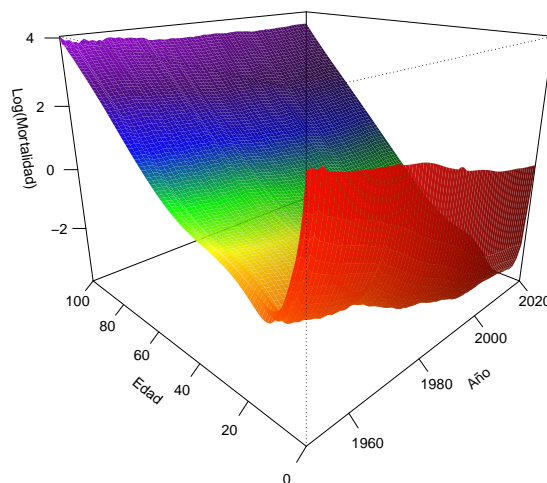
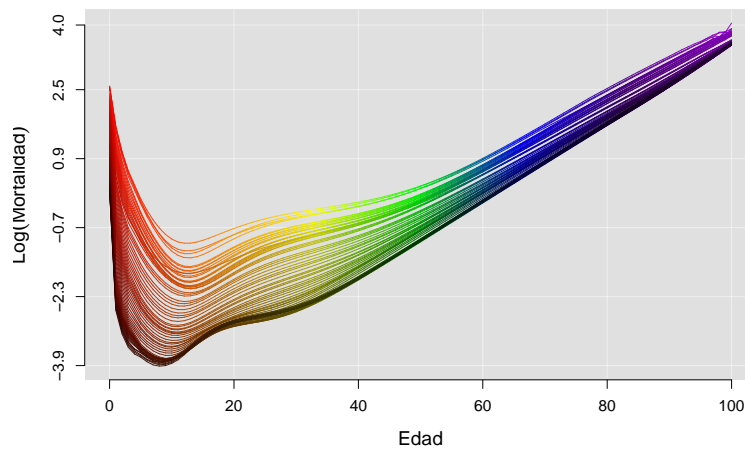


Aplicación del método de Lee-Carter en la población mexicana (1990-2015)

artitzco

México, 2017



Índice general

Introducción	1
1. Medidas demográficas	5
1.1. Tasa	5
1.2. Elaboración de proyecciones	5
1.3. Mortalidad	6
1.4. Tasa Central de mortalidad	6
2. Elementos de inferencia estadística	9
2.1. Introducción	9
2.2. Estimación puntual	11
2.3. Estimación por intervalo	12
2.4. Pruebas de hipótesis	13
3. Series de tiempo	17
3.1. Series de tiempo vistas como procesos estocásticos	17
3.2. Modelos univariados de series de tiempo	19
4. El modelo de Lee-Carter	33
4.1. Tabla de mortalidad	33
4.2. Estimación del modelo de Lee Carter	34

4.2.1. Estimación puntual	35
4.2.2. Método de mínimos cuadrados	35
4.2.3. Máxima verosimilitud	37
4.2.4. Otra forma máximo verosímil	38
4.2.5. Insesgamiento	39
4.3. Pronósticos	40
4.4. Tablas de estimaciones para la población total	47

Introducción

Las proyecciones de población se definen como un escenario hipotético de lo que ocurrirá con la dinámica demográfica de un país o región si determinadas circunstancias económicas, políticas, sociales y demográficas ocurriesen. Representan un instrumento fundamental para todas las acciones y programas de la administración pública en la medida que permiten anticiparnos a las demandas sociales. Son usadas para planear diversos aspectos estratégicos dirigidos a satisfacer las demandas básicas en materia de vivienda, educación, salud y empleo. Asimismo, son necesarias para conocer la demanda futura de alimento, agua, energía, transporte y otros servicios.

En cuestión de educación, es necesario saber cuánta población en edades escolares puede haber en el futuro en cada entidad federativa para determinar la construcción de escuelas y asignar los recursos humanos necesarios de manera adecuada. Con respecto a la vivienda, las proyecciones servirían para indicar qué localidades podrían crecer más y requerir más viviendas [García, 2013].

En cuestión de empleo, representa un esfuerzo para un Estado en vías de desarrollo cuyo modelo económico se encuentra centrado en apertura comercial. La demanda de trabajo empieza a los 15 años, creciendo de manera acelerada a los 20 años, rango de edad de la fuerza de trabajo primaria, alcanza su máximo nivel a los 30-45 años, tornándose declinante después de los 65 años, debido a la mortalidad, inicio de la jubilación e inactividad. Pero, también podría dar una idea sobre la intensidad y duración que tendría el crimen y las actividades económicas que se encuentran fuera de la ley en un contexto de empleo escaso y precario.

En cuestión de salud, las proyecciones demográficas son necesarias para conocer la posible demanda de vacunas, así como para tener una idea de cuántos medicamentos deben ser comprados en caso de una emergencia y para realizar la construcción de hospitales y centros de salud cercanos a los asentamientos poblacionales [García, 2013].

Historicamente las primeras proyecciones fueron las de la población total, pues no se contaba con información sobre natalidad, mortalidad y migración. Los primeros antecedentes de proyección realizados por Gregory King en Inglaterra, a finales del siglo XVII, se basaban en el tiempo de duplicación del número de habitantes. Malthus también hizo proyecciones de población utilizando la tasa geométrica de crecimiento,

aunque advirtió que estas tenían poca probabilidad de ocurrencia debido a la irregularidad de epidemias y hambrunas. Posteriormente, en 1838, Verhulst sugirió la curva logística. Esta función, que modificaron Pearl y Reed, fue objeto de mucha atención entre 1920 y 1930 y luego muy utilizada para reproducir la Teoría de la Transición Demográfica e incluso ha servido para realizar proyecciones de la fecundidad y de la esperanza de vida [García and Ordorica, 2010].

Durante el siglo XX y lo que llevamos de este siglo, se ha utilizado el método de los componentes para realizar proyecciones de población, el cual consiste en aplicar hipótesis sobre fecundidad, mortalidad y migración a la estructura por edades y sexo. Cada grupo se proyecta con probabilidades de supervivencia. La cifra de nacimientos se obtiene a partir de las tasas de fecundidad proyectadas y a la composición por edad resultante se le suman o se le restan los migrantes [García and Ordorica, 2010].

México tiene una historia importante en pronósticos demográficos. La División de Población de las Naciones Unidas publicó para México en 1954 la primera proyección moderna de población y el Consejo Nacional de Población (CONAPO) publicó la última en el año 2006.

En el país, las primeras proyecciones fueron elaboradas por Benítez y Cabrera en 1966 con el fin de determinar principalmente la demanda futura de productos agrícolas [García, 2013]. Posteriormente, a principios de los setenta, las proyecciones de población mostraban que México tendría 151.8 millones de habitantes para el año 2000. En esta etapa se manifestaba un crecimiento acelerado y en 1977, el Consejo Nacional de Población (CONAPO) realizó proyecciones demográficas con las cuales se definió disminuir este crecimiento a 1 % al año 2000. El objetivo era hacer ajustes a los programas para cumplir con las metas establecidas, pero desafortunadamente no se alcanzó la meta de 1 % y a pesar de eso este número fue una guía para la política demográfica de nuestro país en el último cuarto del siglo XX [García and Ordorica, 2010].

Según los resultados del censo de 2000, llegamos a una cifra de aproximadamente 100 millones; alrededor de 50 millones menos a lo proyectado a partir de la hipótesis constante. Esta cifra en el número de personas es equivalente a la población que México tuvo en 1970 [García and Ordorica, 2010].

Las proyecciones de población elaboradas a finales de los setenta sirvieron para definir la meta de 1 % en el ritmo de crecimiento demográfico al año 2000 y las metas intermedias. Las que se elaboraron a partir de 1980 hasta final del siglo XX y las realizadas en los primeros años del siglo XXI, sirvieron para monitorear el avance en las metas planteadas y su efecto sobre la dinámica demográfica [García and Ordorica, 2010].

Se tiene que existen algunos métodos de proyecciones de población estos se adaptan a la información existente y al análisis de escenarios que se desea elaborar sobre el futuro, los cuales se pueden clasificar en: métodos matemáticos, métodos económicos y método de los componentes. Los más sencillos para calcular la población futura son aquellos que utilizan una ecuación que expresa la dinámica de la población como función del tiempo. El método matemático se distingue porque se aplica generalmente al

número total de población. Hay métodos matemáticos que usan polinomios de varios grados, o aquellos que utilizan los métodos de los mínimos cuadrados, ya sea adaptando una recta o un polinomio de grado mayor. Hay métodos que siguen la línea recta, la curva exponencial, la geométrica, la logística u otras funciones [Brenes, 2014].

Actualmente, dentro del marco del método de los componentes para las proyecciones de población se han utilizado distintos procedimientos para la proyección de la mortalidad: métodos de tablas límite [CELADE, 2012], método de Lee-Carter [Fernández, 2013] [Berliard and Williams, 2013], o métodos Bayesianos [Raftery et al., 2013] [Brenes, 2014].

México está atravesando una etapa de la transición demográfica esta etapa se caracteriza por tener a la población concentrada en las edades poslaborales o en retiro. Por lo que las proyecciones demográficas deberían ser utilizadas para analizar los posibles escenarios futuros en materia de seguridad social y sus costos. A partir de las proyecciones será posible analizar los principales cambios en cuanto al número y estructura por edad de la población mexicana, relacionando dichos cambios con los requerimientos de orden social en materia de educación, salud y vivienda.

En virtud de lo anterior, este trabajo tiene por objetivo mostrar los alcances de las proyecciones demográficas para la planeación nacional y las oportunidades que existen si se mejora la metodología tradicionalmente utilizada. En el segundo apartado se analiza qué son y para qué son utilizadas las proyecciones de población. En el tercero, se describe la forma en que fueron calculadas y los supuestos que estuvieron detrás las primeras proyecciones de población realizadas para México, así como el uso que se les dio en ese entonces, y se comparan con las últimas proyecciones de población calculadas por el Consejo Nacional de Población (CONAPO). En el cuarto, se describe una forma alternativa de proyectar la población mexicana y se propone una forma de utilizarla y regularla. Finalmente se concluye con algunas reflexiones finales y sugerencias.

En virtud de lo anterior, este trabajo tiene por objetivo hacer una proyección demográfica a nivel nacional usando el método de Lee Carter que nos permita comparar los resultados obtenidos con las proyecciones de población calculadas por el Consejo Nacional de Población (CONAPO). En el primer apartado se estudiarán algunas definiciones que son utilizados para las proyecciones de población. En el segundo, se hace una breve explicación de teoría estadística la cual es la base de nuestro análisis. En el tercer apartado se describe brevemente el análisis de series de tiempo. En el cuarto, se define y desarrolla el modelo de Lee-Carter En el quinto, se describe la forma en que fueron calculadas y los supuestos que estuvieron detrás de las proyecciones de población realizadas para México, así como el uso que se les da, y se comparan con las últimas proyecciones de población calculadas por el Consejo Nacional de Población (CONAPO). Finalmente se concluye con algunas reflexiones finales y sugerencias.

Capítulo 1

Medidas demográficas

Para el análisis del modelo de Lee Carter se estudiarán algunas medidas demográficas que permitirán contar, de manera adecuada, los acontecimientos de una población y algunos conceptos importantes.

1.1. Tasa

Definición 1.1 *Tasa*

*Una **tasa** mide la frecuencia relativa de la población con la que un evento se presenta dentro de una población o subpoblación en un determinado periodo de tiempo, generalmente un año calendario. Se calcula como un cociente $\frac{n}{d}$, el numerador n es la cantidad de eventos ocurridos en el año de referencia, y el denominador d es la población media (población existente al 30 de Junio del año considerado), el resultado se multiplica por mil o por cien [Rodríguez et al., 1997].*

1.2. Elaboración de proyecciones

Las proyecciones son muy importantes en apoyo a la política de población y son utilizadas como la información futura más “probable” sobre la evolución de la dinámica demográfica. Estas contribuyen de dos maneras: tanto para planificar como para evaluar cuando no hay observaciones o estimaciones disponibles. Es decir: las proyecciones son principalmente utilizadas cuando no se dispone de observaciones o estimaciones demográficas. Cabe mencionar que, las estimaciones demográficas se refieren tanto al presente como al pasado demográfico del país, y las proyecciones se refieren únicamente al futuro [García, 2013].

Cuando los datos disponibles no permiten precisar el valor de una cantidad demográfica, se puede estimar dicho valor con mayor o menor aproximación por medio de diversas técnicas propias del análisis demográfico. La operación correspondiente se llama estimación, y el resultado o valor estimado se llama también estimación. Las estimaciones pueden ser calculadas para datos demográficos del presente o del pasado y requieren de una calidad mínima de la información para que sean confiables [García, 2013] [García and Ordorica, 2010].

1.3. Mortalidad

Durante el siglo XX la humanidad, en general, tuvo importantes transformaciones las cuales han precedido eventos que van desde los naturales hasta los tecnológicos, económicos, políticos y culturales, y que han marcado de forma distinta la dinámica de las poblaciones del mundo.

En materia demográfica, específicamente para la mortalidad, los cambios que se han presentado en la población han traído consigo logros como el descenso en los niveles de mortalidad, lo cual apunta a una mejora sustancial en las condiciones generales de vida de la población [Mina, 2002].

El estudio de la mortalidad se emplea para expresar la acción de la muerte sobre los integrantes de una población. Existen algunas características que debemos señalar para su análisis.

La muerte es un riesgo al que se encuentra expuesto toda la vida y es, además, un hecho que ocurre una sola vez a cada persona.

Por otra parte, los hechos que determinan la mortalidad están estrechamente vinculados a factores biológicos y a la estructura por edad de la población, para nuestro caso se consideraron las defunciones hasta los 100 años de edad, lo que, por una parte, facilita la aplicación de técnicas demográficas y, por otra parte, nos permite exponer de mejor manera los resultados obtenidos del modelo [Rodríguez et al., 1997].

1.4. Tasa Central de mortalidad

Las tasas de mortalidad por edad, también llamadas tasas centrales de mortalidad o tasas específicas de mortalidad, se abrevian con ${}_n m_x^j$ y se calculan de la siguiente manera:

$${}_n m_x^j := \frac{{}_n D_x^j}{{}_n N_x^j}$$

donde:

${}_n m_x^j$:= La mortalidad del grupo de edad x a $x + n - 1$ en el año j .

${}_nD_x^j$:= El número de defunciones ocurridas en el año j de personas con edad cumplida entre x y $x + n - 1$.

${}_nN_x^j$:= La población al 30 de junio del año j perteneciente al grupo de edad x y $x + n - 1$.

Las tasas se expresan por mil y las cifras que se arrojan deben interpretarse como la cantidad de fallecidos de edad x a $x + n - 1$ por cada mil personas de esa misma edad en el año calendario analizado. Para el caso de incidencia de mortalidad, el estudio para las edades se inicia con el cálculo de las tasas de mortalidad por edad. Para el estudio que se llevará a cabo del caso México consideramos la mortalidad en cada edad.

La estimación de la población a la mitad del año (P_{30-06}), es decir, al 30 de junio se calcula de la siguiente manera:

$$P_{30-06} := P_{t_0} (1 + r)^{\frac{N\delta}{360}}$$

donde:

P_k := El tamaño de la población en el k -ésimo censo de la población realizado.

t_0 := Fecha del último censo realizado.

N := Número de días transcurridos entre el 30 de junio del mismo año y el censo t_0 .

$$\delta := \begin{cases} 1 & \text{si el censo es realizado antes del 30 de junio,} \\ -1 & \text{en caso contrario.} \end{cases}$$

$r := \left(\frac{P_{t_0}}{P_{t_0-1}} \right)^{\frac{1}{n}} - 1$, r se llama tasa de crecimiento y n es el número de años transcurridos entre el censo t_0 y $t_0 - 1$ [Mina, 2002].

Capítulo 2

Elementos de inferencia estadística

2.1. Introducción

Una **población** es un conjunto de individuos, objetos o mediciones en los que se tiene un interés particular de estudio. Para realizar estos estudios se consideran, por lo general, subconjuntos representativos de la población a los que se les conoce como **muestras**. Una **variable** es una característica o medición específica que posee cada uno de los individuos de una población, y a los valores que toman ciertas variables de interés en una muestra se les llama **datos**.

La **estadística inferencial** hace uso de técnicas para que a partir de los datos de una muestra se pueda predecir el comportamiento de la población completa. Al obtener los datos x_1, x_2, \dots, x_n de una muestra, correspondiente a cierta variable, se dará por hecho que dichos valores provienen de variables aleatorias X_1, X_2, \dots, X_n , la cuales son independientes y tienen la misma distribución, de esta idea nace la siguiente definición:

Definición 2.1 *Muestra aleatoria (m.a.)*

*Se le conoce como **muestra aleatoria** a una colección de variables aleatorias X_1, X_2, \dots, X_n que cumplen con ser:*

- a) Independientes.*
- b) Idénticamente distribuidas (tienen la misma función de distribución).*

*El entero positivo n se llama **tamaño de la muestra**.*

En general, en el estudio de una muestra aleatoria se incurre en alguna de las dos siguientes posibilidades:

- Se desconoce por completo su distribución.
- Se conoce la distribución de la muestra pero ésta depende de un parámetro desconocido.

Si se tiene X_1, X_2, \dots, X_n una m.a. como en la segunda situación, entonces cada X_i tiene como función de densidad de probabilidad (fdp) o función de masa de probabilidad (fmd) $f(\bullet; \theta)$ para algún $\theta \in \Theta$. Al conjunto Θ se le conoce como **espacio parametral**, es el conjunto de todos los posibles valores de θ y está dictado por la distribución. Algunas veces es de interés estudiar no sólo el parámetro θ sino un valor $\tau(\theta)$, en estos casos se le conoce a τ como **función parametral** de θ .

Definición 2.2 Estadística

Sea X_1, X_2, \dots, X_n una m.a. Una **estadística** (o **estadístico**) de la muestra es una función $T = \xi(X_1, X_2, \dots, X_n)$ que no depende de parámetros desconocidos.

Algunas estadísticas importantes:

Sea X_1, X_2, \dots, X_n una m.a.

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ (media muestral).
- $E^r = \frac{1}{n} \sum_{i=1}^n X_i^r$, $r \in \mathbb{N}$ (r -ésimo momento muestral).
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (X_i^2) - n\bar{X}^2 \right)$ (varianza muestral).
- $S = \sqrt{S^2}$ (desviación estandar muestral).
- $X_{(1)} = \min \{X_1, X_2, \dots, X_n\}$.
- $X_{(n)} = \max \{X_1, X_2, \dots, X_n\}$.
- $R = X_{(n)} - X_{(1)}$ (rango muestral).

2.2. Estimación puntual

Definición 2.3 Estimador puntual

Sea X_1, X_2, \dots, X_n una m.a. con fdp o fmd $f(\bullet; \theta)$. Un **estimador puntual** de θ es una estadística cuyos valores ayudan a encontrar el valor exacto de θ o de $\tau(\theta)$.

Definición 2.4 Estimador sesgado e insesgado

Sea $\hat{\theta}(X_1, X_2, \dots, X_n)$ un estimador de $\tau(\theta)$. Se le llama a $\hat{\theta}$ un estimador **insesgado** de $\tau(\theta)$ cuando:

$$E[\hat{\theta}] = \tau(\theta).$$

En caso contrario $\hat{\theta}$ es un estimador **sesgado** y se define el sesgo de $\tau(\theta)$ como:

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

Particularmente cuando un estimador es sesgado puede suceder que:

$$\lim_{n \rightarrow \infty} E[\hat{\theta}] = \tau(\theta).$$

Entonces se le conoce a $\hat{\theta}$ como estimador **asintóticamente insesgado** de $\tau(\theta)$.

Un estimador insesgado siempre se preferirá a cualquier otro estimador. Cuando la muestra es grande un estimador asintóticamente insesgado es casi tan bueno como el primero.

Sea X_1, X_2, \dots, X_n una m.a. con fdp o de fmp $f(\bullet; \theta)$ con $\theta \in \Theta$. Se le conoce como **función de verosimilitud** a:

$$\mathcal{L} : \Theta \rightarrow \mathbb{R} \text{ tal que } \mathcal{L}(\theta; x_1, x_2, \dots, x_n) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta).$$

donde $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)$ es la densidad conjunta de X_1, X_2, \dots, X_n en x_1, x_2, \dots, x_n dado el parámetro θ .

El método de estimación puntual por máxima verosimilitud consiste en encontrar el máximo de \mathcal{L} . Si se encuentra que $\theta_0 = \mathcal{G}(x_1, x_2, \dots, x_n)$ es máximo para \mathcal{L} entonces:

$$\hat{\theta}_{MV} = \mathcal{G}(X_1, X_2, \dots, X_n)$$

es el estimador **máximo verosimil** de θ .

Teorema 2.1 Principio de invarianza

Sea X_1, X_2, \dots, X_n una m.a. con fdp o fmp $f(\bullet; \theta)$ con $\theta \in \Theta$, sea $\hat{\theta}_{MV}$ el estimador máximo verosimil de θ y τ una función parametral de θ . Si \mathcal{L} está en función de $\tau(\theta)$ entonces el estimador máximo verosimil de $\tau(\theta)$ es $\tau(\hat{\theta}_{MV})$, es decir,

$$\widehat{\tau(\theta)}_{MV} = \tau(\widehat{\theta}_{MV})$$

2.3. Estimación por intervalo

Por lo general, no basta con tener una buena estimación de un parámetro, también importará conocer con qué probabilidad se encuentran contenidos en cierto intervalo.

Definición 2.5 Intervalo de confianza

Sea X_1, X_2, \dots, X_n una m.a. con fdp o fmp $f(\bullet; \theta)$ con $\theta \in \Theta$. Si T_1 y T_2 son estadísticas de la muestra tales que $T_1 \leq T_2$ y además:

$$P[T_1 \leq \tau(\theta) \leq T_2] = 1 - \alpha,$$

se dice que $[T_1, T_2]$ es un **intervalo al $(1 - \alpha)100$ por ciento de confianza** de $\tau(\theta)$, en tal caso también se dice que dicho intervalo tiene una **significancia** de tamaño α .

Existen diversos métodos para construir intervalos de confianza, en general la dificultad para encontrar un intervalo adecuado dependerá del tipo de distribución que se esté manejando, a continuación se presentan tres métodos importantes:

Sea X_1, X_2, \dots, X_n una muestra aleatoria con función de densidad o de probabilidad $f(\bullet; \theta)$ con $\theta \in \Theta$.

•Método de la cantidad pivotal

Sea q_p^X el cuantil p de la distribución cuya variable aleatoria es X , es decir, $P[X \leq q_p^X] = p$.

Se conoce como **cantidad pivotal** a una función Q de la muestra X_1, X_2, \dots, X_n , cuya distribución no depende de parámetros desconocidos. Para el método de la cantidad pivotal se buscará que Q tenga la forma: $Q = \mathcal{G}_{X_1, X_2, \dots, X_n}(\tau(\theta))$ donde \mathcal{G} es una función invertible de $\tau(\theta)$ y así se tiene lo siguiente:

$$\begin{aligned} 1 - \alpha &= P[q_{\alpha_1}^Q \leq Q \leq q_{1-\alpha_2}^Q] \\ &= P[q_{\alpha_1}^Q \leq \mathcal{G}_{X_1, X_2, \dots, X_n}(\tau(\theta)) \leq q_{1-\alpha_2}^Q] \\ &= \begin{cases} P[\mathcal{G}_{X_1, X_2, \dots, X_n}^{-1}(q_{\alpha_1}^Q) \leq \tau(\theta) \leq \mathcal{G}_{X_1, X_2, \dots, X_n}^{-1}(q_{1-\alpha_2}^Q)] & \text{si } \mathcal{G} \text{ es creciente,} \\ P[\mathcal{G}_{X_1, X_2, \dots, X_n}^{-1}(q_{1-\alpha_2}^Q) \leq \tau(\theta) \leq \mathcal{G}_{X_1, X_2, \dots, X_n}^{-1}(q_{\alpha_1}^Q)] & \text{si } \mathcal{G} \text{ es no creciente,} \end{cases} \end{aligned}$$

donde $\alpha \in [0, 1]$ tal que $\alpha = \alpha_1 + \alpha_2$.

En ambos casos se ha encontrado un intervalo de confianza al $(1 - \alpha)100\%$ de confianza de $\tau(\theta)$.

•Distribución asintótica del Teorema Central del Límite

Muchas veces es posible despejar el parámetro desconocido $\tau(\theta)$ de $E[X_i]$, i.e., $E[X_i] = \mathcal{G}(\tau(\theta))$ donde \mathcal{G} es invertible, en tal caso este método es adecuado. Por lo general, la varianza de una muestra aleatoria depende también de parámetros desconocidos, así que se utiliza su estimador máximo verosímil $\widehat{\sigma^2} = S^2$ para hacer uso de ésta. Supongamos que se cumple que $E[X_i] = \mathcal{G}(\tau(\theta))$ con \mathcal{G} invertible. Sea $Z \sim N(0, 1)$, por el **Teorema Central del Límite** se tiene la aproximación:

$$\frac{\bar{X} - \mathcal{G}(\tau(\theta))}{\sqrt{S^2}} \approx Z$$

y entonces:

$$\begin{aligned} 1 - \alpha &= P \left[\left| \frac{\bar{X} - \mathcal{G}(\tau(\theta))}{\sqrt{S^2}} \right| \leq \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \\ &= P \left[-\phi^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \frac{\bar{X} - \mathcal{G}(\tau(\theta))}{\sqrt{S^2}} \leq \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \\ &= P \left[\bar{X} - \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \leq \mathcal{G}(\tau(\theta)) \leq \bar{X} + \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \right] \\ &= \begin{cases} P \left[\mathcal{G}^{-1} \left(\bar{X} - \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \right) \leq \tau(\theta) \leq \mathcal{G}^{-1} \left(\bar{X} + \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \right) \right] & \text{si } \mathcal{G} \text{ es creciente} \\ P \left[\mathcal{G}^{-1} \left(\bar{X} + \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \right) \leq \tau(\theta) \leq \mathcal{G}^{-1} \left(\bar{X} - \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{S^2} \right) \right] & \text{si } \mathcal{G} \text{ es no creciente} \end{cases} \end{aligned}$$

donde ϕ es la función de distribución acumulada asociada a una distribución normal estandar. Con lo anterior se ha encontrado, de manera asintótica, un intervalo de confianza al $(1 - \alpha)100\%$ de confianza para $\tau(\theta)$.

2.4. Pruebas de hipótesis

Una **hipótesis estadística** es una conjetura acerca de la distribución de una o más variables aleatorias. Se le llama **hipótesis simple** cuando se supone por completo la distribución de la o las variables, en caso contrario se le conoce como **hipótesis compuesta**. El esquema general para un **contraste de hipótesis** es tener dos hipótesis estadísticas, comúnmente denotadas por H_0 y H_1 , las cuales serán expuestas a una regla

de decisión para determinar si se considera válida H_0 (**hipótesis nula**) o H_1 (**hipótesis alternativa**).

Definición 2.6 Prueba de hipótesis estadística

Una **prueba de hipótesis estadística** o simplemente **prueba de hipótesis** es una regla o procedimiento para decidir si se acepta la hipótesis nula o se rechaza a favor de una hipótesis alternativa.

Sea X_1, X_2, \dots, X_n m.a., en una prueba de hipótesis para esta muestra existe un conjunto \mathcal{C} , conocido como **región crítica** o **región de rechazo**, el cual se caracteriza por lo siguiente: $(x_1, x_2, \dots, x_n) \in \mathcal{C}$ si y sólo si se rechaza la hipótesis nula a favor de la alternativa, donde (x_1, x_2, \dots, x_n) es una observación de la muestra. Al realizar una prueba de hipótesis, es posible incurrir en dos tipos de errores: se rechaza la hipótesis nula dado que ésta es cierta (**Error tipo I**), o se acepta la hipótesis nula dado que ésta es falsa (**Error tipo II**).

Definición 2.7 Tamaños del error

El **tamaño del error tipo I** o **nivel de significancia** es la probabilidad de cometer el error tipo I (α) y el **tamaño del error tipo II** (β) es la probabilidad de cometer el error tipo II:

$$\alpha = P[\text{Se rechaza } H_0 | H_0 \text{ es cierta}],$$

$$\beta = P[\text{Se acepta } H_0 | H_0 \text{ es falsa}].$$

Definición 2.8 Potencia

Sea X_1, X_2, \dots, X_n una m.a. parametrizada por un valor desconocido y sea H_0 una hipótesis de la distribución. Dada una prueba de hipótesis se define la **función potencia** $\pi(\theta)$ como la probabilidad de rechazar H_0 dado que θ , en efecto, parametriza dicha distribución.

Si se tiene una prueba para las hipótesis:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1,$$

se dice que ésta es la **prueba más potente** de tamaño $\alpha \in (0, 1)$ si:

- $\pi(\theta_0) = \alpha$.
- $\pi(\theta_1) \geq \pi_*(\theta_1)$, donde $\pi_*(\theta)$ es la función potencia para cualquier otra prueba de hipótesis que cumpla que $\pi_*(\theta_0) \leq \alpha$.

Teorema 2.2 De Neyman-Pearson

Sea X_1, X_2, \dots, X_n una m.a. con las siguientes hipótesis:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1.$$

Se define $\lambda(\theta_0, \theta_1; x_1, x_2, \dots, x_n) = \frac{\mathcal{L}(\theta_0; x_1, x_2, \dots, x_n)}{\mathcal{L}(\theta_1; x_1, x_2, \dots, x_n)}$. Se rechaza la hipótesis $H_0 : \theta = \theta_0$ al nivel de significancia α si existe $k \in \mathbb{R}$ tal que:

$$\lambda(\theta_0, \theta_1; x_1, x_2, \dots, x_n) \leq k \text{ y } P[\lambda(\theta_0, \theta_1; X_1, X_2, \dots, X_n) \leq k | H_0 \text{ es cierta}] = \alpha,$$

Entonces ésta es la prueba más potente para las hipótesis H_0 y H_1 .

Este resultado se generaliza para pruebas compuestas. Si se tiene una prueba para las hipótesis:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1,$$

se dice que ésta es la **prueba uniformemente más potente** de tamaño $\alpha \in (0, 1)$ si:

- $\sup_{\theta \in \Theta_0} \pi(\theta) = \alpha$.
- $\pi(\theta) \geq \pi_*(\theta) \forall \theta \in \Theta_1$ donde $\pi_*(\theta)$ es la función potencia para cualquier otra prueba de hipótesis con tamaño de error tipo I menor o igual que α .

Teorema 2.3 Familia Exponencial

Sea X_1, X_2, \dots, X_n una m.a. con fdp o fmp $f(x; \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$, $\theta \in \Theta$ si existe $k \in \mathbb{R}$ tal que

$$\alpha = \begin{cases} P[\sum_{i=1}^n d(X_i) > k] & \text{si } c \text{ es monóticamente decreciente,} \\ P[\sum_{i=1}^n d(X_i) < k] & \text{si } c \text{ es monóticamente creciente,} \end{cases}$$

entonces la prueba con región crítica:

$$\mathcal{C} = \begin{cases} \{(x_1, x_2, \dots, x_n) | \sum_{i=1}^n d(x_i) > k\} & \text{si } c \text{ es monóticamente decreciente,} \\ \{(x_1, x_2, \dots, x_n) | \sum_{i=1}^n d(x_i) < k\} & \text{si } c \text{ es monóticamente creciente,} \end{cases}$$

es la prueba más potente de tamaño α para las hipótesis

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0 \text{ o bien } H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0.$$

Capítulo 3

Series de tiempo

En el presente capítulo se abordará de manera breve la teoría del análisis de series de tiempo.

3.1. Series de tiempo vistas como procesos estocásticos

Definición 3.1

Un proceso estocástico, $\{X_t\}_{t \in T}$, es **estacionario** o **fuertemente estacionario** si la distribución del vector $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ es la misma que la de $(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$, para cualquier $(t_1, t_2, \dots, t_n) \in T$ tal que $(t_1+k, t_2+k, \dots, t_n+k) \in T$.

Por ejemplo, la Figura 3.1 muestra una serie de variables aleatorias (X_1, X_2, X_3, \dots) independientes e idénticamente distribuidas $N(0, 1)$.

Sin embargo, debido a lo restrictivo que resulta el supuesto de estacionariedad fuerte se introduce la siguiente definición.

Definición 3.2

Se dice que un proceso estocástico, $\{X_t\}_{t \in T}$, es **estacionario de segundo orden** o **débilmente estacionario** si $\{X_t\}_{t \in T}$ es un proceso de segundo orden, esto es que $\mathbb{E}[X_t^2] < \infty \forall t \in T$, y además que para cualesquiera $r, s \in T$ tales que $r+k, s+k \in T$ se cumple que:

$$\text{Cov}(X_r, X_{r+k}) = \text{Cov}(X_s, X_{s+k}).$$

Se observa que un proceso de segundo orden fuertemente estacionario es también débilmente estacionario. Por lo general se asume que $T = \mathbb{Z}$, así, la última expresión

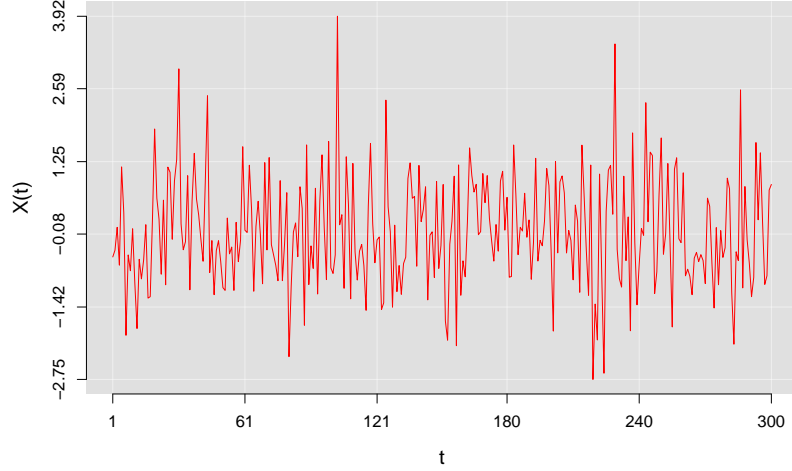


Figura 3.1: Proceso fuertemente estacionario.

define una función para cada $k \in \mathbb{Z}$, llamada función de **autocovarianza** y se denota por γ_k , en consecuencia la función de **autocorrelación** (ACF) se define como $\rho_k := \frac{\gamma_k}{\gamma_0}$ y cumplen las siguientes propiedades:

- $-1 \leq \rho_k \leq 1$
- $\rho_k = \rho_{-k}$
- $\lim_{k \rightarrow \infty} \rho_k = 0$

Al valor k se le conoce como **rezago** (lag).

Más procesos:

Proceso de ruido blanco $WN(\mu, \sigma^2)$

El **ruido blanco** es un proceso, $\{X_t\}_{t \in T}$, de variables aleatorias no correlacionadas tales que $\mathbb{E}[X_t] = \mu$ y $\text{Var}[X_t] = \sigma^2 \quad \forall t \in T$, claramente

$$\gamma_k = \sigma^2 \mathbb{1}_{\{0\}}(k) \text{ y } \rho_k = \mathbb{1}_{\{0\}}(k).$$

Por ejemplo, se puede probar que si $\{W_t\}_{t \in \mathbb{Z}}$ y $\{Z_t\}_{t \in \mathbb{Z}}$ son procesos independientes de variables aleatorias independientes con distribución:

$$f_{W_t}(x) = \begin{cases} \frac{1}{2} & , \quad x = 0, 1 \\ 0 & , \quad e.o.c. \end{cases}, \quad f_{Z_t}(x) = \begin{cases} \frac{1}{2} & , \quad x = -1, 1 \\ 0 & , \quad e.o.c. \end{cases}.$$

Entonces, el proceso de la Figura 3.2 definido como $X_t = W_t (1 - W_{t-1})$ es un proceso de ruido blanco.

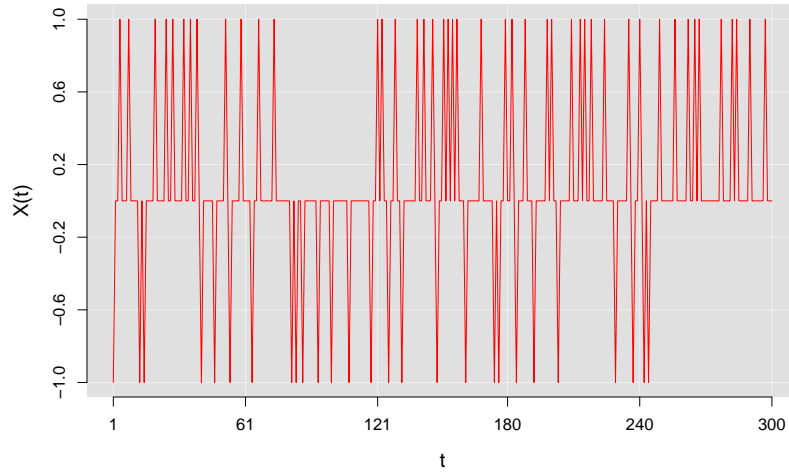


Figura 3.2: Proceso de ruido blanco.

3.2. Modelos univariados de series de tiempo

Proceso de promedios móviles de orden q $MA(q)$

Un proceso de **promedios móviles** de orden q es de la forma:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

donde $\{\varepsilon_t\}_{t \in T}$ es un proceso de ruido blanco, entonces $\{X_t\}_{t \in T}$ es un proceso estacio-

nario de segundo orden, $E[X_t] = 0$ y la función de autocorrelación está dada por:

$$\rho_k = \begin{cases} \frac{\sum_{j=0}^{q-|k|} \theta_j \theta_{|k|+j}}{\sum_{j=0}^q \theta_j^2} & , \quad |k| \leq q \\ 0 & , \quad e.o.c. \end{cases} \quad (3.1)$$

En la Figura 3.3 se muestra un proceso de promedio móvil de orden 1 con $\theta = 0.5$.

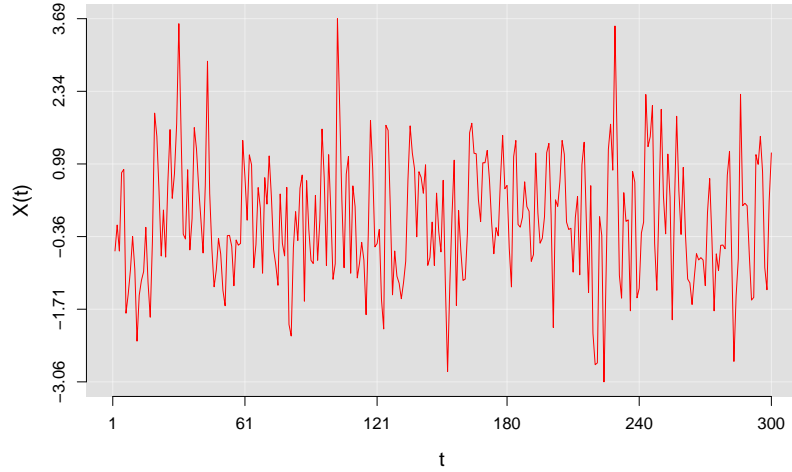


Figura 3.3: Proceso de promedio móvil $MA(1)$ con $\theta = 0.5$.

Proceso autorregresivo de orden p $AR(p)$

Un proceso **autorregresivo** de orden p tiene la forma:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t,$$

donde $\{\varepsilon_t\}_{t \in T}$ es un proceso de ruido blanco, no cualquier proceso de la misma forma es estacionario de segundo orden, por lo que se introduce la siguiente definición y en consecuencia un teorema.

Definición 3.3 Causalidad Un proceso estocástico, $\{X_t\}_{t \in T}$ es **causal** si puede ser escrito como:

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \phi_j \varepsilon_{t-j}$$

donde $\{\varepsilon_t\}_{t \in T}$ es un proceso de ruido blanco.

Teorema 3.1 Condición de causalidad

Sea $\{X_t\}_{t \in T}$ un proceso estocástico de la forma

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t.$$

$\{X_t\}_{t \in T}$ es causal si y sólo si $|z| > 0 \quad \forall z \in \{x \mid 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0\}$, donde $|\bullet|$ es el módulo complejo.

En consecuencia, cualquier proceso estocástico causal es estacionario de segundo orden, volviendo al proceso autorregresivo, se tiene que éste tiene media cero y la función de autocorrelación cumple con lo siguiente:

$$\begin{aligned} \rho_k &= \text{Cov}[X_t, X_{t-k}] = \mathbb{E}[X_t X_{t-k}] - \mathbb{E}[X_t] \mathbb{E}[X_{t-k}] = \mathbb{E}[X_t X_{t-k}] \\ &= \mathbb{E}[(X_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t) X_{t-k}] \\ &= \phi_1 \mathbb{E}[X_{t-1} X_{t-k}] + \phi_2 \mathbb{E}[X_{t-2} X_{t-k}] + \dots + \phi_p \mathbb{E}[X_{t-p} X_{t-k}] + \mathbb{E}[\varepsilon_t X_{t-k}] \\ &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} + \mathbb{E}[\varepsilon_t X_{t-k}] \\ &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \end{aligned}$$

la última igualdad se debe a la causalidad ($k > 0$).

$$\therefore \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}.$$

La expresión anterior permite formular sistemas de ecuaciones que permite encontrar por completo la función de autocorrelación.

En la Figura 3.4 se muestra un proceso autorregresivo de orden 1 con $\phi = .9$.

**Proceso autorregresivo de media móvil
de orden (p,q) ARMA(p,q)**

Un proceso **autorregresivo de media móvil** de orden (p, q) es de la forma:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \dots + \theta_q \varepsilon_{t-q},$$

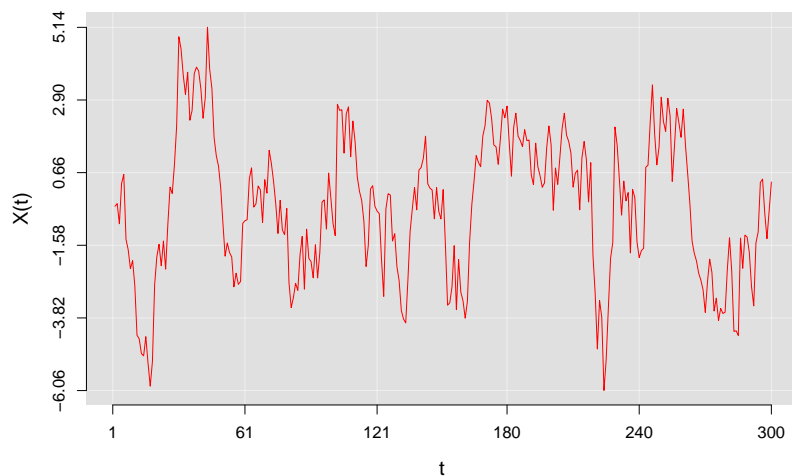


Figura 3.4: Proceso autorregresivo $AR(1)$ con $\phi = 0.9$.

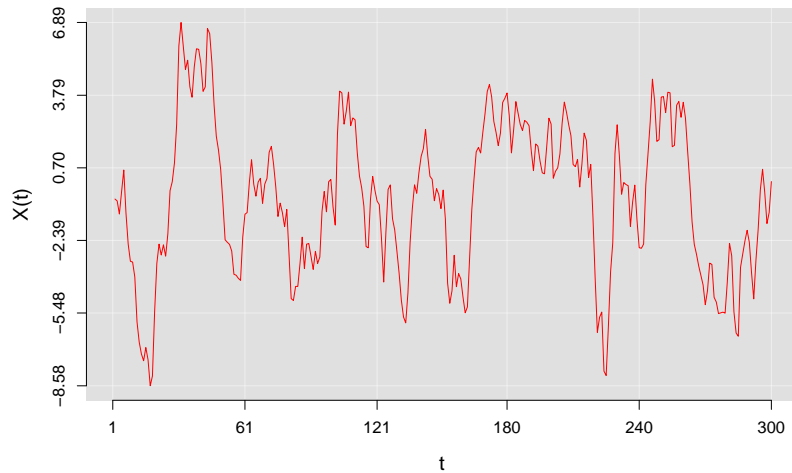


Figura 3.5: Proceso autorregresivo de media móvil.

donde $\{\varepsilon_t\}_{t \in T}$ es un proceso de ruido blanco, como el proceso consta de una parte autorregresiva, las condiciones de estacionariedad son las mismas que las de un proceso autorregresivo, en la Figura 3.5 se muestra un proceso autorregresivo de media móvil de orden $(1, 1)$ con $\theta = .5$ y $\phi = .9$.

Se definen los siguiente polinomios

$$\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q,$$

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p,$$

y los operadores lineales para un proceso $\{X_t\}_{t \in \mathbb{Z}}$

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

$$B(X_t) = X_{t-1} \text{ (Retraso).}$$

Así, es posible escribir las transformaciones de los polinomios Θ y Φ en términos del operador B :

$$\Theta(B) = \mathbb{I} + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q,$$

$$\Phi(B) = \mathbb{I} - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

donde $B^n(Y_t) := B(B^{n-1}(Y_t)) = Y_{t-n}$.

De acuerdo con la notación anterior, los procesos $MA(q)$, $AR(p)$ y $ARMA(p, q)$ se escriben como:

$$X_t = \Theta(B)[\varepsilon_t],$$

$$\Phi(B)[X_t] = \varepsilon_t,$$

$$\Phi(B)[X_t] = \Theta(B)[\varepsilon_t],$$

respectivamente.

Proceso autorregresivo integrado de media móvil de orden (p,d,q) $ARIMA(p, d, q)$

Un proceso **autorregresivo integrado de media móvil** de orden (p, d, q) es aquel que cumple que $\nabla^d(X_t)$ es un proceso $ARMA(p, q)$, es decir:

$$\begin{aligned} \nabla^d(X_t) - \phi_1 \nabla^d(X_{t-1}) - \phi_2 \nabla^d(X_{t-2}) - \dots - \phi_p \nabla^d(X_{t-p}) \\ = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

o bien

$$\Phi(B)[Y_t] = \Theta(B)[\varepsilon_t],$$

donde $\{\varepsilon_t\}_{t \in T}$ es un proceso de ruido blanco, $Y_t = \nabla^d(X_t)$ y $\nabla^d(X_t) := \nabla(\nabla^{d-1}(X_t))$.

Un proceso $ARIMA(p, d, q)$ no necesariamente es estacionario, no obstante se vuelve estacionario al aplicar el operador ∇^d , por ejemplo, en la Figura 3.6 se muestra un proceso $ARIMA$ de orden $(1, 1, 1)$ con $\theta = 0.6$ y $\phi = 0.8$, posteriormente se le aplica el operador diferencia para obtener el proceso $ARMA$ de orden $(1, 1)$ con $\theta = 0.6$ y $\phi = 0.8$ Figura 3.7.

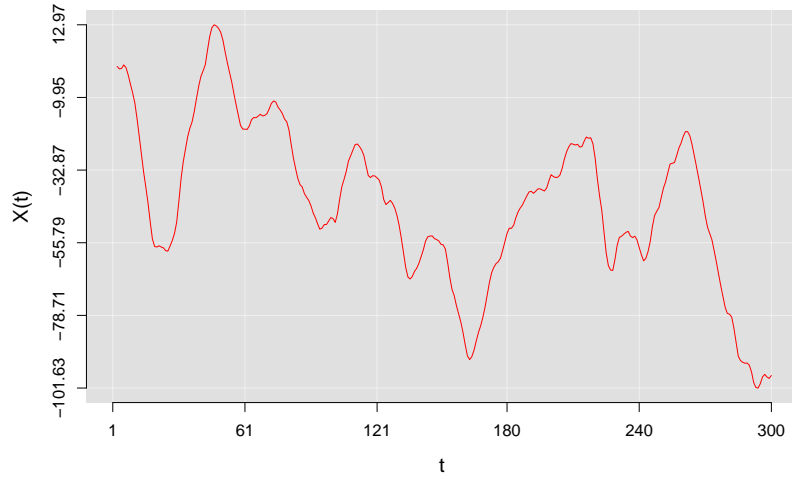


Figura 3.6: Proceso autorregresivo integrado de media móvil $ARIMA(1, 1, 1)$ con $\theta = 0.6$ y $\phi = 0.8$.

Dada la muestra $\{X_1, X_2, \dots, X_n\}$, una forma de estimar el orden de la diferencia necesaria para estacionarizar una serie es hallar el valor d que minimice la desviación estándar muestral de las series $\{\nabla^d(X_t)\}$, es decir d es tal que

$$S_{\nabla^d(X_t)} = \min \{S_{\nabla^r(X_t)} \mid r \in \{0, 1, 2, 3\}\}, \quad (3.2)$$

donde

$$S_{\nabla^r(X_t)} = \sqrt{\frac{1}{n-r-1} \sum_{t=r+1}^n \left(\nabla^r(X_t) - \sum_{j=r+1}^n \frac{\nabla^r(X_j)}{n-r} \right)^2}.$$

En la expresión (3.2) el rango de r puede extenderse a $r \geq 0$, no obstante lo más recomendable es tomar únicamente $r \in \{0, 1, 2, 3\}$, debido a las dificultades que pueden resultar de trabajar con diferencias de orden “grande” [Guerrero, 2003].

A partir de una muestra $\{X_1, X_2, \dots, X_n\}$ de un proceso estacionario con media μ la

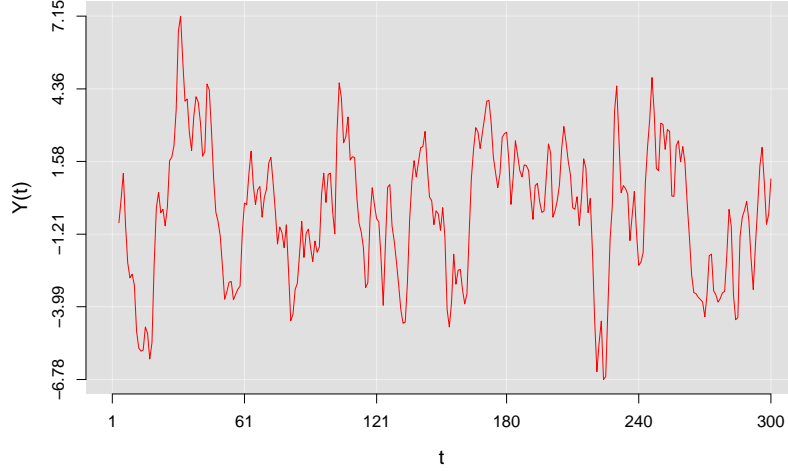


Figura 3.7: Diferencia de primer orden, $Y_t = \nabla(X_t) = X_t - X_{t-1}$.

función de autocovarianza puede estimarse como:

$$\hat{\gamma}_k = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) & \text{si } 0 \leq |k| \leq n-1 \\ 0 & \text{e.o.c.} \end{cases} \quad (3.3)$$

La función (3.3) es un estimador asintóticamente insesgado de $\gamma(k)$ y \bar{X} tiene una distribución asintótica cuando $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$:

$$\bar{X} \simeq N\left(\mu, \frac{1}{n} \sum_{k=-\infty}^{\infty} \gamma(k)\right),$$

lo anterior se entiende como que entre mayor sea el tamaño de la muestra, las estimaciones $\hat{\gamma}_k$ y $\hat{\rho}_k$ serán más confiables, donde

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}.$$

En un modelo de promedio móvil de orden q se tiene que la función $ACF(k)$ vale cero cuando $k > q$ (3.1), si por ejemplo, se tiene una muestra de un modelo de promedio móvil, la función $\hat{\rho}$ podría darnos evidencia del orden del proceso. De manera

general $\hat{\rho}(k)$ define un intervalo al $(1 - \alpha) 100\%$ de confianza para ρ_k de la siguiente forma:

$$\mathbb{I}_k = \left[\hat{\rho}_k - \frac{Z_{1-\alpha/2}}{\sqrt{n}}, \hat{\rho}_k + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \right],$$

donde $Z_{1-\alpha/2}$ es el cuantil $1 - \alpha/2$ de una distribución normal estándar. Por lo que se puede construir la prueba

$$H_0 : \rho_k = 0 \quad \text{vs.} \quad H_1 : \rho_k \neq 0,$$

que se rechaza cuando $0 \notin \mathbb{I}_k$, es decir $|\hat{\rho}_k| > \frac{Z_{1-\alpha/2}}{\sqrt{n}}$, en las Figuras 3.8 y 3.9 se muestran ejemplos de la función de autocorrelación estimada de procesos de promedio móvil de orden 4 y 7 respectivamente, en las cuales se observa que los valores de ACF para los rezagos mayores que el orden se encuentran contenidos en las bandas de confianza $|y| \leq \frac{Z_{1-\alpha/2}}{\sqrt{n}}$.

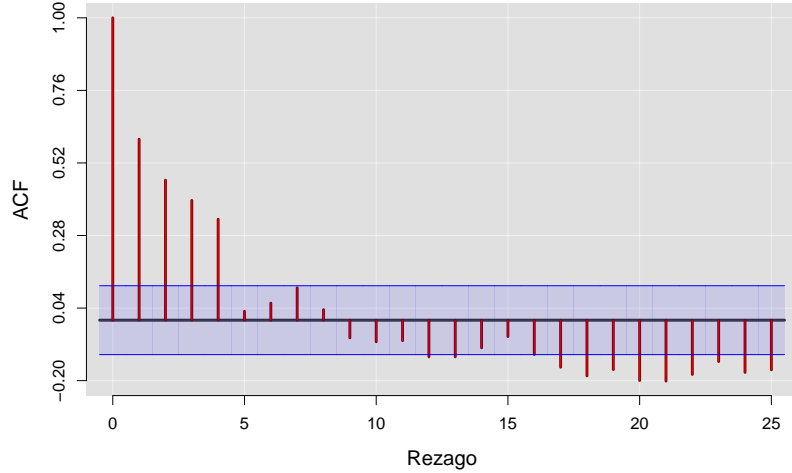
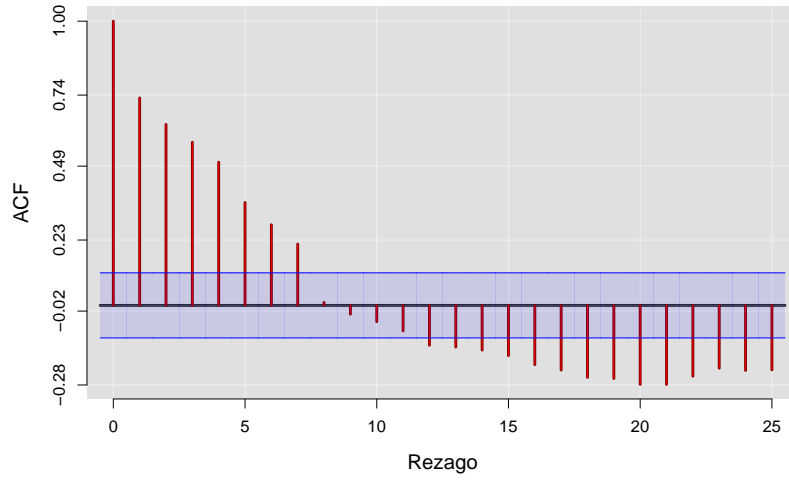
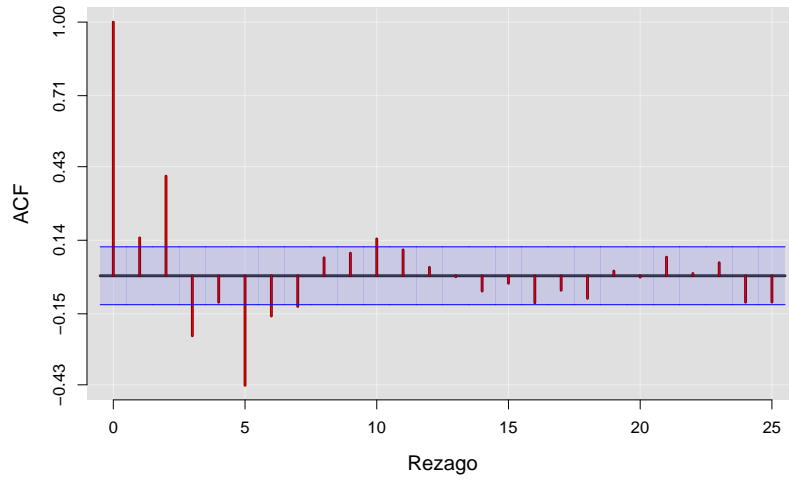


Figura 3.8: ACF estimada de un proceso $MA(4)$

Por otro lado la función $\hat{\rho}$ no proporciona información acerca del orden de un proceso con una parte autorregresiva, pero si podría evidenciar la presencia de dicho comportamiento autorregresivo, ya que en estos procesos el decaimiento de ρ es progresivo a diferencia de en un promedio móvil, en las Figuras 3.10 y 3.11 se muestra la función de autocorrelación estimada para un $AR(5)$ y un $ARMA(3, 1)$ respectivamente.

Para resolver el problema de identificar un proceso estacionario de segundo orden con un componente autorregresivo se construye la función de **autocorrelación parcial** (PACF), para un proceso de media cero se define

$$\hat{X}_{t+n+1} = \phi_{n1}X_{t+n} + \phi_{n2}X_{t+n-1} + \dots + \phi_{nn}X_{t+1} \quad (3.4)$$

Figura 3.9: ACF estimada de un proceso $MA(7)$ Figura 3.10: ACF estimada de un proceso $AR(5)$.

donde $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ son aquellos que minimizan el error

$$\mathbb{E} \left[(X_{t+n+1} - \phi_{n1}X_{t+n} - \phi_{n2}X_{t+n-1} - \dots - \phi_{nn}X_{t+1})^2 \right], \quad (3.5)$$

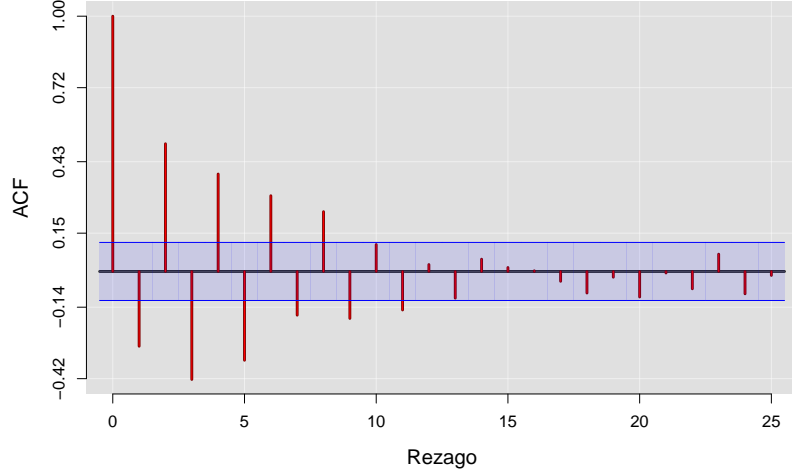


Figura 3.11: ACF estimada de un proceso $ARMA(3, 1)$.

debido a la estacionariedad del proceso, las expresiones (3.4) y (3.5) no dependen de t , por lo que conviene escribir simplemente:

$$\hat{X}_{n+1} = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \dots + \phi_{nm}X_1. \quad (3.6)$$

El problema anterior define el sistema de ecuaciones

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \phi_{n3} \\ \vdots \\ \phi_{nm} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_n \end{pmatrix}$$

o bien

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \phi_{n3} \\ \vdots \\ \phi_{nm} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_n \end{pmatrix}.$$

Debido a las dificultades que resultan de invertir matrices de gran tamaño, de manera alternativa, se utiliza el algoritmo Durbin-Levison [Stoffer, 2006] para hallar dichas soluciones:

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \quad j < k,$$

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j}. \quad (3.7)$$

Así, la función de autocorrelación parcial de un proceso débilmente estacionario se define como $(\phi_{11}, \phi_{22}, \phi_{33}, \dots)$, en una muestra $\{X_1, X_2, \dots, X_n\}$ un intervalo al $(1 - \alpha)$ 100% de confianza para ϕ_{kk} es

$$\mathbb{I}_{kk} = \left[\hat{\phi}_{kk} - \frac{Z_{1-\alpha/2}}{\sqrt{n}}, \hat{\phi}_{kk} + \frac{Z_{1-\alpha/2}}{\sqrt{n}} \right],$$

el cual plantea la prueba de hipótesis

$$H_0 : \phi_k = 0 \quad \text{vs.} \quad H_1 : \phi_k \neq 0,$$

que se rechaza cuando $0 \notin \mathbb{I}_{kk}$, es decir $|\hat{\phi}_k| > \frac{Z_{1-\alpha/2}}{\sqrt{n}}$.¹ El criterio para evidenciar el orden de la parte autorregresiva (p) en un modelo $AR(p)$ o $ARMA(p, q)$, será que la función PACF estimada quede contenida en las bandas de confianza para rezagos mayores a p . En las Figuras 3.12 y 3.13 se muestran ejemplos de PACF estimadas para un proceso autorregresivo y un proceso autorregresivo de media móvil, respectivamente.

En una muestra de tamaño n de un proceso de media cero, la ecuación (3.6) permite obtener la mejor proyección lineal a un paso como

$$\hat{X}_{n+1} = \hat{\phi}_{n1}X_n + \hat{\phi}_{n2}X_{n-1} + \dots + \hat{\phi}_{nn}X_1.$$

De manera general, puede obtenerse la mejor predicción lineal en k pasos para un proceso con parte autorregresiva de media cero, encontrando los valores

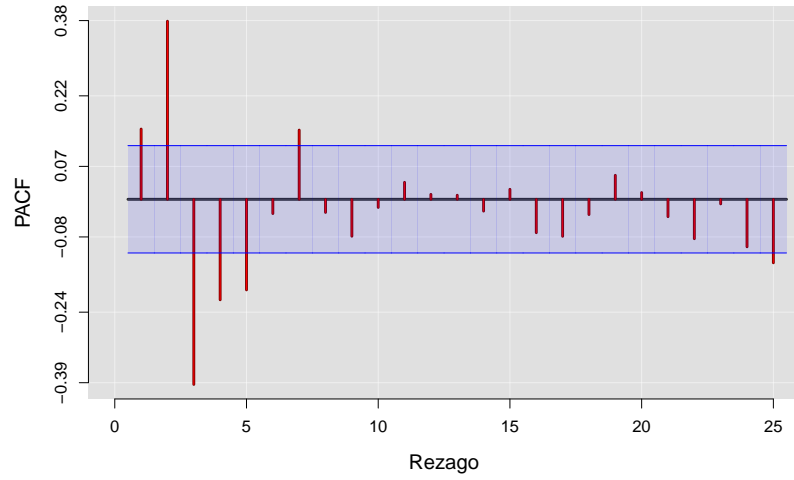
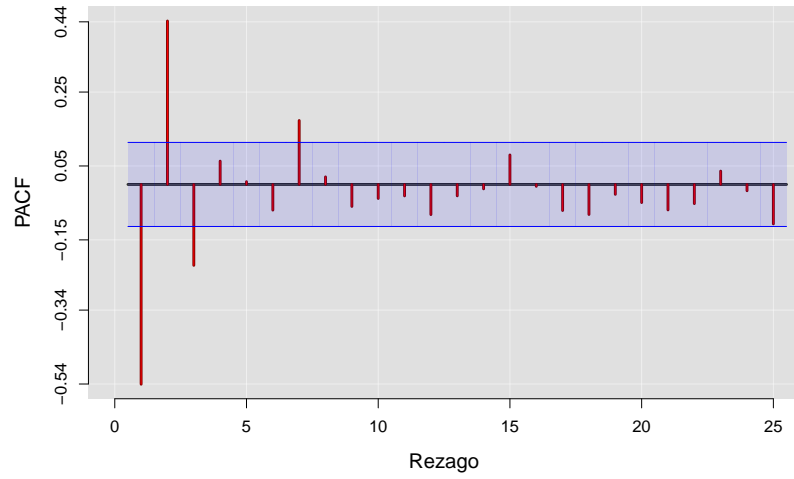
${}_k\phi_{n1}, {}_k\phi_{n2}, \dots, {}_k\phi_{nn}$ que minimizan

$$\mathbb{E} \left[(X_{n+k} - {}_k\phi_{n1}X_n - {}_k\phi_{n2}X_{n-1} - \dots - {}_k\phi_{nn}X_1)^2 \right],$$

lo que da origen al sistema

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} {}_k\phi_{n1} \\ {}_k\phi_{n2} \\ {}_k\phi_{n3} \\ \vdots \\ {}_k\phi_{nn} \end{pmatrix} = \begin{pmatrix} \gamma_k \\ \gamma_{k+1} \\ \gamma_{k+2} \\ \vdots \\ \gamma_{n+k-1} \end{pmatrix}$$

¹ $\hat{\phi}_{kk}$ se obtiene de sustituir ρ_k por $\hat{\rho}_k$ en (3.7).

Figura 3.12: PACF estimada de un proceso $AR(5)$ Figura 3.13: PACF estimada de un proceso $ARMA(3,1)$

o bien

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} {}_k\phi_{n1} \\ {}_k\phi_{n2} \\ {}_k\phi_{n3} \\ \vdots \\ {}_k\phi_{nn} \end{pmatrix} = \begin{pmatrix} \rho_k \\ \rho_{k+1} \\ \rho_{k+2} \\ \vdots \\ \rho_{n+k-1} \end{pmatrix}.$$

[Douglas et al., 2015]

Para resolverlo, es necesario haber establecido un modelo con parte autorregresiva a partir de la muestra X_1, X_2, \dots, X_n , haciendo uso de softwares estadísticos, esta tarea puede realizarse de manera sencilla como se ilustrará en el capítulo 4.

Capítulo 4

El modelo de Lee-Carter

El modelo de Lee-Carter propuesto en 1992 es referenciado como el “*Modelo estadístico líder para el pronóstico de la mortalidad en la literatura demográfica*” [Deaton and Paxson, 2004], tiene como precedentes los modelos de Gompertz (1825) y de Heligman y Pollard (1980). El método de Lee-Carter tiene como objetivo pronosticar la mortalidad utilizando únicamente las variables de edad y tiempo, es decir, subestima avances tecnológicos, cuestiones sociales, fenómenos naturales, etc, hace uso de la estadística paramétrica y de las series de tiempo y su esencia probabilística permite la creación de intervalos de confianza para sus proyecciones [Giroi and King, 2007].

4.1. Tabla de mortalidad

La tabla de mortalidad (o tabla de vida) es un instrumento que resume la mortalidad de una cohorte y permite conocer datos y tendencias importantes de la población como la esperanza de vida, la tabla de mortalidad tiene la siguiente estructura:

x	l_x	d_x	q_x	p_x
-----	-------	-------	-------	-------

donde,

x := edad de los miembros de la cohorte. Se asume que $0 \leq x \leq \omega$, donde ω es una edad máxima.

l_x := número de individuos que llegan con vida a la edad x .

$d_x := l_x - l_{x+1}$ es el número de individuos que mueren entre las edades x y $x + 1$.

$q_x := \frac{d_x}{l_x}$ es la probabilidad de morir entre las edades x y $x+1$.

$p_x := \frac{l_{x+1}}{l_x}$ es la probabilidad de sobrevivir a la edad x .

4.2. Estimación del modelo de Lee Carter

Supóngase que se construyen las tablas de vida de una cohorte de cierta población para cada año en el intervalo $[t_1, t_n]$, por lo que cada variable de la tabla de vida correspondiente al año t es indexada con j . Lo que plantearon Lee y Carter [Lee and Carter, 1992] fue estimar $q_{x,t}$, la probabilidad de que una persona muera entre la edad x y $x+1$ en el año t , proponiendo el siguiente modelo log-lineal:

$$\log(q_{x,t}) = a_x + b_x k_t + \mathcal{E}_{x,j}, \quad (4.1)$$

o bien

$$q_{x,t} = e^{a_x + b_x k_t + \mathcal{E}_{x,j}} = e^{a_x} e^{b_x k_t} e^{\mathcal{E}_{x,j}},$$

Con las siguientes restricciones de unicidad:

$$\sum_{x=0}^{\omega} b_x = 1, \quad (4.2)$$

$$\sum_{t=1}^n k_t = 0. \quad (4.3)$$

A los coeficientes a_x se les conoce como parámetros de **forma** y a los b_x parámetros de **cambio**, ambos dependen únicamente de la edad, los coeficientes k_t son **índices de mortalidad** y no están determinados por la edad, sino por el tiempo, finalmente, $\mathcal{E}_{x,j}$ es un proceso de ruido blanco con media cero y varianza $\sigma_{\mathcal{E}}^2$ ¹, en esta serie se ven reflejadas todas aquellas variables excluidas por el modelo, es decir, la mortalidad causada por agentes no naturales.

El coeficiente e^{a_x} describe de manera general el comportamiento de la mortalidad, b_x explica la proporción en la que cambia la velocidad con la que crece o decrece $q_{x,t}$ con respecto a k_t , es decir, $\frac{d \log(q_{x,t})}{dt} = b_x \frac{dk_t}{dt}$, y k_t es la tendencia de la mortalidad a lo largo del tiempo.

¹ Se observa que:

- $E[\log(q_{x,t})] = a_x + b_x k_t$
- $\text{Var}[\log(q_{x,t})] = \sigma_{\mathcal{E}}^2$

4.2.1. Estimación puntual

De las condiciones (4.2) y (4.3), propuestas por Lee y Carter, es posible obtener los estimadores para a_x y k_t para todo $x \in \{0, 1, \dots, \omega\}$ y $j \in \{1, 2, \dots, n\}$. Sumando (4.1) sobre todos los valores de t :

$$\sum_{t=1}^n \log(q_{x,t}) = \sum_{t=1}^n a_x + b_x k_t + \mathcal{E}_{x,j} = na_x + b_x \sum_{t=1}^n k_t + \sum_{t=1}^n \mathcal{E}_{x,j} = na_x + \sum_{t=1}^n \mathcal{E}_{x,j}.$$

La última igualdad se debe a (4.3)

$$\Rightarrow \sum_{t=1}^n \log(q_{x,t}) = n\hat{a}_x \quad \Rightarrow \quad \hat{a}_x = \frac{1}{n} \sum_{t=1}^n \log(q_{x,t}).$$

Sumando (4.1) sobre todos los valores de x :

$$\begin{aligned} \sum_{x=0}^{\omega} \log(q_{x,t}) &= \sum_{x=0}^{\omega} (a_x + b_x k_t + \mathcal{E}_{x,j}) = \sum_{x=0}^{\omega} a_x + k_t \sum_{x=0}^{\omega} b_x + \sum_{x=0}^{\omega} \mathcal{E}_{x,j} \\ &= \sum_{x=0}^{\omega} a_x + k_t + \sum_{x=0}^{\omega} \mathcal{E}_{x,j} \end{aligned}$$

La última igualdad se debe a (4.2)

$$\Rightarrow \sum_{x=0}^{\omega} \log(q_{x,t}) = \sum_{x=0}^{\omega} \hat{a}_x + \hat{k}_t \quad \Rightarrow \quad \hat{k}_t = \sum_{x=0}^{\omega} (\log(q_{x,t}) - \hat{a}_x)$$

Finalmente, resta hallar las estimaciones para los b_x , se presentan, a continuación, dos formas de hacerlo las cuales convergen a los mismos resultados:

4.2.2. Método de mínimos cuadrados

El método de mínimos cuadrados consiste en realizar un ajuste de una función conjeturada a un conjunto de datos de forma óptima, es decir, minimizando el error entre la función estimada y los datos.

Considérense el conjunto de datos $\{y_1, y_2, \dots, y_n\} \subseteq \mathbb{R}$ donde cada y_i está en función de valores en \mathbb{R}^m , para alguna $m \in \mathbb{N}$, es decir,

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{im}),$$

y dada la función $G_{\theta_1, \theta_2, \dots, \theta_p} : B \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$, entonces, la mejor parametrización para G , $G_{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p}$, será tal que el vector $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$ sea un máximo de la función $\mathcal{E} : A \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$, tal que $\mathcal{E}(\theta_1, \theta_2, \dots, \theta_p) = \sum_{i=1}^n [y_i - G_{\theta_1, \theta_2, \dots, \theta_p}(x_{i1}, x_{i2}, \dots, x_{im})]^2$.

De acuerdo con el procedimiento anterior, las estimaciones para los parámetros de forma $\{b_x\}_{x=0}^\omega$ están dadas por los valores que minimizan

$$\mathcal{E} = \sum_{x=0}^\omega \sum_{t=1}^n [\log(q_{x,t}) - \hat{a}_x - b_x \hat{k}_t]^2, \quad \text{donde } A = \left\{ (b_0, b_1, \dots, b_\omega) \mid \sum_{x=0}^\omega b_x = 1 \right\}.$$

Se procede a optimizar la función \mathcal{E} :

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial b_i} &= 2 \sum_{t=1}^n [\log(q_{i,t}) - a_i - b_i \hat{k}_t] (-\hat{k}_t) = 2 \left[\sum_{t=1}^n -\hat{k}_t \log(q_{i,t}) + \hat{k}_t a_i + b_i \hat{k}_t^2 \right] \\ &= 2 \left[-\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) + a_i \sum_{t=1}^n \hat{k}_t + b_i \sum_{t=1}^n \hat{k}_t^2 \right] = 2 \left[-\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) + b_i \sum_{t=1}^n \hat{k}_t^2 \right], \end{aligned}$$

la última igualdad debe a (4.3), por lo que

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial b_i} = 0 &\Leftrightarrow -\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) + b_i \sum_{t=1}^n \hat{k}_t^2 = 0, \\ \Rightarrow b_i^* &= \frac{\sum_{t=1}^n \hat{k}_t \log(q_{i,t})}{\sum_{t=1}^n \hat{k}_t^2}, \end{aligned}$$

en consecuencia, se tiene que $(b_0^*, b_1^*, \dots, b_\omega^*)$ es punto crítico de \mathcal{E} , para caracterizarlo se debe calcular la matriz Hessiana ($H_{\mathcal{E}}$).

$$\frac{\partial^2 \mathcal{E}}{\partial b_l \partial b_i} = \frac{\partial}{\partial b_l \partial b_i} 2 \left[-\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) + b_i \sum_{t=1}^n \hat{k}_t^2 \right] = \begin{cases} \sum_{t=1}^n \hat{k}_t^2 & \text{si } i = l \\ 0 & \text{si } i \neq l \end{cases}.$$

$$\Rightarrow H_{\mathcal{E}}(b_0, b_1, \dots, b_\omega) = \begin{pmatrix} \sum_{t=1}^n \hat{k}_t^2 & 0 & \cdots & 0 \\ 0 & \sum_{t=1}^n \hat{k}_t^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{t=1}^n \hat{k}_t^2 \end{pmatrix} = \left(\sum_{t=1}^n \hat{k}_t^2 \right) \mathbb{I}_{(n \times n)}$$

$$\therefore \det(H_{\mathcal{E}}) = \left(\sum_{t=1}^n \hat{k}_t^2 \right)^n > 0 \text{ y } \frac{\partial^2}{\partial b_0^2}(b_0, b_1, \dots, b_\omega) \sum_{t=1}^n \hat{k}_t^2 > 0,$$

por lo que b^* es mínimo relativo de \mathcal{E} . Finalmente, se debe verificar que $\sum_{i=0}^\omega b_i^* = 1$, en efecto:

$$\begin{aligned} \sum_{i=0}^\omega b_i^* &= \sum_{i=0}^\omega \frac{\sum_{t=1}^n \hat{k}_t \log(q_{i,t})}{\sum_{t=1}^n \hat{k}_t^2} = \sum_{i=0}^\omega \frac{\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) - \hat{a}_i 0}{\sum_{t=1}^n \hat{k}_t^2} \\ &= \sum_{i=0}^\omega \frac{\sum_{t=1}^n \hat{k}_t \log(q_{i,t}) - \hat{a}_i \sum_{t=1}^n \hat{k}_t}{\sum_{t=1}^n \hat{k}_t^2} = \sum_{i=0}^\omega \frac{\sum_{t=1}^n \hat{k}_t (\log(q_{i,t}) - \hat{a}_i)}{\sum_{t=1}^n \hat{k}_t^2} \\ &= \frac{\sum_{t=1}^n \hat{k}_t \sum_{i=0}^\omega (\log(q_{i,t}) - \hat{a}_i)}{\sum_{t=1}^n \hat{k}_t^2} = \frac{\sum_{t=1}^n \hat{k}_t \hat{k}_t}{\sum_{t=1}^n \hat{k}_t^2} = 1. \end{aligned}$$

$$\therefore \hat{b}_i = \frac{\sum_{t=1}^n \hat{k}_t \log(q_{i,t})}{\sum_{t=1}^n \hat{k}_t^2}$$

4.2.3. Máxima verosimilitud

Otra forma de encontrar la estimación de los parámetros, $\{b_x\}_{x=0}^\omega$, es mediante un procedimiento de máxima verosimilitud, para esto se asume en (4.1) que $\mathcal{E}_{x,j}$ tiene una distribución normal con media cero y varianza $\sigma_{\mathcal{E}}^2$, $\mathcal{E}_{x,j} \sim N(0, \sigma_{\mathcal{E}}^2)$, por lo que

$$\log(q_{x,t}) = a_x + b_x k_t + \mathcal{E}_{x,j} \sim N(a_x + b_x k_t, \sigma_{\mathcal{E}}^2),$$

y así, la función de verosimilitud de (b_0, \dots, b_ω) es

$$\begin{aligned} \mathcal{L} &= \prod_{x=0}^\omega \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_{\mathcal{E}}^2}} \exp \left\{ -\frac{1}{2\sigma_{\mathcal{E}}^2} [\log(q_{x,t}) - (a_x + b_x \hat{k}_t)]^2 \right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_{\mathcal{E}}^2}} \right)^{n(\omega+1)} \exp \left\{ \sum_{x=0}^\omega \sum_{j=1}^n -\frac{1}{2\sigma_{\mathcal{E}}^2} [\log(q_{x,t}) - (a_x + b_x \hat{k}_t)]^2 \right\} \\ &= \left(\sqrt{2\pi\sigma_{\mathcal{E}}^2} \right)^{-n(\omega+1)} \exp \left\{ -\frac{1}{2\sigma_{\mathcal{E}}^2} \sum_{x=0}^\omega \sum_{j=1}^n [\log(q_{x,t}) - a_x - b_x \hat{k}_t]^2 \right\}, \end{aligned}$$

si se considera la función $g(y) := \left(\sqrt{2\pi\sigma_\varepsilon^2}\right)^{-n(\omega+1)} e^{\frac{1}{2\sigma_\varepsilon^2}y}$, se tiene que

$$\mathcal{L}(b_0, \dots, b_\omega) = g\left(-\sum_{x=0}^{\omega} \sum_{j=1}^n \left[\log(q_{x,t}) - a_x - b_x \hat{k}_t\right]\right),$$

entonces, como g es creciente, maximizar $\mathcal{L}(b_0, \dots, b_\omega)$ es equivalente a maximizar

$$g^{-1}(\mathcal{L}(b_0, \dots, b_\omega)) = -\sum_{x=0}^{\omega} \sum_{j=1}^n \left[\log(q_{x,t}) - a_x - b_x \hat{k}_t\right]^2.$$

Por lo anterior, para maximizar la función de verosimilitud de (b_0, \dots, b_ω) , $\mathcal{L}(b_0, \dots, b_\omega) = \prod_{x=0}^{\omega} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \left[\log(q_{x,t}) - \hat{a}_x - b_x \hat{k}_t\right]\right\}$, es suficiente minimizar

$$\sum_{x=0}^{\omega} \sum_{j=1}^n \left[\log(q_{x,t}) - \hat{a}_x - b_x \hat{k}_t\right]^2,$$

y así, los estimadores máximo verosímil de (b_0, \dots, b_ω) son los mismos que se obtuvieron por el método de mínimos cuadrados.

4.2.4. Otra forma máximo verosímil

Si se considera la variable aleatoria $D_{x,t}$, definida como el número de defunciones de individuos de edad x en el año t , $D_{x,t}$ tiene una distribución aproximada Poisson con media $l_x q_{x,t}$,

$$D_{x,t} \simeq \text{Poisson}(l_x q_{x,t}) \approx \text{Poisson}(l_x e^{a_x + b_x k_t}),$$

por lo que las estimaciones de $\{a_0, \dots, a_\omega, b_0, \dots, b_\omega, k_{t_1}, \dots, k_{t_n}\}$ pueden ser obtenidas por el método de máxima verosimilitud:

$$\begin{aligned} f_{D_{x,t}}(d_{x,t}) &= \frac{1}{d_{x,t}!} e^{-l_x e^{a_x + b_x k_t}} \left(l_x e^{a_x + b_x k_t}\right)^{d_{x,t}} \\ \Rightarrow \mathcal{L} &= \prod_{x=0}^{\omega} \prod_{t=1}^n \frac{1}{d_{x,t}!} e^{-l_x e^{a_x + b_x k_t}} \left(l_x e^{a_x + b_x k_t}\right)^{d_{x,t}} \end{aligned}$$

Maximizar esta función de verosimilitud puede hacerse por los métodos: Quasi-Newton y Simplex como se plantea en la bibliografía [Aburto, 2012].

4.2.5. Insesgamiento

Los estimadores:

$$\widehat{a}_x = \frac{1}{n} \sum_{t=1}^n \log(q_{x,t}) \quad (4.4)$$

$$\widehat{b}_x = \frac{\sum_{t=1}^n \widehat{k}_t \log(q_{x,t})}{\sum_{t=1}^n \widehat{k}_t^2} \quad (4.5)$$

$$\widehat{k}_t = \sum_{x=0}^{\omega} \log(q_{x,t}) - \widehat{a}_x \quad (4.6)$$

son insesgados:

De (4.4)

$$\begin{aligned} \mathbb{E}[\widehat{a}_x] &= \mathbb{E} \left[\frac{1}{n} \sum_{t=1}^n \log(q_{x,t}) \right] = \frac{1}{n} \sum_{t=1}^n \mathbb{E}[\log(q_{x,t})] = \frac{1}{n} \sum_{t=1}^n (a_x + b_x k_t) \\ &= \frac{1}{n} \left(n a_x + b_x \sum_{t=1}^n k_t \right) = \frac{1}{n} (n a_x) = a_x, \end{aligned}$$

la penúltima igualdad se debe a (4.3).

$$\therefore \mathbb{E}[\widehat{a}_x] = a_x$$

De (4.6)

$$\begin{aligned} \mathbb{E}[\widehat{k}_t] &= \mathbb{E} \left[\sum_{x=0}^{\omega} (\log(q_{x,t}) - \widehat{a}_x) \right] = \sum_{x=0}^{\omega} (\mathbb{E}[\log(q_{x,t})] - \mathbb{E}[\widehat{a}_x]) \\ &= \left(\sum_{x=0}^{\omega} a_x + b_x k_t - a_x \right) = \sum_{x=0}^{\omega} b_x k_t = k_t \sum_{x=0}^{\omega} b_x = k_t \end{aligned}$$

la última igualdad se debe a (4.2).

$$\therefore \mathbb{E}[\widehat{k}_t] = k_t$$

De (4.5)

$$\mathbb{E}[\widehat{b}_x] = \mathbb{E} \left[\frac{\sum_{t=1}^n \widehat{k}_t \log(q_{x,t})}{\sum_{t=1}^n \widehat{k}_t^2} \right] = \mathbb{E} \left[\frac{\sum_{t=1}^n \widehat{k}_t \log(q_{x,t})}{\sum_{t=1}^n \widehat{k}_t^2} \right]$$

$$\begin{aligned}
& \sim \frac{\sum_{t=1}^n k_t \mathbb{E}[\log(q_{x,t})]}{\sum_{t=1}^n k_t^2} = \frac{\sum_{t=1}^n k_t (a_x + b_x k_t)}{\sum_{t=1}^n k_t^2} \\
& = \frac{a_x \sum_{t=1}^n k_t + b_x \sum_{t=1}^n k_t^2}{\sum_{t=1}^n k_t^2} = \frac{b_x \sum_{t=1}^n k_t^2}{\sum_{t=1}^n k_t^2} = b_x,
\end{aligned}$$

la penúltima igualdad se debe a (4.3).

$$\therefore \mathbb{E}[\widehat{b}_x] \sim b_x$$

4.3. Pronósticos

Para poder hacer pronósticos de la mortalidad, una vez realizadas las estimaciones pertinentes, se procede a modelar el comportamiento de los índices de mortalidad (k_t) esto puede hacerse viendo a $\{k_t\}_t$ como un proceso estocástico² de la forma

$$k_t = k + k_{t-1} + e_t,$$

[García and Ordorica, 2012]

o bien, el desarrollado en este trabajo, como un un proceso ARIMA(p, d, q) de series de tiempo, para el primero, k es una constante y los e_t son variables aleatorias independientes de distribución $N(0, \sigma_e^2)$, se puede estimar el parámetro k como

$$\widehat{k} = \frac{1}{n-1} \sum_{t=2}^n (k_t - k_{t-1}),$$

y la varianza σ_e^2 como

$$\widehat{\sigma_e^2} = \sqrt{\frac{1}{n-2} \sum_{t=2}^n (\widehat{e}_t - \widehat{k})^2}.$$

Una vez establecido un modelo para el proceso $\{k_t\}$ se realizan las estimaciones futuras de los valores $\widehat{k_{n+i}}$, para finalmente encontrar las proyecciones de las tasas centrales de mortalidad a edad x en el año $t_n + i$, como

$$\widehat{q_{x,t_n+i}} = e^{\widehat{a}_x + \widehat{b}_x \widehat{k_{n+i}}}, \quad i \in (1, \dots, h). \quad (4.7)$$

²Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizadas por un conjunto T , llamado espacio parametral, en donde las variables toman valores en un conjunto S llamado espacio de estados.[Rincón, 2012]

En el presente el año inicial de observación, t_1 , es 1990, la muestra es de tamaño 25, es decir $n = 25$, por lo que el último año de observación, t_n , es 2014, y se hará una predicción de la información para los próximos 30 años, es decir $h = 30$.

Como ejemplo utilizaremos el caso de la población total mexicana, haciendo uso de (4.4), (4.5) y (4.6) se encuentran las estimaciones \hat{a}_x , \hat{b}_x así como los valores de la serie \hat{k}_t , las cuales pueden verse en las Tablas 4.1, 4.2 y 4.3.

A continuación, se procede a analizar el comportamiento de la serie $\{k_t\}_j$, cuya gráfica puede verse en la Figura 4.1, que al no presentar un comportamiento estacionario se le aplicarán diferencias, haciendo uso del criterio (3.2) se determina que es necesario aplicar una única diferencia: $Y_t := \nabla k_t = k_t - k_{t-1}$ ver Figura 4.2.

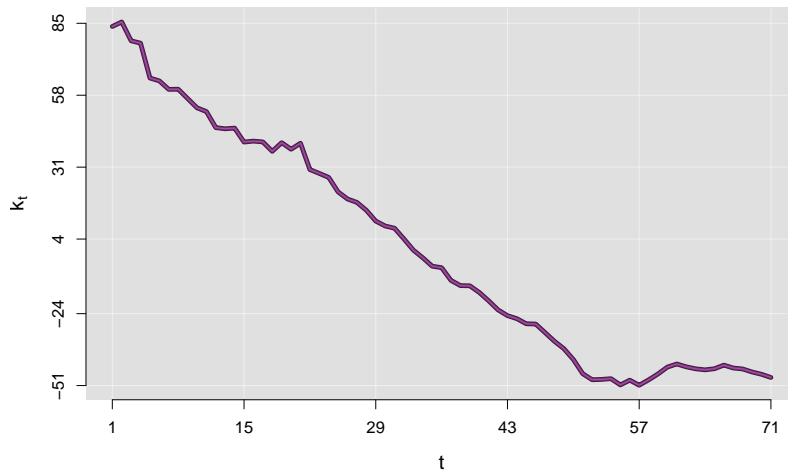


Figura 4.1: Serie k_t .

Observando la función ACF estimada de la serie Figura 4.3, se deduce que Y_t consta de un componente autorregresivo, debido a su decremento (en valor absoluto) progresivo, por otro lado la función PACF estimada sugiere que podría tratarse de un proceso autorregresivo de orden 5 Figura 4.4.

Después de realizar varias pruebas se obtiene que un ARMA(4,3) es un buen modelo para la serie Y_t , en terminos de la varianza de los residuales y cumplimiento de normalidad de estos, es decir que la serie k_t (centrada en cero) puede ser modelada mediante un ARIMA(4,1,3), en la Figura 4.5 puede verse que la función ACF estimada de los residuales es congruente con la de un proceso de ruido blanco.

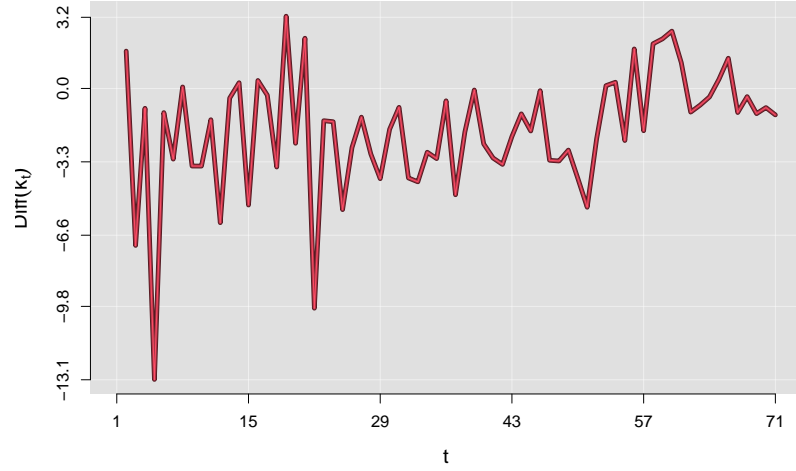


Figura 4.2: Primera diferencia de la serie k_t .

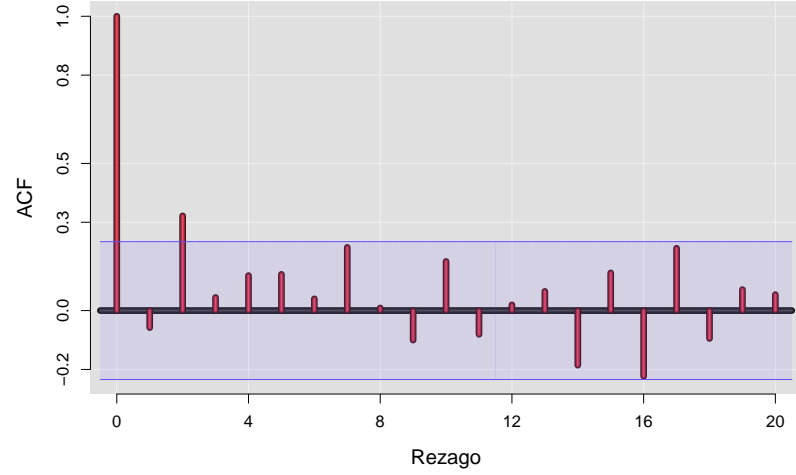


Figura 4.3: Estimación $\hat{\rho}_j$ de la serie ∇k_t centrada en cero.

El modelo ARMA(4,3) obtenido para la serie Y_t es el siguiente

$$Y_t + 0.9423Y_{t-1} + 0.8711Y_{t-2} + 0.8868Y_{t-3} + 0.8371Y_{t-4} \\ = -0.9409 + 0.6239\varepsilon_t + 0.9646\varepsilon_{t-1} + 0.6763\varepsilon_{t-2}.$$

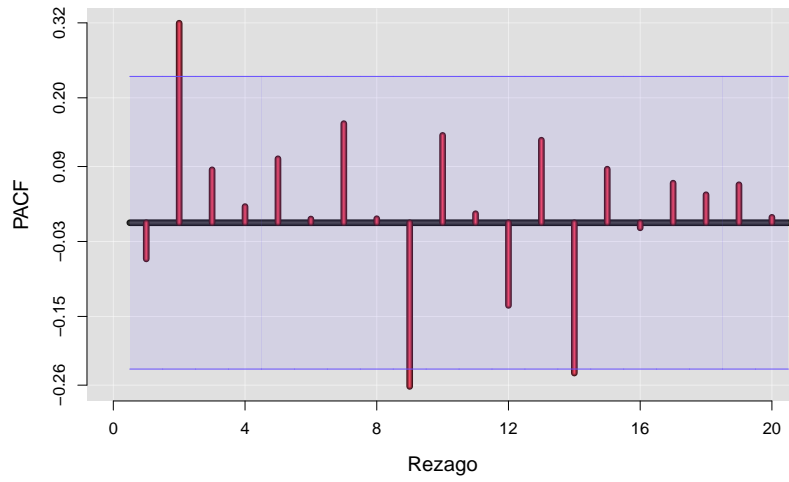


Figura 4.4: Estimación $\widehat{\phi}_{jj}$ de la serie ∇k_t centrada en cero.

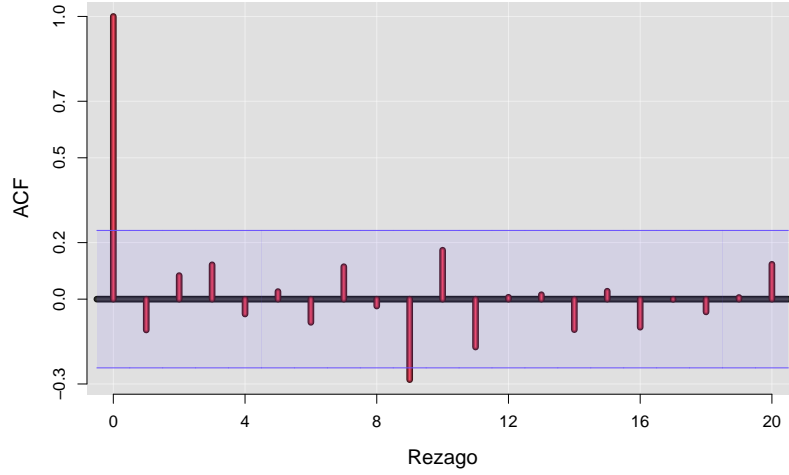


Figura 4.5: Función ACF estimada de los residuales de la serie Y_t .

Lo que puede traducirse a

$$\begin{aligned}
 k_t - k_{t-1} &+ 0.9423(k_{t-1} - k_{t-2}) + 0.8711(k_{t-2} - k_{t-3}) \\
 &+ 0.8868(k_{t-3} - k_{t-4}) + 0.8371(k_{t-4} - k_{t-5}) \\
 &= -0.9409 + 0.6239\varepsilon_t + 0.9646\varepsilon_{t-1} + 0.6763\varepsilon_{t-2}.
 \end{aligned}$$

Haciendo uso de la función `predict` de R se realiza el pronóstico de la serie Y_t en 30 periodos, con sus respectivos intervalos de confianza, como se ve en la Figura 4.6.

Para obtener los pronósticos de la serie original k_t , se hace uso de la equivalencia:

$$k_t = Y_t + k_{t-1}, \text{ Figura 4.7.}$$

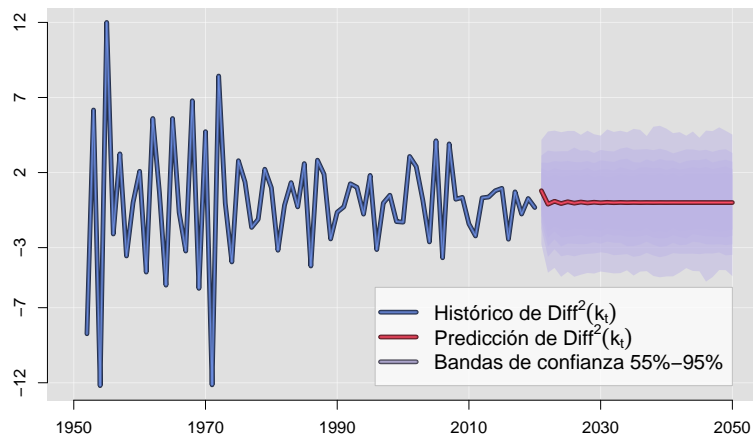


Figura 4.6: Pronóstico de la serie Y_t con intervalos al 95 % de confianza.

También es necesario verificar el supuesto de normalidad de los residuales del modelo creado para la serie Y_t , dicho supuesto es fundamental por la forma en la que se construyen los intervalos de confianza de los pronósticos, para probar lo anterior se utiliza la función `shapiro.test` de R:

```
> nortest::ad.test(residuals)
```

Anderson-Darling normality test

data: residuals

A = 0.9257, p-value = 0.01772

```
> nortest::lillie.test(residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

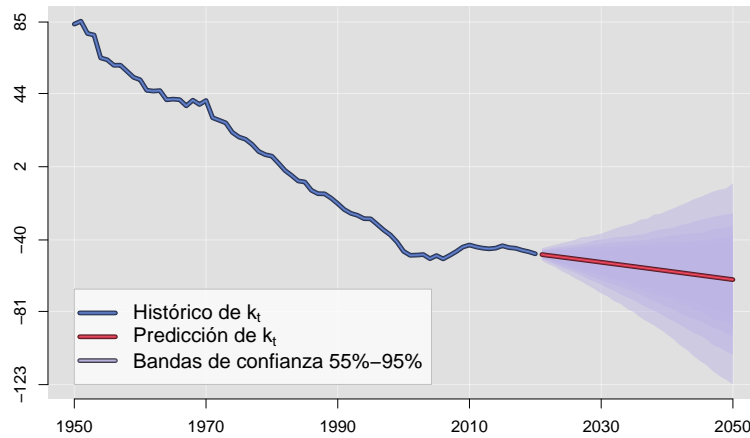


Figura 4.7: Pronóstico de la serie k_t con intervalos al 95 % de confianza.

```
data: residuals
D = 0.10416, p-value = 0.05418
```

El resultado de la prueba indica que no se rechaza la hipótesis de normalidad de los residuales con un p-valor de 0.9156, de forma adicional se revisa el histograma de los residuales así como la gráfica Normal Q-Q plot y no se observan incongruencias, Figuras 4.8 y 4.9.

Hasta este momento se tienen las condiciones necesarias para realizar los pronósticos de las tasas centrales de mortalidad. Por ejemplo, si se considera a la población total de edad 30 las proyecciones de dichas tasas se calculan según (4.7), considerando que $\hat{a}_{30} = -6.373778$ y $\hat{b}_{30} = 0.009527$, el gráfico de lo anterior se muestra en la Figura 4.10 junto a las respectivas bandas de confianza.

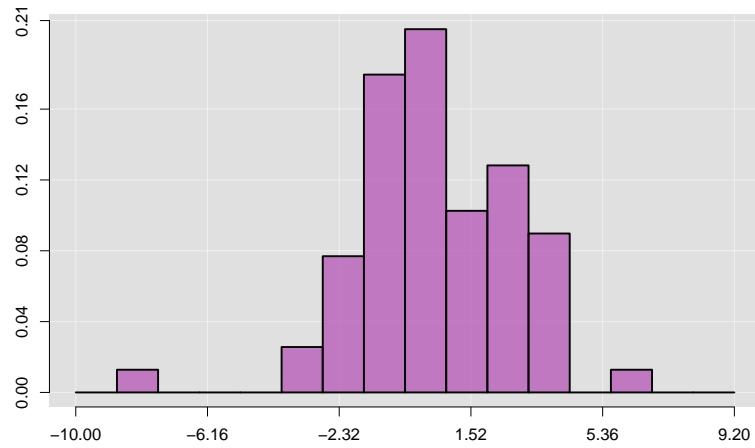


Figura 4.8: Histograma de la serie de residuales.

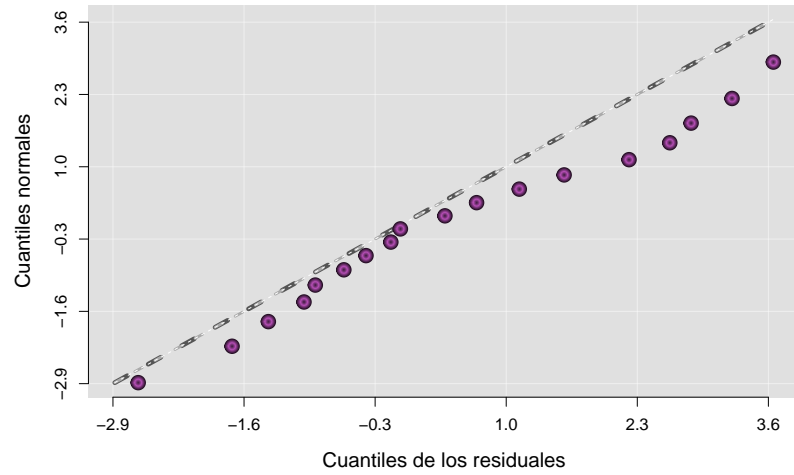


Figura 4.9: Gráfica Normal Q-Q plot de los residuales.

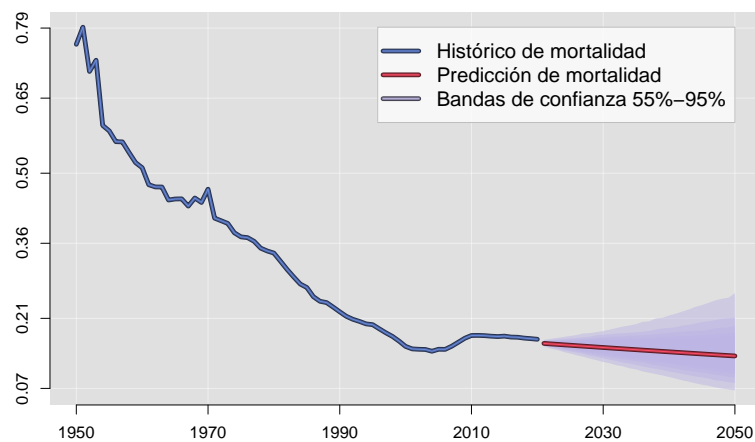


Figura 4.10: Proyección de tasas centrales de mortalidad para la población total de 30 años.

4.4. Tablas de estimaciones para la población total

\hat{a}_0	-4.085707	\hat{a}_{26}	-6.609096	\hat{a}_{52}	-5.070978	\hat{a}_{78}	-2.985795
\hat{a}_1	-6.410902	\hat{a}_{27}	-6.572588	\hat{a}_{53}	-5.008783	\hat{a}_{79}	-2.884034
\hat{a}_2	-7.081841	\hat{a}_{28}	-6.515053	\hat{a}_{54}	-4.917493	\hat{a}_{80}	-2.811155
\hat{a}_3	-7.526144	\hat{a}_{29}	-6.525840	\hat{a}_{55}	-4.834244	\hat{a}_{81}	-2.674357
\hat{a}_4	-7.794908	\hat{a}_{30}	-6.373778	\hat{a}_{56}	-4.755934	\hat{a}_{82}	-2.588432
\hat{a}_5	-7.948460	\hat{a}_{31}	-6.466726	\hat{a}_{57}	-4.674411	\hat{a}_{83}	-2.514935
\hat{a}_6	-8.043872	\hat{a}_{32}	-6.357913	\hat{a}_{58}	-4.612516	\hat{a}_{84}	-2.414808
\hat{a}_7	-8.097876	\hat{a}_{33}	-6.331978	\hat{a}_{59}	-4.548067	\hat{a}_{85}	-2.341993
\hat{a}_8	-8.136525	\hat{a}_{34}	-6.290792	\hat{a}_{60}	-4.465453	\hat{a}_{86}	-2.257176
\hat{a}_9	-8.147216	\hat{a}_{35}	-6.143937	\hat{a}_{61}	-4.370872	\hat{a}_{87}	-2.193071
\hat{a}_{10}	-8.145950	\hat{a}_{36}	-6.182625	\hat{a}_{62}	-4.285181	\hat{a}_{88}	-2.114726
\hat{a}_{11}	-8.112747	\hat{a}_{37}	-6.131851	\hat{a}_{63}	-4.208145	\hat{a}_{89}	-2.046255
\hat{a}_{12}	-8.022422	\hat{a}_{38}	-6.043019	\hat{a}_{64}	-4.131157	\hat{a}_{90}	-2.006458
\hat{a}_{13}	-7.881431	\hat{a}_{39}	-6.038062	\hat{a}_{65}	-4.043192	\hat{a}_{91}	-1.915150
\hat{a}_{14}	-7.687423	\hat{a}_{40}	-5.908985	\hat{a}_{66}	-3.983971	\hat{a}_{92}	-1.865466
\hat{a}_{15}	-7.509025	\hat{a}_{41}	-5.920834	\hat{a}_{67}	-3.902637	\hat{a}_{93}	-1.817173
\hat{a}_{16}	-7.271249	\hat{a}_{42}	-5.813795	\hat{a}_{68}	-3.828082	\hat{a}_{94}	-1.720744
\hat{a}_{17}	-7.089572	\hat{a}_{43}	-5.783278	\hat{a}_{69}	-3.753897	\hat{a}_{95}	-1.746738
\hat{a}_{18}	-6.955653	\hat{a}_{44}	-5.688076	\hat{a}_{70}	-3.685703	\hat{a}_{96}	-1.714811
\hat{a}_{19}	-6.894676	\hat{a}_{45}	-5.579825	\hat{a}_{71}	-3.569248	\hat{a}_{97}	-1.644931
\hat{a}_{20}	-6.826558	\hat{a}_{46}	-5.552263	\hat{a}_{72}	-3.479762	\hat{a}_{98}	-1.621776
\hat{a}_{21}	-6.808870	\hat{a}_{47}	-5.463387	\hat{a}_{73}	-3.413068	\hat{a}_{99}	-1.616337
\hat{a}_{22}	-6.711255	\hat{a}_{48}	-5.377396	\hat{a}_{74}	-3.327508	\hat{a}_{100}	-1.820010
\hat{a}_{23}	-6.697342	\hat{a}_{49}	-5.332540	\hat{a}_{75}	-3.246692		
\hat{a}_{24}	-6.671367	\hat{a}_{50}	-5.238182	\hat{a}_{76}	-3.156003		
\hat{a}_{25}	-6.584568	\hat{a}_{51}	-5.167071	\hat{a}_{77}	-3.077699		

Tabla 4.1: Valores estimados de los parámetros a_x del modelo de Lee-Carter para población total en el periodo 1990-2014.

\hat{b}_0	0.028635	\hat{b}_{26}	0.013628	\hat{b}_{52}	0.006962	\hat{b}_{78}	0.003404
\hat{b}_1	0.044581	\hat{b}_{27}	0.013601	\hat{b}_{53}	0.007054	\hat{b}_{79}	0.011745
\hat{b}_2	0.039373	\hat{b}_{28}	0.013905	\hat{b}_{54}	0.008377	\hat{b}_{80}	-0.000281
\hat{b}_3	0.034103	\hat{b}_{29}	0.012490	\hat{b}_{55}	0.006162	\hat{b}_{81}	0.016709
\hat{b}_4	0.031056	\hat{b}_{30}	0.009527	\hat{b}_{56}	0.007280	\hat{b}_{82}	0.009073
\hat{b}_5	0.028007	\hat{b}_{31}	0.014766	\hat{b}_{57}	0.009975	\hat{b}_{83}	0.006931
\hat{b}_6	0.025369	\hat{b}_{32}	0.013020	\hat{b}_{58}	0.006552	\hat{b}_{84}	0.009655
\hat{b}_7	0.025219	\hat{b}_{33}	0.013691	\hat{b}_{59}	0.008618	\hat{b}_{85}	0.003880
\hat{b}_8	0.024620	\hat{b}_{34}	0.014518	\hat{b}_{60}	-0.001642	\hat{b}_{86}	0.003934
\hat{b}_9	0.023238	\hat{b}_{35}	0.014091	\hat{b}_{61}	0.014752	\hat{b}_{87}	0.004395
\hat{b}_{10}	0.018458	\hat{b}_{36}	0.014126	\hat{b}_{62}	0.007291	\hat{b}_{88}	0.002442
\hat{b}_{11}	0.017113	\hat{b}_{37}	0.014713	\hat{b}_{63}	0.005299	\hat{b}_{89}	0.002894
\hat{b}_{12}	0.014036	\hat{b}_{38}	0.012867	\hat{b}_{64}	0.007292	\hat{b}_{90}	-0.010551
\hat{b}_{13}	0.011555	\hat{b}_{39}	0.012085	\hat{b}_{65}	0.003320	\hat{b}_{91}	0.005519
\hat{b}_{14}	0.011260	\hat{b}_{40}	0.009546	\hat{b}_{66}	0.006845	\hat{b}_{92}	-0.000860
\hat{b}_{15}	0.006860	\hat{b}_{41}	0.015439	\hat{b}_{67}	0.007729	\hat{b}_{93}	0.000213
\hat{b}_{16}	0.007530	\hat{b}_{42}	0.011088	\hat{b}_{68}	0.004666	\hat{b}_{94}	0.008729
\hat{b}_{17}	0.009273	\hat{b}_{43}	0.011029	\hat{b}_{69}	0.009752	\hat{b}_{95}	-0.007112
\hat{b}_{18}	0.008704	\hat{b}_{44}	0.014246	\hat{b}_{70}	-0.001638	\hat{b}_{96}	-0.008680
\hat{b}_{19}	0.012118	\hat{b}_{45}	0.009992	\hat{b}_{71}	0.015824	\hat{b}_{97}	0.003845
\hat{b}_{20}	0.009299	\hat{b}_{46}	0.011734	\hat{b}_{72}	0.006709	\hat{b}_{98}	-0.009054
\hat{b}_{21}	0.011738	\hat{b}_{47}	0.012473	\hat{b}_{73}	0.006804	\hat{b}_{99}	-0.007917
\hat{b}_{22}	0.012456	\hat{b}_{48}	0.008597	\hat{b}_{74}	0.007577	\hat{b}_{100}	-0.035334
\hat{b}_{23}	0.011917	\hat{b}_{49}	0.008779	\hat{b}_{75}	0.002615		
\hat{b}_{24}	0.012414	\hat{b}_{50}	0.002769	\hat{b}_{76}	0.005995		
\hat{b}_{25}	0.011600	\hat{b}_{51}	0.013831	\hat{b}_{77}	0.009170		

Tabla 4.2: Valores estimados de los parámetros b_x del modelo de Lee-Carter para población total en el periodo 1990-2014.

\hat{k}_1	26.810117	\hat{k}_8	3.007490	\hat{k}_{15}	-8.420427	\hat{k}_{22}	-2.618301
\hat{k}_2	7.919146	\hat{k}_9	-2.215656	\hat{k}_{16}	-2.446781	\hat{k}_{23}	-4.265987
\hat{k}_3	5.762016	\hat{k}_{10}	-1.625490	\hat{k}_{17}	-8.770231	\hat{k}_{24}	-5.361426
\hat{k}_4	5.566351	\hat{k}_{11}	2.156215	\hat{k}_{18}	-7.480437	\hat{k}_{25}	-6.794329
\hat{k}_5	3.770209	\hat{k}_{12}	-7.122954	\hat{k}_{19}	-4.981409		
\hat{k}_6	14.514955	\hat{k}_{13}	-5.761530	\hat{k}_{20}	-0.701167		
\hat{k}_7	3.583627	\hat{k}_{14}	-5.851943	\hat{k}_{21}	1.327941		

Tabla 4.3: Valores estimados de la serie k_t del modelo de Lee-Carter para población total en el periodo 1990-2014.

Bibliografía

- [Aburto, 2012] Aburto, J. (2012). Estimación de la mortalidad mexicana por el método de lee-carter y por el modelo aditivo doble multiplicativo propuesto por wilmoth. *INEGI*.
- [Berliard and Williams, 2013] Berliard, M. and Williams, I. (2013). Proyección estocástica de la mortalidad. una aplicación de lee-carter en la argentina. *Revista Latinoamericana de Población*, pages 129–148.
- [Brenes, 2014] Brenes, G. (2014). Proyección de la mortalidad de 60 años y más por causas de muerte, usando el método del producto/razón para “proyecciones coherentes”.
- [CELADE, 2012] CELADE (2012). Mortalidad. *Observatorio Demográfico*.
- [Deaton and Paxson, 2004] Deaton, A. and Paxson, C. (2004). Mortality, income, and income inequality over time in the britain and the united states. *National Office of Economic Research Cambridge*.
- [Douglas et al., 2015] Douglas, M., Cheryl, J., and Murat, K. (2015). *Time series analysis and forecasting*. Wiley.
- [Fernández, 2013] Fernández, A. (2013). Estimación y proyección de la mortalidad para costa rica con la aplicación del método lee-carter con dos variantes. *Población y Salud en Mesoamérica*.
- [García and Ordorica, 2010] García, B. and Ordorica, M. (2010). *Población*. El colegio de México.
- [García, 2013] García, V. (2013). *Las estimaciones y las proyecciones demográficas como herramientas para el diseño de las políticas de población*. Instituto de Investigaciones Jurídicas UNAM.
- [García and Ordorica, 2012] García, V. and Ordorica, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de lee-carter. *Estudios Demográficos y Urbanos*, 27(2):409–448.

- [Giroi and King, 2007] Giroi, F. and King, G. (2007). Understanding the lee-carter mortality forecasting method.
- [Guerrero, 2003] Guerrero, V. (2003). *Análisis estadístico de series de tiempo económicas*. Thomson.
- [Lee and Carter, 1992] Lee, R. and Carter, L. (1992). Modeling and forecasting u.s. mortality. *Taylor & Francis*.
- [Mina, 2002] Mina, A. (2002). La demografía en la formación del actuario. Material didáctico.
- [Raftery et al., 2013] Raftery, A., Chunn, J., Gerland, P., and Sevcikova, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, pages 777–801.
- [Rincón, 2012] Rincón, L. (2012). *Introducción a los Procesos Estocásticos*. Las prensas de ciencias.
- [Rodríguez et al., 1997] Rodríguez, J., Chackiel, J., Guzmán, J., Faijer, D., Macció, G., Pujol, J., and Villa, M. (1997). *Demografía I*. PROLAP-IISUNAM.
- [Stoffer, 2006] Stoffer, R. S. . D. (2006). *Time series analysis and Its applications*. Springer.