

*Synergy Vision*

---

# *Series de Tiempo en R*



---

## ***Índice general***

---

<b>Índice de cuadros</b>	<b>V</b>
<b>Índice de figuras</b>	<b>VII</b>
<b>Prefacio</b>	<b>IX</b>
Información sobre los programas y convenciones . . . . .	IX
<b>Acerca del Autor</b>	<b>XI</b>
<b>Introducción</b>	<b>XIII</b>
0.1. Conceptos financieros básicos . . . . .	XIX
0.2. Conceptos básicos . . . . .	XXIV
0.3. Ejemplos . . . . .	XXV
0.3.1. Clasificación de las series de tiempo . . . . .	XXXII
0.4. Componentes de una serie de tiempo . . . . .	XXXIV
0.4.1. El Modelo Aditivo de Componentes de Series de Tiempo . . . . .	XXXVI
0.4.2. El Modelo Multiplicativo de Componentes de Series de Tiempo . . . . .	XXXIX
<b>Características de series de tiempo</b>	<b>XLI</b>
0.5. Medidas de dependencia para series de tiempo . . .	XLI
0.6. Estimación de la Tendencia . . . . .	LI
0.6.1. Estimación de la tendencia en ausencia de estacionalidad . . . . .	LV
0.6.2. Estimación de la tendencia y la estacionali- dad . . . . .	LXVI
0.7. Estimación de la tendencia por regresión clásica . .	LXXV
0.7.1. Regresión Clásica . . . . .	LXXV
<b>Modelos de series de tiempo</b>	<b>LXXXI</b>

0.8. Modelos Estocásticos . . . . .	LXXXI
0.8.1. Procesos Estocásticos . . . . .	LXXXI
0.8.2. Momentos, Varianza, Covarianza y Correlación . . . . .	LXXXII
0.8.3. Variación de un proceso . . . . .	LXXXII
0.8.4. Martingalas . . . . .	LXXXIII
0.8.5. Propiedad de Markov . . . . .	LXXXVIII
0.9. Modelos lineales . . . . .	LXXXIX
0.9.1. Proceso de Ruido Blanco . . . . .	LXXXIX

**Modelos AR****XCVII**

0.10. Modelo AR(1) . . . . .	XCVIII
0.11. Modelo AR(2) . . . . .	CII
0.12. Procesos AR(p) . . . . .	CVII
0.13. Función de Autocorrelación Parcial . . . . .	CIX
0.14. Criterios de Información . . . . .	CXI
0.15. Estimación de Parámetros. . . . .	CXV
0.16. Predicciones con modelos AR . . . . .	CXVI
0.16.1. Predicción de un paso . . . . .	CXVI
0.16.2. Predicción de dos pasos . . . . .	CXVII
0.16.3. Predicción de múltiples pasos . . . . .	CXVIII

**Modelos MA****CXXI**

0.17. Propiedades de los modelos MA . . . . .	CXXIV
0.17.1. Estacionaridad . . . . .	CXXIV
0.17.2. Función de autocorrelación (ACF) . . . . .	CXXVI
0.18. Identificación del orden de un MA . . . . .	CXXIX
0.19. Estimación . . . . .	CXXX

# Índice de cuadros



## Índice de figuras

1.	Beneficios por acción trimestrales para la compañía Johnson y Johnson . . . . .	XXV
2.	Índice Dow-Jones desde enero 1930 hasta octubre 2017 . . . . .	XXVIII
3.	Porcentaje de cambio de la bolsa de New York . . .	XXIX
4.	Evolución anual del EURIBOR (1999-2018) . . . .	XXX
5.	Evolución mensual del EURIBOR (2007-2018) . . .	XXXI
6.	Histórico de cambio del USD vs. EUR (01/05/2017 al 26/04/2018) . . . . .	XXXII
7.	Modelo aditivo de series de tiempo . . . . .	XXXVII
8.	Descomposición aditiva de la serie Johnson y Johnson . . . . .	XXXVIII
9.	Descomposición de la serie Johnson y Johnson usando la descomposición de Loess (STL) . . . . .	XXXVIII
10.	Descomposición multiplicativa de la serie Johnson y Johnson . . . . .	XL
11.	Gráficos de caminos aleatorios: con tendencia (negro), sin tendencia (rojo) . . . . .	XCVI
12.	Simulaciones de procesos autoregresivos de orden 1, AR(1), para distintos valores de phi . . . . .	CI
13.	Funciones de autocovarianzas para las series AR(1) simuladas . . . . .	CII
14.	Simulación de dos modelos MA(1) . . . . .	CXXV
15.	ACF para dos modelos MA(1) . . . . .	CXXVII
16.	Serie de tiempo y función de autocorrelación de los porcentajes de cambio diario de la Bolsa de Valores de New York, desde el 2 de febrero de 1984 hasta el 31 de diciembre de 1991 . . . . .	CXXIX





---

## ***Prefacio***

---



La versión en línea de este libro se comparte bajo la licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License<sup>1</sup>.

---

### **¿Por qué leer este libro?**

---

---

### **Estructura del libro**

---

---

### **Información sobre los programas y convenciones**

Este libro es posible gracias a una gran cantidad de desarrolladores que contribuyen en la construcción de herramientas para generar documentos enriquecidos e interactivos. En particular al autor de los paquetes Yihui Xie xie2015.

---

<sup>1</sup><http://creativecommons.org/licenses/by-nc-sa/4.0/>

---

## Prácticas interactivas con R

Vamos a utilizar el paquete Datacamp Tutorial<sup>2</sup> que utiliza la librería en JavaScript Datacamp Light<sup>3</sup> para crear ejercicios y prácticas con R. De esta forma el libro es completamente interactivo y con prácticas incluidas. De esta forma estamos creando una experiencia única de aprendizaje en línea.

eyJ5YW5ndWFnZSI6InIiLCJwcmVfZXhlcmlNpc2VfY29kZSI6ImIgPC0gNSIsInNhbXBsZSI6Il

---

## Agradecimientos

Synergy Vision, Caracas, Venezuela

---

<sup>2</sup><https://github.com/datacamp/tutorial>

<sup>3</sup><https://github.com/datacamp/datacamp-light>

---

## *Acerca del Autor*

---

Este material es un esfuerzo de equipo en Synergy Vision, (<http://synergy.vision/nosotros/>).

El propósito de este material es ofrecer una experiencia de aprendizaje distinta y enfocada en el estudiante. El propósito es que realmente aprenda y practique con mucha intensidad. La idea es cambiar el modelo de clases magistrales y ofrecer una experiencia más centrada en el estudiante y menos centrado en el profesor. Para los temas más técnicos y avanzados es necesario trabajar de la mano con el estudiante y asistirlo en el proceso de aprendizaje con prácticas guiadas, material en línea e interactivo, videos, evaluación continua de brechas y entendimiento, entre otros, para procurar el dominio de la materia.

Nuestro foco es la Ciencia de los Datos Financieros y para ello se desarrollará material sobre: **Probabilidad y Estadística Matemática en R, Programación Científica en R, Mercados, Inversiones y Trading, Datos y Modelos Financieros en R, Renta Fija, Inmunización de Carteras de Renta Fija, Teoría de Riesgo en R, Finanzas Cuantitativas, Ingeniería Financiera, Procesos Estocásticos en R, Series de Tiempo en R, Ciencia de los Datos, Ciencia de los Datos Financieros, Simulación en R, Desarrollo de Aplicaciones Interactivas en R, Minería de Datos, Aprendizaje Estadístico, Estadística Multivariante, Riesgo de Crédito, Riesgo de Liquidez, Riesgo de Mercado, Riesgo Operacional, Riesgo de Cambio, Análisis Técnico, Inversión Visual, Finanzas, Finanzas Corporativas, Valoración, Teoría de Portafolio**, entre otros.

Nuestra cuenta de Twitter es (<https://twitter.com/>

[bysynergyvision](https://github.com/synergyvision)) y nuestros repositorios están en GitHub (<https://github.com/synergyvision>).

**Somos Científicos de Datos Financieros**

# 0

---

## *Introducción*

---

Las series de tiempo ya han desempeñado un papel importante en las primeras ciencias naturales. La astronomía babilónica utilizó series de tiempo de las posiciones relativas de estrellas y planetas para predecir eventos astronómicos. Las observaciones de los movimientos de los planetas formaron la base de las leyes que Johannes Kepler descubrió. El análisis de las series de tiempo ayuda a detectar las regularidades en las observaciones de una variable y a derivar “leyes” de ellas, y/o explotar toda la información incluida en esta variable para predecir mejor los desarrollos futuros. La idea metodológica básica detrás de estos procedimientos, que también eran válidos para los babilonios, es que es posible descomponer series de tiempos en un número finito de componentes independientes pero no directamente observables que se desarrollan regularmente y que por lo tanto pueden ser calculados de antemano. Para este procedimiento es necesario que existan diferentes factores independientes que incidan en la variable.

A mediados del siglo XIX, este enfoque metodológico de la astronomía fue asumido por los economistas Charles Babbage y William Stanley Jevons. La descomposición en componentes no observados que dependen de diferentes factores causales, como suele emplearse en el análisis clásico de series de tiempo, fue desarrollada por Warren M. Persons (1919). Distinguía cuatro componentes diferentes:

- Desarrollo a largo plazo, tendencia,
- Componente cíclico con períodos de más de un año, el ciclo económico,
- Componente que contiene los altibajos dentro de un año, el ciclo estacional, y

- Componente que contiene todos los movimientos que no pertenecen ni a la tendencia ni al ciclo económico ni al componente estacional, el residual.

Suponiendo que los diferentes factores no observables son independientes, su recubrimiento aditivo genera las series de tiempo que, sin embargo, sólo podemos observar en su conjunto. Para obtener información sobre el proceso de generación de datos, tenemos que hacer suposiciones sobre sus componentes no observados. El análisis clásico de series de tiempo supone que los componentes sistemáticos, es decir, la tendencia, el ciclo económico y el ciclo estacional, no están influenciados por perturbaciones estocásticas y, por lo tanto, pueden representarse mediante funciones determinísticas del tiempo. El impacto estocástico se limita a los residuos, que, por otra parte, no contienen movimientos sistemáticos. Por lo tanto, se modela como una serie de variables aleatorias independientes o no correlacionadas con esperanza cero y varianza constante, es decir, como un proceso aleatorio puro.

Este enfoque cambió desde la presentación de los trabajos de George E. P. Box and Gwilym M. Jenkins, “Time Series Analysis: Forecasting and Control”, en los años 70 del siglo XX. Se abandonaron los procedimientos puramente descriptivos del análisis clásico de series de tiempo y, en su lugar, se han utilizado los resultados y métodos de la teoría de la probabilidad y las estadísticas matemáticas. Desde ese entonces, el análisis de series ha tenido un desarrollo creciente. Se han presentado una gran variedad de libros sobre este tópico, cada uno de ellos influenciado principalmente por la orientación de las series que se discuten en sus contenidos. Una gran parte de la literatura está dirigida a exponer los aspectos teóricos alrededor de las series de tiempo, siendo en muchos casos, rigurosamente desarrollados y descritos, sin embargo poco de ellos presentan implementaciones de las técnicas estudiadas y su comprensión en ejemplos reales lo que a veces puede dificultar su comprensión en especial para aquellos que no posean una apropiada formación matemática.

Los primeros intentos de estudiar el comportamiento de las series de tiempo financieras fueron realizados por profesionales financieros y periodistas en lugar de por académicos. De hecho, esto parece

haberse convertido en una tradición de larga data, ya que, incluso hoy en día, gran parte de la investigación y el desarrollo empíricos todavía se originan en la propia industria financiera. Esto puede explicarse por el carácter práctico de los problemas, la necesidad de datos especializados y las posibles ventajas de dicho análisis. El primer y más conocido ejemplo de la investigación publicada sobre series de tiempo financieras es el legendario Charles Dow, como se expresa en sus editoriales en el *Wall Street Times* entre 1900 y 1902. Estos escritos formaron la base de la “teoría del Dow” e influyeron en lo que más tarde se conoció como análisis técnico y carisma. Aunque Dow no coleccionó y publicó sus editoriales por separado, esto fue hecho póstumamente por su seguidor Samuel Nelson (Nelson, 1902). Las ideas originales de Dow fueron posteriormente interpretadas y ampliadas por Hamilton (1922) y Rhea (1932). Estas ideas gozaron de cierto reconocimiento entre los académicos de la época: por ejemplo, Hamilton fue elegido miembro de la *Royal Statistical Society*.

Aunque Dow y sus seguidores discutieron muchas de las ideas que encontramos en el análisis moderno de finanzas y series de tiempo, incluyendo estacionalidad, eficiencia del mercado, correlación entre rendimiento de activos e índices, diversificación e imprevisibilidad, no hicieron ningún esfuerzo serio para adoptar métodos estadísticos formales. La mayor parte del análisis empírico consistió en la interpretación minuciosa de gráficos detallados de las medias bursátiles sectoriales, formando así los famosos Índices Dow-Jones. Se argumentó que estos índices descuentan toda la información necesaria y proporcionan el mejor pronóstico de eventos futuros. Una idea fundamental, muy relevante para la teoría de los ciclos de Stanley Jevons y la metodología de descomposición de tendencias de la “curva Harvard A-B-C” de Warren Persons, fue que las variaciones de precios del mercado consistían en tres movimientos primarios: diarios, a medio y largo plazo.

La investigación empírica más temprana que utiliza métodos estadísticos formales se remonta a los documentos de Working (1934), Cowles (1933,1944) y Cowles and Jones (1937). El trabajo centró la atención en una característica previamente señalada de los

precios de las materias primas y las acciones: que se asemejan a la acumulación de cambios puramente aleatorios. Alfred Cowles 3rd, analista financiero cuantitativamente entrenado y fundador de Econometric Society and the Cowles Foundation, investigó la habilidad de los analistas de mercado y servicios financieros para predecir los futuros cambios de precios, encontrando que había pocas pruebas de que pudieran hacerlo. Cowles y Jones reportaron evidencia de correlación positiva entre sucesivas variaciones de precios, pero, como posteriormente Cowles (1960) comentó, esto fue probablemente debido a que tomaron promedios mensuales de precios diarios o semanales antes de computar los cambios: un fenómeno de “correlación espuria”, analizado por Working (1960).

La previsibilidad de los cambios de precios se ha convertido desde entonces en un tema importante de la investigación financiera, pero, sorprendentemente, poco más se publicó hasta el estudio de Kendall (1953), en el que encontró que los cambios semanales en una amplia variedad de precios financieros no podían predecirse ni a partir de los cambios pasados en las series ni a partir de los cambios pasados en otras series de precios. Este parece haber sido el primer informe explícito de esta propiedad de los precios financieros a menudo citada, aunque la investigación sobre la previsibilidad de los precios sólo se vio impulsada por la publicación de los documentos de Roberts (1959) y Osborne (1959). El primero presenta un argumento en gran medida heurístico sobre por qué las sucesivas variaciones de precios deben ser independientes, mientras que el segundo desarrolla la proposición de que no se trata de cambios absolutos de precios, sino de cambios logarítmicos de precios independientes entre sí. Con la suposición auxiliar de que las propias modificaciones se distribuyen normalmente, esto implica que los precios se generan como movimiento Browniano.

El análisis de series de tiempo desempeña un papel importante en el análisis requerido para el pronóstico de eventos futuros. Existen varias formas o métodos de calcular cual va a ser la tendencia del comportamiento del proceso en estudio.

Un **pronóstico** es una predicción de algún evento o eventos futuros. Como sugirió Neils Bohr, hacer buenas predicciones no siempre es



fácil. Los pronósticos famosamente “malos” incluyen lo siguiente del libro *“Malas Predicciones”*:

- “La población es de tamaño constante y se mantendrá hasta el fin de la humanidad.” La Enciclopedia, 1756.
- “1930 será un espléndido año de empleo.” Departamento de Trabajo de los EE.UU., pronóstico de Año Nuevo en 1929, justo antes de que el mercado se desplomara el 29 de octubre.
- “Las computadoras se multiplican a un ritmo rápido. Para el cambio de siglo habrá 220,000 en los EE.UU.” Wall Street Journal, 1966.

Algunos ejemplos donde se puede utilizar y hacer precciones con series de tiempo:

- 1) **Dirección de Operaciones.** Las organizaciones empresariales utilizan habitualmente las previsiones de ventas de productos o la demanda de servicios para programar la producción, controlar los inventarios, gestionar la cadena de suministro, determinar las necesidades de personal y planificar la capacidad. Las previsiones también pueden utilizarse para determinar la combinación de productos o servicios que deben ofrecerse y las ubicaciones en las que deben fabricarse los productos.
- 2) **Marketing.** La previsión es importante en muchas decisiones de marketing. Las previsiones de respuesta de las ventas a los gastos publicitarios, las nuevas promociones o los cambios en las políticas de precios permiten a las empresas evaluar su eficacia, determinar si se están alcanzando los objetivos y realizar ajustes.
- 3) **Finanzas y Gestión de Riesgos.** Los inversores en activos financieros están interesados en pronosticar los rendimientos de sus inversiones. Estos activos incluyen, pero no se limitan a acciones, bonos y materias primas; otras decisiones de inversión se pueden tomar en relación con las previsiones de tasas de interés, opciones y tipos de cambio.

La gestión del riesgo financiero requiere previsiones de la volatilidad de la rentabilidad de los activos para que se puedan evaluar y asegurar los riesgos asociados a las carteras de inversión, y para que los derivados financieros puedan cotizarse adecuadamente.

- 4) **Economía.** Los gobiernos, las instituciones financieras y las organizaciones de política requieren pronósticos de las principales variables económicas, como el producto interno bruto, el crecimiento demográfico, el desempleo, las tasas de interés, la inflación, el crecimiento del empleo, la producción y el consumo. Estas previsiones son parte integrante de la orientación de la política monetaria y fiscal, así como de los planes y decisiones presupuestarias adoptadas por los gobiernos. También son fundamentales en las decisiones de planificación estratégica tomadas por organizaciones empresariales e instituciones financieras.
- 5) **Control de Procesos Industriales.** Las previsiones de los valores futuros de las características de calidad crítica de un proceso de producción pueden ayudar a determinar cuándo deben cambiarse las variables controlables importantes del proceso, o si el proceso debe detenerse y revisarse. Los esquemas de retroalimentación y control feedforward son ampliamente utilizados en el monitoreo y ajuste de procesos industriales, y las predicciones de la producción del proceso son una parte integral de estos esquemas.
- 6) **Demografía.** Las previsiones de población por país y región se realizan de manera rutinaria, a menudo estratificadas por variables como el género, la edad y la raza. Los demógrafos también pronostican nacimientos, muertes y patrones migratorios de las poblaciones. Los gobiernos utilizan estas previsiones para planificar políticas y acciones de servicio social, como el gasto en atención médica, programas de jubilación y programas de lucha contra la pobreza. Muchas empresas utilizan pronósticos de poblaciones por grupos de edad para hacer planes estratégicos

en relación con el desarrollo de nuevas líneas de productos o tipos de servicios que será ofrecido.

### 0.1. Conceptos financieros básicos

La mayoría de los estudios financieros y económicos implican rendimiento, en lugar de precios de los activos. Existen dos buenas razones para ello. Primero, para los inversores medios, el rendimiento de un activo es un resumen completo y libre de escala de la oportunidad de inversión. Segundo, las series de rendimiento son más fáciles de manejar que las series de precios porque las primeras tienen propiedades estadísticas más atractivas. Sin embargo, existen varias definiciones de rendimiento de activos. Sea  $P_t$  el precio de un activo en tiempo  $t$ . Discutiremos algunas definiciones de rendimiento que utilizaremos a lo largo del libro. Supongamos por el momento que el activo no paga dividendos.

**Definición 0.1** (Rendimiento simple de un periodo). Mantener el activo fijo durante un periodo a partir de tiempo  $t - 1$  hasta tiempo  $t$  da lugar a una **rentabilidad bruta simple**

$$1 + R_t = \frac{P_t}{P_{t-1}} \quad \text{o} \quad P_t = P_{t-1}(1 + R_t) \quad (0.1)$$

El correspondiente **rendimiento neto simple** de un periodo o **rendimiento simple** es

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (0.2)$$

**Definición 0.2** (Rendimiento simple multiperiodo). Mantener el activo fijo durante  $k$  periodos entre los tiempos  $t - k$  y  $t$  da un **rendimiento bruto simple de periodo  $k$**

$$\begin{aligned}
1 + R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} \\
&= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\
&= \prod_{j=0}^{k-1} (1 + R_{t-j})
\end{aligned}$$

De la definición, se tiene que la rentabilidad bruta simple de periodo  $k$  es el producto de las  $k$  rentabilidades brutas simples de un periodo. Esto se llama **rendimiento compuesto**. El **rendimiento neto simple de periodo  $k$**  es  $R_t[k] = (P_t - P_{t-k})/P_{t-k}$ .

En la práctica, el intervalo de tiempo real es importante para discutir y comparar los rendimientos (por ejemplo, rendimiento mensual o rendimiento anual). Si no se indica el intervalo de tiempo, se asume implícitamente que es de un año. Si el activo se mantuvo durante  $k$  años, entonces el rendimiento anualizado se define como

$$\text{Anualizado}\{R_t[k]\} = \left[ \prod_{j=0}^{k-1} (1 + R_{t-j}) \right]^{1/k} - 1.$$

Esta es una media geométrica de los  $k$  rendimientos brutos simple de un periodo y lo podemos calcular por

$$\text{Anualizado}\{R_t[k]\} = \exp \left[ \frac{1}{k} \sum_{j=0}^{k-1} \ln(1 + R_{t-j}) \right] - 1.$$

Debido a que es más fácil calcular el promedio aritmético que la media geométrica y los rendimientos de un periodo tienden a ser pequeños, podemos utilizar el desarrollo de Taylor de primer orden para aproximar el rendimiento anualizado y obtener

$$\text{Anualizado}\{R_t[k]\} \approx \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j} \quad (0.3)$$

Sin embargo, la exactitud de la aproximación en la ecuación (0.3) puede no ser suficiente en algunas aplicaciones.

Otra definición útil es la de **rendimiento compuesto continuo**, pero antes de discutir tales definiciones, discutamos el efecto de la capitalización. Supongamos que la tasa de interés de un depósito bancario es del 10 % anual y el depósito inicial es de \$1,00. Si el banco paga intereses una vez al año, entonces el valor neto del depósito se convierte en  $\$1,00(1 + 0,10) = \$1,10$  un año después. Si el banco paga intereses semestralmente, el tipo de interés a 6 meses es  $10\%/2 = 5\%$  y el valor del depósito será  $\$1,00(1 + 0,10/2)^2 = \$1,1025$  después del primer año. En general, si el banco para intereses  $m$  veces al año, entonces la tasa de interés para cada pago es  $10\%/m$  y el valor neto del depósito se convierte en  $\$1,00(1 + 0,1/m)^m$  un año después. La tabla siguiente da los resultados para algunos intervalos de tiempo comúnmente usados. En particular, el valor neto se aproxima a \$1,1052, que se obtiene con  $\exp(0,1)$  y se refiere al resultado de la capitalización continua.

Tipo	Número de pagos	Tasa de interés por periodo	Valor neto
Anual	1	0.1	\$1.10000
Semestral	2	0.05	\$1.10250
Trimestral	4	0.025	\$1.10831
Mensual	12	0.0083	\$1.10471
Semanal	52	0.1/52	\$1.10506
Diario	365	0.1/365	\$1.10516
Continuo	Inf.		\$1.10517

Tabla. Ilustración de los efectos de la combinación. El intervalo de tiempo es de 1 año y la tasa de interés es del 10 % anual.

En general, el valor liquidativo  $A$  de la capitalización continua es

$$A = C \exp(r \times n) \quad (0.4)$$

donde  $r$  es el tipo de interés anual,  $C$  es el capital inicial y  $n$  es el número de años. A partir de la ecuación (0.4), tenemos

$$C = A \exp(-r \times n) \quad (0.5)$$

el cual se refiere como el valor presente de un activo que vale  $A$  dolares  $n$  años a partir de ahora, asumiendo que la tasa de interés compuesta continua es  $r$  por año.

**Definición 0.3** (Rendimiento compuesto continuo). El logaritmo natural de rendimiento bruto simple de un activo se denomina **rendimiento compuesto continuo** o **rendimiento logarítmico**

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = p_t - o_{t-1} \quad (0.6)$$

donde  $p_t = \ln(P_t)$ .

Los rendimientos compuestos continuos disfrutan de algunas ventajas sobre los rendimientos netos simples  $R_t$ . En primer lugar, consideremos los rendimientos multiperiodos. Tenemos

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) = \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \cdots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1}. \end{aligned}$$

Por lo tanto, el rendimiento multiperiodo compuesto continuo es simplemente la suma de los rendimientos compuesto continuo de un periodo involucrados. En segundo lugar, las propiedades estadísticas de los logaritmos de los rendimientos son más manejables.

**Definición 0.4** (Rentabilidad de la cartera). El rendimiento neto simple de una cartera de inversión compuesta por  $N$  activos es una media ponderada de los rendimientos netos simples de los activos en cuestión, en la que la ponderación de cada activo es el porcentaje del valor de la cartera invertido en ese activo. Sea  $p$  un portafolio que ponga peso con el activo  $i$ , entonces el rendimiento simple  $p$  en el tiempo  $t$  es

$$R_{p,t} = \sum_{i=1}^N \omega_i R_{it}$$

donde  $R_{it}$  es el rendimiento simple del activo  $i$ .

Los rendimientos compuestos continuos de una cartera, sin embargo, no tienen la propiedad conveniente anterior. Si los rendimientos simples  $R_t$  son todos pequeños en magnitud, entonces tenemos

$$r_{p,t} \approx \sum_{i=1}^N \omega_i r_{it}$$

donde  $r_{p,t}$  es el rendimiento compuesto continuo de la cartera en el momento  $t$ . Esta aproximación se utiliza a menudo para estudiar los rendimientos de las carteras.

**Definición 0.5** (Pago de dividendos). Si un activo paga dividendos periódicamente, debemos modificar las definiciones de rendimientos de activos.

Sea  $D_t$  el pago de dividendos de un activo entre los tiempos  $t - 1$  y  $t$  y sea  $P_t$  el precio del activo al final del periodo  $t$ . Entonces el dividendo no se incluye en  $P_t$ . Entonces el rendimiento neto simple y el rendimiento compuesto continuo en el tiempo  $t$  están dados por

$$\begin{aligned} R_t &= \frac{P_t + D_t}{P_{t-1}} - 1 \\ r_t &= \ln(P_t + D_t) - \ln(P_{t-1}) \end{aligned}$$

**Definición 0.6** (Exceso de rendimiento). El **rendimiento excesivo** de un activo en el momento  $t$  es la diferencia entre el rendimiento del activo y el rendimiento de algún activo de referencia.

A menudo se considera que el activo de referencia no tiene riesgo, como una devolución de letras del Tesoro de EE.UU. a corto plazo. El exceso de rentabilidad simple y el logaritmo de exceso de rentabilidad de un activo se definen como

$$Z_t = R_t - R_{0t}; \quad z_t = r_t - r_{0t} \quad (0.7)$$

donde  $R_{0t}$  y  $r_{0t}$  son los rendimientos simples y logarítmicos del activo de referencia, respectivamente.

En la literatura financiera, el exceso de rentabilidad se considera

como el pago de una cartera de arbitraje que va larga en un activo y corta en el activo de referencia sin inversión inicial neta.

**\*\* Resumen de la relación\*\***

Las relaciones entre el rendimiento simple  $R_t$  y el rendimiento compuesto continuo (o logarítmico)  $r_t$  son

$$r_t = \ln(1 + R_t), \quad R_t = e^{r_t} - 1$$

La agregación temporal de los rendimientos produce

$$\begin{aligned} 1 + R_t[k] &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\ r_t[k] &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$

Si el tipo de interés compuesto continuo es por año, entonces la relación entre los valores presentes y futuros de un activo fijo es

$$A = C \exp(r \times n), \quad C = A \exp(-r \times n).$$


---

## 0.2. Conceptos básicos

Una serie tiempo es una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí. El principal objetivo de una serie de tiempo es su análisis para hacer pronóstico. Formalmente se tiene la siguiente definición.

**Definición 0.7.** Una **serie de tiempo** es un conjunto de observaciones  $x_t$ , cada una registrada a un tiempo específico  $t$ .

**Definición 0.8.** Un **modelo de series de tiempo** para los datos observados  $\{x_t\}$  es una especificación de una distribución conjunta (o posiblemente solo de medias y covarianzas) de una sucesión de variables aleatorias  $\{X_t\}$  de las cuales  $\{x_t\}$  es una realización.

A continuación presentaremos una serie de ejemplos que demuestran la utilidad y lo cotidiano de las series de tiempo, también se



mostrarán los códigos en **R** para cargar los archivos de datos y graficar las respectivas series de tiempo.

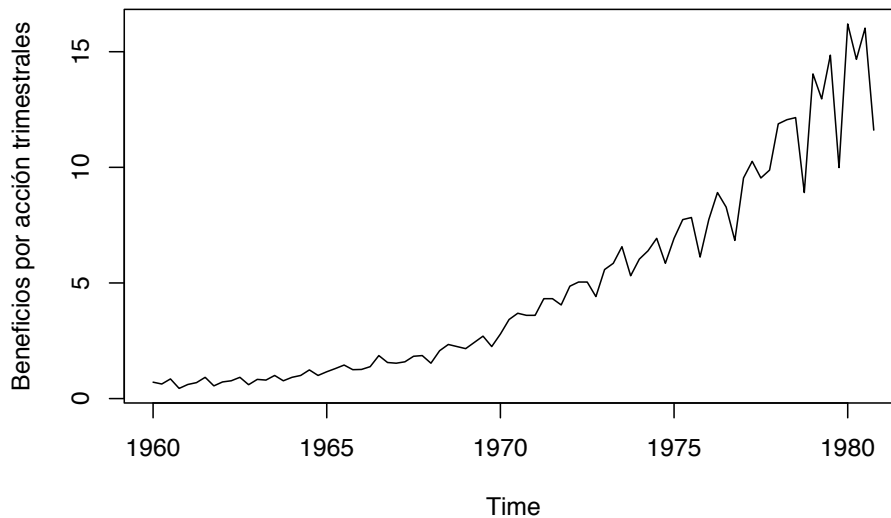
---

### 0.3. Ejemplos

**Ejemplo 0.1** (Beneficios de acciones). Beneficios por acción trimestrales para la compañía Johnson y Johnson. Se tienen 84 trimestres iniciando el primer trimestre de 1960 hasta el último trimestre de 1980. Los métodos para analizar tales datos se verán en el Tema 3 usando técnicas de regresión. El archivo es “*jj.txt*”.

Los comandos en R para cargar el archivo y graficar la serie de tiempo son los siguientes:

```
jj=ts(scan("data/jj.txt"),start=1960,freq=4)
plot(jj, type="l",ylab="Beneficios por acción trimestrales")
```



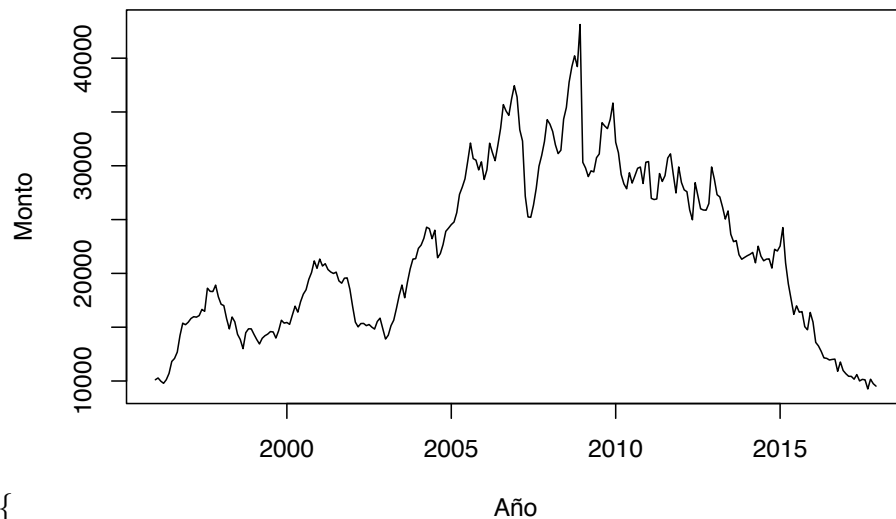
**Figura 1:** Beneficios por acción trimestrales para la compañía Johnson y Johnson

**Ejemplo 0.2.** El archivo “*ReservasInternacionales.xlsx*”, contiene el registro mensual de Reservas Internacionales Venezolanas en millones de dólares (\$), iniciando en el mes de enero de 1996 hasta el mes de diciembre de 2017

```
library(readxl)
reservas <- read_excel("data/ReservasInternacionales.xlsx")
reservas=ts(reservas,start = 1996,frequency = 12)
plot.ts(reservas[,2], xlab="Año",ylab="Monto",
        main="Reservas Internacionales de Venezuela (millones $)")
```

```
\begin{figure}
```

### Reservas Internacionales de Venezuela (millones \$)



```
{
```

```
}
```

```
\caption{Reservas Internacionales de Venezuela (millones $)
1996-2017} \end{figure}
```

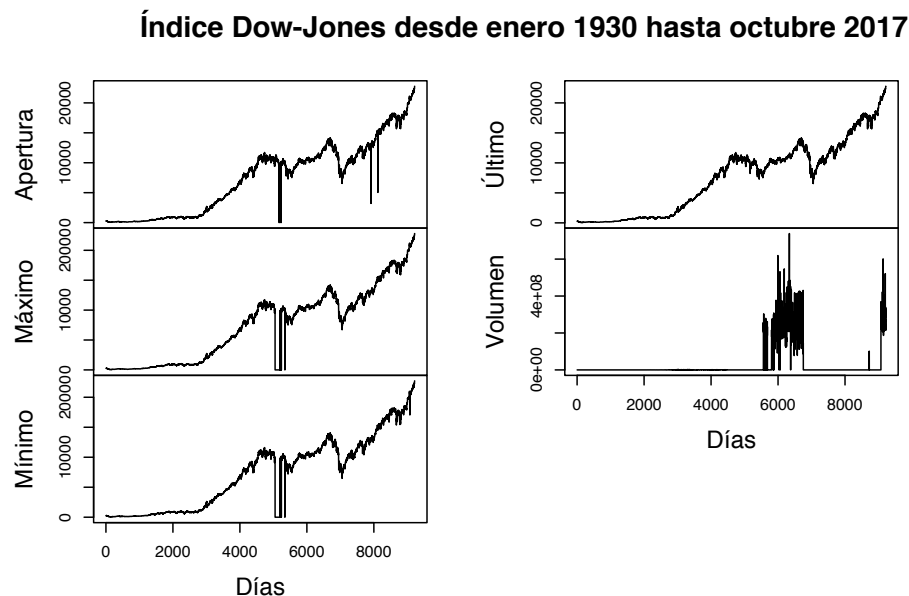
**Ejemplo 0.3.** El archivo “*PreciosPetroleo Vzla.xlsx*” contiene el precio promedio mensual de venta para el petróleo venezolano (en dólares) desde enero 2006 hasta noviembre 2017

```
library(readxl)
petroleo <- read_excel("data/PreciosPetroleoVzla.xlsx")
petroleo=ts(petroleo,start = 2006,frequency = 12)
plot.ts(petroleo[,2], xlab="Año",ylab="Monto",
        main="Precio promedio del petróleo venezolano (en dolares $)")
```



**Ejemplo 0.4.** El archivo “*IndiceDowJones.xlsx*” contiene los valores histórico del Índice Dow-Jones desde enero de 1930 hasta octubre de 2017. En el archivo podems notar que desde enero de 1930 hasta diciembre de 1994, los registros son el promedio semanal, a partir de enero de 1995, los registros son diarios. La primera columna es la fecha, la segunda columna es el valor de apertura, la tercera columna el valor máximo, la cuarta el valor mínimo, la quinta el último valor del índice o valor de cierre y la sexta columna es el volumen de acciones.

```
DJ=read_excel("data/IndiceDowJones.xlsx")
DJ=ts(DJ)
plot.ts(DJ[, -1], xlab="Días",
        main="Índice Dow-Jones desde enero 1930 hasta octubre 2017")
```

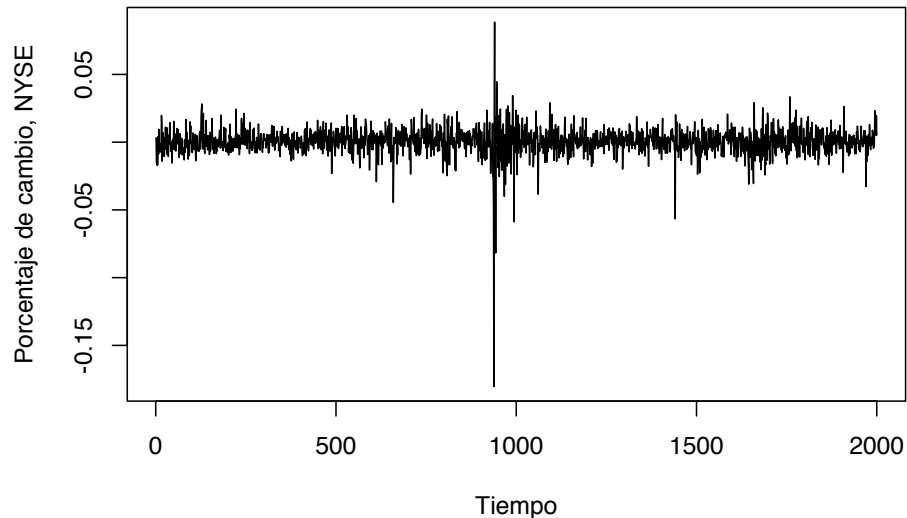


**Figura 2:** Índice Dow-Jones desde enero 1930 hasta octubre 2017

**Ejemplo 0.5.** La figura siguiente muestra los porcentajes de cambio diario de la Bolsa de Valores de New York desde el 2 de febrero de 1984 hasta el 31 de diciembre de 1991. Como se ve hay una caída fuerte, esta ocurrió el 19 de octubre de 1987 en  $t = 938$ . El archivo de datos es “nyse.txt”.

```
NYSE=ts(scan("data/nyse.txt"))
plot(NYSE,xlab="Tiempo",ylab="Porcentaje de cambio, NYSE")
```

**Ejemplo 0.6.** La evolución del EURIBOR es algo que fluctúa a diario. Se entiende por EURIBOR (Euro Interbank Offered Rate) el tipo de interés, promovido por el Instituto Europeo de Mercados

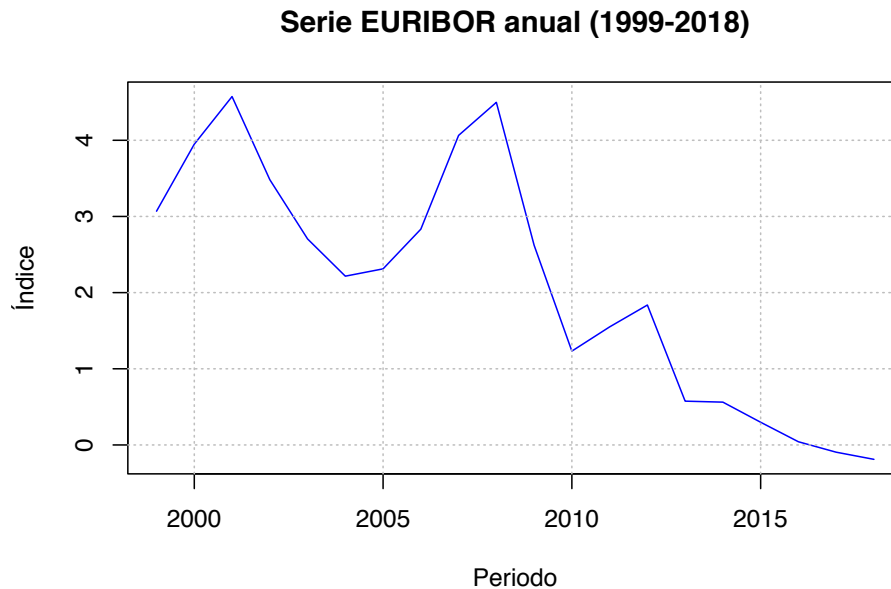


**Figura 3:** Porcentaje de cambio de la bolsa de New York

Monetarios (EMMI), consistente en la media aritmética simple de los valores diarios con días de mercado para operaciones de depósitos en euros a plazo de uno/tres/seis/doce meses y referido al día quince del mes anterior al comienzo de cada período de interés o al día siguiente hábil si aquel no lo fuese, calculado a partir del ofertado por una muestra de Bancos para operaciones entre entidades de similar calificación.

A continuación mostramos dos series del EURIBOR. La primera es la evolución histórica anual del EURIBOR desde su implantación en 1999 hasta 2018, los datos se corresponden al mes de enero de cada año. La segunda es la evolución mensual desde enero de 2007 hasta marzo de 2018.

```
EURIBORa<-read_excel("data/EURIBOR-anual.xlsx")
plot(EURIBORa,type="l", col = "blue", xlab = "Periodo",
     main="Serie EURIBOR anual (1999-2018)")
grid(col = "gray")
```

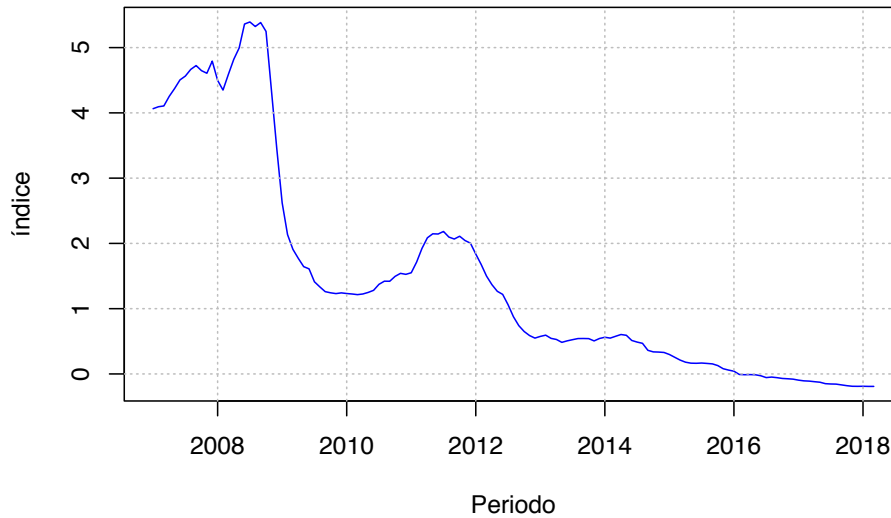


**Figura 4:** Evolución anual del EURIBOR (1999-2018)

```
EURIBORM<-read_excel("data/EURIBOR-mensual.xlsx")
EURITS<-ts(EURIBORM[,2],start = 2007, frequency = 12)
plot.ts(EURITS,xlab = "Periodo", col = "blue",
        main="Serie EURIBOR mensual (enero 2007- marzo 2018)")
grid(col = "gray")
```

**Ejemplo 0.7.** El archivo *Cambio-EUR-USD.xlsx* contiene el histórico de la cotización dolar estadounidense versus el euro desde el 01/05/2017 hasta el 26/04/2018. En la primera columna se muestra la fecha, la segunda columna el precio de apertura, la tercera el precio de cierre, la cuarta la diferencia en %, la quinta el precio máximo del día, la sexta el precio mínimo y la última el volumen de transacciones. A continuación presentamos los gráficos de apertura, cierre, máximo y mínimo.

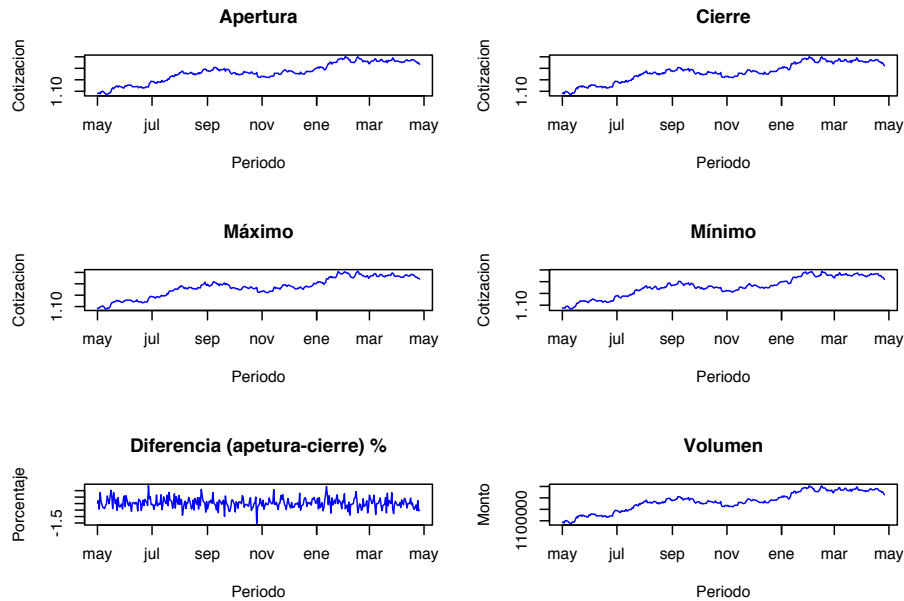
```
Cambio<-read_excel("data/Cambio-EUR-USD.xlsx")
par(mfrow=c(3,2))
```

**Serie EURIBOR mensual (enero 2007- marzo 2018)****Figura 5:** Evolución mensual del EURIBOR (2007-2018)

```

plot(Cambio$Fecha,Cambio$Apertura, col="blue", type = "l",
     xlab = "Periodo", ylab = "Cotizacion",
     main = "Apertura")
plot(Cambio$Fecha,Cambio$Cierre, col="blue", type = "l",
     xlab = "Periodo", ylab = "Cotizacion",
     main = "Cierre")
plot(Cambio$Fecha,Cambio$Máximo, col="blue", type = "l",
     xlab = "Periodo", ylab = "Cotizacion",
     main = "Máximo")
plot(Cambio$Fecha,Cambio$Mínimo, col="blue", type = "l",
     xlab = "Periodo", ylab = "Cotizacion",
     main = "Mínimo")
plot(Cambio$Fecha,Cambio$`Dif.%`, col="blue", type = "l",
     xlab = "Periodo", ylab = "Porcentaje",
     main = "Diferencia (apertura-cierre) %")
plot(Cambio$Fecha,Cambio$Volumen, col="blue", type = "l",
     xlab = "Periodo", ylab = "Monto",
     main = "Volumen")

```



**Figura 6:** Histórico de cambio del USD vs. EUR (01/05/2017 al 26/04/2018)

### 0.3.1. Clasificación de las series de tiempo

Como se ha mostrado en los ejemplos anteriores, hay una amplia variedad de series de tiempo que pueden clasificarse en varias categorías desde varios puntos de vista.

- **Serie de tiempo continuas y discretas.** Los datos registrados continuamente, por ejemplo, por un dispositivo analógico, se denominan series de tiempo continuas. Por otra parte, los datos observados en ciertos intervalos de tiempo, como la presión atmosférica medida cada hora, se denominan series de tiempo discretas. Existen dos tipos de series de tiempo discretas: una en la que las observaciones de los datos se realizan a intervalos de igual espaciamiento y otra en la que las observaciones de los datos se realizan a intervalos de espaciamiento desigual. Aunque las series



de tiempo mostradas en los ejemplos anteriores están conectadas continuamente por líneas sólidas, todas ellas son series de tiempo discretas. A partir de ahora en este libro, consideramos sólo series de tiempo discretas registradas a intervalos igualmente espaciados, porque las series de tiempo que analizamos en ordenadores digitales son generalmente series de tiempo discretas.

- **Series de tiempo univariadas y multivariadas.** Las series de tiempo que consisten en una sola observación en cada punto temporal, como se muestran en los ejemplos 1.1, 1.2, 1.3 y 1.5, se denominan series de tiempo univariadas. Por otra parte, las series de tiempo que se obtienen grabando simultáneamente dos o más fenómenos como los ilustrados en el ejemplo 1.4 se denominan series de tiempo multivariadas. Sin embargo, puede ser difícil distinguir entre series de tiempo univariadas y multivariadas desde su naturaleza; más bien, la distinción se hace desde el punto de vista del analista y por varios otros factores, como la restricción de la medición y los conocimientos empíricos o teóricos sobre el tema. Desde el punto de vista del modelado estadístico, la selección de variables en sí misma es un problema importante en el análisis de series de tiempo.
- **Series de tiempo estacionarias y no estacionarias.** Una serie de tiempo es un registro de un fenómeno que varía irregularmente con el tiempo. En el análisis de series de tiempo, las series de tiempo de variación irregular se expresan generalmente mediante modelos estocásticos. En algunos casos, un fenómeno aleatorio puede ser considerado como la realización de un modelo estocástico con una estructura de variación temporal. Estas series de tiempo se denominan series de tiempo estacionarias. El ejemplo 1.5 es un ejemplo típico de una serie de tiempo estacionaria. Por otra parte, si la estructura estocástica de una serie de tiempo cambia con el tiempo, se denomina serie de tiempo no estacionaria. Como ejemplos típicos de series de tiempo no estacionarias, considere la serie en los ejemplos 1.1 a 1.4. Se puede observar que los valores medios cambian a lo largo del tiempo.
- **Series de tiempo gaussianas y no gaussianas.** Cuando una distribución de una serie de tiempo sigue una distribución normal,

la serie de tiempo se denomina serie de tiempo gaussiana; de lo contrario, se denomina serie de tiempo no gaussiana. La mayoría de los modelos considerados en este libro son modelos gaussianos, asumiendo que las series de tiempo siguen distribuciones gaussianas. Al igual que en el caso del ejemplo 1.3, el patrón de las series de tiempo es a veces asimétrico, de modo que la distribución marginal no puede considerarse gaussiana. Incluso en tal situación, podemos obtener una serie de tiempo gaussiana aproximada mediante una transformación de datos apropiada.

- **Series de tiempo lineales y no lineales.** Una serie de tiempo expresable como la salida de un modelo lineal se denomina serie de tiempo lineal. Por el contrario, la salida de un modelo no lineal se denomina serie de tiempo no lineal.
- **Datos faltantes y valores atípicos.** En el modelado de series de tiempo de problemas del mundo real, a veces necesitamos tratar con observaciones faltante y valores atípicos. Algunos valores de las series de tiempo que no se han registrado por algunas razones se denominan observaciones que faltan en las series de tiempo. Los valores atípicos (observaciones exteriores) pueden ocurrir debido al comportamiento extraordinario del objeto, mal funcionamiento del dispositivo de observación o errores en el registro. En los datos de los ejemplos 1.4 y 1.5 se pueden observar datos atípicos. En el ejemplo 1.4 podemos notar caídas en los índices del DowJones y en el ejemplo 1.4 podemos notar una fuerte caída en el porcentaje de cambio de diario ocurrido el 19 de octubre de 1987.

---

#### 0.4. Componentes de una serie de tiempo

El análisis clásico de las series de tiempo se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de tres componentes, cuya actuación conjunta da como resultado los valores medidos, estos componentes son:

- 1) **Componente de tendencia.** Se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, o el cambio a largo plazo de la media. La tendencia se identifica con un movimiento suave de la serie a largo plazo.
- 2) **Componente estacional.** Muchas series de tiempo presentan cierta periodicidad o dicho de otro modo, variación de cierto período (semestral, mensual, etc.). Por ejemplo las Ventas al Detalle en Puerto Rico aumentan por los meses de noviembre y diciembre por las festividades navideñas. Estos efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar de la serie de datos, a este proceso se le llama desestacionalización de la serie.
- 3) **Componente aleatoria.** Esta componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

De los tres componentes anteriores los dos primeros son componentes determinísticos, mientras que la última es aleatoria.

Los modelos que se utilizan con más frecuencia son:

- **Modelo aditivo:**  $X_t = T_t + E_t + \epsilon_t$
- **Modelos multiplicativos:**
  - *Puro:*  $X_t = T_t \times E_t \times \epsilon_t$
  - *Mixto:*  $X_t = T_t \times E_t + \epsilon_t$

La elección de uno de estos modelos se hará de manera que el modelo seleccionado sea capaz de agrupar las principales características observadas en el gráfico de la serie en estudio.

#### 0.4.1. El Modelo Aditivo de Componentes de Series de Tiempo

Dada una serie  $X_t, t = 1, \dots, n$ , el *Modelo Aditivo de Componentes* consiste en asumir que  $X_t$  se puede descomponer en tres componentes:

$$X_t = T_t + E_t + \epsilon_t \quad (0.8)$$

donde  $T_t$  es la componente de tendencia,  $E_t$  es la componente estacional y  $\epsilon_t$  es la componente aleatoria o de errores. Las componentes  $T_t$  y  $E_t$  son funciones de  $t$  determinísticas. Su evolución es perfectamente predecible.

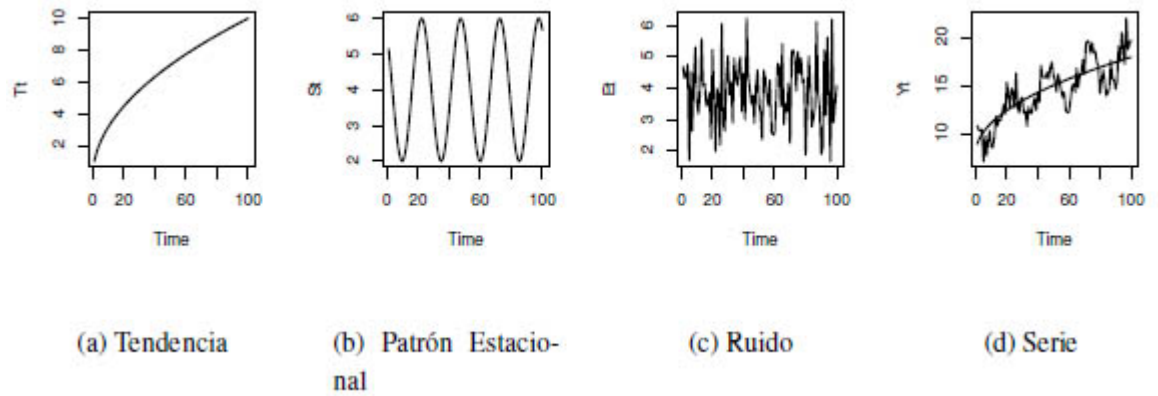
Este modelo es apropiado cuando la magnitud de la fluctuaciones estacionales de la serie no varía al hacerlo la tendencia.

La componente  $T_t$  en algunos casos también puede ser una componente estacional, pero de baja frecuencia, o, equivalentemente, una componente con período muy grande. Por ejemplo, en una serie diaria,  $E_t$  puede tener período 30 días, y  $T_t$  período 360 días.

En la Figura ?? se muestra la idea de la descomposición. Al superponer las series en los gráficos (a), (b) y (c) se obtiene la serie en el gráfico (d).

Asumiendo el modelo aditivo, el análisis de series de tiempo consiste en modelar y estimar  $T_t$  y  $E_t$  y luego extraerlas de  $X_t$  para obtener  $\hat{\epsilon}_t = X_t - \hat{T}_t - \hat{E}_t$ . La serie  $\hat{\epsilon}_t$  se modela y estima para finalmente reconstruir  $X_t$ ,  $\hat{X}_t = \hat{T}_t + \hat{E}_t + \hat{\epsilon}_t$ , y poder realizar el pronóstico  $\hat{X}_{t+h} = \hat{T}_{t+h} + \hat{E}_{t+h} + \hat{\epsilon}_{t+h}$ , utilizando la información disponible  $X_t, \dots, X_n$  con  $h = 1, 2, \dots, m$ . Sin embargo, puede suceder que la serie  $\hat{\epsilon}_t$  sea incorrelacionada, es decir,  $Corr(\hat{\epsilon}_t, \hat{\epsilon}_{t+s}) = 0$ , para  $s \neq 0$ . En este caso  $\hat{\epsilon}_{t+h} = 0$  para todo  $h > 0$ .

En **R** podemos descomponer una serie de tiempo usando la función *stl()* o la función *decompose()*. Retomando la serie de beneficios trimestrales de las acciones de Johnson y Johnson



**Figura 7:** Modelo aditivo de series de tiempo

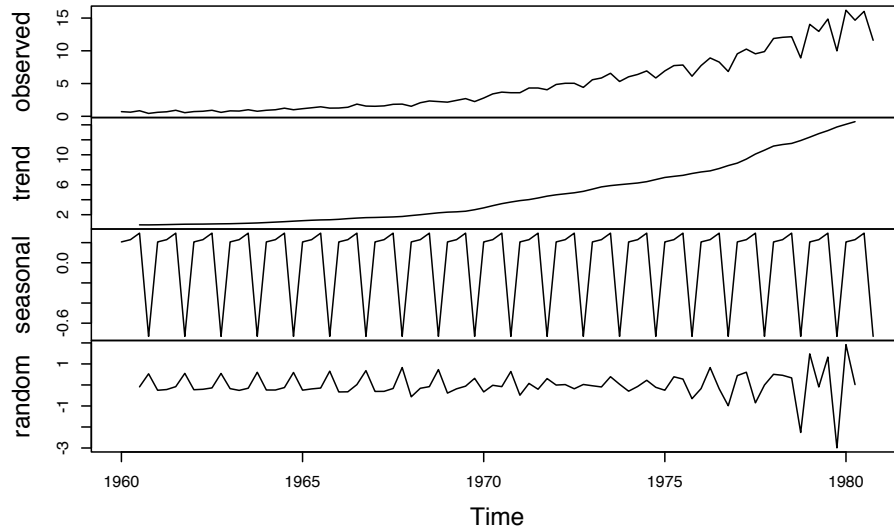
(Ejemplo 0.1) podemos observar la descomposición de la misma. En la parte superior de la gráfica se observa la serie original, en el gráfico siguiente la estacionalidad, en el tercero la tendencia y en el gráfico inferior los residuales.

```
plot(decompose(jj, type = "additive", filter = NULL))
```

La función *stl()* es más sofisticada que *decompose()*, la misma usa la descomposición de estacionalidad y tendencia de Loess (Seasonal and Trend decomposition using Loess) el cual es un método robusto y versátil para la descomposición de series de tiempo. El método STL fue desarrollado por Cleveland et al. (1990). A continuación mostramos la misma serie de beneficios de acciones de Johnson y Johnson usando esta función.

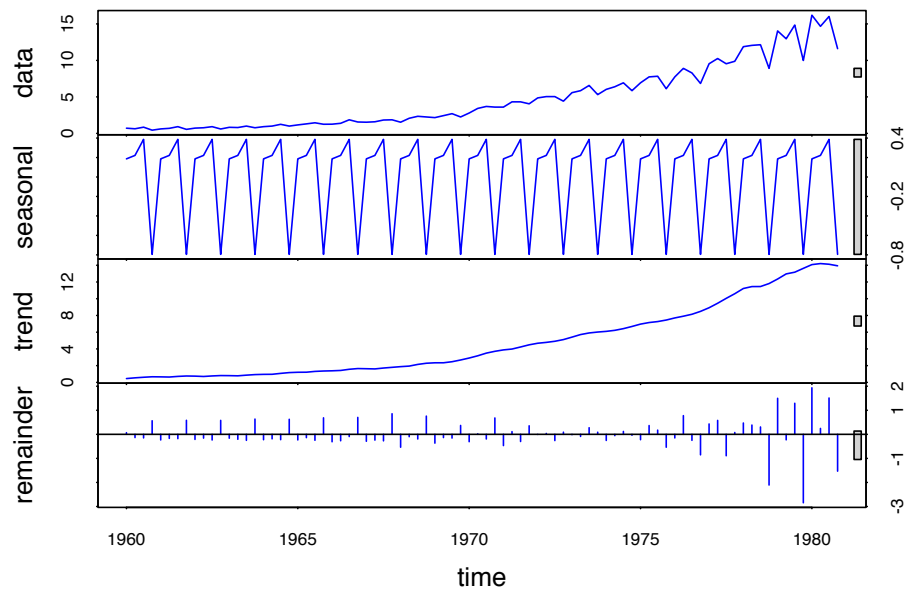
```
plot(stl(jj,s.window="periodic"), col="blue",
     main="Descomposicion de la serie Johnson y Johnson")
```

### Decomposition of additive time series



**Figura 8:** Descomposición aditiva de la serie Johnson y Johnson

### Descomposicion de la serie Johnson y Johnson



**Figura 9:** Descomposición de la serie Johnson y Johnson usando la descomposición de Loess (STL)

### 0.4.2. El Modelo Multiplicativo de Componentes de Series de Tiempo

Dada una serie de tiempo  $X_t, t = 1, \dots, n$ , el *Modelo Multiplicativo de Componentes* consiste en asumir que  $X_t$  se puede descomponer de una de las siguientes maneras:

- *Puro*:

$$X_t = T_t \times E_t \times \epsilon_t \quad (0.9)$$

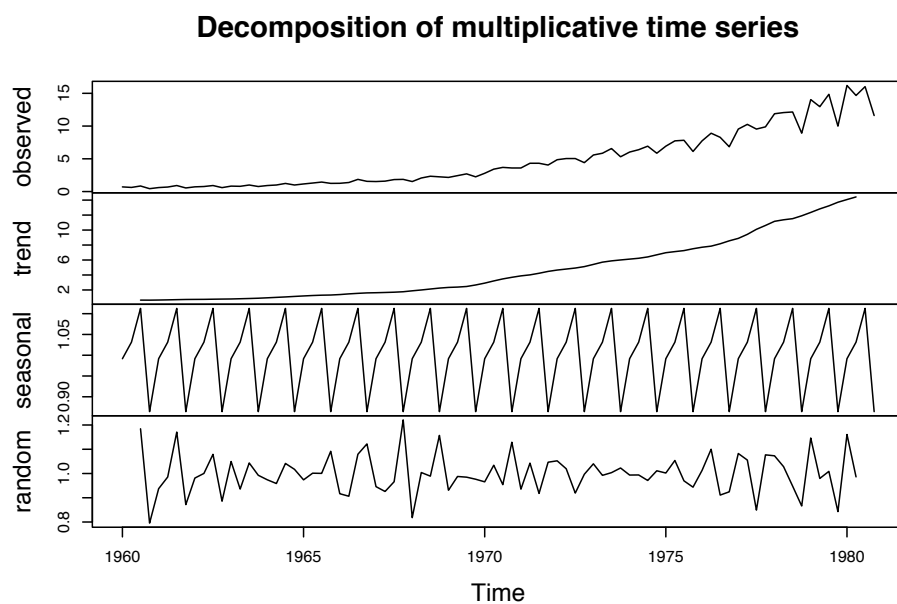
- *Mixto*:

$$X_t = T_t \times E_t + \epsilon_t \quad (0.10)$$

donde  $T_t$  es la componente de tendencia,  $E_t$  es la componente estacional y  $\epsilon_t$  es la componente aleatoria o de errores. Estos modelos son apropiados cuando la magnitud de las fluctuaciones estacionales de la serie crece y decrece proporcionalmente con los crecimientos y decrecimientos de la tendencia respectivamente.

Usamos la misma función *decompose()* para realizar la descomposición multiplicativa de la serie de tiempo, para ello en ‘type’ cambiamos “additive” por “multiplicative”

```
plot(decompose(jj, type = "multiplicative", filter = NULL))
```



**Figura 10:** Descomposición multiplicativa de la serie Johnson y Johnson



---

# 0

---

## *Características de series de tiempo*

---

El objetivo primario en el análisis de Series de Tiempo es desarrollar modelos matemáticos que provean una descripción apropiada para los datos muestrales, como los vistos en los ejemplos del capítulo anterior. Así, lo primero que hacemos es utilizar la definición 0.7, para tener un soporte estadístico. En este capítulo daremos algunas definiciones que serán de uso general en todo el resto del libro, también se describirán algunos métodos para el análisis exploratorio de las series de tiempo

---

### 0.5. Medidas de dependencia para series de tiempo

**Definición 0.9.** Un **proceso estocástico** es una familia de variables aleatorias indexadas  $x(\omega, t)$  ó  $x_t(\omega)$  donde  $t$  pertenece a un conjunto de índices  $T$  y  $\omega$  pertenece a un espacio muestral  $\Omega$ . Si  $t = t^*$  fijo,  $x(\omega, t^*)$  es una variable aleatoria. Si  $\omega = \omega^*$  fijo,  $x(\omega^*, t)$  es una función de  $t$ , y se llama una realización del proceso. Una **serie de tiempo** es la realización de un proceso estocástico.

Una descripción completa de una serie de tiempo, observada como una colección de  $n$  variables aleatorias en puntos de tiempo enteros arbitrarios  $t_1, t_2, \dots, t_n$ , para cada entero positivo  $n$ , es proporcionada por la función de distribución conjunta, evaluada como la probabilidad de que los valores de la serie sean conjuntamente menor que  $n$  constantes  $c_1, c_2, \dots, c_n$ , esto es

$$F(c_1, c_2, \dots, c_n) = P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_n} \leq c_n). \quad (0.11)$$

Desafortunadamente, la función de distribución multidimensional usualmente no se puede escribir fácilmente a menos que las variables aleatorias tengan distribución normal conjunta, en cuyo caso, la ecuación (0.11) llega a ser la distribución normal multivariada usual.

Un caso particular en la cual la función de distribución multidimensional es fácil de escribir, será en el caso de variables aleatorias normal estándar independientes e idénticamente distribuidas, para lo cual la función de distribución se puede expresar como el producto de las distribuciones marginales, es decir,

$$F(c_1, c_2, \dots, c_n) = \prod_{t_1}^n \Phi(c_t) \quad (0.12)$$

donde

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \mathbb{E}xp \left\{ -\frac{z^2}{2} \right\} dz \quad (0.13)$$

es la función de distribución normal estándar acumulada.

Aunque la función de distribución multidimensional describa los datos completamente, esto es un instrumento poco manejable para mostrar y analizar datos de series de tiempo. La función de distribución (0.11) debe ser evaluada como una función de  $n$  argumentos, entonces cualquier graficación de las correspondientes funciones de densidad multivariante es prácticamente imposible.

La función de distribución unidimensional

$$F_t(x) = P\{x_t \leq x\}$$

o la correspondiente función de densidad unidimensional

$$f_t(x) = \frac{\partial F_t(x)}{\partial x},$$

cuando existen, a menudo son más útiles para determinar si una

coordenada en particular de la serie de tiempo tiene una función de densidad conocida, como la distribución normal (gaussiana), por ejemplo.

**Definición 0.10.** La **función de media** es definida como

$$\mu_{xt} = \mathbb{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx, \quad (0.14)$$

en caso de que exista, donde  $\mathbb{E}$  denota el operador usual de esperanza. Cuando no haya confusión sobre a que serie de tiempo nos referimos, escribiremos  $\mu_{xt}$  como  $\mu_t$ .

Lo importante de comprender sobre  $\mu_t$  consiste en que es una media teórica para la serie de tiempo en un punto particular, donde la media se asume o calcula sobre todos los posibles eventos que podrían haber producido  $x_t$ .

**Definición 0.11.** La **función de autocovarianza** es definida como producto del segundo momento

$$\gamma_x(s, t) = \mathbb{E}[(x_s - \mu_s)(x_t - \mu_t)], \quad (0.15)$$

para todo  $t$  y  $s$ . cuando no haya confusión en la existencia sobre a que serie nos referimos, escribiremos  $\gamma_x(s, t) = \gamma(s, t)$ .

Note que  $\gamma_x(s, t) = \gamma_x(t, s)$  para todo los puntos  $s$  y  $t$ . La función de autocovarianza mide la dependencia lineal entre dos puntos de la misma serie en diferentes tiempos. La autocovarianza (0.15) es el promedio de los productos cruzados relacionado con la densidad conjunta  $F(x_s, x_t)$ . Es claro que, para  $s = t$ , la autocovarianza se reduce a la varianza (en el caso finito), dado que

$$\gamma_x(t, t) = \mathbb{E}[(x_t - \mu_t)^2] \quad (0.16)$$

Otro función de medida de tendencia importante es la *función de autocorrelación*.

**Definición 0.12.** La **función de autocorrelación (ACF)** (ACF, siglas en ingles: Autocorrelation Function) se define como

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}} \quad (0.17)$$

La *ACF* mide la predictibilidad lineal de una serie de tiempo en tiempo  $t$ , digamos  $x_t$  usando solo el valor  $x_s$ . Es fácil de demostrar que  $-1 \leq \rho(s, t) \leq 1$  usando la desigualdad de Cauchy-Schwarz <sup>4</sup>

Si podemos predecir  $x_t$  exactamente de  $x_s$  a través de la relación lineal  $x_t = \beta_0 + \beta_1 x_s$  entonces la correlación será 1 cuando  $\beta_1 > 0$  y  $-1$  cuando  $\beta_1 < 0$ .

**Definición 0.13.** La **función de covarianza cruzada** entre dos series  $x_t$  y  $y_t$  se define como

$$\gamma_{xy}(s, t) = \mathbb{E}[(x_s - \mu_{xs})(y_t - \mu_{yt})] \quad (0.18)$$

**Definición 0.14.** La **función de correlación cruzada (CCF)** (CCF, siglas en ingles: Cross Correlation Function) es definida como

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}} \quad (0.19)$$

Las definiciones anteriores de funciones de media y varianza son completamente generales. Aunque nosotros no hayamos hecho ninguna suposición especial sobre el comportamiento de las series de tiempo, muchos de los ejemplos precedentes han insinuado que puede existir una especie de regularidad en el comportamiento de las mismas. Introducimos la noción de regularidad que usa el concepto de *estacionaridad*, que ya hemos introducido empíricamente en el apartado 1.2.1 “Clasificación de las series de tiempo”

Formalmente tenemos las siguientes definiciones de estacionaridad

---

<sup>4</sup>Note que la desigualdad de Cauchy-Schwartz implica  $|\gamma(s, t)|^2 \leq \gamma(s, s)\gamma(t, t)$ .

**Definición 0.15.** Una serie de tiempo **estrictamente estacionaria** es una serie para la cual el comportamiento probabilístico de cada sucesión de valores

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$$

es idéntico a la serie trasladada en el tiempo

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$$

Esto es,

$$P[X_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k] = P[X_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k] \quad (0.20)$$

para todo  $k = 1, 2, \dots$ , todo puntos de tiempos  $t_1, t_2, \dots, t_k$  y números  $c_1, c_2, \dots, c_k$  y todo salto  $h = \pm 0, \pm 1, \pm 2, \dots$

Esta definición de estacionaridad es muy fuerte para la mayoría de las aplicaciones prácticas. Por ello necesitamos una versión menos fuerte que imponga menos condiciones sobre las distribuciones de probabilidad, ya que si observamos bien la ecuación (0.20), lo que nos dice la misma es que todas las posibles distribuciones de probabilidad deben ser iguales, lo que como ya indicamos en la práctica es muy difícil de comprobar aún para conjuntos de datos sencillos. La siguiente versión de estacionaridad solo impone condiciones sobre los dos primeros momentos de la serie

**Definición 0.16.** Una serie de tiempo **débilmente estacionaria**  $x_t$ , es un proceso de varianza finita tal que

- 1) la función de media  $\mu_t$  es constante y no depende del tiempo  $t$ ,
- 2) la función de covarianza  $\gamma(t, s)$  depende solo de las diferencias de  $s$  y  $t$ ,  $|t - s|$ .

Por consiguiente, usaremos el término **estacionaridad** para referirnos a estacionaridad débil; si un proceso es estacionario en el sentido estricto usaremos el término *estrictamente estacionario*.

*Nota.* 1) Si una serie de tiempo es estrictamente estacionaria, entonces todas las funciones de distribución multivariadas para subconjuntos de variables deben coincidir con sus contrapartes en el conjunto trasladado, para todos los valores del parámetro  $h$ . Por ejemplo para  $k = 1$  La ecuación (0.20) implica que

$$P\{x_s \leq c\} = P\{x_t \leq c\} \quad (0.21)$$

para cada puntos  $s$  y  $t$ .

Esta declaración implica, por ejemplo, que si la probabilidad de un valor de una serie de tiempo muestreada cada hora es negativa a la 1:00a.m, la probabilidad a la 10:00a.m. es la misma. Además, si la función de media,  $\mu_t$  de la serie  $x_t$  existe, (0.21) implica que  $\mu_s = \mu_t$  para todo  $s$  y  $t$ , y por consiguiente  $\mu_t$  debe ser constante.

2) Cuando  $k = 2$ , podemos escribir la ecuación (0.20) como

$$P\{x_s \leq c_1, x_t \leq c_2\} = P\{x_{s+h} \leq c_1, x_{t+h} \leq c_2\} \quad (0.22)$$

para cada par de puntos  $s$  y  $t$  y salto  $h$ . Entonces, si la función de varianza del proceso existe, (0.22) implica que la función de autocovarianza de la serie  $x_t$  satisface  $\gamma(s, t) = \gamma(s + h, t + h)$  para todos  $s$  y  $t$  y salto  $h$ .

Podemos interpretar este resultado diciendo que la función de autocovarianza del proceso depende sólo de las diferencias de tiempo entre  $s$  y  $t$ , y no del tiempo actual.

Es claro de la definición 0.15 de serie estrictamente estacionaria, que una serie de tiempo estrictamente estacionaria con varianza finita, también es una serie estacionaria. El recíproco no es cierto a menos que impongamos condiciones adicionales. Un importante caso donde estacionaridad implica estricta estacionaridad es el caso de series de tiempo gaussianas.

Ya que la función de media  $\mathbb{E}(x_t) = \mu_t$  de una serie de tiempo estacionaria es independiente del tiempo  $t$ , escribimos

$$\mu_t = \mu \quad (0.23)$$

Debido a que la función de covarianza de una serie de tiempo estacionaria,  $\gamma(s, t)$  en tiempos  $s$  y  $t$  depende sólo de la diferencia  $|s - t|$ , podemos simplificar la notación. Sea  $s = t + h$ , donde  $h$  representa el tiempo de traslación o salto, entonces

$$\begin{aligned} \gamma(s, t) &= \mathbb{E}[(x_{t+h} - \mu)(x_t - \mu)] \\ &= \mathbb{E}[(x_h - \mu)(x_0 - \mu)] \\ &= \gamma(h, 0) \end{aligned} \quad (0.24)$$

no depende del argumento de tiempo  $t$ ; asumiendo que  $\text{Var}(x_t) = \gamma(0, 0) < \infty$ . De ahora en adelante, por conveniencia, prescindiremos del segundo argumento de  $\gamma(h, 0)$ , es decir, la función de covarianza se denotará  $\gamma(h)$ .

**Definición 0.17.** La **función de autocovarianza de una serie de tiempo estacionaria** se escribirá como

$$\gamma(h) = \mathbb{E}[(x_{t+h} - \mu)(x_t - \mu)] \quad (0.25)$$

**Definición 0.18.** La **función de autocorrelación (ACF) de una serie de tiempo estacionaria** será escrita, usando (0.17) como

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)} \quad (0.26)$$

La desigualdad de Cauchy-Schwartz muestra nuevamente que  $-1 \leq \rho(h) \leq 1$  para todo  $h$ .

**\*\* Propiedades de la función de covarianza\*\***

1) Para el valor en  $h = 0$ , la función de autocovarianza

$$\gamma(0) = \mathbb{E}[(x_t - \mu)^2] \quad (0.27)$$

es la varianza de la serie de tiempo; note que la desigualdad de Cauchy-Schwartz implica que  $|\gamma(h)| \leq \gamma(0)$ .

- 2) La autocovarianza de una serie estacionaria es simétrica respecto al origen, esto es

$$\gamma(h) = \gamma(-h) \quad (0.28)$$

para todo  $h$ . Esta propiedad se debe a que trasladar la serie por  $h$  significa que

$$\begin{aligned} \gamma(h) &= \gamma(t+h-t) \\ &= \mathbb{E}[(x_{t+h} - \mu)(x_t - \mu)] \\ &= \mathbb{E}[(x_t - \mu)(x_{t+h} - \mu)] \\ &= \gamma(t - (t+h)) \\ &= \gamma(-h) \end{aligned}$$

lo cual muestra como usar la notación para demostrar el resultado.

**Definición 0.19.** Dos series de tiempo  $x_t$  y  $x_s$  se dice que son **conjuntamente estacionarias** si cada una de ellas es estacionaria y la función de correlación cruzada

$$\gamma_{xy}(h) = \mathbb{E}[(x_{t+h} - \mu_x)(y_t - \mu_y)] \quad (0.29)$$

es una función sólo del salto  $h$ .

**Definición 0.20.** La **función de correlación cruzada (CCF)** de dos series conjuntamente estacionarias  $x_t$  y  $y_t$  se define como

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}} \quad (0.30)$$

De nuevo, tenemos el resultado  $-1 \leq \rho_{xy}(h) \leq 1$  lo cual nos permite comparar los valores extremos -1 y 1 cuando vemos la relación entre  $x_{t+h}$  y  $y_t$ . La función de correlación cruzada satisface

$$\rho_{xy}(h) = \rho_{yx}(-h) \quad (0.31)$$

lo cual se puede demostrar de manera similar que para (0.28).



**Ejemplo 0.8** (Estacionaridad conjunta). Considere las series  $x_t$  y  $y_t$  formadas por las sumas y diferencias de dos valores sucesivos de un ruido blanco respectivamente, esto es

$$x_t = w_t + w_{t-1}$$

y

$$y_t = w_t - w_{t-1}$$

donde  $w_t$  son variables aleatorias independientes con media cero y varianza  $\sigma_w^2$ . Es fácil demostrar que  $\gamma_x(0) = \gamma_y(0) = 2\sigma_w^2$  y  $\gamma_x(1) = \gamma_x(-1) = \sigma_w^2$ ,  $\gamma_y(1) = \gamma_y(-1) = -\sigma_w^2$ . También

$$\begin{aligned}\gamma_{xy}(1) &= \mathbb{E}[(x_{t+1} - 0)(y_t - 0)] \\ &= \mathbb{E}[(w_{t+1} + w_t)(w_t - w_{t-1})] \\ &= \sigma_w^2\end{aligned}$$

porque solo uno de los productos es distinto de cero. Similarmente,  $\gamma_{xy}(0) = 0$ ,  $\gamma_{xy}(-1) = -\sigma_w^2$ . Usando (??), obtenemos

$$\rho_{xy}(h) = \begin{cases} 0, & h = 0 \\ 1/2, & h = 1 \\ -1/2, & h = -1 \\ 0, & |h| \geq 2 \end{cases}.$$

Claramente, las funciones de autocovarianza y correlación cruzada dependen solo del salto  $h$ , por lo tanto las series son conjuntamente estacionarias.

El concepto de estacionaridad débil forma la base para muchos de los análisis realizados con series de tiempo. Las propiedades fundamentales de la media (0.23) y la función de covarianza (0.25) son satisfechas por muchos modelos teóricos que aparecen para generar realizaciones muestrales apropiadas.

**Definición 0.21.** Un **proceso lineal**  $x_t$  se define como una combinación lineal de variables aleatorias de ruido blanco  $w_t$ , y está dado por

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j} \quad (0.32)$$

donde los coeficientes satisfacen

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (0.33)$$

Para un proceso lineal, podemos demostrar que la función de autocovarianza está dada por

$$\gamma(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h} \psi_j \quad (0.34)$$

para todo  $h \geq 0$ ; recuerde que  $\gamma(-h) = \gamma(h)$ . Finalmente como mencionamos anteriormente, un caso importante en el cual una serie débilmente estacionaria es también estrictamente estacionaria es la serie normal o gaussiana.

**Definición 0.22.** Un proceso  $\{x_t\}$ , se dice que es un **proceso gaussiano** si el  $k$ -ésimo vector dimensional  $\hat{x} = (x_{t_1}, x_{t_2}, \dots, x_{t_k})$ , para cada conjunto de puntos  $t_1, t_2, \dots, t_k$  y cada entero positivo  $k$  tiene distribución normal multivariada.

Definiendo  $k \times 1$  vector de medias  $\hat{\mu} = (\mu_{t_1}, \mu_{t_2}, \dots, \mu_{t_k})'$  y la  $k \times k$  matriz de covarianza positiva como  $\Gamma = \{\gamma(t_i, t_j); i, j = 1, \dots, k\}$ , la función de densidad normal multivariada se puede escribir como

$$f(\hat{x}) = (2\pi)^{-k/2} |\Gamma|^{-1/2} \exp \left\{ -\frac{1}{2} (\hat{x} - \hat{\mu})' \Gamma^{-1} (\hat{x} - \hat{\mu}) \right\} \quad (0.35)$$

donde  $|\cdot|$  denota el determinante. Esta distribución forma la base para resolver problemas que envuelven inferencia estadística para

series de tiempo. Si una serie de tiempo gaussiana  $\{x_t\}$  es débilmente estacionaria, entonces  $\mu_t = \mu$  y  $\gamma(t_i, t_j) = \gamma(|t_i - t_j|)$ , de modo que el vector  $\hat{\mu}$  y la matriz  $\Gamma$  son independientes del tiempo. Este hecho implica que todas las distribuciones finitas, (0.35) de la serie  $\{x_t\}$  dependen sólo del salto de tiempo y no del tiempo actual, y por consiguiente la serie debe ser estrictamente estacionaria.

---

## 0.6. Estimación de la Tendencia

En esta sección introducimos la estimación de la tendencia. En esencia, existen dos métodos para estimar la tendencia y la componente estacional de una serie de tiempo:

- **Método paramétrico:** Se basa en
  - Proponer modelos paramétricos para expresar la relación que guardan la tendencia y la componente estacional con el tiempo.
  - Ajustar dichos modelos a la serie de tiempo (por ejemplo, a través del método de mínimos cuadrados).
  - Aislar la tendencia y la componente estacional por medio de los modelos ajustados.
- **Método no paramétrico:** Se basa en
  - Asumir “suavidad” en la relación que guardan la tendencia y la componente estacional con el tiempo.
  - Aislar la tendencia y la componente estacional a través de la suavización del gráfico de la serie (aplicando, por ejemplo, filtros de promedios móviles).

Hay otros métodos que no consideraremos en este libro, por ejemplo, *wavelets*. En ocasiones la expresión “suavizar una serie” es equivalente a “extracción de la tendencia de una serie”, y ambas equivalen a la estimación de la tendencia.

A continuación presentamos una lista de posibles modelos para la tendencia  $T_t$ :

- Lineal

$$T_t = \beta_0 + \beta_1 t \quad (0.36)$$

- Cuadrático

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad (0.37)$$

- Cúbico

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \quad (0.38)$$

- Exponencial

$$T_t = \exp(\beta_0 + \beta_1 t) \quad (0.39)$$

- Logístico

$$T_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)} \quad (0.40)$$

En la tendencia cuadrática podemos observar:

- Si  $\beta_1, \beta_2 > 0$ ,  $T_t$  es monótona creciente.
- Si  $\beta_1, \beta_2 < 0$ ,  $T_t$  es monótona decreciente.
- Si  $\beta_1 > 0$  y  $\beta_2 < 0$ ,  $T_t$  es cóncava.
- Si  $\beta_1 < 0$  y  $\beta_2 > 0$ ,  $T_t$  es convexa.

Otro modelo propuesto para la tendencia es el dado por la siguiente definición.

**Definición 0.23.** El modelo **Logarítmico Lineal** o **Log-Lineal** se define como

$$\ln X_t = \beta_0 + \beta_1 t + \epsilon_t \quad (0.41)$$

El modelo anterior corresponde a un modelo con tendencia lineal para el logaritmo de  $X_t$ . En (0.41) al tomar exponencial se tiene  $X_t = \exp(\beta_0 + \beta_1 t + \epsilon_t)$ , que es similar al modelo con tendencia exponencial (0.39). Sin embargo, son modelos diferentes y se estiman por métodos diferentes.

Para la estimación de los parámetros  $\beta_0, \beta_1, \beta_2$  en los modelos lineales (0.36), (0.37), (0.38) y (0.41) utilizaremos el método de

mínimos cuadrados clásico (MCC). En este método los parámetros estimados son aquellos que producen el valor mínimo de la suma de errores cuadrados. Para los modelos (0.39) y (0.40) se usa el método de mínimos cuadrados no lineales, que también minimiza la suma de errores cuadrados.

El modelo Log-Lineal (0.41) es equivalente, algebraicamente, a

$$X_t = \exp(\beta_0 + \beta_1 t + \epsilon_t).$$

Sin embargo, este último modelo es no lineal y no coincide con el modelo exponencial, (0.39),  $X_t = \exp(\beta_0 + \beta_1 t) + \epsilon_t$ . Es posible estimar por mínimos cuadrados ordinarios el modelo Log-Lineal y utilizar los parámetros estimados  $\hat{\beta}_0, \hat{\beta}_1$  como valores iniciales en la estimación del modelo exponencial por mínimos cuadrados no lineales. Pero los parámetros estimados en ambos modelos no necesariamente coinciden.

Aunque la serie tenga una componente estacional  $E_t$ ,  $X_t = T_t + E_t + \epsilon_t$ , solamente consideramos un modelo de regresión entre  $X_t$  y  $T_t$ , tal que  $X_t = T_t + \eta_t$ , donde  $\eta_t$  es el término de error, de forma que  $\eta_t = E_t + \epsilon_t$ . Por ejemplo,

1. En el caso lineal  $T_t = \beta_0 + \beta_1 t$ , ajustamos el modelo de regresión lineal:  $X_t = \beta_0 + \beta_1 t + \eta_t$ .
2. En el caso cuadrático  $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ , ajustamos el modelo de regresión cuadrático  $X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \eta_t$ . Nótese que en este caso hay que definir una variable explicativa adicional  $t^2$ .

En general, para que datos de series de tiempo sean estacionarias, es necesario hacer un promedio de productos en el tiempo. Como para datos de serie de tiempo es importante medir la dependencia entre los valores de la serie; al menos, debemos ser capaces de estimar las autocorrelaciones con precisión. Será difícil medir la dependencia de estos valores si la estructura de dependencia no es regular o si cambia en el tiempo. De ahí, que para realizar cualquier análisis estadístico significativo de datos de series de

tiempo, será crucial que las funciones de media y autocovarianza satisfagan las condiciones de estacionaridad dadas en la Definición 0.16. A menudo, este no es el caso, y en esta sección daremos algunos métodos para lidiar con los efectos de no-estacionaridad sobre las propiedades estacionarias de las series a estudiar.

Quizás la forma más fácil de trabajar con series no-estacionarias es el modelo de tendencia estacionaria donde el proceso tiene comportamiento estacionario alrededor de una tendencia.

Podemos escribir este tipo de modelos como

$$X_t = T_t + Y_t \quad (0.42)$$

donde  $X_t$  son las observaciones,  $T_t$  denota la tendencia y  $Y_t$  es un proceso estacionario.

Por lo general, una tendencia fuerte  $T_t$  puede oscurecer el comportamiento del proceso estacionario  $Y_t$ , como veremos en ejemplos posteriores.

De aquí, será una ventaja el que podamos remover la tendencia como un primer paso para un análisis exploratorio de los datos.

Los pasos envuelven obtener un estimador razonable del componente de tendencia, llamémoslo  $\hat{T}_t$  y entonces trabajar con el residual

$$\hat{Y}_t = X_t - \hat{T}_t. \quad (0.43)$$

El primer paso en el análisis de cualquier tipo de serie es un gráfico de los datos.

- Si existe alguna aparente discontinuidad en la serie, tal como un cambio súbito en el nivel de la serie, esto puede darnos una idea para el análisis de la serie, un primer paso sería dividir la serie en segmentos homogéneos.
- Si existen observaciones o datos “*outliers*”, estos deben ser estudiados con cuidado para verificar si existe alguna justificación para descartar estas observaciones, como por ejemplo si una observación ha sido registrada de algún otro proceso por error.

- La inspección del gráfico también podría sugerir la representación de los datos como una realización de un proceso, como el modelo clásico de descomposición dado por (0.8).

Si la componente estacional y la componente aleatoria o ruido parecen incrementarse con el nivel del proceso entonces una transformación preliminar de los datos es a menudo usada para hacer que los datos transformados sean compatibles con el modelo (0.8). En esta sección discutiremos algunas técnicas para identificar y eliminar las componentes en (0.8).

Nuestro objetivo es estimar y extraer las componentes determinísticas  $T_t$  y  $E_t$  con la esperanza de que el residual o la componente aleatoria  $\epsilon_t$  llegue a ser un proceso estacionario. Entonces podremos usar la teoría de tales procesos para hallar un modelo probabilístico satisfactorio para el proceso  $\epsilon_t$ , analizar sus propiedades y usarlo en conjunto con  $T_t$  y  $E_t$  para hacer pronósticos y control de  $X_t$ .

Los dos enfoques para la eliminación de las componentes de tendencia y estacional son:

1. Estimación de  $T_t$  y  $E_t$  en el modelo (0.8),
2. Diferencia de los datos  $X_t$ .

Ilustraremos ambos enfoque con varios ejemplos

#### 0.6.1. Estimación de la tendencia en ausencia de estacionalidad

Si tenemos una serie de tiempo para la cual está ausente la componente estacional  $E_t$  el modelo (0.8) llega ser

$$X_t = T_t + \epsilon_t, \quad t = 1, \dots, n \quad (0.44)$$

donde, sin pérdida de generalidad, podemos suponer que  $\mathbb{E}(\epsilon_t) = 0$ . A continuación vamos a describir tres métodos para estimar la tendencia  $T_t$ .

1. **Método T1: Estimación de  $T_t$  por mínimos cuadrados.** El objetivo de este método es intentar ajustar una familia paramétrica de funciones como las vistas en las ecuaciones (0.36) a ??, a los datos eligiendo los parámetros que minimicen  $\sum_t (X_t - T_t)^2$ . Esto es, asumiendo que  $\mathbb{E}(\epsilon_t) = 0$ , se tiene  $\mathbb{E}(X_t) = T_t = f(t)$

Una suposición común es que la función  $f$  depende de ciertos parámetros (desconocidos)  $\beta_1, \dots, \beta_p$ , es decir,

$$f(t) = f(t; \beta_1, \dots, \beta_p) \quad (0.45)$$

Sin embargo, el *tipo* de función es conocida. Los parámetros  $\beta_1, \dots, \beta_p$  serán estimados a partir de una realización  $x_t$  de la variable aleatoria  $X_t$ . La aproximación por *estimación de mínimos cuadrados*  $\hat{\beta}_1, \dots, \hat{\beta}_p$  debe satisfacer

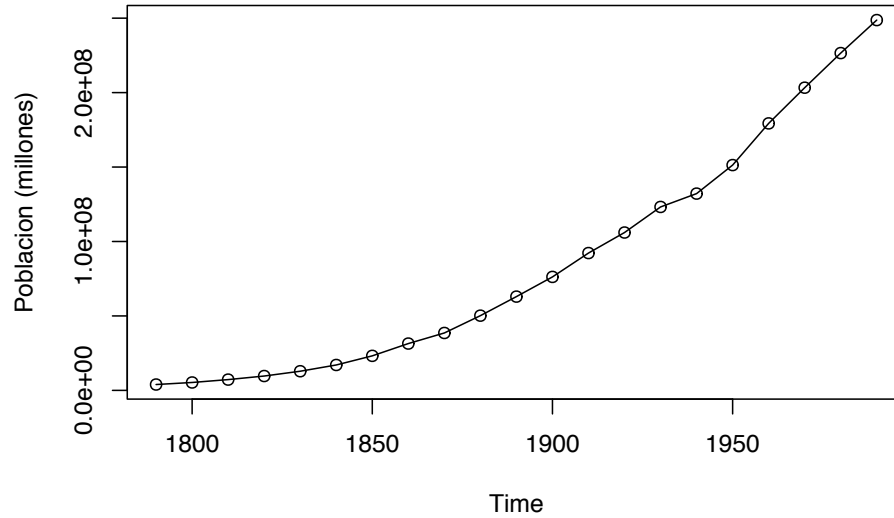
$$\sum_t (x_t - f(t; \hat{\beta}_1, \dots, \hat{\beta}_p))^2 = \min_{\beta_1, \dots, \beta_p} \sum_t (x_t - f(t; \beta_1, \dots, \beta_p))^2 \quad (0.46)$$

cuya solución, si existe, es un problema numérico. El valor  $\hat{x}_t = f(t; \hat{\beta}_1, \dots, \hat{\beta}_p)$  servirá como una *predicción* de futuros valores  $x_t$ . Las diferencias observadas  $x_t - \hat{x}_t$  son llamadas *residuales*. Ellas contienen información sobre la bondad de ajuste del modelo a los datos.

**Ejemplo 0.9.** El archivo “USPOP.txt” contiene la información de la población de Estados Unidos de América desde 1780 hasta 1980 segun el censo poblacional cada 10 años. En el gráfico podemos observar que no existe estacionalidad, por lo que podemos aplicar el método descrito para ajustar la tendencia.

```
uspop=ts(scan("data/USPOP.txt"),frequency=1/10,start=1790)
pop=window(uspop,start=1790)
plot(pop,type="o",ylab="Poblacion (millones)")
```





Podemos notar del gráfico que la tendencia es creciente y parece tener un comportamiento cuadrático, por lo que ajustando una función de la forma (0.37) para la población de los datos USPOP para  $1790 \leq t \leq 1980$  nos da los parámetros estimados

$$\hat{a}_0 = 2,101 \times 10^{10}; \quad \hat{a}_1 = -2,338 \times 10^7; \quad \hat{a}_2 = 6,506 \times 10^3$$

En el gráfico siguiente se puede observar la curva ajustada y los datos originales. Los valores estimados del proceso de ruido  $\epsilon_t$ ,  $1790 \leq t \leq 1980$ , son los residuales obtenidos por sustracción de  $\hat{T}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2$  de la serie  $X_t$ . La componente de tendencia  $T_t$  nos proporciona un predictor natural de los valores futuros de  $X_t$ .

Por ejemplo si deseamos estimar  $T_{1990}$  por su valor medio, obtenemos

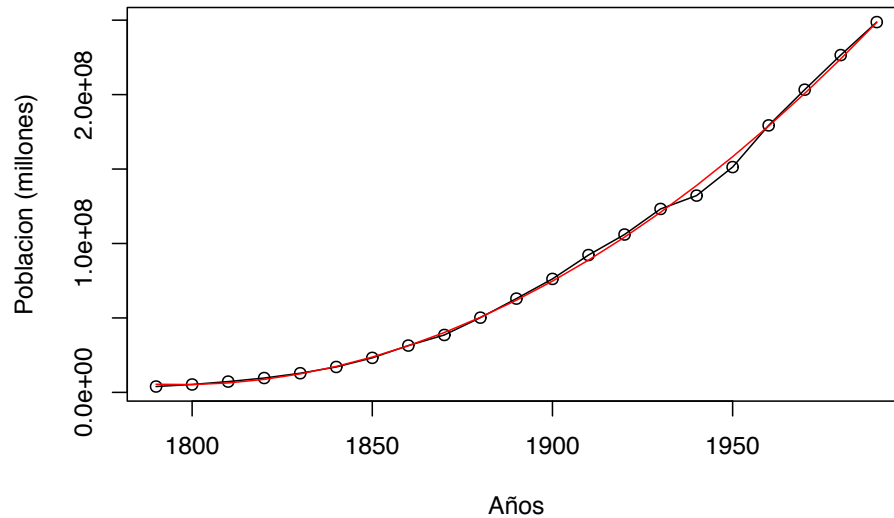
$$T_{1990} = 2,4853 \times 10^8$$

para la población de EE.UU en 1990. Sin embargo si los residuales  $\hat{\epsilon}_t$  están altamente correlacionados podemos ser capaces de usar esos valores para dar una mejor estimación de  $T_{1990}$  y por consiguiente de  $X_{1990}$ .

```
x=time(pop)
reg=lm(pop~x+I(x^2),na.action=NULL)
summary(reg)

##
## Call:
## lm(formula = pop ~ x + I(x^2), na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6947521 -358167  436285  1481410  3391761
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.10e+10   6.59e+08   31.9   <2e-16 ***
## x           -2.34e+07   6.98e+05  -33.5   <2e-16 ***
## I(x^2)        6.51e+03   1.85e+02   35.2   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2770000 on 18 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:  0.999
## F-statistic: 8.05e+03 on 2 and 18 DF, p-value: <2e-16

plot(pop,type="o",xlab="Años",ylab="Poblacion (millones)")
lines(reg$fitted.values,col="red")
```



**Ejemplo 0.10.** El archivo “Population-North-Rhine-Westphalia.txt” contiene la población de la región North-Rhine-Westphalia (Alemania) en millones cada 5 años desde 1935 hasta 1980. Observando el gráfico podemos suponer que la tendencia se puede ajustar por el modelo cúbico (0.38), esto es

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

El código en R para el gráfico y el ajuste es

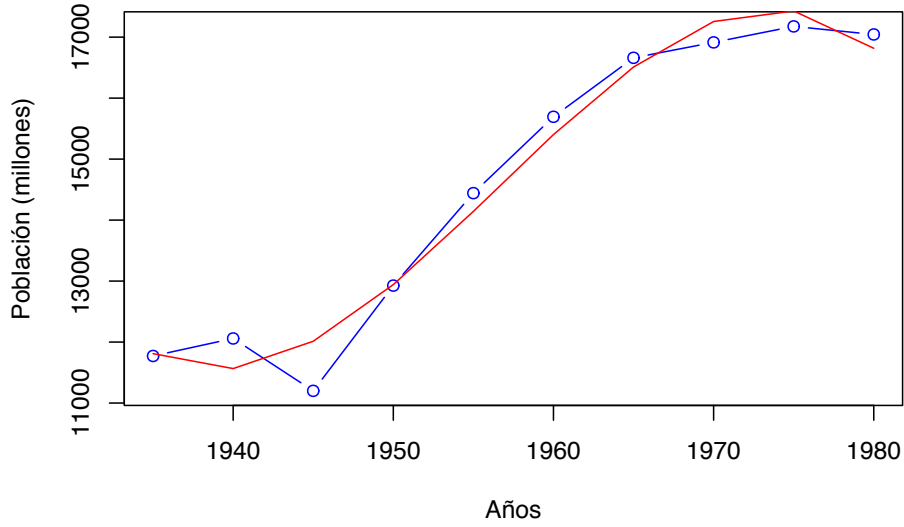
```
NRWpop=read.table("data/Population-North-Rhine-Westphalia.txt",
                  header = TRUE)
knitr::kable(head(NRWpop,booktabs=TRUE,
                  caption="Población (en millones) de North-Rhine-Westphalia")
```

Year	Population
1935	11772
1940	12059
1945	11200
1950	12926
1955	14442
1960	15694

```
plot(NRWpop, type = "b", col="blue", xlab = "Años", ylab = "Población (millones)")
# Modelo cúbico
t=NRWpop[,1]
pob=NRWpop[,2]
modelo=lm(pob~t+I(t^2)+I(t^3), na.action = NULL)
summary(modelo)
```

```
##
## Call:
## lm(formula = pob ~ t + I(t^2) + I(t^3), na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -813.0 -199.2   67.1  275.6  493.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.11e+09   5.10e+08   4.13   0.0061 **
## t           -3.23e+06   7.81e+05  -4.14   0.0061 **
## I(t^2)        1.65e+03   3.99e+02   4.14   0.0061 **
## I(t^3)       -2.81e-01   6.79e-02  -4.14   0.0061 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 472 on 6 degrees of freedom
## Multiple R-squared:  0.974, Adjusted R-squared:  0.962
## F-statistic: 76.2 on 3 and 6 DF, p-value: 3.63e-05
```

```
lines(t,modelo$fitted.values,col="red")
```



La curva punteada en azul corresponde a los datos originales, la curva en rojo corresponde al ajuste mediante el modelo cúbico.

2. **Método T2: Suavizado por medio de un promedio móvil.** Sea  $q$  un entero no negativo y consideremos un promedio móvil de la forma

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t+j} \quad (0.47)$$

de un proceso  $\{X_t\}$  definido por (0.44). Entonces para  $q+1 \leq t \leq n-q$ ,

$$\begin{aligned} W_t &= \frac{1}{2q+1} \sum_{j=-q}^q T_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^q \epsilon_{t+j} \\ &\simeq T_t \end{aligned} \quad (0.48)$$

suponiendo que  $T_t$  es aproximadamente lineal sobre el intervalo

$[t - q, t + q]$  y que el promedio del término de error sobre este intervalo es cercano a cero.

El promedio móvil entonces nos provee con el estimador

$$\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t+j}, \quad q+1 \leq t \leq n-q. \quad (0.49)$$

Dado que  $X_t$  es no observado para  $t \leq 0$  o  $t \geq n$  no podemos usar (0.49) para  $t \leq q$  o  $t > n - q$ . Una forma de resolver este problema es haciendo  $X_t = X_1$  para  $t < 1$  y  $X_t = X_n$  para  $t > n$ . A continuación presentamos un ejemplo

---

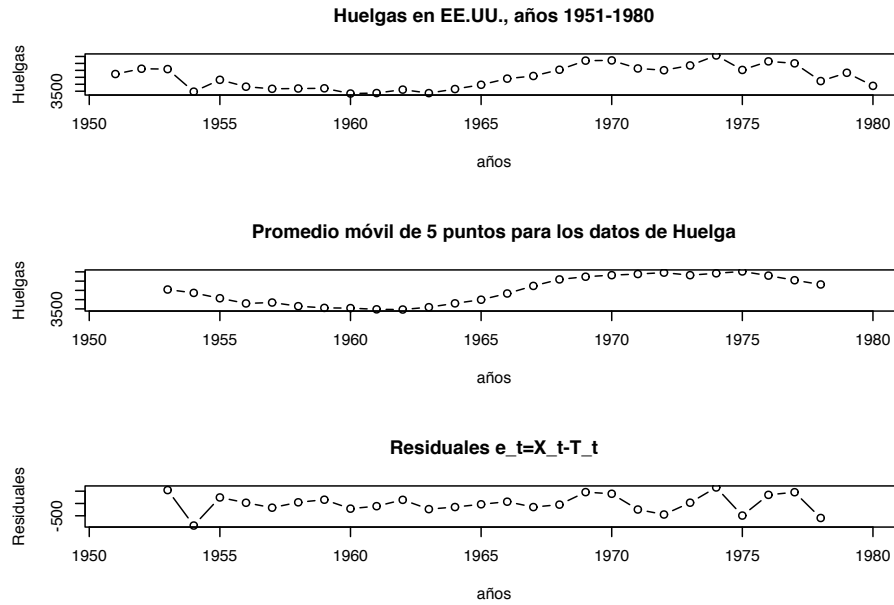
**Ejemplo 0.11.** El gráfico siguiente muestra las huelgas ocurridas en EE.UU, de 1951 a 1980, según la Oficina de Estadísticas Laborales del Departamento de Trabajo de los EE.UU.

A estos datos le aplicamos un promedio móvil de 5 puntos, la Figura muestra la serie suavizada y el término de error estimado  $\hat{\epsilon}_t = X_t - \hat{T}_t$  se muestra en la Figura ?? . Como era de esperarse ellos no presentan una tendencia clara.

Las instrucciones en R para el suavizado y los gráficos son los siguientes:

```
H=read.table("data/Huelgas.txt")
# Promedio móvil por medio de la función "filter"
W=filter(H[,2],sides=2,rep(1/5,5))
# Residuales de  $X_t$ 
y=H[,2]-W
# Graficos
par(mfrow=c(3,1))
plot(H,xlab="años",ylab="Huelgas",type='b',
      main = "Huelgas en EE.UU., años 1951-1980")
plot(H[,1],W,xlab="años",ylab="Huelgas",type='b',
      main = "Promedio móvil de 5 puntos para los datos de Huelga")
```

```
plot(H[,1],y,xlab="años",ylab="Residuales",type='b',
     main = "Residuales e_t=X_t-T_t")
```



Para cada valor fijo  $a \in [0, 1]$ , el promedio móvil de un lado

$\hat{T}_t, t = 1, \dots, n$ , definido por la recursión

$$\hat{T}_t = aX_t + (1 - a)\hat{T}_{t-1}, \quad t = 2, \dots, n \quad (0.50)$$

y

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

se puede calcular usando la opción *sides=1* en la función *filter* de R.

Es usual pensar como aplicación de la ecuación (0.50) como un suavizado exponencial, dado que se sigue de la recursión que para  $t \leq 2, \hat{T}_t = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$ , es un promedio

móvil con peso de  $X_t, X_{t-1}, \dots$ , con pesos decreciendo exponencialmente (excepto para el último término).

Es útil pensar en  $\{\hat{T}_t\}$  en (*filter*) como un proceso obtenido de  $\{X_t\}$  por aplicación de un operador lineal o filtro lineal  $\hat{T}_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}$  con pesos  $a_j = (2q+1)^{-1}$ ,  $-q \leq j \leq q$ , y  $a_j = 0$ ,  $|j| > q$ . Este filtro particular es un filtro de “paso-bajo” ya que toma los datos  $\{X_t\}$  y remueve la componente de rápida fluctuación (o de alta frecuencia)  $\{\hat{\epsilon}_t\}$ , para dejar el término de la tendencia estimada de lenta variación  $\{\hat{T}_t\}$ .

3. **Método T3: Diferenciación para generar datos estacionarios.** En lugar de intentar remover el ruido por suavizado como en el Método T2, ahora intentaremos eliminar la tendencia por diferenciación. Definamos primero el operador diferencia  $\nabla$  por

$$\nabla x_t = x_t - x_{t-1} = (1 - B)x_t, \quad (0.51)$$

donde  $B$  es el operador de desplazamiento hacia atrás (*backward shift operator* en inglés),

$$Bx_t = x_{t-1}. \quad (0.52)$$

Las potencias de los operadores  $B$  y  $\nabla$  se definen de manera obvia, esto es,  $B^j(x_t) = x_{t-j}$  y  $\nabla^j(x_t) = \nabla(\nabla^{j-1}(x_t))$ ,  $j \geq 1$  con  $\nabla^0(x_t) = x_t$ . Los polinomios en  $B$  y  $\nabla$  se manipulan de la misma manera que las funciones polinómicas de variables reales. Por ejemplo

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

Si el operador  $\nabla$  se aplica a una función con tendencia lineal  $T_t = at + b$ , entonces obtenemos la función constante  $\nabla T_t = a$ . De



la misma manera cada tendencia polinomial de grado  $k$  se puede reducir a una constante por aplicación del operador  $\nabla^k$ .

Iniciando entonces con el modelo  $X_t = T_t + \epsilon_t$ , donde  $T_t = \sum_{j=0}^k a_j t^j$  y  $\epsilon_t$  es estacionario con media cero, obtenemos

$$\nabla^k X_t = k!a_k + \nabla^k \epsilon_t,$$

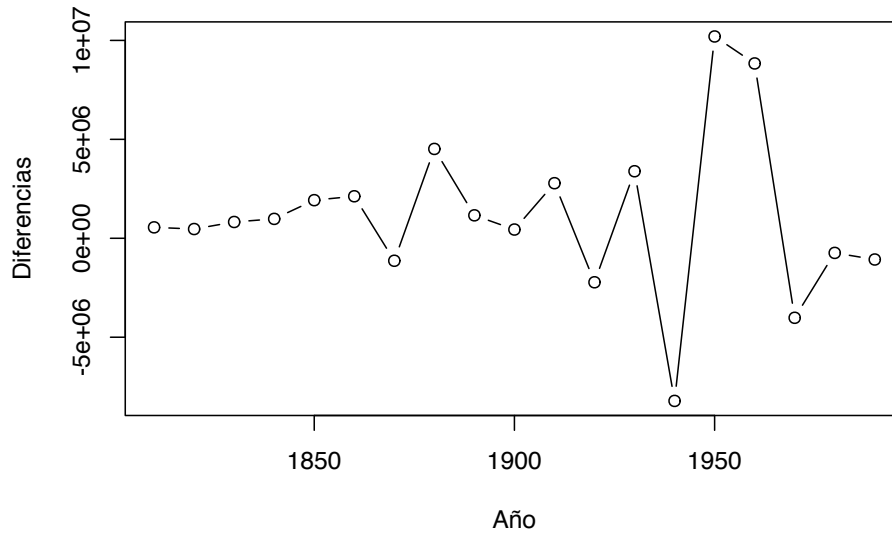
un proceso estacionario con media  $k!a_k$ . Esta consideración sugiere la posibilidad, dada una sucesión  $\{X_t\}$  de datos, de aplicar el operador  $\nabla$  repetidamente hasta conseguir una sucesión  $\{\nabla^k X_t\}$  la cual puede ser apropiadamente modelada como una realización de un proceso estacionario. Se encuentra a menudo en la práctica que el orden  $k$  de diferenciación es bastante pequeño, frecuentemente uno o dos.<sup>5</sup>

**Ejemplo 0.12.** Aplicando esta técnica al ejemplo 0.9 de población de los EE.UU, hallamos que dos operaciones de diferenciación son suficientes para producir una serie sin aparente tendencia. Los datos diferenciados se muestran en la Figura. Note que la magnitud de las fluctuaciones en  $\nabla^2 X_n$  se incrementa con el valor de  $n$ . Este efecto se puede suprimir tomando primero logaritmo natural,  $y_n = \ln X_n$  y entonces aplicando el operador  $\nabla^2$  a la serie  $\{y_n\}$ .

Las instrucciones en R son las siguientes

```
Dx=diff(uspop,difference=2)
plot(Dx,type="b",xlab="Año", ylab="Diferencias")
```

<sup>5</sup>Esto depende del hecho de que muchas funciones pueden ser aproximadas bastante bien, en un intervalo de longitud finita, por un polinomio de grado razonablemente bajo.



### 0.6.2. Estimación de la tendencia y la estacionalidad

Los métodos descritos para estimar y remover la tendencia pueden ser adaptados de manera natural para estimar tanto la tendencia como la estacionalidad en el modelo general

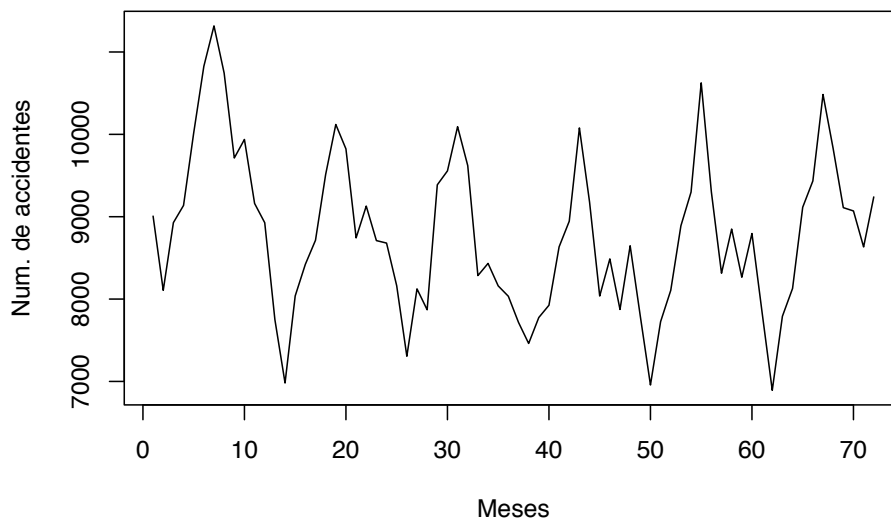
$$X_t = T_t + E_t + \epsilon_t \quad (0.53)$$

donde  $\mathbb{E}(\epsilon_t) = 0$ ,  $E_{t+d} = E_t$  y  $\sum_{j=1}^d E_t = 0$ . Ilustraremos estos métodos con referencia al siguiente ejemplo de accidentes. El archivo “Accidentes3.txt” muestra el número de accidentes mortales de automóviles mensual ocurridos en EE.UU., entre los años 1973 y 1978. En la tabla siguiente se muestran los datos

```
X<-read.table("data/Accidentes3.txt", header = TRUE)
```

Mes	X1973	X1974	X1975	X1976	X1977	X1978
Ene	9007	7750	8162	7717	7792	7836
Feb	8106	6981	7306	7461	6957	6892
Mar	8928	8038	8124	7776	7726	7791
Abr	9137	8422	7870	7925	8106	8129
May	10017	8714	9387	8634	8890	9115
Jun	10826	9512	9556	8945	9299	9434

En la figura podemos observar que los datos presentan claramente una componente estacional con periodo  $d = 12$ .



Será conveniente para el primer método indexar los datos por mes y año. Entonces  $X_{j,k}$ ,  $j = 1, \dots, 12$ ,  $k = 1, \dots, 6$  denotará el número de muertes accidentales reportados para el  $j$ -ésimo mes del  $k$ -ésimo año,  $(1972 + k)$ . En otras palabras, definimos

$$X_{j,k} = X_{j+12(k-1)}, \quad j = 1, \dots, 12; k = 1, \dots, 6.$$

- 1) **Método E1: Método de la tendencia pequeña.** Si la tendencia es pequeña (como en los datos de accidentes) no es irrazonable suponer que el término de la tendencia es constante, digamos  $T_k$  para el año  $k$ . Dado que  $\sum_{j=1}^{12} E_j = 0$ , nos lleva al estimador insesgado natural para la tendencia

$$\hat{T}_k = \frac{1}{12} \sum_{j=1}^{12} X_{j,k}, \quad (0.54)$$

mientras que para la estacionalidad  $E_j, j = 1, \dots, 12$  tenemos el estimador

$$\hat{E}_j = \frac{1}{6} \sum_{k=1}^6 (X_{j,k} - \hat{T}_k), \quad (0.55)$$

el cual automáticamente satisface el requisito de que  $\sum_{j=1}^{12} \hat{E}_j = 0$ .

El término de error estimado para el mes  $j$  del año  $k$  es por supuesto

$$\hat{\epsilon}_{j,k} = X_{j,k} - \hat{T}_k - \hat{E}_j, \quad j = 1, \dots, 12; k = 1, \dots, 6. \quad (0.56)$$

La generalización de (0.54) a (0.56) para datos con estacionalidad con un periodo distinto de 12 es bastante sencillo de realizar, simplemente cambiamos 12 por el correspondiente valor de  $d$ . Así, en general, si tenemos  $n$  años (meses, semanas, días, etc.) y estacionalidad con periodo  $d$ , los estimadores serán:

Para la tendencia  $T_k$ :

$$\hat{T}_k = \frac{1}{d} \sum_{j=1}^d X_{j,k} \quad (0.57)$$

Para la estacionalidad  $E_j$ :

$$\hat{E}_j = \frac{1}{n} \sum_{k=1}^n (X_{j,k} - \hat{T}_k), \quad j = 1, \dots, d \quad (0.58)$$

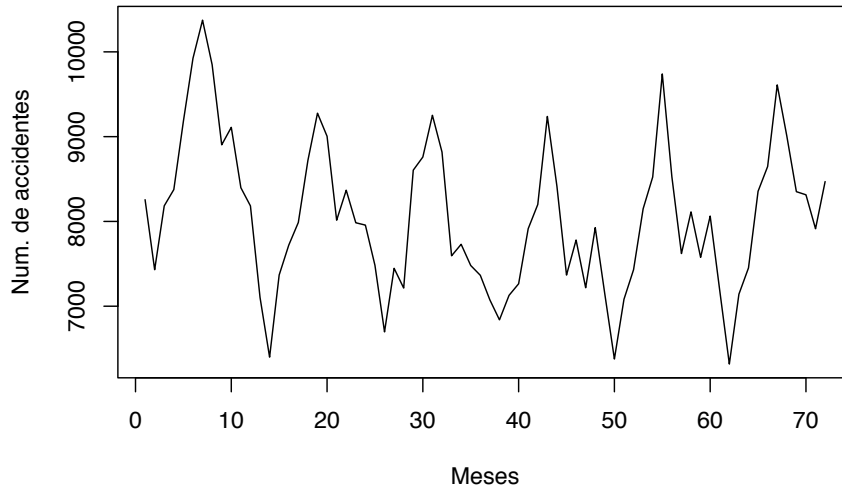
Para el error

$$\hat{\epsilon}_{j,k} = X_{j,k} - \hat{T}_k - \hat{E}_j, \quad k = 1, \dots, n; j = 1, \dots, d. \quad (0.59)$$

Las Figuras siguientes muestran respectivamente las observaciones con la tendencia removida  $X_{j,k} - \hat{T}_k$ , la componente estacional estimada  $\hat{E}_j$  y las observaciones con la tendencia y la estacionalidad removida  $\hat{e}_{j,k} = X_{j,k} - \hat{T}_k - \hat{E}_j$ . En la última no se observa una aparente tendencia o estacionalidad.

```
# Estimacion de la tendencia
Tk=numeric(n*d)
for(k in 1:n)
{
  for(j in 1:d)
  {
    Tk[(k-1)*d+j]=Tk[(k-1)*d+j]+(1/12)*X[j,k+1]
  }
}
# Grafico con la tendencia removida
plot(V-Tk,type = "l",xlab = "Meses",ylab = "Num. de accidentes",
     main = "Accidentes mortales mensuales con la tendencia T_k removida")
```

**Accidentes mortales mensuales con la tendencia T\_k removida**



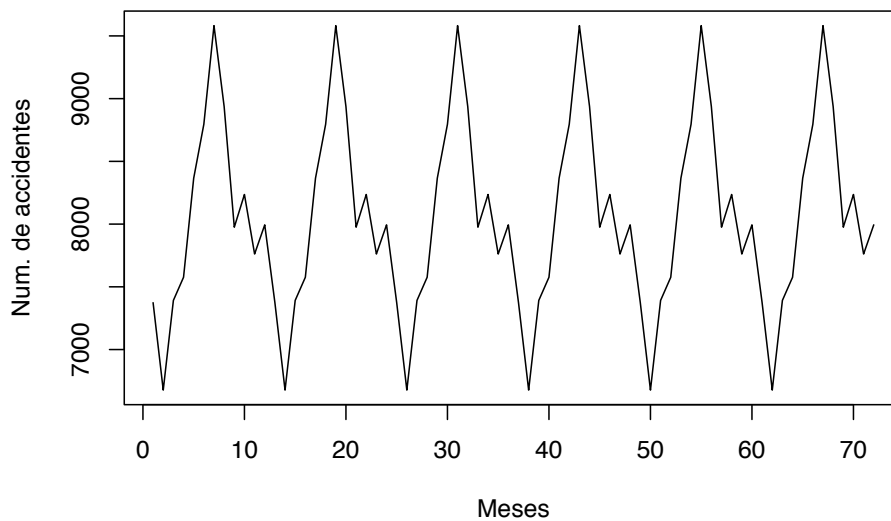
```
# Estimacion de la estacionalidad
Ej=numeric(n*d)
```

```

for(j in 1:d)
{
  aux=0
  for(k in 1:n)
  {
    aux=aux+(X[j,k+1]-Tk[(k-1)*d+j])
  }
  for(k in 1:n)
  {
    Ej[(k-1)*d+j]=(1/n)*aux
  }
}
# Grafico de la estacionalidad
plot(Ej,type = "l",xlab = "Meses",ylab = "Num. de accidentes",
     main = "Estacionalidad de los accidentes mortales mensuales")

```

**Estacionalidad de los accidentes mortales mensuales**

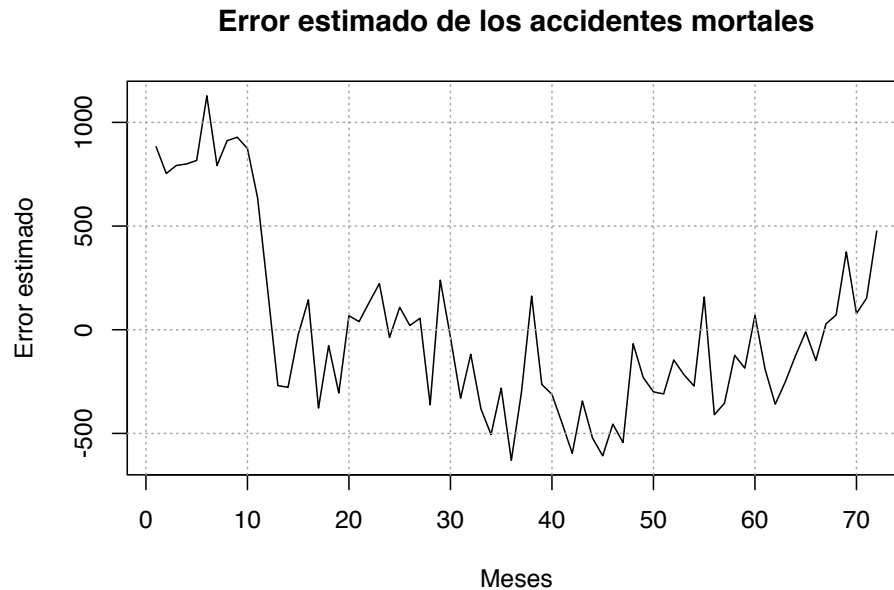


```

# Estimacion del error
error=V-Tk-Ej
# Grafico del error estimado

```

```
plot(error,type = "l",xlab = "Meses",ylab = "Error estimado",
     main = "Error estimado de los accidentes mortales")
grid(col = "darkgray")
```



- 2) **Método E2: Estimación por promedio móvil.** La siguiente técnica es preferible al Método E1 ya que no se basa en la suposición de que  $T_t$  es casi constante sobre cada ciclo estacional.

Suponga que tenemos las observaciones  $\{x_1, \dots, x_n\}$ . Se estima primero la tendencia aplicando un filtro de promedio móvil especialmente elegido para eliminar la componente estacional y para amortiguar el ruido. Si el periodo  $d$  es par, digamos  $d = 2q$ , entonces usamos

$$\hat{T}_t = (0,5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0,5x_{t+q})/d, q < t \leq n - q. \quad (0.60)$$

Si el periodo es impar, digamos  $d = 2q + 1$ , entonces usamos el promedio móvil simple (0.49). La Figura~?? muestra la tendencia

estimada  $\hat{T}_t$  para los datos de accidentes mortales obtenido de (0.60). También muestra la tendencia constante a trozos obtenida por el Método S1.

El segundo paso, es estimar la componente estacional. Para cada  $k = 1, \dots, d$ , calculamos el promedio  $w_k$  de las desviaciones  $\{(X_{k+jd} - \hat{T}_{k+jd}) : q < k + jd \leq n - q\}$ . Dado que este promedio de desviaciones no necesariamente suma cero, estimamos la componente estacional  $E_k$  como

$$\hat{E}_k = w_k - \frac{1}{d} \sum_{i=1}^d w_i, \quad i = 1, \dots, d, \quad (0.61)$$

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

Los datos sin la componente estacional se definen entonces como la serie original con la componente estacional removida, es decir,

$$d_t = X_t - \hat{E}_t, \quad t = 1, \dots, n. \quad (0.62)$$

Finalmente, reestimamos la tendencia de  $\{d_t\}$  aplicando un filtro de promedio móvil como se describió para los datos no estacionales o fijando un polinomio a la serie  $\{d_t\}$ . El término del ruido estimado llega a ser entonces

$$\hat{\epsilon}_t = X_t - \hat{E}_t - \hat{E}_t, \quad t = 1, \dots, n.$$

Los resultados de aplicar los Métodos S1 y S2 a los datos de accidentes mortales son casi iguales, dado que en este caso la constante a trozos y el promedio móvil de  $T_t$  están razonablemente cercanos.

Una comparación de los valores estimados de  $E_k, k = 1, \dots, 12$ , obtenido con ambos métodos se muestra en la Tabla~??

k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{E}_t(S1)$	-7434	-1504	-724	-523	338	808	1665	961	-87	197	-321	-67
$\hat{E}_t(S2)$	-804	-1522	-737	-526	343	746	1680	987	-109	258	-259	-57



Componentes estacional estimadas para los datos de accidentes mortales

- **Método E3: Diferenciación a paso d.** La técnica de diferenciación la cual aplicamos antes a datos no estacionales se pueden adaptar para lidiar con el caso estacional de periodo  $d$  introduciendo el operador de diferencia de paso  $d$   $\nabla_d$  definido por

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t. \quad (0.63)$$

Este operador no debe confundirse con el operador  $\nabla^d = (1 - B)^d$  definido por (??).

Aplicando el operador  $\nabla_d$  al modelo

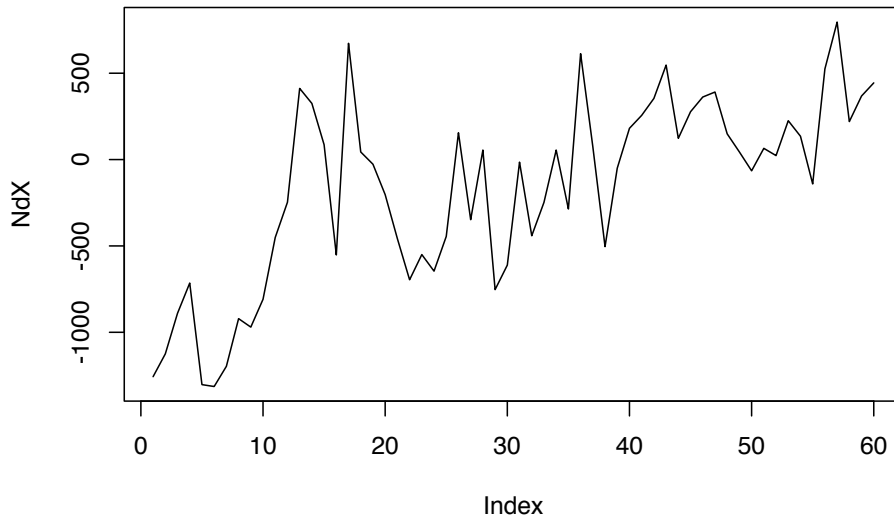
$$X_t = T_t + E_t + \epsilon_t,$$

donde  $\{E_t\}$  tiene periodo  $d$ , obtenemos

$$\nabla_d X_t = T_t - T_{t-d} + \epsilon_t - \epsilon_{t-d},$$

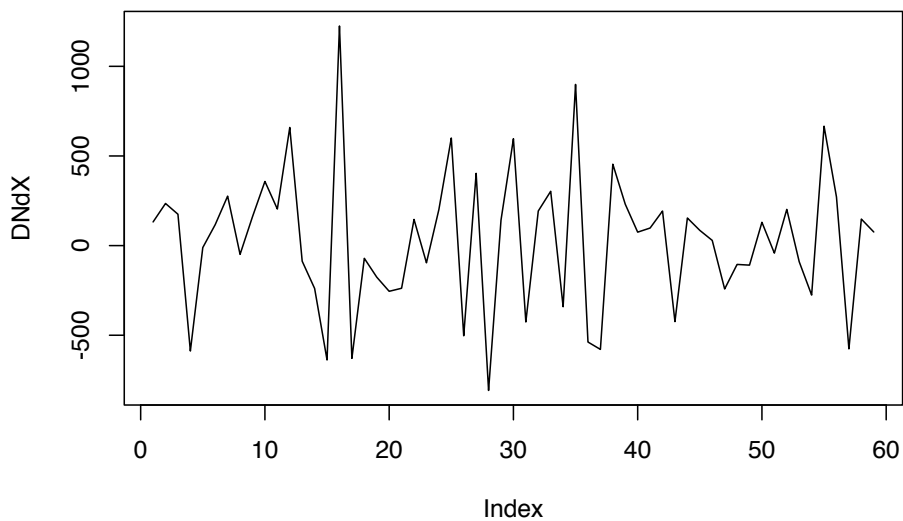
lo cual nos da una descomposición de la diferencia  $\nabla_d X_t$  en una componente de tendencia  $(T_t - T_{t-d})$  y un término de ruido  $(\epsilon_t - \epsilon_{t-d})$ . La tendencia  $(T_t - T_{t-d})$  se puede eliminar usando los métodos ya descritos, por ejemplo, aplicando alguna potencia del operador  $\nabla$ . La figura siguiente muestra el resultado de aplicar el operador  $\nabla_{12}$  a los datos de accidentes mortales.

```
# Operador Nabla_d, usamos la funcion diff con lag=12
NdX=diff(V,lag=12)
plot(NdX,type = "l")
```



La componente estacional evidente en la Figura~?? está ausente en la Figura de  $\nabla_{12}X_t$ ,  $13 \leq t \leq 72$ . Sin embargo todavía parece haber una tendencia decreciente. Si ahora aplicamos el operador  $\nabla$  a  $\nabla_{12}X_t$  y graficamos las diferencias  $\nabla\nabla_{12}X_t$ ,  $t = 14, \dots, 72$  obtenemos el gráfico mostrado en la Figura~??, los cuales no tienen una aparente tendencia o componente estacional.

```
DNdX=diff(NdX)
plot(DNdX,type = "l")
```



## 0.7. Estimación de la tendencia por regresión clásica

Los modelos de regresión son importantes para modelos en el dominio de tiempo y de frecuencia que discutiremos posteriormente. La idea principal depende de poder expresar una serie respuesta  $X_t$  como una combinación lineal de entradas  $z_{t_1}, z_{t_2}, \dots, z_{t_q}$ . La estimación de los coeficientes  $\beta_1, \beta_2, \dots, \beta_q$  de la combinación por mínimos cuadrados proporciona un método para modelar  $X_t$  en términos de las entradas.

### 0.7.1. Regresión Clásica

Supongamos que tenemos  $X_t$ , para  $t = 1, 2, \dots, n$  influenciada por una colección de series independientes  $z_{t_1}, z_{t_2}, \dots, z_{t_q}$ , donde consideraremos primero que las entradas son fijas y conocidas.

Podemos expresar esta relación como

$$X_t = \beta_1 z_{t_1} + \beta_2 z_{t_2} + \dots + \beta_q z_{t_q} + w_t \quad (0.64)$$

donde  $\beta_1, \beta_2, \dots, \beta_q$  son los coeficientes de regresión fijos y desconocidos,  $\{w_t\}$  es un error aleatorio o un proceso de ruido consistente de variables normales iid con media cero y varianza  $\sigma_w^2$ .

El modelo lineal descrito en (0.64) se puede escribir de forma más general por medio de definir los vectores columna

$\mathbf{z}_t = (z_{t_1}, z_{t_2}, \dots, z_{t_q})^t$  y  $\beta = (\beta_1, \beta_2, \dots, \beta_q)^t$  donde  $t$  denota la traspuesta, así (0.64) será

$$X_t = \beta^t \mathbf{z}_t + w_t \quad (0.65)$$

donde  $w_t \sim iid(0, \sigma_w^2)$ . Es natural considerar la estimación de los coeficientes del vector  $\beta$  minimizando la suma residual de cuadrados

$$RSS = \sum_{t=1}^n (X_t - \beta^t \mathbf{z}_t)^2 \quad (0.66)$$

con respecto a  $\beta_1, \beta_2, \dots, \beta_q$ . Minimizando RSS obtenemos el estimador común de mínimos cuadrados. Esta minimización se puede hacer por diferenciación de (0.66) con respecto al vector  $\beta$  o usando las propiedades de proyección. En la notación anterior, obtenemos la ecuación normal

$$\left( \sum_{t=1}^n \mathbf{z}_t \mathbf{z}_t^t \right) \hat{\beta} = \sum_{t=1}^n \mathbf{z}_t X_t \quad (0.67)$$

Definiendo la matriz  $Z = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)^t$  como una matriz  $n \times q$  compuesta de  $n$  muestras de las variables de entradas y el vector observado  $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$  se puede hacer una simplificación de la notación. Esta identificación nos lleva a

$$(Z^t Z) \hat{\beta} = Z^t \mathbf{x} \quad (0.68)$$

y la solución es

$$\hat{\beta} = (Z^t Z)^{-1} Z^t \mathbf{x}, \quad (0.69)$$

cuando la matriz  $Z^t Z$  es de rango  $q$ . El residual minimizado de suma de cuadrados (0.66) tiene la forma matricial equivalente

$$\begin{aligned} RSS &= (\mathbf{x} - Z \hat{\beta})^t (\mathbf{x} - Z \hat{\beta}) \\ &= \mathbf{x}^t \mathbf{x} - \hat{\beta}^t Z^t \mathbf{x} \\ &= \mathbf{x}^t \mathbf{x} - \mathbf{x}^t Z (Z^t Z)^{-1} Z^t \mathbf{x}. \end{aligned} \quad (0.70)$$

El estimador común de mínimos cuadrados es insesgado, esto es,  $\mathbb{E}(\hat{\beta}) = \beta$ , y tiene la menor varianza de todos los estimadores insesgados lineales.

Si los errores  $w_t$  son normalmente distribuidos (Gaussianos),  $\hat{\beta}$  es

también el estimador de máxima verosimilitud para  $\beta$  y es normalmente distribuido con

$$\begin{aligned}\text{cov}(\hat{\beta}) &= \sigma_w^2 \left( \sum_{t=1}^n \mathbf{z}_t \mathbf{z}_t^t \right)^{-1} \\ &= \sigma_w^2 (Z^t Z)^{-1} \\ &= \sigma_w^2 C,\end{aligned}\tag{0.71}$$

donde

$$C = (Z^t Z)^{-1}.\tag{0.72}$$

Un estimador insesgado para la varianza  $\sigma_w^2$  es

$$s_w^2 = \frac{RSS}{n - q}\tag{0.73}$$

contrastado con el estimador de máxima verosimilitud  $\hat{\sigma}_w^2 = RSS/n$  el cual tiene divisor  $n$ . Bajo la suposición de que  $s_w^2$  tiene distribución proporcional a una variable aleatoria chi-cuadrado con  $n - q$  grados de libertad,  $\chi_{n-q}^2$ , e independiente de  $\hat{\beta}$ , se sigue que

$$t_{n-q} = \frac{(\hat{\beta}_i - \beta_i)}{s_w \sqrt{c_{ii}}}\tag{0.74}$$

tiene distribución  $t$ -de Student con  $n - q$  grados de libertad;  $c_{ii}$  denota el  $i$ -ésimo elemento de la diagonal de  $C$ , como se definió en (0.72).

Hay varios modelos que podemos utilizar de manera de seleccionar el mejor subconjunto de variables independientes. Suponga que tenemos un modelo que sólo considera un subconjunto  $q_1 < q$  de variables independientes  $\mathbf{z}_{1t} = (z_{t1}, z_{t2}, \dots, z_{tq_1})^t$  que influyen a la variable  $X_t$ , así el modelo

$$X_t = \beta_1^t \mathbf{z}_{1t} + w_t \quad (0.75)$$

llega a ser la hipótesis nula, donde  $\beta_1 = (\beta_1, \beta_2, \dots, \beta_{q1})^t$  es un subconjunto de los coeficientes de las  $q$  variables originales. Podemos contrastar el modelo reducido (0.75) contra el modelo completo (0.65) comparando el residual de la suma de cuadrados bajo los dos modelos usando el estadístico  $F$

$$F_{q-q1, n-q} = \frac{RSS_1 - RSS}{RSS} \frac{n-q}{q-q1} \quad (0.76)$$

el cual tiene distribución  $F$  con  $q - q1$  y  $n - q$  grados de libertad cuando (0.76) es el modelo correcto. La información envuelta en la prueba se resume en una tabla de Análisis de Varianza (ANOVA) como la mostrada en la Tabla siguiente para este caso particular. La diferencia en el numerador es llamada regresión de la suma de cuadrados.

Fuente	g.l	Suma de cuadrados	Medias Cuadrados
$z_{t,q1+1}, \dots, z_{t,q}$	$q - q1$	$SS_{reg} = RSS_1 - RSS$	$MS_{reg} = SS_{reg}/(q - q1)$
Error	$n - q$	$RSS$	$s_e^2 = RSS/(n - q)$
Total	$n - q1$	$RSS_1$	

En términos de la Tabla, por convención escribimos el estadístico  $F$  dado en (0.76) como el ratio de dos medias cuadrados, obteniéndose

$$F_{q-q1, n-q} = \frac{MS_{reg}}{s_w^2}. \quad (0.77)$$

Un caso de especial interés es para  $q1 = 1$  y  $z_{1t} = 1$ , así el modelo

en (0.75) es  

$$\bar{X}_t = \beta_1 + w_t$$

y la proporción de variación explicada por las otras variables es

$$R_{xz}^2 = \frac{RSS_0 - RSS}{RSS_0}, \quad (0.78)$$

donde la suma residual de cuadrados bajo el modelo reducido  
dada por

$$RSS_0 = \sum_{t=1}^n (X_t - \bar{X})^2 \quad (0.79)$$

es precisamente la suma de desviaciones al cuadrado de la media  
 $\bar{X}$ . La medida  $R_{xz}^2$  es la correlación múltiple cuadrado entre  $X_t$  y

$$z_{t2}, z_{t3}, \dots, z_{tq}.$$

Las técnicas discutidas se pueden usar para hacer comparación  
entre varios modelos. Estas pruebas han sido usadas en el pasado  
en una manera gradual, donde las variables son añadidas o  
suprimidas cuando los valores de la prueba  $F$  exceden o fallan en  
exceder algunos niveles predeterminados. El procedimiento,  
llamado regresión múltiple por pasos, es útil para conseguir un  
conjunto de variables que sea de utilidad. Una manera alternativa  
es enfocándose en un procedimiento para selección del modelo que  
no sea secuencial, sino simplemente evaluar cada modelo en base a  
sus propios méritos.

Suponga que consideramos un modelo de regresión con  $k$   
coeficientes y denotemos el estimador de máxima verosimilitud  
para la varianza como

$$\hat{\sigma}_k^2 = \frac{RSS_k}{n} \quad (0.80)$$

donde  $RSS_k$  denota la suma residual de cuadrados bajo el modelo  
con  $k$  coeficientes de regresión. Entonces, Akaike (1969, 1973,  
1974) sugirió medir la bondad del ajuste para este modelo en

particular equilibrando el error del ajuste contra el número de parámetros en el modelo; definiendo lo siguiente

**Definición 0.24** (Criterio de Información de Akaike (AIC)). El Criterio de Información de Akaike se define como

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n + 2k}{n} \quad (0.81)$$

donde  $\hat{\sigma}_k^2$  está dado por (0.80) y  $k$  es el número de parámetros en el modelo

---

El *criterio de información de Akaike* (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo. El valor de  $k$  que minimiza  $AIC$  especifica el mejor modelo. La idea es que la minimización de  $\hat{\sigma}_k^2$  sea razonablemente objetiva, excepto que decrezca monótonamente cuando  $k$  crece.

Por lo tanto, debemos penalizar la variación del error por un término proporcional al número de parámetros. La elección del término de penalización dado por (0.108) no es único.



---

# 0

---

## *Modelos de series de tiempo*

---

Como indicamos en el capítulo anterior el objetivo principal en el análisis de series de tiempo es desarrollar modelos matemáticos que provean una descripción apropiada para los datos muestrales.

Recordando las definiciones 0.7 y 0.9 podemos describir los modelos generales útiles para la descripción de series de tiempo

---

### 0.8. Modelos Estocásticos

#### 0.8.1. Procesos Estocásticos

De la definición de procesos estocásticos (Definición 0.9), las variables aleatorias de la familia (medibles para todo  $t \in T$ ) son funciones de la forma  
$$x(\omega, t) : \Omega \times T \rightarrow \mathbb{R}$$

Para  $T = \mathbb{N}$ , tenemos un proceso en *tiempo discreto* y para  $T \subset \mathbb{R}$  tenemos un proceso en *tiempo continuo*. En lo que respecta a este libro, consideraremos como subconjunto de índices  $T = (0, \infty)$ .

Como ya indicamos, usaremos la notación  $X_t$  para denotar la realización de un proceso estocástico  $x_t(\omega^*)$  cuando no haya lugar a confusión. De esta forma, adoptaremos sin pérdida de generalidad, el conjunto de índices habitual de las series de tiempo en el ámbito de las finanzas y economía  $I = (1, T)$ .

De lo anterior, se tiene que los procesos estocásticos suelen ser descritos mediante su distribución conjunta de probabilidades, de manera que la relación que existe entre una realización y un

proceso estocástico es análoga a la existente entre la muestra y la población en el análisis estadístico clásico.

### 0.8.2. Momentos, Varianza, Covarianza y Correlación

**Definición 0.25.** El **valor esperado** y **varianza** de un proceso estocástico están dados por

$$\mathbb{E}(x_t) = \int_{\Omega} x(\omega, t) dP(\omega), \quad t \in [0, T] \quad (0.82)$$

y

$$Var(x_t) = \mathbb{E}(x_t - \mathbb{E}(x_t))^2, \quad t \in [0, T] \quad (0.83)$$

siempre que las integrales existan y sean finitas.

**Definición 0.26.** El  $k$ -ésimo **momento** de  $x_t$ , con  $k \geq 1$ , se define como  $\mathbb{E}(x_t^k)$  para todo  $t \in [0, T]$ .

**Definición 0.27.** La **función de covarianza** del proceso para dos instantes de tiempo  $t$  y  $s$  está dada por

$$\gamma(t, s) = Cov(x_t, x_s) = \mathbb{E}[(x_t - \mathbb{E}(x_t))(x_s - \mathbb{E}(x_s))]$$

La cantidad  $x_t - x_s$  es llamada el proceso de *incrementos* desde  $s$  a  $t$ , con  $s < t$ .

### 0.8.3. Variación de un proceso

Sea  $P_n = \{0 = t_0 < t_1 < \dots < t_i < \dots < t_n = t\}$  una partición cualquiera del intervalos  $[0, t]$  en  $n$  subintervalos y denotemos por

$$\|P_n\| = \max\{j = 0, 1, \dots, n-1(t_{j+1} - t_j)\}$$

el tamaño de paso máximo de discretización de la partición  $P_n$ .

**Definición 0.28.** La **variación** del proceso  $x$  se define como

$$V_t(x) = p - \lim_{\|P_n\|} \sum_{k=0}^{n-1} |x_{t_{k+1}} - x_{t_k}| \quad (0.84)$$

Si  $x$  es diferenciable, entonces  $V_t(x) = \int_0^t |x'(u)| du$ . Si  $V_t(X) < \infty$ , entonces decimos que  $x$  es de *variación acotada* en  $[0, t]$ . Si es cierto para todo  $t \geq 0$ , entonces decimos que  $x$  tiene *variación acotada*.

**Definición 0.29.** La **variación cuadrática** de un proceso estocástico  $x$ , denotada por  $[x, x]_t$ , se define como

$$[x, x]_t = p - \lim_{\|P_n\|} \sum_{k=0}^{n-1} |x_{t_{k+1}} - x_{t_k}|^2 \quad (0.85)$$

Para procesos estocásticos con trayectorias continua, el límite existe, y en dicho caso usamos la notación  $\langle x, x \rangle_t$  y podemos definirla alternativamente como

$$\langle x, x \rangle_t = p - \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left( x_{\min(t, k/2^n)} - x_{\min(t, (k-1)/2^n)} \right)^2 \quad (0.86)$$

Si  $x$  es continuo y tiene variación acotada cuadrática finita, entonces su variación total es infinita. Note que  $V_t(x)$  y  $[x, x]_t$  son también procesos estocásticos.

#### 0.8.4. Martingalas

En teoría de probabilidad, un proceso estocástico de tipo **martingala** (galicismo de *martingale*) es todo proceso caracterizado por no tener deriva. Este tipo de procesos estocásticos reciben su nombre de la estrategia de la martingala, un método de apuestas que tuvo cierta fama en el siglo XVIII. La estrategia de la martingala consiste en volver a apostar por el total perdido al momento de incurrir en una pérdida en un juego de azar. En la nueva apuesta, el jugador tiene la posibilidad de recobrar todas sus pérdidas, por lo que podría parecer que a largo plazo la esperanza de ganancia con esta estrategia se mantienen constantes y a favor del jugador. De hecho, estadísticamente es así: el capital medio del jugador (esto es, el dinero que el jugador tiene a su disposición para jugar) se mantiene constante. El problema reside en que, al incurrir en sucesivas pérdidas, el jugador que siga

la estrategia de la martingala se ve obligado a apostar de nuevo cantidades cada vez mayores (las pérdidas acumuladas), que tienden a crecer exponencialmente. Al cabo de unos pocos ciclos de apuestas, el jugador, cuyos recursos son habitualmente muy inferiores a los de la banca, se ve arruinado al ser incapaz de apostar de nuevo por el total de sus pérdidas. Evitar jugadores que intenten seguir la estrategia de la martingala es de todos modos una de las razones por las que los casinos actuales establecen límites máximos de apuesta.

La estrategia de la martingala se popularizó en el siglo XVIII con fama de ser una estrategia ingenua y propia de mentes simples, puesto que aunque en apariencia es infalible, sin embargo, está abocada a arruinar al jugador. Recibe el nombre de los habitantes de la localidad francesa de *Martigues* (martingales en francés), situada en las cercanías de Marsella, que por aquel entonces tenían fama de ser ingenuos y simplones.

El concepto de la martingala en la teoría de probabilidades fue introducido por *Paul Pierre Lévy*, y una gran parte del desarrollo original de la teoría la realizó *Joseph Leo Doob*. Parte de la motivación para ese esfuerzo era demostrar la inexistencia de estrategias de juego infalibles.

El concepto fue inmediatamente aplicado al análisis de procesos bursátiles. Uno de los resultados más importantes de la matemática financiera es, precisamente, que un mercado perfecto sin posibilidades de arbitraje es una martingala.

**Definición 0.30.** Sea  $(\omega, \mathcal{F}, P)$  un espacio de probabilidad. Una **filtración**  $\{\mathcal{F}_t, t \geq 0\}$  es una familia creciente de sub- $\sigma$ -álgebras de  $\mathcal{F}$  indexadas por  $t \geq 0$ ; es decir, para cada  $s, t > 0$  tal que  $s < t$ , se tiene  $\mathcal{F}_s \subset \mathcal{F}_t$  con  $\mathcal{F}_0 = \{\Omega, \emptyset\}$ .

Para cada proceso estocástico  $\{x_t\}_{t \geq 0}$  y para cada  $t$ , podemos asociar una  $\sigma$ -álgebra denotada por  $\mathcal{F}_t = \sigma\{x_s : 0 \leq s \leq t\}$ , y que además es la  $\sigma$ -álgebra generada por  $x$ ; es decir, la  $\sigma$ -álgebra más pequeña (minimal) de  $\mathcal{F}$  que hace a  $x(s, \omega)$  medible para cada  $0 \leq s \leq t$ .

**Definición 0.31.** Dado un proceso estocástico  $\{X_t\}_{t \geq 0}$  y una filtración  $\{\mathcal{F}_t, t \geq 0\}$  (no necesariamente la que genera  $X$ ), el proceso  $X$  se denomina **adaptado** a  $\{\mathcal{F}_t, t \geq 0\}$  ( $\mathcal{F}_t$ -adaptado) si para cada  $t \geq 0$ ,  $X(t)$  es  $\mathcal{F}_t$ -medible.

En otras palabras  $X = \{X_t\}_{t \geq 0}$  es  $\mathcal{F}_t$ -adaptado cuando el valor de  $X_t$  en el tiempo  $t$  solo depende de la información contenida en la realización hasta el instante  $t$ .

Dado un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  y una filtración  $\{\mathcal{F}_t, t \geq 0\}$ , entonces definimos el **espacio de probabilidad filtrado** a la cuaterna  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ .

**Definición 0.32.** Sea  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$  un espacio de probabilidad filtrado. Un proceso  $X_t$  con  $t \in T$ ,  $T \subseteq \mathcal{R}$  un conjunto de índices, es una **martingala** relativo a la filtración  $\{\mathcal{F}_t, t \geq 0\}$ , si

- 1)  $X_t$  es adaptado a la filtración  $\{\mathcal{F}_t, t \geq 0\}$
- 2)  $X_t$  es integrable, es decir,  $\mathbb{E}|X_t| < \infty$ ,
- 3) Para cualesquieras  $s$  y  $t$  con  $s < t$ ,  $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$  c.s.

Decimos que el proceso es una **submartingala** si  $\mathbb{E}(X_t|\mathcal{F}_s) \geq X_s$  c.s.

Decimos que es una **supermartingala** si  $\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s$  c.s.

**Ejemplo 0.13.** Sean  $X_0, X_1, \dots, X_n$  variables aleatorias iid tal que  $\mathbb{E}(X_1) = \mu$  y sean

$$\begin{aligned} M_0 &= X_0 \\ M_1 &= X_0 + X_1 \\ &\vdots \\ M_n &= X_0 + X_1 + \dots + X_n \end{aligned}$$

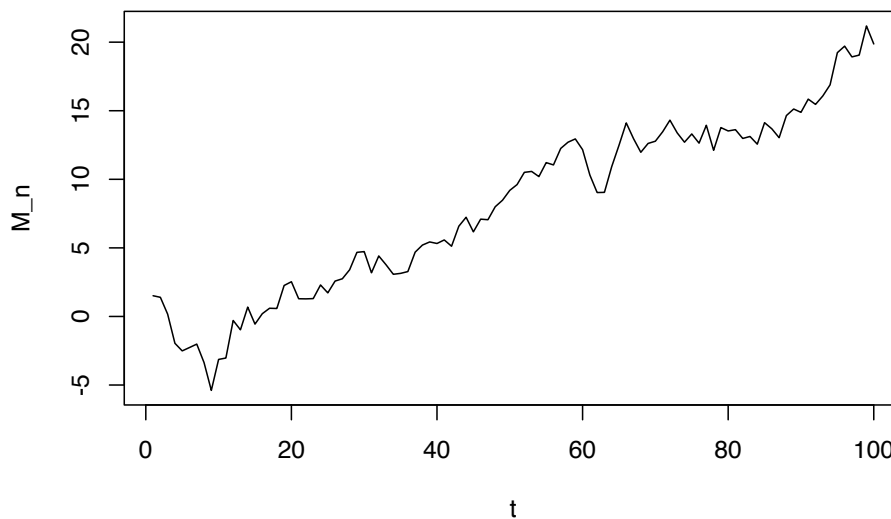
La sucesión de variables aleatorias  $M_n$  se llama **paseo aleatorio** y es una supermartingala si  $\mu \leq 0$ , una martingala si  $\mu = 0$  y una submartingala si  $\mu \geq 0$ .

Es fácil demostrarlo, sencillamente usamos el hecho de que

$$M_{n+1} = M_n + X_{n+1}$$

y que  $M_n$  y  $X_{n+1}$  son independientes. Podemos generar tal proceso en **R**.

```
n=100
mu=0
sigma=1
X=rnorm(n,mu,sigma)
M=cumsum(X)
plot(M,type = "l",xlab = "t",ylab = "M_n")
```



**Ejemplo 0.14** (Precio de acciones). Sean  $Y_0, Y_1, \dots, Y_n$  variables aleatorias independientes y positivas. Supongamos que una acción tiene precio  $M_0$  a tiempo  $t = 0$ .

Un modelo común para modelar el precio de la acción en tiempo  $t = n$  es

$$M_{n+1} = M_n Y_n$$

donde  $(Y_n - 1) \times 100$  representa (en porcentaje) la variabilidad de

la acción. Usando las propiedades de esperanza condicional (Apéndice), es muy sencillo demostrar que

$$\mathbb{E}(M_{n+1}|M_0, \dots, M_n) = M_n \mathbb{E}(Y_n)$$

En particular, si  $Y_1, \dots, Y_n$  son idénticamente distribuidas con  $\mathbb{E}(Y_1) = \mu$  tenemos que  $M_n$  es

- Una **martingala** si  $\mu = 1$
- Una **submartingala** si  $\mu > 1$
- Una **supermartingala** si  $\mu < 1$ .

Dos modelos bien conocidos de lo anterior son

1) **Modelo Black-Scholes discreto.**

Sean  $Y_1, \dots, Y_n$  definidas por

$$Y_n = e^{Z_n}$$

donde  $Z_1, \dots, Z_n$  son variables aleatorias independientes normales  $N(\mu, \sigma^2)$ .

2) **Modelo Binomial.**

Sean  $Y_1, \dots, Y_n$  definidas por

$$P(Y_i = (1+t)e^{-r}) = p \quad \text{y} \quad P(Y_i = (1+t)^{-1}e^{-r}) = 1-p$$

La constante  $r$  es la tasa de interés y los factores  $(1+t)$  y  $(1+t)^{-1}$  modelan las variaciones del mercado y garantizan que el precio tiene la forma  $M_0(1+t)^ye^{-nr}$ , con  $|y| \leq n$ . La volatilidad del precio está asociada a  $p$ .

**Definición 0.33.** Una variable aleatoria  $X$  es **cuadrado integrable** si  $\mathbb{E}(X^2) < \infty$ . Un proceso estocástico  $X_t$  en el intervalo  $[0, T]$ , donde  $T$  puede ser infinito, es **cuadrado integrable** si

$$\sup_{t \in [0, T]} \mathbb{E}(X_t^2) < \infty \quad (0.87)$$

es decir, si sus segundos momentos son acotados.

**Definición 0.34.** Un proceso estocástico  $X_t, 0 \leq t \leq T$  se dice que es **uniformemente integrable** si

$$\mathbb{E}(|X_t| \mathbf{1}_{\{|X_t| > n\}})$$

converge a 0 cuando  $n \rightarrow \infty$  uniformemente en  $t$ .

### 0.8.5. Propiedad de Markov

La propiedad de Markov establece que si conocemos el estado actual de un proceso estocástico, entonces el comportamiento futuro de dicho proceso es independiente de su pasado. Un proceso  $X_t$  tiene la *propiedad de Markov* si la distribución condicional del proceso  $X_t$  dado el proceso en el instante  $X_t = x$ , no depende de los valores pasados.

**Definición 0.35.**  $X$  es un **proceso de Markov** si para cualquier  $t$  y  $s > 0$ ,

$$P(X_{t+s} \leq y | \mathcal{F}_t) = P(X_{t+s} \leq y | X_t) \text{ c.s.}$$

donde  $\mathcal{F}_t$  es la  $\sigma$ -álgebra generada por el proceso hasta el tiempo  $t$ .

**Definición 0.36.** La **función de transición de probabilidad** de un proceso  $X$  se define como

$$P(y, t, x, s) = P(X_y \leq y | X_s \leq x)$$

la función de distribución condicional del proceso en el instante  $t$ , dado que éste está en el punto  $x$  en el instante  $s < t$ .

La propiedad de Markov implica una expresión que resulta muy útil en términos de la esperanza condicional por la  $\sigma$ -álgebra de eventos, la cual es válida tanto para procesos en tiempo discreto como en tiempo continuo.

Las definiciones y propiedades anteriores son temas de estudio de gran importancia y con una amplia teoría matemática que está fuera del alcance de este libro, pero lo que hemos descrito es suficiente para el objetivo del mismo.



## 0.9. Modelos lineales

Los modelos lineales proporcionan un enfoque natural que permite analizar el comportamiento de los procesos estocásticos o series de tiempo y en especial a lo referente a finanzas y economía. En esta sección discutiremos la estructura de dependencia, autocorrelación, modelización y predicción de los modelos lineales teóricos, con los correspondientes comandos en **R** para generar y analizar dichos procesos.

### 0.9.1. Proceso de Ruido Blanco

**Definición 0.37.** Un proceso  $\{w_t\}$  se denomina **ruido blanco** (white noise) de media 0 y varianza  $\sigma^2$  si satisface

$$\begin{aligned}\mathbb{E}(w_t) &= 0, & \text{Var}(w_t) &= \sigma_w^2 < \infty \\ \text{Cov}(w_t, w_{t-k}) &= 0, & \forall k &\neq 0\end{aligned}$$

Las series de tiempo generadas de esta manera son muy usadas como modelos para ruido en aplicaciones de ingeniería. La designación “*blanco*” se origina de la analogía con la luz blanca e indica que todos los posibles períodos de oscilación están presentes con igual intensidad.

En particular, una sucesión de variables aleatorias iid con media 0 y varianza  $\sigma_w^2$  representa un caso especial de un proceso de ruido blanco. Este proceso lo denotaremos por  $w_t \sim WN(0, \sigma_w^2)$ . Un muy usado ruido blanco es el **ruido blanco gaussiano**, donde las  $w_t$  son variables aleatorias normales o gaussianas con media 0 y varianza  $\sigma_w^2$  y denotadas como  $w_t \sim iidN(0, \sigma_w^2)$ .

La función de media de un ruido blanco es trivial, es decir  $\mu_w = \mathbb{E}(w_t) = 0$ .

Calculemos la función de autocovarianza de  $w_t$

$$\begin{aligned}
\gamma_w(s, t) &= \mathbb{E}[(w_s - \mu_s)(w_t - \mu_t)] \\
&= \mathbb{E}[w_s w_t] \\
&= \begin{cases} \sigma_w^2, & \text{si } s = t \\ 0, & \text{si } s \neq t \end{cases}
\end{aligned}$$

La última igualdad se sigue del hecho de que  $w_s$  y  $w_t$  son no-correlacionados para  $s \neq t$  por lo que  $\mathbb{E}(w_s w_t) = \mathbb{E}(w_s)\mathbb{E}(w_t) = 0$ .

**Ejemplo 0.15** (Estacionaridad de un ruido blanco). La función de autocovarianza de un ruido blanco es fácil de evaluar como

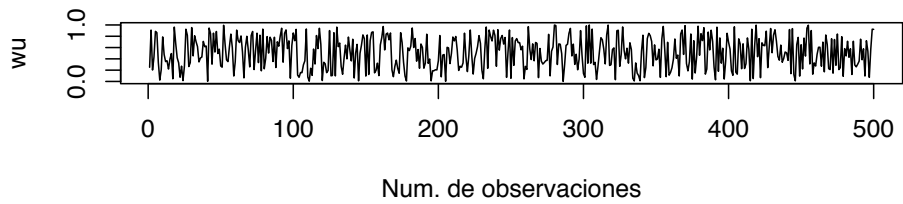
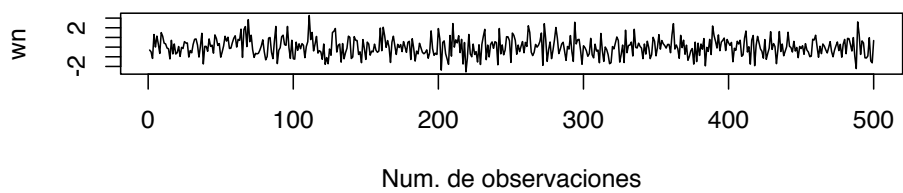
$$\gamma_w(h) = \mathbb{E}(w_{t+h} w_t) = \begin{cases} \sigma_w^2, & \text{si } h = 0 \\ 0, & \text{si } h \neq 0 \end{cases}$$

donde  $\sigma_w^2$  es la varianza del ruido blanco. Esto significa que la serie es *débilmente estacionaria* o *estacionaria*. Si las variables de ruido blanco también son gaussianas, el proceso es *estrictamente estacionario*, como se puede ver evaluando (2.10) usando la relación (2.2).

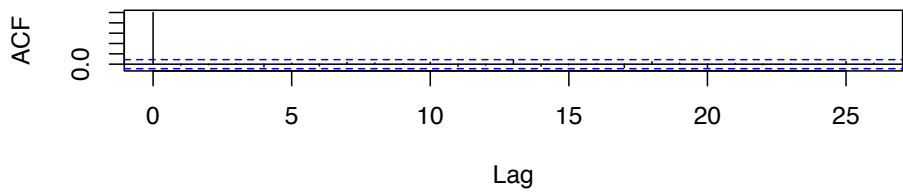
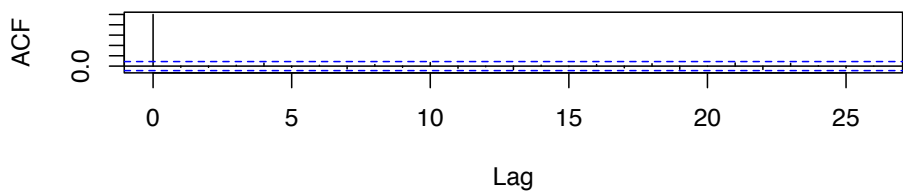
```

#-----
# Ruidos blancos
#-----
# Uniforme [0,1]
wu=runif(500,0,1)
# Gaussiano
wn=rnorm(500,0,1)
# Graficos
par(mfrow=c(2,1))
plot(wu,type = "l",xlab = "Num. de observaciones",
     main = "Ruido blanco uniforme en [0,1]")
plot(wn,type = "l",xlab = "Num. de observaciones",
     main = "Ruido blanco gaussiano")

```

**Ruido blanco uniforme en [0,1]****Ruido blanco gaussiano**

```
# Funciones de autocovarianza (ACF)  
acf(wu)  
acf(wn)
```

**Series wu****Series wn**

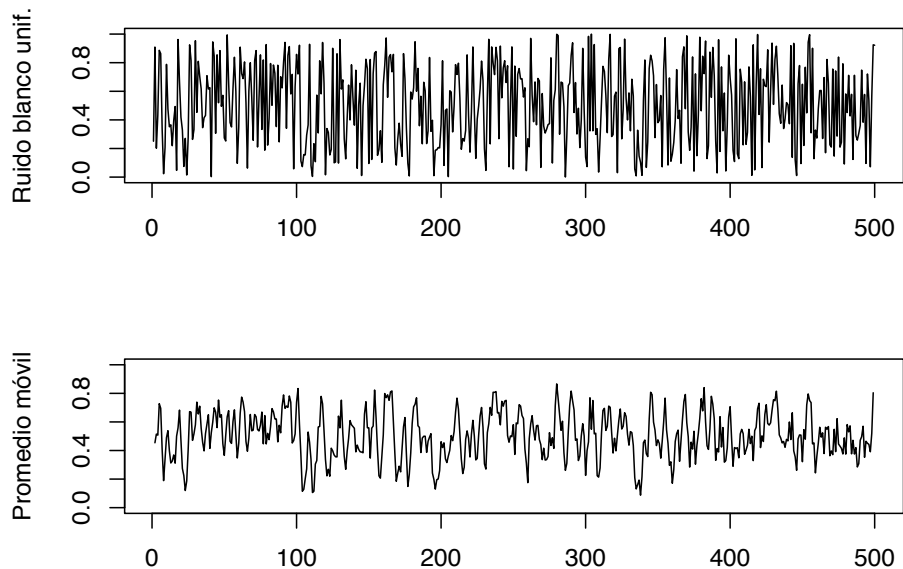
**Ejemplo 0.16.** Podemos reemplazar las series de ruido blanco  $w_t$

por un promedio móvil que suavice la serie. Por ejemplo, consideremos la serie  $w_t$  en la ecuación ( ) y reemplacémosla por un promedio móvil de 3 puntos, dado por

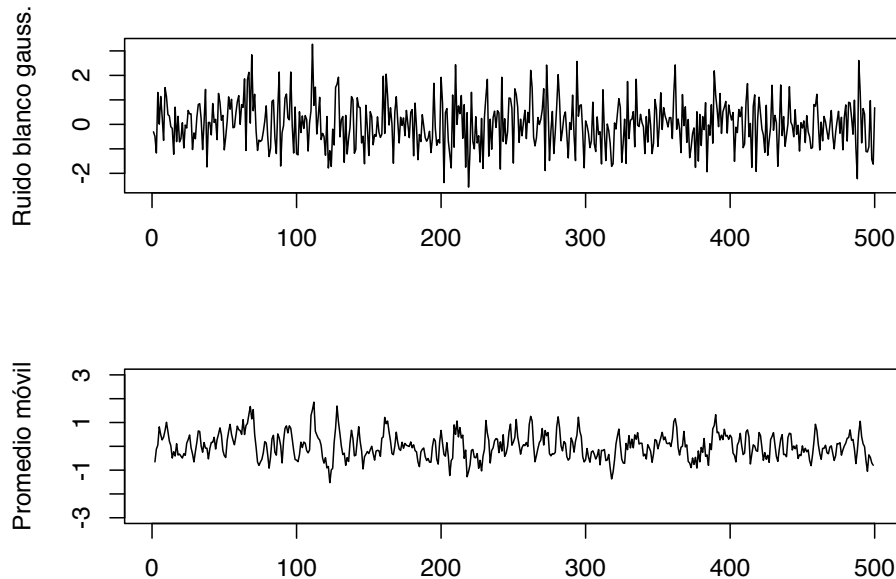
$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1}) \quad (0.88)$$

lo cual nos da una serie suavizada. Tomando la serie del ejemplo anterior y usando la función ‘filter’ de **R** se obtienen los gráficos siguientes:

```
#-----  
# Promedio movil  
#-----  
# Uniforme  
vu=filter(wu,sides = 2,rep(1/3,3))  
par(mfrow=c(2,1),mar=c(3,4,3,2))#  
plot.ts(wu,xlab=" ",ylab="Ruido blanco unif.")  
plot.ts(vu,ylim=c(0,1),ylab="Promedio móvil")
```



```
# Gaussiano
vn=filter(wn,sides = 2,rep(1/3,3))
par(mfrow=c(2,1),mar=c(3,4,3,2))
plot.ts(wn,xlab=" ",ylab="Ruido blanco gauss.")
plot.ts(vn,ylim=c(-3,3),ylab="Promedio móvil")
```



En la parte superior de cada uno se observan los ruidos blancos y en la parte inferior los respectivos promedios móviles. Podemos notar que las series de promedio móvil suavizan el comportamiento de las series originales, si tomamos más puntos en el promedio mayor será el suavizado.

**Ejemplo 0.17** (Función de media de un promedio móvil). Si  $w_t$  denota una serie de ruido blanco, entonces  $\mu_{wt} = \mathbb{E}(w_t) = 0$  para todo  $t$ . Luego para el promedio móvil de 3 puntos se tiene  $\mu_{wt} = \mathbb{E}(v_t) = \frac{1}{3}\mathbb{E}(w_{t-1} + w_t + w_{t+1}) = \frac{1}{3}(\mathbb{E}(w_{t-1}) + \mathbb{E}(w_t) + \mathbb{E}(w_{t+1})) = 0$ .

**Ejemplo 0.18** (Autocovarianza de un promedio móvil). Consideremos el promedio móvil de 3 puntos del ejemplo anterior y calculemos su función de autocovarianza

$$\begin{aligned}
\gamma_v(s, t) &= \mathbb{E}[(v_s - \mu_s)(v_t - \mu_t)] \\
&= \mathbb{E}[(v_s - o)(v_t - 0)] \\
&= \frac{1}{9}\mathbb{E}[(w_{s-1} + w_s + w_{s+1})(w_{t-1} + w_t + w_{t+1})]
\end{aligned}$$

Consideremos  $s - t = h$ , para  $h = 0, \pm 1, \pm 2, \dots$ . Entonces, tenemos para  $h = 0$

$$\begin{aligned}
\gamma_v(t, t) &= \frac{1}{9}\mathbb{E}[(w_{t-1} + w_t + w_{t+1})(w_{t-1} + w_t + w_{t+1})] \\
&= \frac{1}{9}[\mathbb{E}(w_{t-1}w_{t-1}) + \mathbb{E}(w_t w_t) + \mathbb{E}(w_{t+1}w_{t+1})] \\
&= \frac{3}{9}
\end{aligned}$$

Para  $h = 1$ , tenemos

$$\begin{aligned}
\gamma_v(t+1, t) &= \frac{1}{9}\mathbb{E}[(w_t + w_{t+1} + w_{t+2})(w_{t-1} + w_t + w_{t+1})] \\
&= \frac{1}{9}[\mathbb{E}(w_t w_t) + \mathbb{E}(w_{t+1}w_{t+1})] \\
&= \frac{2}{9}
\end{aligned}$$

Usando el hecho de que  $\mathbb{E}(w_t w_s) = 0$  si  $s \neq t$ . Cálculos similares nos dan  $\gamma_v(t-1, t) = 2/9$ ,  $\gamma_v(t+2, t) = \gamma_v(t-2, t) = 1/9$  y 0 para  $h \geq 3$ . Resumiendo se tiene

$$\gamma_v(s, t) = \begin{cases} 3/9, & \text{si } s = t \\ 2/9, & \text{si } |s - t| = 1 \\ 1/9, & \text{si } |s - t| = 2 \\ 0, & \text{si } |s - t| \geq 3 \end{cases} \quad (0.89)$$

**Ejemplo 0.19** (Estacionaridad de un promedio móvil). El proceso de promedio móvil usado en los ejemplos 0.16 y 0.17 es estacionario ya que podemos escribir la función de autocovarianza obtenida en

(0.89) como

$$\gamma_v(h) = \begin{cases} 3/9, & \text{si } h = 0 \\ 2/9, & \text{si } h = \pm 1 \\ 1/9, & \text{si } h = \pm 2 \\ 0, & \text{si } |h| \geq 3 \end{cases}$$

**Ejemplo 0.20.** Un modelo para analizar tendencias es el camino aleatorio con tendencia dado por

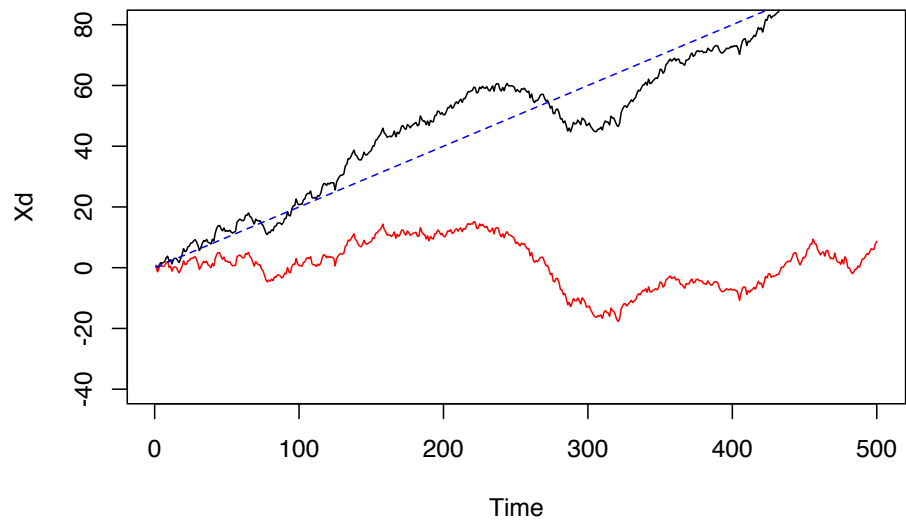
$$X_t = \delta + X_{t-1} + w_t \quad (0.90)$$

para  $t = 1, 2, \dots$ , con condición inicial  $X_0 = 0$ , y donde  $w_t$  es un ruido blanco. La constante  $\delta$  es llamada *tendencia*, y cuando  $\delta = 0$ , (0.90) es llamado simplemente *camino aleatorio*. El término camino aleatorio viene del hecho de que cuando  $\delta = 0$  el valor de la serie de tiempo en tiempo  $t$  es el valor de la serie de tiempo al tiempo  $t - 1$  más un movimiento completamente aleatorio determinado por  $w_t$ . La expresión (0.90) la podemos reescribir como una suma de variables de ruido blanco, esto es,

$$X_t = \delta t + \sum_{j=1}^N w_j \quad (0.91)$$

para  $t = 1, 2, \dots$ . A continuación generaremos un camino aleatorio usando **R**

```
set.seed(154)
w=rnorm(500,0,1)
X=cumsum(w)
wd=w+0.2; Xd=cumsum(wd)
plot.ts(Xd,ylim=c(-40,80))
lines(X,col="red")
lines(0.2*(1:500),lty="dashed",col="blue")
```



**Figura 11:** Gráficos de caminos aleatorios: con tendencia (negro), sin tendencia (rojo)



# 0

## Modelos AR

Los modelos autoregresivos están basados en la idea de que el valor actual de la serie  $x_t$  se puede explicar como una función de  $p$  valores pasados  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  donde  $p$  determina el número de pasos necesarios para predecir el valor actual. Una parte de las series de tiempo económicas y financieras suelen ser caracterizadas por los modelos autorregresivos. Entre los principales ejemplos de las finanzas tenemos valoración de precios y de dividendos, las tasas reales de cambio, tasas de interés y los diferenciales de tipos de interés (spreads).

**Definición 0.38.** Un *modelo autoregresivo de orden  $p$* , abreviado  $AR(p)$  es de la forma

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t, \quad (0.92)$$

donde  $x_t$  es estacionario,  $\phi_1, \phi_2, \dots, \phi_p$  son constantes ( $\phi_p \neq 0$ ). A menos que se declare lo contrario, se asume que  $w_t$  es un ruido blanco gaussiano de media cero y varianza  $\sigma_w^2$ . La media de  $x_t$  es (0.92) es cero. Si la media  $\mu$  de  $x_t$  no es cero, reemplazamos  $x_t$  por  $x_t - \mu$  en (0.92), es decir

$$x_t - \mu = \phi_1 (x_{t-1} - \mu) + \phi_2 (x_{t-2} - \mu) + \dots + \phi_p (x_{t-p} - \mu) + w_t$$

o escribimos

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t, \quad (0.93)$$

donde  $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ .

Note que (0.93) es similar al modelo de regresión dado en (0.64) y por consiguiente el término *autoregresión*. Sin embargo, se presentan algunas dificultades técnicas para la aplicación de este modelo, porque los regresores  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  son aleatorios, mientras que  $x_t$  se asume fijo. Una forma más útil se deriva de usar el siguiente operador de cambio dado por (0.52). Para escribir el modelo  $AR(p)$  como

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = w_t \quad (0.94)$$

o más conciso como

$$\phi(B)x_t = w_t. \quad (0.95)$$

Las propiedades de  $\phi(B)$  son importantes para resolver (0.95).

Esto nos lleva a la siguiente definición.

**Definición 0.39.** El *operador autoregresivo* de orden  $p$  se define como

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (0.96)$$

---

## 0.10. Modelo AR(1)

Iniciaremos el estudio de los modelos  $AR$  considerando el modelo de primer orden  $AR(1)$ , dado por  $x_t = \phi x_{t-1} + w_t$ . Iterando el operador de cambio  $k$  veces, obtenemos

$$\begin{aligned} x_t &= \phi x_{t-1} + w_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t \\ &= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\ &\vdots \\ &= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}. \end{aligned}$$

Este método sugiere que por iteración continua del operador de cambio, siempre que  $|\phi| < 1$  y  $x_t$  sea estacionario, podemos representar un modelo  $AR(1)$  como un proceso lineal dado por <sup>6</sup>

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} \quad (0.97)$$

El proceso  $AR(1)$  definido en (0.97) es estacionario con media

$$\mathbb{E}(x_t) = \sum_{j=0}^{\infty} \phi^j \mathbb{E}(w_{t-j}) = 0,$$

y la función de autocovarianza es

$$\begin{aligned} \gamma(h) &= Cov(x_{t+h}, x_t) \\ &= \mathbb{E} \left[ \left( \sum_{j=0}^{\infty} \phi^j w_{t+h-j} \right) \left( \sum_{k=0}^{\infty} \phi^k w_{t-k} \right) \right] \\ &= \sigma_w^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} \\ &= \frac{\sigma_w^2 \phi^h}{1 - \phi^2}, h > 0 \end{aligned} \quad (0.98)$$

Recuerde que  $\gamma(h) = \gamma(-h)$  de modo que basta presentar la función de autocovarianza para  $h \geq 0$ .

Si en (0.98), hacemos  $h = 0$ , obtenemos la varianza del proceso  $AR(1)$ , siendo esta

$$Var(x_t) = \frac{\sigma_w^2}{1 - \phi^2},$$

asumiendo que  $\phi_1^2 < 1$ . El requisito de que  $\phi_1^2 < 1$  resulta del hecho de que la varianza de una variable aleatoria es acotada y no

---

<sup>6</sup>Note que  $\lim_{k \rightarrow \infty} \mathbb{E}(x_t - \sum_{j=0}^{\infty} \phi^j w_{t-j})^2 = \lim_{k \rightarrow \infty} \phi^{2k} \mathbb{E}(x_{t-k}^2) = 0$ , de modo que (0.97) existe en el sentido de media cuadrado.

negativa. Por consiguiente, la estacionaridad de un modelo  $AR(1)$  implica que  $-1 < \phi_1 < 1$ . Pero si  $-1 < \phi_1 < 1$ , entonces por (0.97) y la independencia de  $\{w_t\}$  se puede demostrar que la media y la varianza de  $x_t$  son finitas. Además, por la desigualdad de Cauchy-Schwartz todas las autocovarianzas de  $x_t$  son finitas. Por lo tanto, el modelo  $AR(1)$  es estacionario. En resumen, una condición necesaria y suficiente para que un proceso  $AR(1)$  sea estacionario es  $|\phi_1| < 1$ .

De (0.98) la ACF de un modelo  $AR(1)$  es

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h > 0 \quad (0.99)$$

y  $\rho(h)$  satisface la recursión

$$\rho(h) = \phi \rho(h-1), \text{ para } h = 1, 2, \dots \quad (0.100)$$

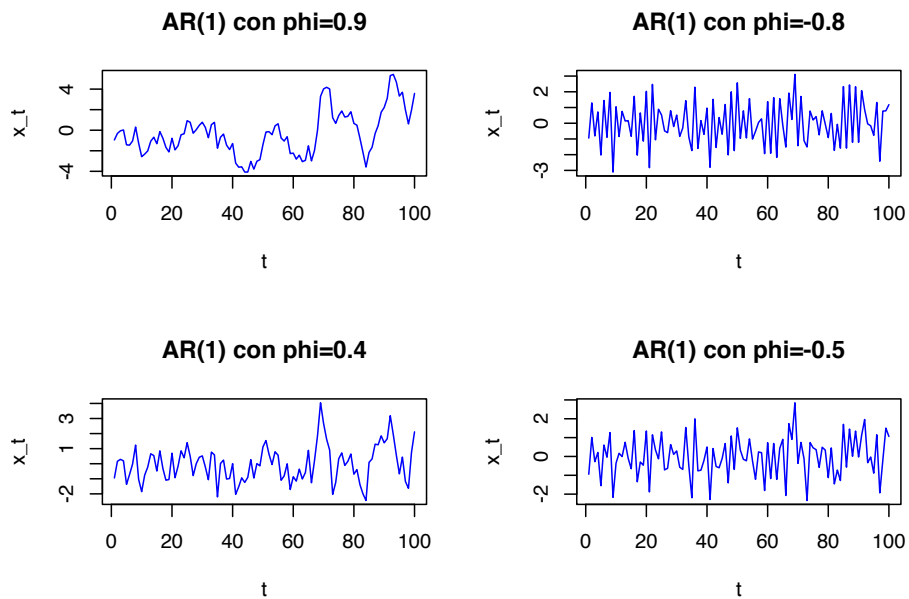
Las ecuaciones (0.99) y (0.100) indican que la ACF de un modelo  $AR(1)$  estacionario tiene un decaimiento exponencial con tasa igual a  $\phi_1$ . Si  $\phi_1 > 0$  el decaimiento es constante. Si por el contrario,  $\phi_1 < 0$  entonces el decaimiento es compuesto y se presenta de forma alternante con tasa  $\phi_1^2$ . Para tener una idea de esto, consideremos los modelos autoregresivos de orden 1 simulados, para distintos valores de  $\phi_1$ .

```
# Coeficientes phi
phi1=0.9
phi2=-0.8
phi3=0.4
phi4=-0.5
# Ruido blanco gaussiano
w=rnorm(100,0,1)
# Series AR(1)
ar1_1=filter(w,filter = phi1,method = "recursive")
ar1_2=filter(w,filter = phi2,method = "recursive")
ar1_3=filter(w,filter = phi3,method = "recursive")
```

```

ar1_4=filter(w,filter = phi4,method = "recursive")
# Graficos
par(mfrow=c(2,2))
plot.ts(ar1_1, col="blue",type = "l",
        main = "AR(1) con phi=0.9",xlab="t",ylab="x_t")
plot.ts(ar1_2, col="blue",type = "l",
        main = "AR(1) con phi=-0.8",xlab="t",ylab="x_t")
plot.ts(ar1_3, col="blue",type = "l",
        main = "AR(1) con phi=0.4",xlab="t",ylab="x_t")
plot.ts(ar1_4, col="blue",type = "l",
        main = "AR(1) con phi=-0.5",xlab="t",ylab="x_t")

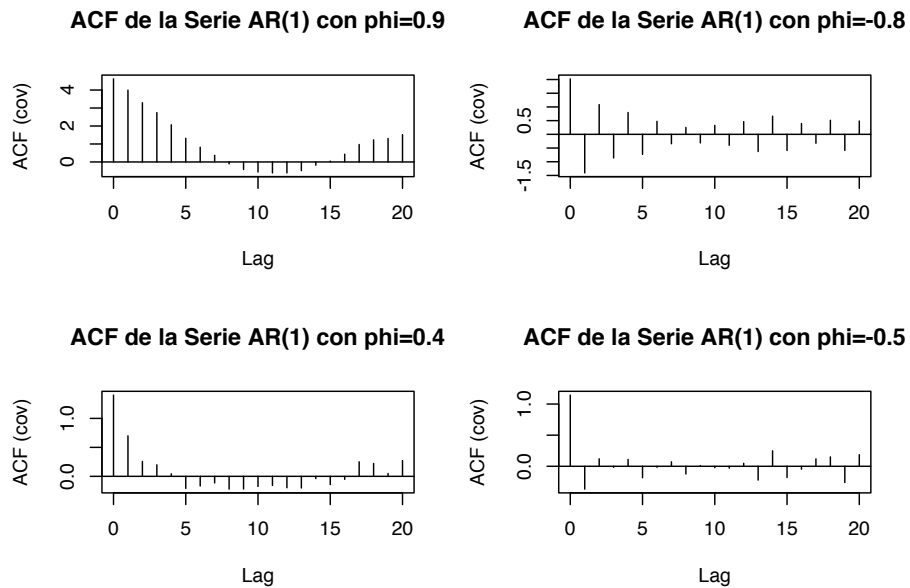
```



**Figura 12:** Simulaciones de procesos autoregresivos de orden 1,  $AR(1)$ , para distintos valores de  $\phi$

A continuación mostramos las funciones de autocovarianzas de las series  $AR(1)$  simuladas anteriormente

```
par(mfrow=c(2,2))
acf(ar1_1,type = "covariance", main="ACF de la Serie AR(1) con phi=0.9")
acf(ar1_2,type = "covariance", main="ACF de la Serie AR(1) con phi=-0.8")
acf(ar1_3,type = "covariance", main="ACF de la Serie AR(1) con phi=0.4")
acf(ar1_4,type = "covariance", main="ACF de la Serie AR(1) con phi=-0.5")
```



**Figura 13:** Funciones de autocovarianzas para las series AR(1) simuladas

### 0.11. Modelo AR(2)

Un proceso  $AR(2)$  tiene la forma general

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t \quad (0.101)$$

siendo  $\alpha = \mu(1 - \phi_1 - \phi_2)$ , con  $\phi_2 \neq 0$ . Podemos calcular su función de media

$$\begin{aligned}\mathbb{E}(x_t) &= \mathbb{E}(\alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t) \\ &= \alpha + \phi_1 \mathbb{E}(x_{t-1}) + \phi_2 \mathbb{E}(x_{t-2})\end{aligned}$$

Por estacionalidad, se tiene que  $\mathbb{E}(x_t) = \mathbb{E}(x_{t-1}) = \mathbb{E}(x_{t-2})$ , luego

$$\mathbb{E}(x_t)(1 - \phi_1 - \phi_2) = \alpha$$

Así,  $\mathbb{E}(x_t) = \frac{\alpha}{1 - \phi_1 - \phi_2}$ , siempre que  $\phi_1 + \phi_2 \neq 1$ . Usando  $\alpha = (1 - \phi_1 - \phi_2)\mu$  podemos reescribir el proceso  $AR(2)$  como

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + w_t.$$

Multiplicando por  $x_{t-h} - \mu$ , tenemos

$$(x_{t-h} - \mu)(x_t - \mu) = \phi_1(x_{t-h} - \mu)(x_{t-1} - \mu) + \phi_2(x_{t-h} - \mu)(x_{t-2} - \mu) + (x_{t-h} - \mu)w_t.$$

Tomando valor esperado y usando el hecho de que

$$\mathbb{E}[(x_{t-h} - \mu)w_t] = 0, \text{ para } h > 0, \text{ obtenemos}$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2), \text{ para } h > 0. \quad (0.102)$$

Este último resultado se conoce como la **ecuación de momentos** de un proceso estacionario  $AR(2)$ . Dividiendo (0.102) por  $\gamma(0)$ , tenemos la propiedad

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2), \text{ para } h > 0 \quad (0.103)$$

para la ACF de  $x_t$ . En particular, para paso 1 ( $h = 1$ ) la ACF satisface

$$\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(-1) = \phi_1 + \phi_2 \rho(1)$$

Por lo tanto, para un proceso  $AR(2)$  estacionario  $x_t$ , tenemos

$$\begin{aligned}
\rho(0) &= 1 \\
\rho(1) &= \frac{\phi_1}{1 - \phi_2} \\
\rho(h) &= \phi_1 \rho(h-1) + \phi_2 \rho(h-2), \quad h \geq 2
\end{aligned}$$

El resultado de la ecuación (0.103) nos dice que la ACF de un proceso estacionario  $AR(2)$  satisface la ecuación en diferencias de segundo orden

$$(1 - \phi_1 B - \phi_2 B^2)\rho(h) = 0 \quad (0.104)$$

donde  $B$  es el operador definido en (0.52). La ecuación (0.104) determina las propiedades de la ACF de un proceso  $AR(2)$  estacionario. También determina el comportamiento de los pronósticos de  $x_t$ . Correspondiendo a la ecuación en diferencias anterior, existe una ecuación polinómica de segundo orden

$$x^2 - \phi_1 x - \phi_2 = 0 \quad (0.105)$$

Las soluciones de esta ecuación son las raíces características de un proceso  $AR(2)$  y estas son

$$x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

Denotamos las dos raíces por  $r_1$  y  $r_2$ . Si ambos son reales, entonces la ecuación en diferencias de segundo orden la podemos factorizar como

$$(1 - r_1 B)(1 - r_2 B)$$

y el proceso  $AR(2)$  lo podemos considerar como un proceso  $AR(1)$  operando sobre otro proceso  $AR(1)$ .

La ACF de  $x_t$  es entonces una mezcla de dos decaimientos



exponenciales. Pero si  $\phi_1^2 + 4\phi_2 < 0$ , entonces  $r_1$  y  $r_2$  son raíces complejas conjugadas, y el gráfico de la ACF de  $x_t$  mostrará un amortiguamiento de senos y cosenos.

En aplicaciones financieras y económicas, las raíces características complejas son importantes. Dan lugar al comportamiento de los ciclos económicos. Por lo tanto, es común que los modelos económicos de series de tiempo tengan raíces características de valor complejo. Para un proceso  $AR(2)$  dado por (0.101) con raíces características complejas, la longitud *promedio* de un ciclo estocástico es

$$k = \frac{360^\circ}{\arccos(\phi_1/2\sqrt{-\phi_2})},$$

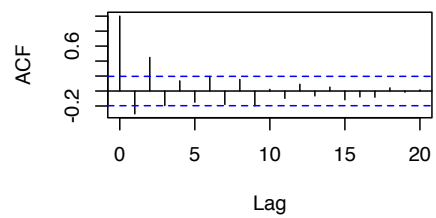
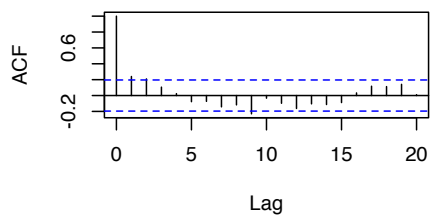
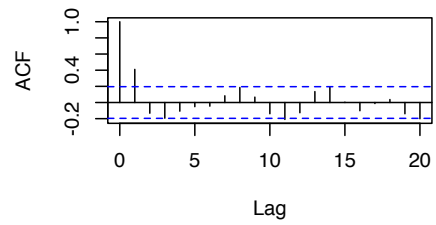
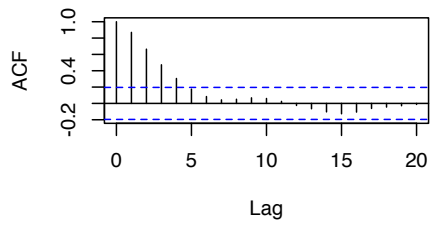
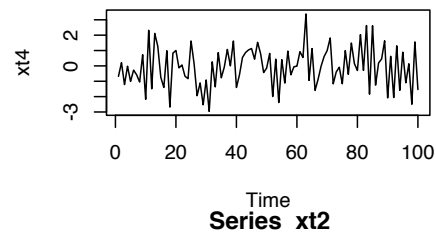
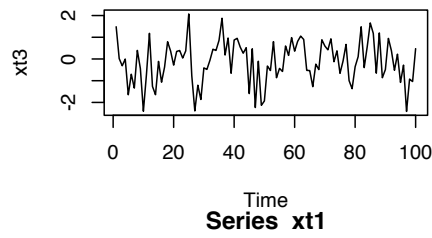
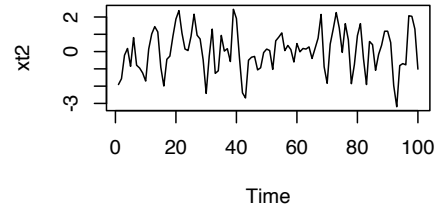
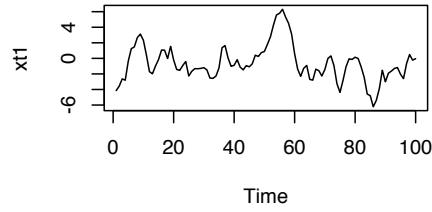
donde el arcocoseno está expresado en grados.

La figura siguiente muestra la ACF de 4 procesos estacionarios  $AR(2)$ . Los procesos  $AR(2)$  mostrados son:

- a)  $x_t = 1,2x_{t-1} - 0,35x_{t-2} + w_t$
- b)  $x_t = 0,6x_{t-1} - 0,4x_{t-2} + w_t$
- c)  $x_t = 0,2x_{t-1} + 0,35x_{t-2} + w_t$
- d)  $x_t = -0,2x_{t-1} + 0,35x_{t-2} + w_t$

CVI

*Modelos AR*



La serie (b) tiene raíces características complejas, en efecto

$$\phi_1^2 + 4\phi_2 = (0,6)^2 + 4 \times (-0,4) = 1,24 < 0$$

Se puede notar que en el gráfico de la ACF que este exhibe un comportamiento de ondas de senos y cosenos. Los otros 3 procesos  $AR(2)$  tienen raíces características reales, por lo que las ACF

decaen exponencialmente. La condición de estacionaridad de un proceso  $AR(2)$  es que los valores absolutos de sus raíces características sean menor que uno, esto es  $|\phi_1| < 1, |\phi_2| < 1$ . Bajo esta condición, la ecuación recursiva (0.103) asegura que la ACF del proceso converge a cero cuando el salto  $h$  crece. Esta propiedad de convergencia es una condición necesaria para una serie de tiempo estacionaria. De hecho, la condición también aplica para un proceso  $AR(1)$  donde la ecuación polinómica es  $x - \phi_1 = 0$ . La raíz característica es  $x = \phi_1$ , la cual debe ser menor que uno en módulo para que  $x_t$  sea estacionario. Como mostramos antes, para un proceso estacionario  $AR(1)$  la ACF es  $\rho(h) = \phi^h$ , (0.99).

Así, la condición  $|\phi| < 1$ , asegura que  $\rho(h) = \phi^h \rightarrow 0$ , cuando  $h \rightarrow \infty$ .

---

## 0.12. Procesos $AR(p)$

Los resultados de los procesos  $AR(1)$  y  $AR(2)$ , los podemos generalizar a procesos  $AR(p)$ . Así, la función de media del proceso  $AR(p)$  estacionario será

$$\mathbb{E}(x_t) = \frac{\alpha}{1 - \phi_1 - \dots - \phi_p} \quad (0.106)$$

siempre que el denominador sea distinto de cero. La ecuación polinómica asociada al modelo es

$$x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \dots - \phi_p = 0 \quad (0.107)$$

la cual nos referimos como la *ecuación característica* del modelo.

Si todas las *raíces características* de esta ecuación son menores que uno en módulo, esto es  $|r_j| < 1$ , con  $j = 1, \dots, p$ , entonces la serie  $x_t$  es estacionaria. Para un proceso  $AR(p)$  estacionario, la ACF satisface la ecuación en diferencias

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) \rho(h) = 0, \text{ para } h > 0.$$

El gráfico de la ACF de un proceso  $AR(p)$  estacionario mostrará una mezcla de ondas de senos y cosenos con decaimientos exponenciales dependiendo de la naturaleza de sus raíces características.

**Ejemplo 0.21.** Consideremos el modelo  $AR(3)$  de la forma

$$x_t = 0,0047 + 0,35x_{t-1} + 0,18x_{t-2} - 0,14x_{t-3} + w_t.$$

Reescribiendo el proceso como

$$x_t - 0,35x_{t-1} - 0,18x_{t-2} + 0,14x_{t-3} = 0,0047 + w_t$$

obtenemos la correspondiente ecuación en diferencias de orden 3,

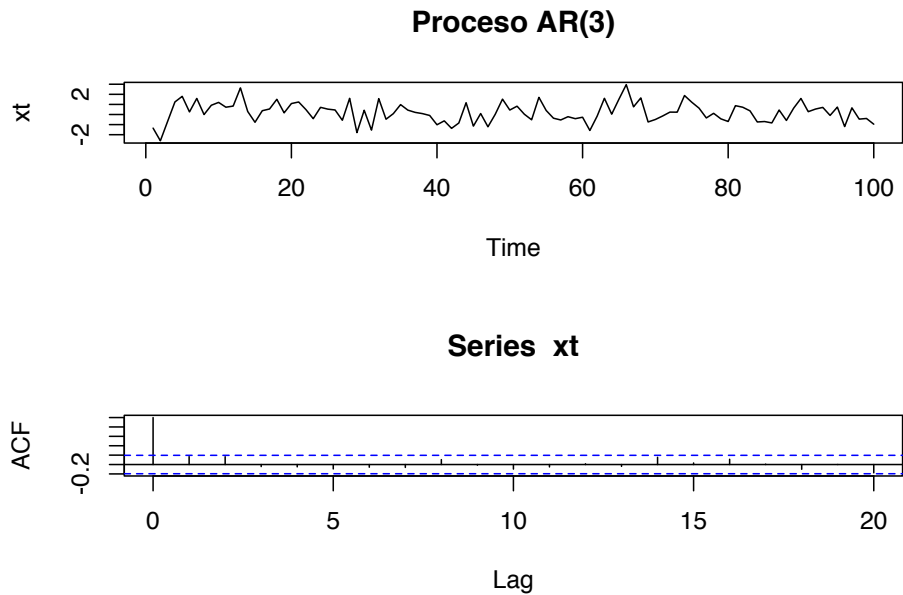
$$(1 - 0,35B - 0,18B^2 + 0,14B^3) = 0$$

la cual podemos factorizar como

$$(1 + 0,52B)(1 - 0,87B + 0,27B^2) = 0$$

El primer factor  $(1 + 0,52B) = 0$ , muestra un decaimiento exponencial en la ACF. Veamos ahora el segundo factor  $(1 - 0,87B - (-0,27)B^2) = 0$ , tenemos que  $\phi_1^2 + 4\phi_2 = (0,87)^2 + 4(-0,27) = -0,3231 < 0$ . Por consiguiente la ACF mostrará un comportamiento en ondas de senos y cosenos.

```
xt<-arima.sim(list(order=c(3,0,0),ar=c(0.35,0.18,-0.14)),n=100)
par(mfrow=c(2,1))
plot(xt,type="l",main="Proceso AR(3)")
acf(xt)
```



En aplicaciones, se desconoce el orden  $p$  de una serie de tiempo autoregresiva, por lo que debe especificarse empíricamente. Esto se conoce como la determinación del orden de los modelos  $AR$ . Existen dos enfoques generales para determinar el valor de  $p$ . El primero es utilizar la función de autocorrelación parcial, y el segundo utilizar alguna función de criterio de información. En **R** existe una función para determinar el orden  $p$  de un proceso  $AR$ , sin embargo vamos a estudiar primero los dos enfoques mencionados de manera de poder entender como funciona **R** al respecto.

### 0.13. Función de Autocorrelación Parcial

La función de autocorrelación parcial PACF (siglas en inglés: Partial Autocorrelation Function) de una serie de tiempo es una función de su ACF y es una herramienta útil para determinar el orden  $p$  de un modelo autoregresivo. Una manera simple pero

efectiva de introducir la PACF es considerando los siguientes modelos *AR* en órdenes consecutivos:

$$\begin{aligned}x_t &= \phi_{0,1} + \phi_{1,1}x_{t-1} + w_{1t} \\x_t &= \phi_{0,2} + \phi_{1,2}x_{t-1} + \phi_{2,2}x_{t-2} + w_{2t} \\x_t &= \phi_{0,3} + \phi_{1,3}x_{t-1} + \phi_{2,3}x_{t-2} + \phi_{3,3}x_{t-3} + w_{3t} \\&\vdots\end{aligned}$$

donde  $\phi_{0,j}$ ,  $\phi_{i,j}$ ,  $\{w_{jt}\}$  son respectivamente, el término constante, el coeficiente de  $x_{t-j}$  y el término de error del modelo  $AR(j)$ . Estos modelos están en la forma de regresión lineal múltiple y se pueden estimar por mínimos cuadrados.

La estimación de  $\hat{\phi}_{1,1}$  de la primera ecuación se llama *PACF muestral de paso 1* de  $x_t$ . La estimación de  $\hat{\phi}_{2,2}$  de la segunda ecuación es la PACF muestral de paso 2 de  $x_t$ . La estimación  $\hat{\phi}_{3,3}$  de la tercera ecuación es la PACF muestral de paso 3 de  $x_t$ , y así sucesivamente.

De la definición, la PACF de paso 2 muestra la contribución añadida de  $x_{t-2}$  a  $x_t$  sobre el modelo  $AR(1)$   $x_t = \phi_0 + \phi_1x_{t-1} + w_{1t}$ . La PACF de paso 3 muestra la contribución añadida de  $x_{t-3}$  a  $x_t$  sobre un modelo  $AR(2)$ , etc. Por lo tanto, para un modelo  $AR(p)$ , la PACF de paso  $p$  no debería ser cero, sino que  $\hat{\phi}_{j,j}$  debería ser cercano a cero para todo  $j > p$ . Esta propiedad será útil para determinar el orden  $p$ . De hecho, bajo ciertas condiciones de regularidad, se puede demostrar que la PACF muestral de un proceso  $AR(p)$  tiene las siguientes propiedades:

- $\hat{\phi}_{p,p}$  converge a  $\phi_p$  cuando el tamaño  $N$  de la muestra tiende a infinito.
- $\hat{\phi}_{l,l}$  converge a cero para todo  $l > p$ .
- La varianza asintótica de  $\hat{\phi}_{l,l}$  es  $1/N$  para  $l > p$ .

Estos resultados dicen que para una serie  $AR(p)$  la PACF muestral se corta en paso o salto  $p$ .

Para concluir esta sección, destaquemos que tanto la función de autocorrelación (ACF) como la función de autocorrelación parcial (PACF) nos permiten determinar el orden de un modelo  $AR(p)$ . El cuadro siguiente indica como usarlas

	ACF	PACF
AR(p)	Disminución gradual	Corte en paso p

### 0.14. Criterios de Información

Existen diversos criterios de información disponibles para determinar el orden  $p$  de un proceso autoregresivo. Todos ellos están basados en verosimilitud. El primer criterio ya lo definimos en 0.24

**Definición 0.40.** El Criterio de Información de Akaike se define como

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n + 2k}{n} \tag{0.108}$$

donde  $\hat{\sigma}_k^2$  es el estimador de máxima verosimilitud para la varianza,  $k$  es el número de parámetros en el modelo y  $n$  es el tamaño de la muestra <sup>7</sup>

El **Criterio de Información de Akaike** (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la

<sup>7</sup>Formalmente, el AIC se define como

$$-2 \ln L_k + 2k$$

donde  $L_k$  es la verosimilitud maximizada y  $k$  es el número de parámetros del modelo.

selección del modelo. El AIC maneja un trade-off <sup>8</sup> entre la bondad de ajuste del modelo y la complejidad del modelo. Se basa en la entropía de información: se ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos.

El AIC no proporciona una prueba de un modelo en el sentido de probar una hipótesis nula, es decir el AIC puede decir nada acerca de la calidad del modelo en un sentido absoluto. Si todos los modelos candidatos encaja mal, el AIC notará ningún aviso de ello. El valor de  $k$  que minimiza el AIC especifica el mejor modelo. La idea es que la minimización de  $\hat{\sigma}_k^2$  sea razonablemente objetiva, excepto que decrezca monótonamente cuando  $k$  crece. Por lo tanto, debemos penalizar la variación del error por un término proporcional al número de parámetros. La elección del término de penalización dado por (0.108) no es único.

En el criterio AIC definido en la ecuación (0.108), el sesgo es aproximado por el número de parámetros los cuales son constantes y no tienen variabilidad. Para el modelo de regresión, la corrección del sesgo el logaritmo de la verosimilitud se define como

**Definición 0.41.**

$$AICc = \ln \hat{\sigma}_k^2 + \frac{n + k}{n - k - 2} \quad (0.109)$$

donde  $\hat{\sigma}_k^2$  es el estimador de máxima verosimilitud de la varianza,  $k$  es el número de parámetros en el modelo y  $n$  es el tamaño de la muestra.

---

<sup>8</sup>*Trade-off* o simplemente *tradeoff*, en algunas ocasiones traducido al español como *sacrificio*, es un anglicismo que describe una situación en la cual se debe perder cierta cualidad a cambio de otra cualidad. Un tradeoff se puede dar por varias razones, entre ellas por simples limitaciones de la física (dentro de una cantidad de espacio dada se pueden meter muchos objetos pequeños o una menor cantidad de objetos grandes). La idea de un tradeoff como una decisión por lo general implica que ésta es realizada con una comprensión total de las ventajas y desventajas de la decisión en particular, como por ejemplo, es el caso cuando una persona decide invertir en acciones de una empresa (una inversión más riesgosa pero con mayor potencial) sobre bonos (por lo general más seguros pero con menor potencial de ganancias).



Este modelo fue propuesto originalmente por *N. Sugiura* en el artículo “**Further analysis of the data by Akaike’s information criterion and the finite corrections**”. *Communications in Statistics, Theory and Methods*, Vol. 7, No. 1, pp. 13-26, 1978. Se tiene que el AICc es esencialmente el AIC con un término de penalización adicional para el número de parámetros. Nótese que cuando  $n \rightarrow \infty$ , el término de penalización adicional converge a 0, y por lo tanto el AICc converge al AIC. De manera similar que en el AIC, se selecciona el modelo con el menor valor AICc.

**Definición 0.42.**

$$SIC = \ln \hat{\sigma}_k^2 + \frac{k \ln n}{n} \quad (0.110)$$

usando la misma notación que en la definición 0.41.

SIC también llamado *Criterio Bayesiano de Información* (BIC) es un criterio para la selección de modelos. Se basa, en parte, en la función de probabilidad y está estrechamente relacionado con el Criterio de Información de Akaike (AIC). Cuando se hace ajuste de modelos, es posible aumentar la probabilidad mediante la adición de parámetros, pero esto puede resultar en sobreajuste.

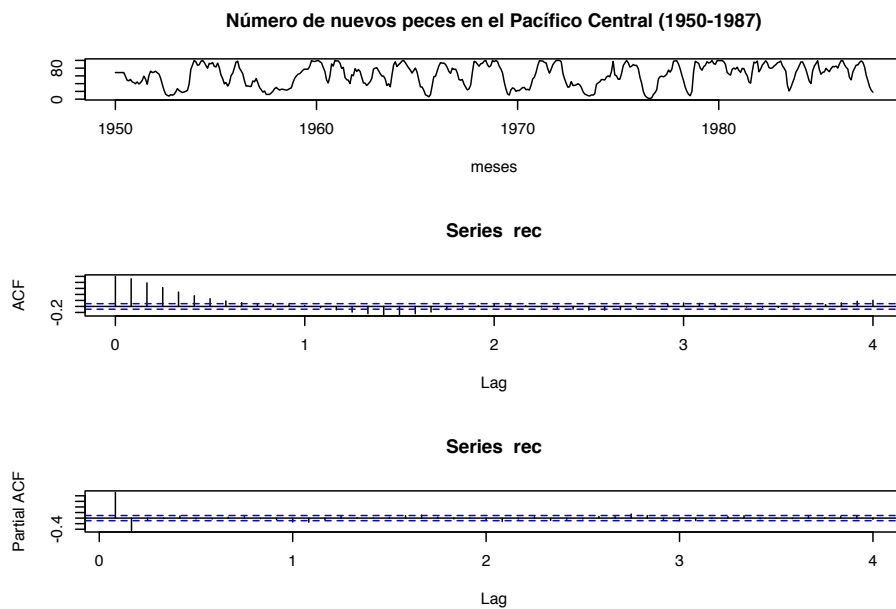
Tanto BIC como AIC resuelven este problema mediante la introducción de un término de penalización para el número de parámetros en el modelo, el término de penalización es mayor en el BIC que en el AIC.

El BIC fue desarrollado por *Gideon E. Schwarz*, quien dio un argumento bayesiano a favor de su adopción. Akaike también desarrolló su propio formalismo bayesiano, que ahora se conoce como la ABIC (siglas en inglés) por Criterio de Información Bayesiano de Akaike.

**Ejemplo 0.22.** Consideremos el archivo de datos “recruit.txt”, consistente del número de nuevos peces en el Pacífico Central relacionado con el Índice de Oscilación del Sur (SOI) que mide

los cambios en la presión del aire relativo a la temperatura de la superficie del mar. Son 453 registros de 1950 a 1987 tomados mensualmente. La ACF y la PACF indican que el modelo que mejor se ajusta es un  $AR(2)$ . La ACF tiene ciclos correspondientes a aproximadamente períodos de 12 meses y la PACF tiene valores grandes para  $h = 1, 2$  y es esencialmente cero para paso de orden mayor.

```
rec=ts(scan("data/recruit.txt"),start = 1950, frequency = 12)
par(mfrow=c(3,1))
plot(rec,type = "l", ylab = "", xlab = "meses", main = "Número de nuevos peces en el Pacífico Central (1950-1987)")
acf(rec,48)
pacf(rec,48)
```



### 0.15. Estimación de Parámetros.

Para un modelo  $AR(p)$  especificado por (0.92), el método de mínimos cuadrados condicional, el cual inicia con  $p + 1$  observaciones, se usa a menudo para estimar los parámetros. Específicamente, condicionando sobre las primeras  $p$  observaciones, tenemos

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t, \quad t = p + 1, \dots, N,$$

el cual se puede estimar por mínimos cuadrados. Denotemos los estimadores de  $\phi_i$  por  $\hat{\phi}_i$ . El *modelo ajustado* será

$$\hat{x}_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \cdots + \hat{\phi}_p x_{t-p}$$

y el residual asociado es

$$\hat{w}_t = x_t - \hat{x}_t$$

La serie  $\{\hat{w}_t\}$  se llama *serie residual*, de la cual obtenemos

$$\hat{\sigma}_w^2 = \frac{\sum_{t=p+1}^N \hat{w}_t^2}{N - 2p - 1}$$

Debemos examinar con cuidado un modelo ajustado para verificar si es adecuado o no. Si el modelo es adecuado, entonces la serie residual debe comportarse como un ruido blanco. La ACF y el estadístico de Ljung-Box <sup>9</sup> de los residuos son útiles para

<sup>9</sup>El estadístico de Ljung-Box está dado por

$$Q(m) = N(N+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2}{N-i}$$

donde  $\hat{\rho}_1, \hat{\rho}_2, \dots$ , son las *funciones de autocorrelación muestral* (ACF) de  $x_t$ ,  $N$  el tamaño de la muestra.

comprobar la cercanía de  $\hat{w}_t$  a un ruido blanco. Para un modelo  $AR(p)$ , el estadístico de Ljung-Box  $Q(m)$  se distribuye asintóticamente como una Chi-Cuadrado con  $m - p$  grados de libertad, para significar que los  $p$  coeficientes del proceso AR son estimados. Si descubrimos que un modelo ajustado es inadecuado, debemos refinarlo.

---

### 0.16. Predicciones con modelos AR

La predicción es una aplicación importante del análisis de series de tiempo. Para el modelo  $AR(p)$  en la ecuación (0.92), supongamos que estamos en tiempo  $m$  y estamos interesados en predecir  $x_{m+h}$ , donde  $h \geq 1$ . El tiempo  $m$  se llama el *origen de predicción* y el entero  $h$  es el *horizonte de predicción*. Sea  $\hat{x}_m(h)$  la predicción de  $x_{m+h}$  usando la función de pérdida de error mínimo al cuadrado.

En otras palabras, escogemos la predicción  $\hat{x}_m(h)$  tal que

$$\mathbb{E}(x_{m+h} - \hat{x}_m(h)) \leq \min_g \mathbb{E}(x_{m+h} - g)^2$$

donde  $g$  es una función de información disponible pen tiempo  $m$ . Nos referimos a  $\hat{x}_m(h)$  como la predicción de  $h$  pasos de  $x_t$  con origen de predicción  $m$ .

#### 0.16.1. Predicción de un paso

Del modelo  $AR(p)$  tenemos

$$x_{m+1} = \phi_0 + \phi_1 x_m + \cdots + \phi_p x_{m+1-p} + w_{m+1}$$

Bajo la función de pérdida de error mínimo al cuadrado, la predicción puntual  $x_{m+1}$  dado el modelo y las observaciones hasta tiempo  $m$ , es el valor esperado condicional

$$\hat{x}_m(1) = \mathbb{E}(x_{m+1}|x_m, x_{m-1}, \dots) = \phi_0 + \sum_{i=1}^p \phi_i x_{m+1-i}$$

y el error de predicción asociado es

$$e_m(1) = x_{m+1} - \hat{x}_m(1) = w_{m+1}$$

Consecuentemente, la varianza del error de predicción de un paso es  $Var(e_m(1)) = Var(w_{m+1}) = \sigma_w^2$ . Si  $w_t$  se distribuye como una normal, entonces un intervalo de predicción de un paso del 95 % de confianza está dado por

$$\hat{x}_m(1) \pm 1,96\sigma_w^2.$$

### 0.16.2. Predicción de dos pasos

Ahora consideremos la predicción de  $x_{m+2}$  con origen de predicción  $m$ . De un modelo  $AR(p)$ , tenemos

$$x_{m+2} = \phi_0 + \phi_1 x_{m+1} + \dots + \phi_p x_{m+2-p} + w_{m+2}$$

Tomando valor esperado condicional, tenemos

$$\begin{aligned} \hat{x}_m(2) &= \mathbb{E}(x_{m+2}|x_m, x_{m-1}, \dots) \\ &= \phi_0 + \phi_1 \hat{x}_m(1) + \phi_2 x_m + \dots + \phi_p x_{m+2-p} \end{aligned}$$

y el error de predicción asociado es

$$\begin{aligned} e_m(2) &= x_{m+2} - \hat{x}_m(2) = \phi_1 [x_{m+1} - \hat{x}_m(1)] + w_{m+2} \\ &= w_{m+2} - \phi_1 w_{m+1} \end{aligned}$$

La varianza del error de predicción es  $Var(e_m(2)) = (1 + \phi_1^2)\sigma_w^2$ . Un intervalo de predicción para  $x_{m+2}$  lo podemos calcular de la misma manera que para  $x_{m+1}$ . Es interesante notar que  $Var(e_m(2)) \geq Var(e_m(1))$ , lo que indica que cuando el horizonte de predicción crece la incertidumbre de la predicción también crece.

### 0.16.3. Predicción de múltiples pasos

En general, tenemos

$$x_{m+h} = \phi_0 + \phi_1 x_{m+h-1} + \cdots + \phi_p x_{m+h-p} + w_{m+h}$$

La predicción de  $h$  pasos basado en la función de pérdida de error mínimo al cuadrado es el valor esperado condicional de  $x_{m+h}$  dado  $\{x_{m-i}\}_{i=0}^{\infty}$ , el cual podemos escribir como

$$\hat{x}_m(h) = \phi_0 + \sum_{i=1}^p \phi_i \hat{x}_m(h-i)$$

donde se entiende que  $\hat{x}_m(i) = x_{m+i}$  si  $i < 0$ . Esta predicción se puede calcular recursivamente usando las predicciones  $\hat{x}_m(i)$  para  $i = 1, \dots, h-1$ . El error de predicción de paso  $h$  es  $e_m(h) = x_{m+h} - \hat{x}_m(h)$ . Se puede demostrar que para un proceso  $AR(p)$  estacionario,  $\hat{x}_m(h)$  converge a  $\mathbb{E}(x_t)$  cuando  $h \rightarrow \infty$ , esto es, para una serie  $AR(p)$  estacionaria, la predicción a largo plazo se aproxima a su media incondicional. Esta propiedad se conoce como la *reversión media* en la literatura financiera. La desviación del error de predicción se aproxima entonces a la desviación incondicional de  $x_t$ .

**Ejemplo 0.23.** Realicemos una predicción para la serie de nuevos peces dado en el ejemplo 0.22. Para ello usaremos la función “predict” de R. Vamos a hacer una predicción de 24 meses. Del ejemplo 0.22, pudimos notar en las ACF y PACF que un modelo que se ajusta a esta serie es un proceso  $AR(2)$ , así que lo primero que hacemos es ajustar el modelo, y luego calculamos la predicción.

Al realizar el gráfico de la serie y la predicción notamos que después de 12 meses, la predicción converge a la media de la serie tal como se describió previamente.

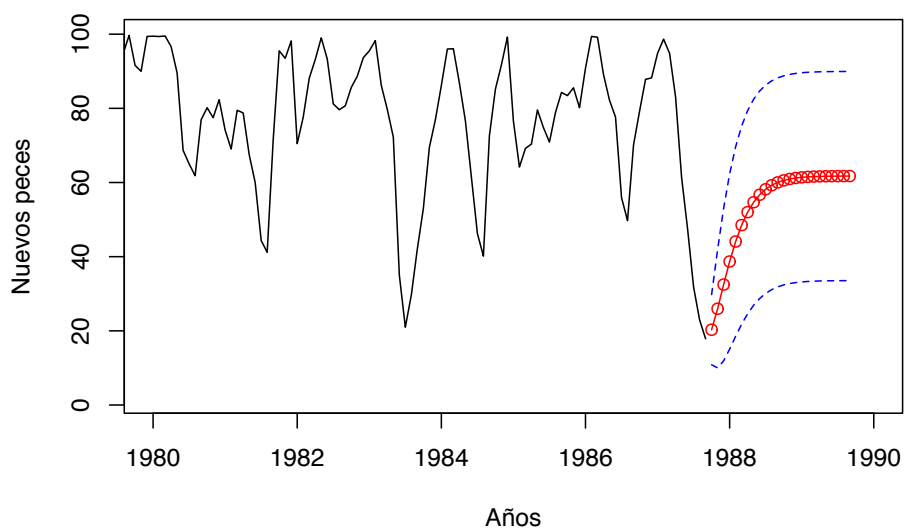
```
# Ajuste AR(2)
regr=ar.ols(rec,order=2,demean = FALSE, intercept = TRUE)
regr$asy.se.coef
```

```
## $x.mean
## [1] 1.111
##
## $ar
## [1] 0.04179 0.04188
```

```
# Prediccion
fore=predict(regr, n.ahead=24)
```

```
## Warning in object$var.pred * vars: Recycling array of length 1 in array-v
## Use c() or as.vector() instead.
```

```
ts.plot(rec,fore$pred,col=1:2,xlim=c(1980,1990),
        xlab="Años", ylab="Nuevos peces")
lines(fore$pred,type = "p",col=2)
lines(fore$pred+fore$se,lty = "dashed",col=4)
lines(fore$pred-fore$se,lty = "dashed",col=4)
```







# 0

## Modelos $MA$

En este capítulo describiremos otra clase de modelos simples que también son útiles en el modelado de series de retorno en finanzas. Estos modelos se denominan modelos de promedio móvil ( $MA$ , siglas en inglés: Moving Average). Hay varias maneras de introducir los modelos  $MA$ . Un enfoque es tratar el modelo como una extensión de una serie de ruido blanco; alternatively a la representación autoregresiva en la cual  $x_t$  del lado izquierdo se asume como una combinación lineal, en los modelos de promedio móvil e orden  $q$   $MA(q)$ , asumimos el ruido blanco  $w_t$  del lado derecho de la ecuación que los define como una combinación lineal de los datos observados.

**Definición 0.43.** El **modelo de promedio móvil de orden  $q$**  o modelo  $MA(q)$ , se define como

$$x_t = w_t + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_q x_{t-q} \quad (0.111)$$

donde hay  $q$  pasos o saltos en el promedio móvil y  $\theta_1, \theta_2, \dots, \theta_q$  ( $\theta_q \neq 0$ ) son parámetros.<sup>10</sup> El ruido  $w_t$  se asume como un ruido blanco gaussiano.

---

Podemos escribir también el modelo  $MA(q)$  en la forma equivalente

---

<sup>10</sup> Algunos libros y algunos paquetes estadísticos escriben el modelo  $MA$  con coeficientes negativos, esto es

$$x_t = w_t - \theta_1 x_{t-1} - \theta_2 x_{t-2} - \cdots - \theta_q x_{t-q}$$

$$x_t = \theta(B)w_t \quad (0.112)$$

usando la siguiente definición.

**Definición 0.44.** El **operador de promedio móvil** se define como

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \quad (0.113)$$

---

A diferencia del proceso autoregresivo  $AR(p)$ , el proceso de promedio móvil  $MA(q)$  es estacionario para cada valor de los parámetros  $\theta_1, \theta_2, \dots, \theta_q$ .

Otro enfoque para introducir los proceso de promedio móvil es tratar el modelo como un modelo  $AR$  de orden infinito con algunas restricciones. No hay una razón en particular, pero por simplicidad y sin pérdida de generalidad, asumiremos *a priori* que el orden del modelo es finito. Podemos suponer, al menos en teoría, un modelo  $AR$  con orden infinito como

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + w_t$$

Sin embargo, tal modelo  $AR$  no es realista porque tiene infinitos parámetros. Una forma de hacer práctico el modelo es asumir que los coeficientes  $\phi_i$  satisfacen algunas restricciones para que sean determinados por un número finito de parámetros. Un caso especial de esta idea es

$$x_t = \phi_0 - \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \cdots + w_t \quad (0.114)$$

donde los coeficientes dependen de un simple parámetro  $\theta_1$  via  $\phi_i = -\theta_1^i$  para  $i > 1$ . Para que el modelo en la ecuación (0.114) sea estacionario,  $\theta_1$  debe ser menor que uno en valor absoluto, esto es,  $|\theta_1| < 1$ ; de otra forma  $\theta_1^i$  y la serie serán explosivas, porque los valores de la serie de tiempo se hacen grande en magnitud

rápidamente. Claramente, porque  $|\phi_i|^j$  crece sin acotación cuando  $j \rightarrow \infty$ .

Dado que  $|\theta_1| < 1$ , tenemos que  $\theta_1^i \rightarrow 0$  cuando  $i \rightarrow \infty$ . Por lo tanto, la contribución de  $x_{t-i}$  a  $x_t$  decae exponencialmente cuando  $i$  crece. Esto es razonable, ya que la dependencia de una serie estacionaria  $x_t$  de su valor  $x_{t-i}$ , si existe, debería decaer con el tiempo.

El modelo en la ecuación (0.114) lo podemos escribir en forma más compacta. Para ello, reescribamos el modelo como

$$x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \cdots = \phi_0 + w_t \quad (0.115)$$

El modelo para  $x_{t-1}$  es entonces

$$x_{t-1} + \theta_1 x_{t-2} + \theta_1^2 x_{t-3} + \cdots = \phi_0 + w_{t-1} \quad (0.116)$$

Multiplicando la ecuación (0.116) por  $\theta_1$  y restando el resultado a la ecuación (0.115), obtenemos

$$x_t = \phi_0(1 - \theta_1) + w_t - \theta_1 w_{t-1}$$

que dice que excepto para el término constante,  $x_t$  es un promedio ponderado de  $w_t$  y  $w_{t-1}$ . Por lo tanto, el modelo se denomina *MA* de orden 1 o modelo *MA*(1). La forma general de un modelo *MA*(1) es

$$x_t = c_0 + w_t - \theta_1 w_{t-1} \quad (0.117)$$

donde  $c_0$  es constante y  $\{w_t\}$  es una serie de ruido blanco. De manera similar, un modelo *MA*(2) es de la forma

$$x_t = c_0 + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} \quad (0.118)$$

y un modelo *MA*( $q$ ) es de la forma

$$x_t = c_0 + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - \cdots - \theta_q w_{t-q} \quad (0.119)$$

con  $q > 0$ .

### 0.17. Propiedades de los modelos MA

Al igual que con los modelos  $AR$ , daremos las propiedades para los modelos  $MA(1)$  y  $MA(2)$ , y luego generalizaremos a los modelos  $MA(q)$ .

#### 0.17.1. Estacionaridad

Los modelos  $MA$  son siempre débilmente estacionarios porque son combinaciones lineales finitas de una sucesión de ruido blanco para el cual los primeros dos momentos son invariantes en el tiempo. Por ejemplo, consideremos el modelo  $MA(1)$  dado en (0.117). Tomando el valor esperado obtenemos

$$\begin{aligned} \mathbb{E}(x_t) &= \mathbb{E}(c_0 + w_t - \theta_1 w_{t-1}) \\ &= \mathbb{E}(c_0) = c_0 \end{aligned}$$

el cual es invariante. Tomando la varianza en la misma ecuación (0.117), obtenemos

$$\begin{aligned} Var(x_t) &= Var(c_0 + w_t - \theta_1 w_{t-1}) \\ &= Var(c_0) + Var(w_t) + \theta_1^2 Var(w_{t-1}) \\ &= \sigma_w^2 + \theta_1^2 \sigma_w^2 \\ &= (1 + \theta_1^2) \sigma_w^2 \end{aligned}$$

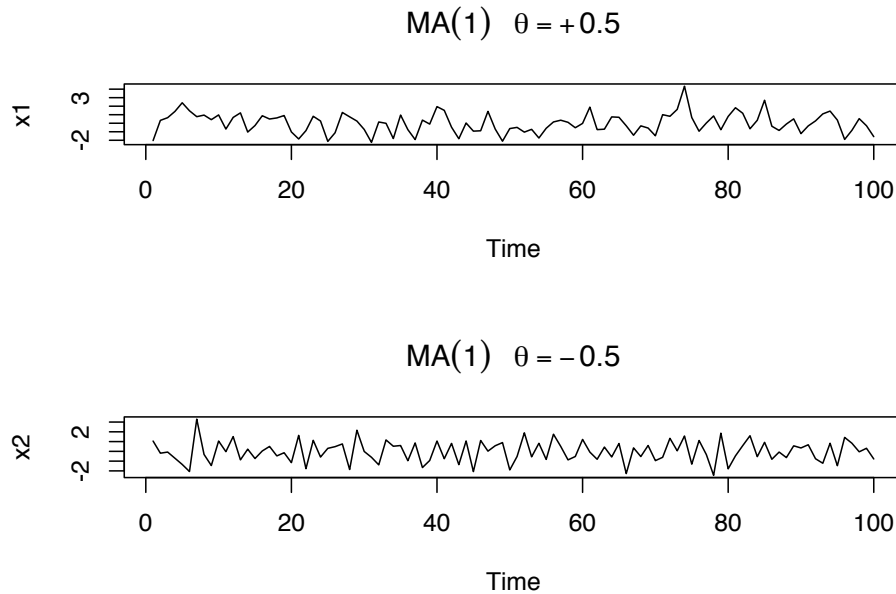
donde usamos el hecho de que  $w_t$  y  $w_{t-1}$  son no correlacionados. Nuevamente,  $Var(x_t)$  es invariante. Los cálculos anteriores los

podemos aplicar a un modelo  $MA(q)$  general, de donde obtenemos dos propiedades generales:

- 1) El término constante de un modelo  $MA(q)$  es la media de la serie, es decir,  $\mathbb{E}(x_t) = c_0$
- 2) La varianza de un modelo  $MA(q)$  es  $Var(x_t) = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma_w^2$

La serie en la figura siguiente corresponde a dos procesos  $MA(1)$ , uno con  $\theta_1 = 0,5$  y el otro con  $\theta_1 = -0,5$ , en ambos casos  $\sigma_w^2 = 1$ .

```
par(mfrow=c(2,1))
plot(arima.sim(list(order=c(0,0,1),ma=0.5),n=100),ylab="x1",
     main=expression(MA(1)~~~theta==+0.5))
plot(arima.sim(list(order=c(0,0,1),ma=-0.5),n=100),ylab="x2",
     main=expression(MA(1)~~~theta== -0.5))
```



**Figura 14:** Simulación de dos modelos  $MA(1)$

### 0.17.2. Función de autocorrelación (ACF)

Para simplificar podemos asumir que  $c_0 = 0$ , para un modelo  $MA(1)$ . Multiplicando el modelo por  $x_{t-h}$ , obtenemos

$$x_{t-h}x_t = x_{t-h}w_t - \theta_1 x_{t-h}w_{t-1}$$

Tomando el valor esperado, obtenemos

$$\begin{aligned}\gamma(h) &= \mathbb{E}(x_{t-h}x_t) = \mathbb{E}(x_{t-h}w_t - \theta_1 x_{t-h}w_{t-1}) \\ &= \begin{cases} -\theta_1 \sigma_w^2, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases} \end{aligned} \quad (0.120)$$

Usando el hecho de que  $Var(x_t) = (1 + \theta_1^2)\sigma_w^2$ , obtenemos la función de autocorrelacion ACF para un modelo  $MA(1)$ , a saber

$$\rho(h) = \begin{cases} 1, & \text{si } h = 0 \\ \frac{-\theta_1}{1+\theta_1^2}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases} \quad (0.121)$$

Entonces, para un modelo  $MA(1)$ , el paso 1 de la ACF es distinto de cero, y para orden o paso superior la ACF es cero. En otras palabras, la ACF de un modelo  $MA(1)$  corta en paso 1. Note que

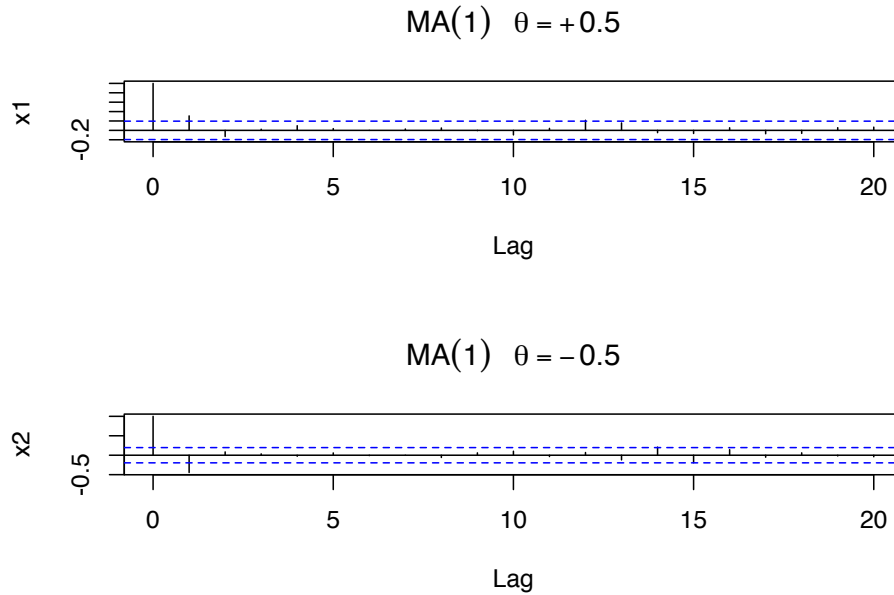
$|\rho(1)| \leq 1/2$  para todo valor de  $\theta_1$ . También,  $x_t$  está correlacionado con  $x_{t-1}$  pero no con  $x_{t-2}, x_{t-3}, \dots$ . Constraste esto con el caso del modelo  $AR(1)$  en el cual la correlación entre  $x_t$  y

$x_{t-k}$  nunca es cero, para  $k > 1$ .

En la figura siguiente podemos observar las funciones de autocorrelacion de los modelos  $MA(1)$  simulados anteriormente

```
par(mfrow=c(2,1))
acf(arima.sim(list(order=c(0,0,1),ma=0.5),n=100),ylab="x1",
    main=expression(MA(1)~~~theta==+0.5))
```

```
acf(arima.sim(list(order=c(0,0,1),ma=-0.5),n=100),ylab="x2",
    main=expression(MA(1)~~~theta=-0.5))
```



**Figura 15:** ACF para dos modelos MA(1)

Podemos notar en cada una de las ACF, que efectivamente tienen un corte en paso 1. Para la serie con  $\theta_1 = 0,5$ , la correlación es positiva y para la serie con  $\theta = -0,5$  la correlación es negativa.

Para un modelo  $MA(2)$  dado por la ecuación (0.118) la función de autocorrelación está dada por

$$\rho(h) = \begin{cases} \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } h = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } h = 2 \\ 0, & \text{si } h > 1 \end{cases} \quad (0.122)$$

En este caso, la ACF corta en paso 2. Esta propiedad la podemos generalizar a los modelos  $MA(q)$ . Así, para un modelo  $MA(q)$  la ACF se corta en paso  $q$ , y vale cero para  $h > q$ .

Consecuentemente, una serie  $MA(q)$  está solo linealmente

relacionada con sus  $q$  primeros valores y por consiguiente es un modelo de “memoria finita”.

**Ejemplo 0.24.** USando las funciones de autocovarianza (ec. (0.120)) y de autocorrelación (ec. (0.121)) de un modelo  $MA(1)$  podemos notar que  $\rho(h)$  es el mismo para  $\theta$  y  $1/\theta$ , probemos, por ejemplo, con  $\theta = 5$  y  $\frac{1}{\theta} = \frac{1}{5}$

$$\begin{aligned}\rho(h) &= \begin{cases} \frac{5}{(1+5^2)}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases} = \begin{cases} \frac{5}{26}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases} \\ \rho(h) &= \begin{cases} \frac{1/5}{(1+(1/5)^2)}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases} = \begin{cases} \frac{5}{26}, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases}\end{aligned}$$

Además, el par  $\sigma_w^2 = 1$  y  $\theta = 5$  llevan a la misma función de autocovarianza que el par  $\sigma_w^2 = 25$  y  $\theta = 1/5$ , esto es

$$\gamma(h) = \begin{cases} 26, & \text{si } h = 0 \\ 5, & \text{si } h = 1 \\ 0, & \text{si } h > 1 \end{cases}$$

Entonces los procesos  $MA(1)$

$$x_t = w_t + \frac{1}{5}w_{t-1}, w_t \sim iidN(0, 25) \text{ y } x_t = v_t + 5v_{t-1}, v_t \sim iidN(0, 1)$$

son los mismos debido a la normalidad, es decir, todas las distribuciones finitas son las mismas.

Para descubrir cual de los modelos es el modelo invertible, podemos invertir los papeles de  $x_t$  y  $w_t$  (porque estamos copiando el caso  $AR$ ) y escribir el modelo  $MA(1)$  como  $w_t = -\theta w_{t-1} + x_t$ . Siguiendo los pasos para (0.97), si  $|\theta| < 1$ , entonces  $w_t = \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}$ , lo cual es la representación del modelo  $AR$  infinito deseado. Por consiguiente, elegimos el modelo con  $\sigma_w^2 = 25$  y  $\theta = 1/5$  ya que este modelo es invertible

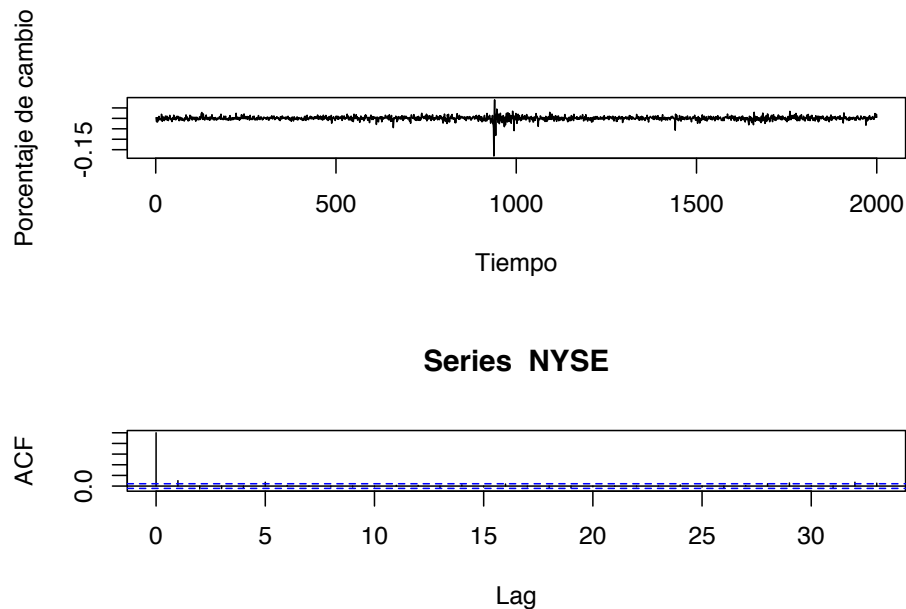


### 0.18. Identificación del orden de un MA

La ACF es muy útil para identificar el orden de un modelo  $MA$ .

Para una serie de tiempo  $x_t$  con ACF  $\rho(h)$ , si  $\rho(q) \neq 0$ , pero  $\rho(h) = 0$  para  $h > q$ , entonces  $x_t$  sigue un modelo  $MA(q)$ . La figura siguiente muestra el gráfico de la ACF para la serie de porcentajes de cambio diario de la Bolsa de Valores de New York del ejemplo 0.5. Las líneas discontinuas en azul en la ACF denotan los dos límites de error estándar. Se puede notar que la serie tiene un ACF significativo en pasos 1, 2 y 5, lo que nos sugiere el

$$x_t = c_0 + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - \theta_5 w_{t-5}$$



**Figura 16:** Serie de tiempo y función de autocorrelación de los porcentajes de cambio diario de la Bolsa de Valores de New York, desde el 2 de febrero de 1984 hasta el 31 de diciembre de 1991

### 0.19. Estimación

La estimación de máxima verosimilitud se usa comúnmente para estimar modelos  $MA$ . Existen dos enfoques para evaluar la función de verosimilitud de un modelo  $MA$ . El primer enfoque asume que los valores iniciales, es decir, para  $t \leq 0$ , son ceros. Como tal, los valores necesarios en el cálculo de la función de verosimilitud se obtiene recursivamente del modelo iniciando con  $w_1 = x_1 - c_0$  y  $w_2 = x_2 - c_0 + \theta w_1$ . Este enfoque se conoce como el *método de verosimilitud condicional* y las estimaciones resultantes son las estimaciones de máxima verosimilitud condicional. El segundo enfoque trata los valores iniciales en  $t \leq 0$  como parámetros adicionales del modelo y los estima conjuntamente con otros parámetros. Este enfoque se conoce como el *método de verosimilitud exacta*. Las estimaciones de verosimilitud exacta son preferibles a las condicionales, pero requieren un cálculo más extenso. Si el tamaño de la muestra es grande, entonces los dos tipos de estimación de máxima verosimilitud están cerca uno del otro.

### 0.20. Predicciones usando modelos MA

Las predicciones con modelos  $MA$  son fáciles de obtener. Como el modelo es de “memoria finita”, la predicción converge rápidamente a la media de la serie. Para ver esto, supongamos que la predicción inicia en  $m$ . Para la predicción de un paso de un proceso  $MA(1)$ , el modelo nos dice que

$$x_{m+1} = c_0 + w_{m+1} - \theta_1 w_m$$

Tomando el valor esperado condicional, tenemos

$$\begin{aligned}\hat{x}_m(1) &= \mathbb{E}(x_{m+1}|x_m, x_{m-1}, \dots) = c_0 - \theta_1 w_m \\ e_m(1) &= x_{m+1} - \hat{x}_m(1) = w_{m+1}\end{aligned}$$

La varianza del error de predicción de un paso es  $\text{Var}[e_m(1)] = \sigma_w^2$ . En la práctica, el valor  $w_m$  lo podemos obtener de varias maneras.

Por ejemplo, suponemos que  $w_0 = 0$ , entonces  $w_1 = x_1 - c_0$  y calculamos  $w_t$  para  $2 \leq t \leq m$  recursivamente usando  $w_t = x_t - c_0 + \theta_1 w_{t-1}$ . Alternativamente, podemos calcularlo usando la representación AR del modelo  $MA(1)$ . (Véase el ejemplo 0.24)

Para la predicción de dos pasos, de la ecuación

$$x_{m+2} = c_0 + w_{m+2} - \theta_1 w_{m+1}$$

obtenemos

$$\begin{aligned}\hat{x}_m(2) &= \mathbb{E}(x_{m+2}|x_m, x_{m-1}, \dots) = c_0 \\ e_m(2) &= x_{m+2} - \hat{x}_m(2) = w_{m+2} - \theta_1 w_{m+1}\end{aligned}$$

La varianza del error de predicción es  $\text{Var}[e_m(2)] = (1 + \theta_1^2)\sigma_m^2$ , la cual es la varianza del proceso  $MA$  y es mayor o igual que la varianza del error de predicción de un paso. Este resultado muestra que para un modelo  $MA(1)$  la predicción de dos pasos de la serie es sencillamente la media incondicional del modelo. Esto es cierto para cualquier predicción iniciando en  $m$ . Más generalmente  $\hat{x}_m(h) = c_0$  para  $h > 2$ . En resumen, para un modelo  $MA(1)$ , la predicción de un paso a futuro con origen en  $m$  es  $c_0 - \theta_1 w_m$  y la predicción de múltiples pasos es  $c_0$ , el cual es la media incondicional del modelo. Si graficamos la predicción  $\hat{x}_m(h)$  versus  $h$  podemos ver que la predicción es una línea horizontal después del paso uno.

De manera similar, para un modelo  $MA(2)$ , tenemos

$$x_{m+h} = c_0 + w_{m+h} - \theta_1 w_{m+h-1} - \theta_2 w_{m+h-2}$$

para el cual obtenemos

$$\begin{aligned}\hat{x}_m(1) &= c_0 - \theta_1 w_m - \theta_2 w_{m-1} \\ \hat{x}_m(2) &= c_0 - \theta_2 w_m \\ \hat{x}_m(h) &= c_0, \text{ para } h > 2\end{aligned}$$

Entonces, la predicción de múltiples pasos de un modelo  $MA(2)$  tiende a la media de la serie después de dos pasos. La varianza del error de predicción converge a la varianza de la serie después de dos pasos. En general, para un modelo  $MA(q)$ , la predicción multipasos converge a la media de la serie después de los primeros  $q$  pasos.

---

\*\* Resumen \*\*

Para concluir daremos un resumen de las propiedades estudiadas para los modelos  $AR$  y  $MA$  en estos dos capítulos:

- Para los modelos  $MA$ , la ACF es útil para especificar el orden dado que la ACF se corta en salto  $q$  para una serie  $MA(q)$ .
- Para un modelo  $AR$ , la PACF es útil para determinar el orden ya que la PACF se corta en salto  $p$  para un proceso  $AR(p)$ .
- Una serie  $MA$  es siempre estacionaria, pero para que una serie  $AR$  sea estacionaria, todas sus raíces características deben ser menor que u1 en módulo.
- Para una serie estacionaria, la predicción de múltiples pasos converge a la media de la serie y la varianza del error de predicción converge a la varianza de la serie.