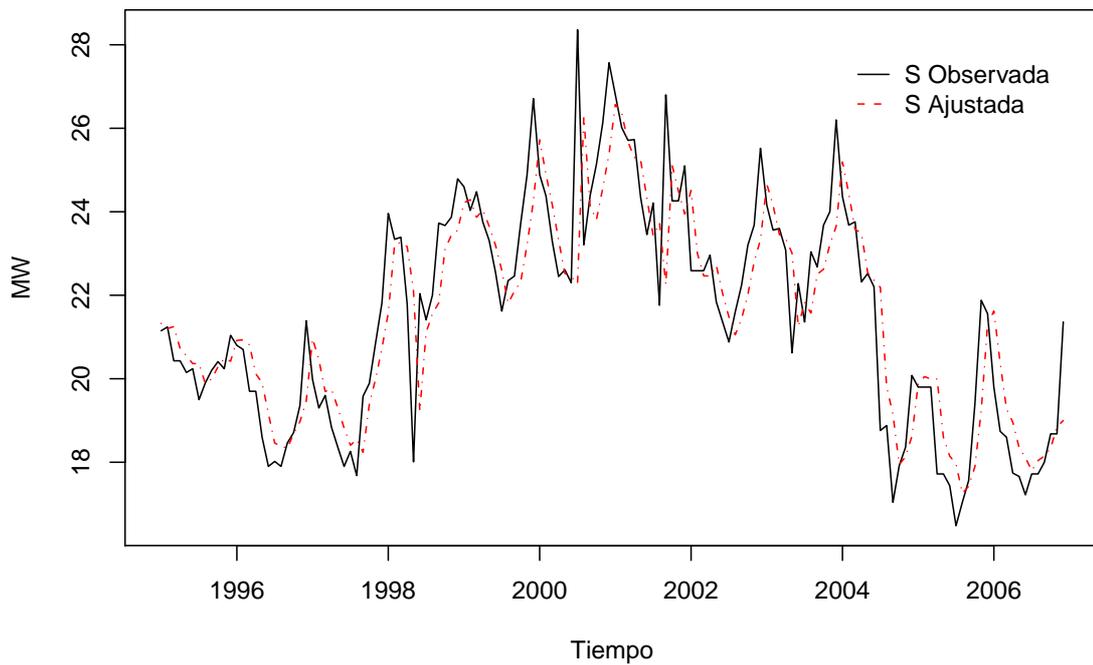


Figura 4.13: Ajuste del modelo a la serie.

Capítulo 5

Conclusiones

En este trabajo se estableció la base teórica de series de tiempo necesaria para llevar a cabo un análisis. Se presentaron aquellas técnicas útiles para modelar los datos de demanda máxima mensual de energía eléctrica. Se hizo además una revisión de los conceptos de probabilidad y estadística útiles para el estudio de las series de tiempo.

Se realizó la aplicación a los datos antes mencionados, para ello, se llevó a cabo una revisión de las FAC y FACP de los datos originales para decidir que técnica sería la que se utilizaría para hacer estacionaria la serie. Una vez que se decidió que lo más conveniente sería estimar las componentes de la serie, se procedió a estimar la componente de tendencia, esto se realizó aplicando un filtro de medias móviles. Después se eliminó esta componente de la serie y se estimó la componente estacional, se volvió a estimar la componente de tendencia pero sin la componente estacional. Luego se eliminó esta última componente de tendencia estimada con lo cual se obtuvo una componente residual de la serie, la cual fue estacionaria y de aquí se pudo obtener el modelo que ajusta a los datos. Recuérdese que no basta con proponer el modelo, por lo que para realizar la verificación se obtuvieron los residuales de la serie y su FAC para confirmar que el modelo realmente era el adecuado. Los resultados obtenidos en este trabajo para los datos de demanda de energía eléctrica mensual fueron satisfactorios, pues se obtuvo un buen ajuste de los datos.

En general, uno de los pasos más difíciles de llevar a cabo en el ajuste de un modelo de serie de tiempo es el de identificar el modelo. Se debe ser muy meticuloso al momento de hacer el análisis, puesto que un error puede conducir al modelo equivocado. Ajustar un modelo de serie de tiempo puede resultar tedioso para datos cuyo comportamiento no permita identificar patrones fácilmente. En este tipo de análisis se deben ajustar varios modelos y al final elegir aquel que mejor ajuste tenga con los datos originales.

Resultaría muy interesante retomar estos resultados previos para estudios futuros, particularmente para llevar a cabo la predicción.

La predicción es otro de los objetivos de las series de tiempo y resulta de gran utilidad en las empresas que deseen hacer planeación sobre su producción principalmente, como en el caso de la CFE.

Apéndice A

Datos

En este apartado se muestran los datos de demanda máxima mensual ocupados en el Capítulo 4.

Año	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio
1995	21.150	21.240	20.430	20.430	20.150	20.240	19.500
1996	20.800	20.700	19.700	19.700	18.600	17.900	18.020
1997	19.980	19.300	19.600	18.840	18.360	17.900	18.260
1998	23.960	23.340	23.390	21.790	18.010	22.040	21.410
1999	24.600	24.032	24.480	23.750	23.310	22.550	21.620
2000	24.890	24.380	23.290	22.448	22.608	22.300	28.360
2001	26.800	26.020	25.710	25.730	24.350	23.454	24.210
2002	22.587	22.587	22.587	22.960	21.840	21.360	20.880
2003	24.160	23.560	23.600	23.080	20.620	22.280	21.360
2004	24.360	23.680	23.760	22.320	22.520	22.200	18.760
2005	19.800	19.800	19.800	17.720	17.720	17.440	16.480
2006	19.800	18.740	18.600	17.740	17.660	17.220	17.720

Año	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1995	19.900	20.200	20.410	20.240	21.040
1996	17.900	18.460	18.710	19.360	21.390
1997	17.680	19.580	19.890	20.870	21.820
1998	22.000	23.730	23.670	23.870	24.788
1999	22.350	22.460	23.730	24.880	26.710
2000	23.210	24.390	25.140	26.130	27.570
2001	21.760	26.800	24.260	24.260	25.100
2002	21.600	22.240	23.200	23.680	25.520
2003	23.040	22.680	23.680	24.000	26.200
2004	18.880	17.040	17.920	18.360	20.080
2005	17.040	17.560	19.400	21.880	21.560
2006	17.720	18.000	18.680	18.680	21.360

Apéndice B

Correlogramas

En este apartado se presentan los comportamientos de las FAC y FACP para los modelos $AR(p)$, $MA(q)$ y $ARMA(p, q)$. A partir de estas tablas se puede estimar el modelo apropiado para los datos disponibles, una vez que se hayan calculado y graficado la FAC y la FACP correspondientes.

La siguiente tabla muestra el comportamiento de la FAC y la FACP de un proceso $AR(1)$ y $MA(1)$.

	ARMA(1,0)	ARMA(0,1)
Comportamiento de ρ_k	decae exponencialmente	solo $\rho_1 \neq 0$
Comportamiento de ϕ_{kk}	solo $\phi_{11} \neq 0$	decae exponencialmente
Estimación de:	$\alpha_1 = \rho_1$	$\rho_1 = \frac{-\beta_1}{1+\beta_1^2}$
Región admisible	$-1 < \alpha_1 < 1$	$-1 < \beta_1 < 1$

Aquí, se presentan los comportamientos de la FAC y la FACP de un proceso $AR(2)$ y $MA(2)$.

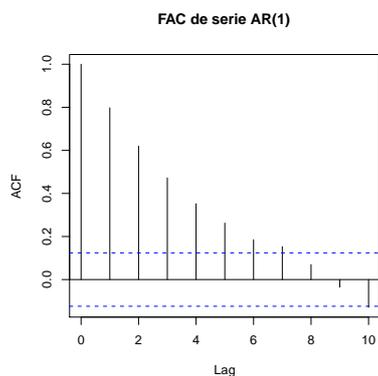
	ARMA(2,0)	ARMA(0,2)
Comportamiento de ρ_k	mezcla de exponenciales u ondas	solo $\rho_1, \rho_2 \neq 0$
Comportamiento de ϕ_{kk}	solo $\phi_{11}, \phi_{22} \neq 0$	mezcla de exponenciales u ondas
Estimación de:	$\alpha_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ $\alpha_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_1 = \frac{-\beta_1(1-\beta_2)}{1+\beta_1^2+\beta_2^2}$ $\rho_2 = \frac{-\beta_2}{1+\beta_1^2+\beta_2^2}$
Región admisible	$-1 < \alpha_2 < 1$ $\alpha_2 + \alpha_1 < 1$ $\alpha_2 - \alpha_1 < 1$	$-1 < \beta_2 < 1$ $\beta_2 + \beta_1 < 1$ $\beta_2 - \beta_1 < 1$

A continuación se resumen las características principales de un proceso $ARMA(1,1)$.

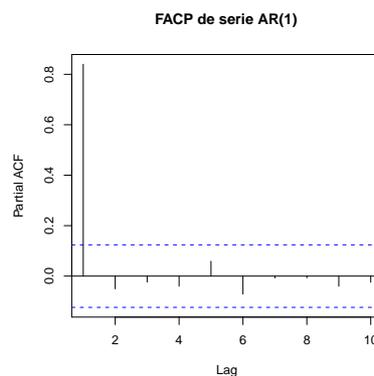
ARMA(1,1)	
Comportamiento de ρ_k	decae exponencialmente
Comportamiento de ϕ_{kk}	decae exponencialmente
Estimación de:	$\rho_1 = \frac{(1-\beta_1\alpha_1)(\alpha_1-\beta_1)}{1+\beta_1^2-2\alpha_1\beta_1}$ $\rho_2 = \rho_1\alpha_1$
Región admisible	$-1 < \alpha_1 < 1$ $-1 < \beta_1 < 1$

Correlogramas para los modelos $AR(1)$, $AR(2)$, $MA(1)$ y $MA(2)$.

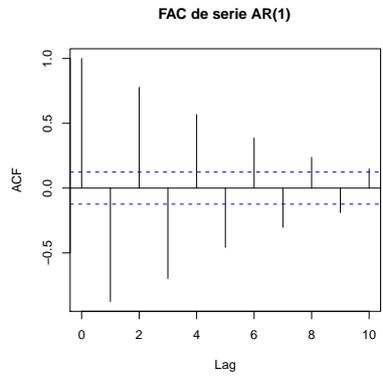
Estos ayudan a identificar de manera más fácil el modelo que se desea ajustar.



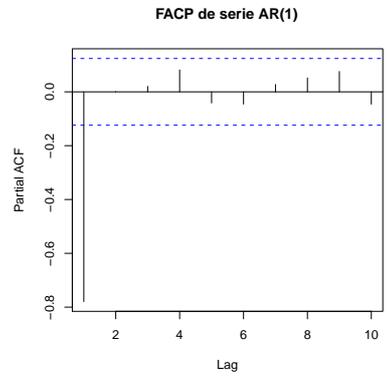
(a) FAC $AR(1)$, $\alpha > 0$.



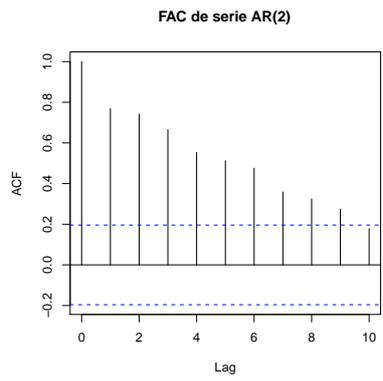
(b) FACP $AR(1)$, $\alpha > 0$.



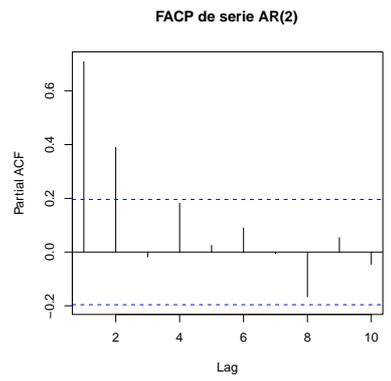
(c) FAC AR(1), $\alpha < 0$.



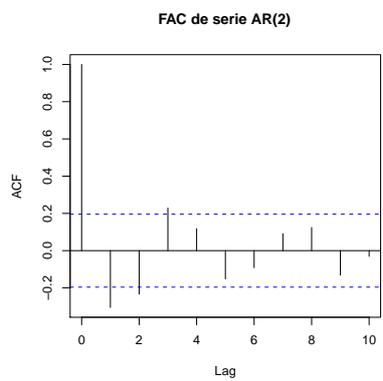
(d) FACP AR(1), $\alpha < 0$.



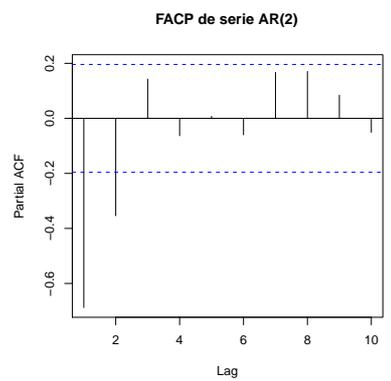
(e) FAC AR(2), $\alpha_1 > 0$ y $\alpha_2 > 0$.



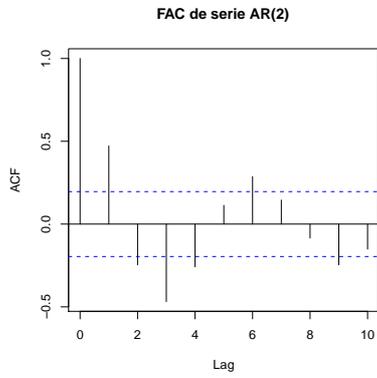
(f) FACP AR(2), $\alpha_1 > 0$ y $\alpha_2 > 0$.



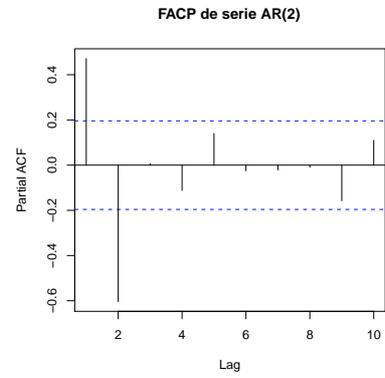
(g) FAC AR(2), $\alpha_1 < 0$ y $\alpha_2 < 0$.



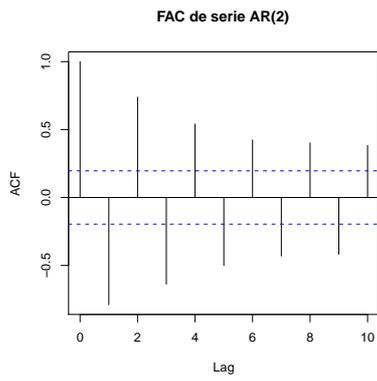
(h) FACP AR(2), $\alpha_1 < 0$ y $\alpha_2 < 0$.



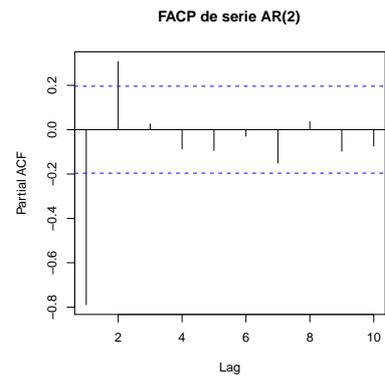
(i) FAC AR(2), $\alpha_1 > 0$ y $\alpha_2 < 0$.



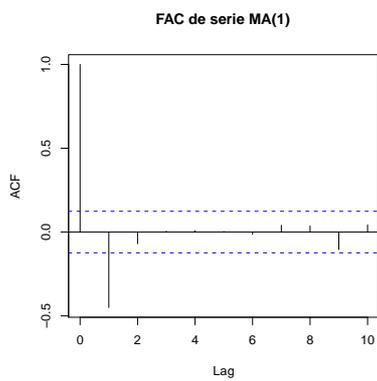
(j) FACP AR(2), $\alpha_1 > 0$ y $\alpha_2 < 0$.



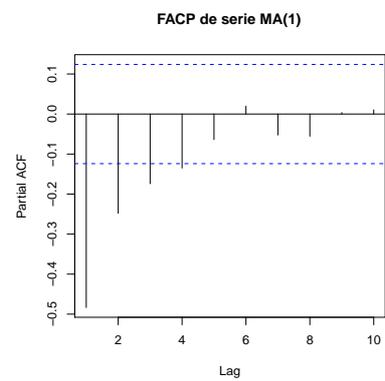
(k) FAC AR(2), $\alpha_1 < 0$ y $\alpha_2 > 0$.



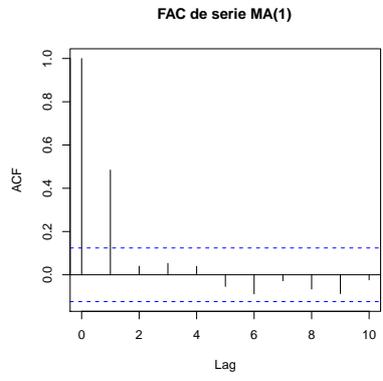
(l) FACP AR(2), $\alpha_1 < 0$ y $\alpha_2 > 0$.



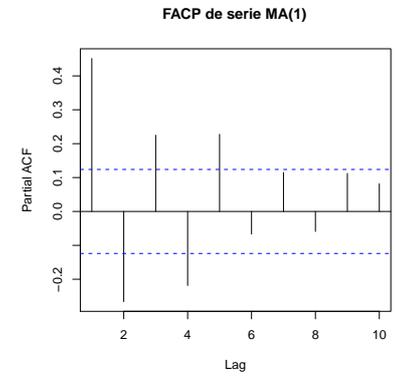
(m) FAC MA(1), $\beta > 0$.



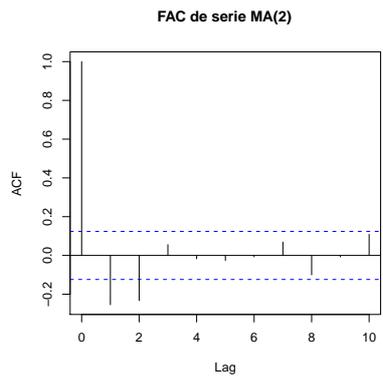
(n) FACP MA(1), $\beta > 0$.



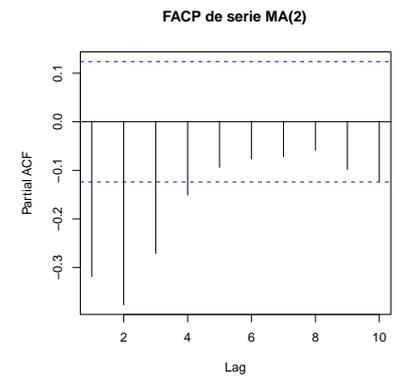
(ñ) FAC MA(1), $\beta < 0$.



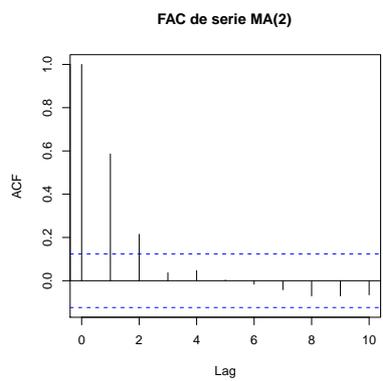
(o) FACP MA(1), $\beta < 0$.



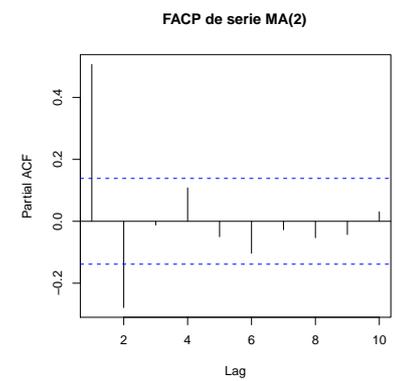
(p) FAC MA(2), $\beta_1 > 0$ y $\beta_2 > 0$.



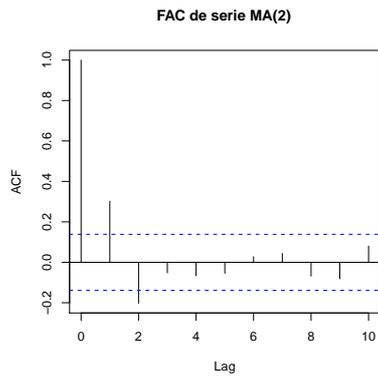
(q) FACP MA(2), $\beta_1 > 0$ y $\beta_2 > 0$.



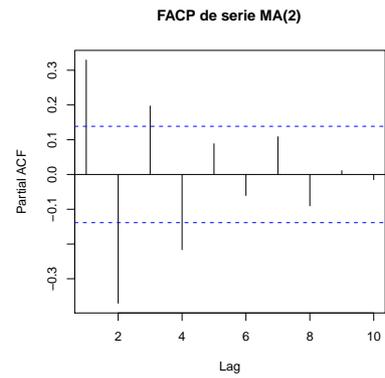
(r) FAC MA(2), $\beta_1 < 0$ y $\beta_2 < 0$.



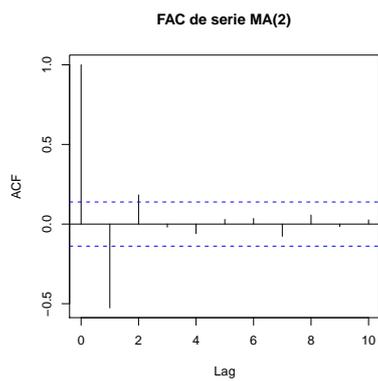
(s) FACP MA(2), $\beta_1 < 0$ y $\beta_2 < 0$.



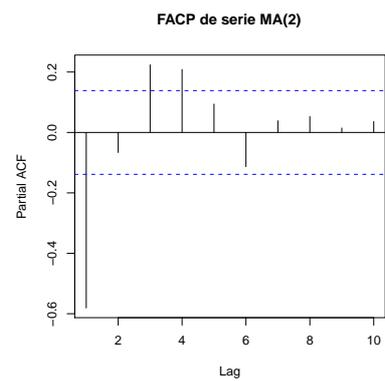
(t) FAC MA(2), $\beta_1 < 0$ y $\beta_2 > 0$.



(u) FACP MA(2), $\beta_1 < 0$ y $\beta_2 > 0$.



(v) FAC MA(2), $\beta_1 > 0$ y $\beta_2 < 0$.



(w) FACP MA(2), $\beta_1 > 0$ y $\beta_2 < 0$.

Bibliografía

- [1] Box-Jenkins. “*Time Series Analysis, Forecasting and Control*”. Holden-day. 1976.
- [2] Brockwell, Peter J. y Richard A. Davis. “*Introduction to Time Series and Forecasting*”. Springer. 1996.
- [3] Canavos, George C. “*Probabilidad y estadística, aplicaciones y métodos*”. Segunda Edición. McGraw-Hill. México. 1988.
- [4] Chatfield Chris. “*The Analysis of Time Series an Introduction*”. Texts in Statistical Science. CHAPMAN & HALL, cuarta edición. 2004.
- [5] DeGroot, Morris H. “*Probability an Statistics*”. Second Edition. Addison Wesley publishing company.
- [6] Kirchgaessner G., Wolters J. “*Introduction to Modern Time Series Analysis*”. Springer. 2007.
- [7] Juan Moreno. “*Control de la demanda de energía eléctrica*”. Física y sociedad. Num. 17, págs. 20-23, (2006)
- [8] Papoulis, Athanasios. “*Probability, Random Variables and Stochastic Processes*”. Third Edition. McGraw-Hill. Hill. 1991.
- [9] R Development Core Team , *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>, 2010.
- [10] Ross, Sheldon M. “*Stochastic Processes*”. Second Edition. Wiley. 1996.
- [11] Taylor, Howard M., Karlin, Samuel. “*An introduction to stochastic modeling*”. Third Edition. Academic Press. 1998.
- [12] Wackerly, Dennis D., William Mendenhall III, Richard L. Scheaffer. “*Estadística matemática con aplicaciones*”. Séptima edición. Cengage learning. 2008.
- [13] William W.S.Wei. “*Time Series Analysis Univariate and Multivariate Methods*”. Addison-Wesley Publishing Company, Inc. 1994.