

Conceptos primarios y análisis descriptivo de una serie de tiempo

El esquema de esta primer parte la damos a continuación. En la sección 1 definimos algunos conceptos básicos sobre series de tiempo los que incluyen la definición de proceso estocástico estacionario así como la descomposición de una serie de tiempo en componentes de tendencia, estacional, cíclica y aleatoria. En la sección 2 damos los pasos a seguir en el modelado de una serie de tiempo. En la sección 3 veremos algunas de las formas más habituales de detectar y estimar la tendencia (cuando exista la misma) entre ellas la de mínimos cuadrados y promedios móviles. En la sección 4 veremos la forma de detectar y calcular la componente estacional. En la sección 5 veremos la función de autocorrelación de una serie de tiempo, su definición teórica y su forma de estimarla. En la sección 6 veremos algunas funciones básicas en R como herramienta para realizar análisis en las series de tiempo. En la sección 7 definimos el operador backward shift (B) y el operador ∇ y veremos cómo en muchos casos puede ser utilizado para quitarle la tendencia a una serie de tiempo. En la sección 8 damos una serie de test de hipótesis para chequear que un conjunto de datos sea estacionario y en las secciones 9 y 10 van algunos comentarios finales y los ejercicios correspondientes a éste capítulo.

1 Primeros conceptos

Una serie de tiempo es un conjunto de observaciones de una variable determinada, medida en determinados instantes de tiempo.

En este curso, indicaremos como x_t al valor de la variable de estudio medida en un instante de tiempo t .

Del punto de vista teórico, las observaciones de una serie de tiempo son las realizaciones de un proceso estocástico, es decir que x_t es la realización de la variable aleatoria X_t .

Definición 1 *Un proceso estocástico indizado en $[0, T]$ es una familia de variables aleatorias $X_t : \Omega \rightarrow \mathbb{R}$ para $t \in [0, T]$, siendo (Ω, \mathcal{A}, P) un espacio de probabilidad.*

En la definición anterior el conjunto $[0, T]$ puede ser de la forma $[0, +\infty)$. También puede estar indizado en los naturales o los enteros. Cuando el proceso está indizado en los naturales o enteros se suele decir que el proceso estocástico es a tiempo discreto mientras que en el caso de estar indizado en un intervalo se llama a tiempo continuo.

En la práctica, las observaciones son siempre una cantidad finita y suelen venir (aunque no siempre) en intervalos espaciados de tiempo. Por ejemplo ventas mensuales de una empresa, velocidad del viento en un determinado punto geográfico medido cada dos horas, producto bruto anual, temperatura máxima o mínima diaria en determinado punto geográfico, etc, etc.

El estudio estadístico de una serie de tiempo apunta a comprender la variabilidad de la variable de estudio en función del tiempo, así como el de realizar predicciones de valores futuros.

Muchas series de tiempo tienen una “tendencia a largo plazo” (por ejemplo si consideramos la variación del precio de un helado en función del tiempo desde 1995 hasta 2020) veremos que es una serie de tiempo que crece), muchas tienen una “componente estacional”, por ejemplo en el caso de los helados, las ventas tienen sus máximos anuales en verano y mínimos anuales en invierno, y eso se repite año a año, muchas series tienen una “componente cíclica” que son como las componentes estacionales pero con duraciones superiores al año, por ejemplo en economía, se habla mucho de “épocas de recesión” o “épocas de prosperidad”. Por otro lado todo valor a futuro de una variable a estudio tiene una componente de incertidumbre.

En la Figura 1 vemos dos ejemplos de observaciones en series de tiempo. En el ejemplo de arriba, tenemos los valores anuales del nivel en pies del lago Huron entre 1872 y 1975. Se observa una tendencia decreciente a lo largo del tiempo. En el ejemplo de abajo se ven los ingresos semestrales (en miles de pesos) de un trabajador independiente desde el primer semestre de 2014 hasta el segundo semestre de 2018. Se observa claramente una componente creciente a lo largo del tiempo y una cíclica semestral, ya que en los primeros semestres de cada año tiene mayores ingresos que en los segundos.

Por lo anteriormente dicho, toda serie de tiempo se la descompone en cuatro componentes: componente de tendencia (T_t), componente estacional (E_t), componente cíclica (C_t) y componente aleatoria o de incertidumbre (I_t) y la relación entre ellas quedará expresada en la próxima definición de acuerdo a que el modelo sea aditivo o multiplicativo.

Definición 2 *Modelo aditivo:* $X_t = T_t + E_t + C_t + I_t$ para todo t .

Modelo multiplicativo: $X_t = T_t E_t C_t I_t$ para todo t .

Suposiciones generales (para ambos modelos)

- Las componentes T_t , E_t y C_t se suponen determinísticas (o sea no aleatorias) de modo que la aleatoriedad de I_t es quien determina la aleatoriedad de X_t .
- La componente aleatoria, no contiene ni tendencia ni componente estacional ni cíclica.

Observación 3 *Muchas veces para modelar una serie de observaciones x_t , lo que se hace es aplicar alguna función invertible a los datos $y_t = f(x_t)$ y modelar el nuevo conjunto de datos y_t (la invertibilidad me permite deshacer en todo momento y volver a las x_t), por ejemplo si puedo hacer una buena predicción de un valor futuro de y_t (por ejemplo \hat{y}_{t+1}), podré hacer una predicción de un futuro valor de x_t simplemente calculando la función inversa (por ejemplo $\hat{x}_{t+1} = f^{-1}(\hat{y}_{t+1})$).*

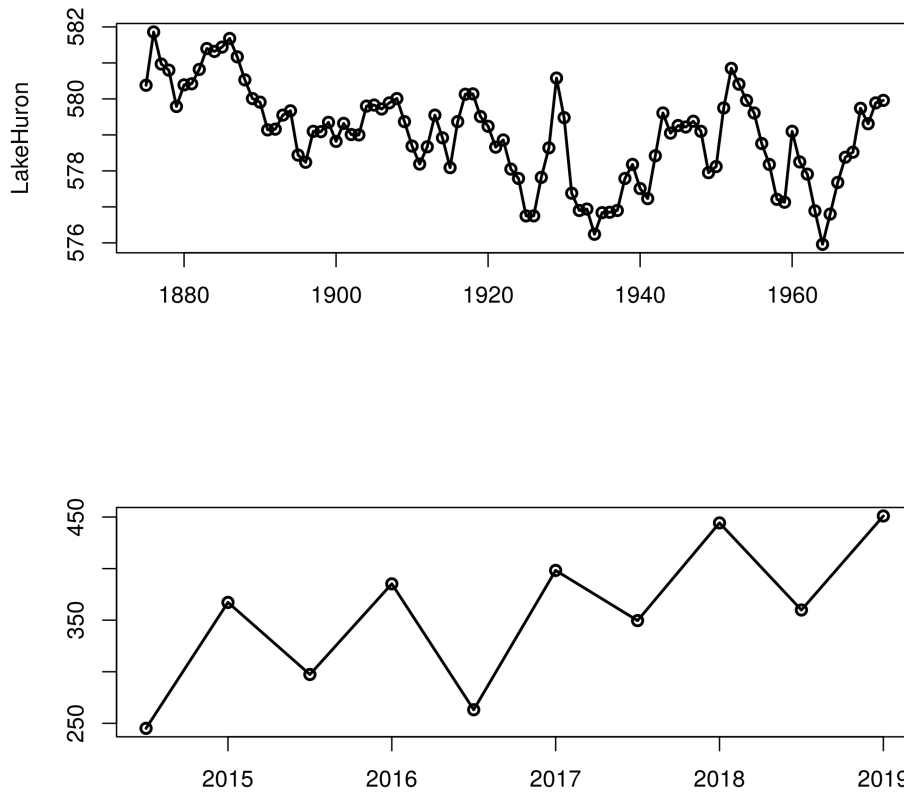


Figura 1: En el gráfico inferior se ve tendencia creciente y componente estacional semestral y en el superior tendencia decreciente.

Observación 4 *El modelo multiplicativo puede ser llevado de manera sencilla a un modelo aditivo por aplicación de la función logaritmo ya que si $X_t = T_t E_t C_t I_t$ entonces $\ln X_t = \ln T_t + \ln E_t + \ln C_t + \ln I_t$.*

En general las componentes cíclicas no suelen venir en los paquetes estadístico de R, porque son difíciles de detectar y sus períodos y duraciones en el tiempo pueden ser variables, de modo que en cada caso particular, el investigador deberá ver si su serie de datos tiene una componente cíclica y quitarla de forma similar a (como veremos) se quita la componente estacional.

Por lo tanto, este curso se realizará suponiendo que $C_t = 0$ si el modelo es aditivo o $C_t = 1$ si el modelo es multiplicativo. De modo que tendremos $X_t = T_t + E_t + I_t$ o bien $X_t = T_t E_t I_t$.

En la Figura 2 vemos un ejemplo de serie de tiempo con tendencia y con componente estacional claramente marcadas.

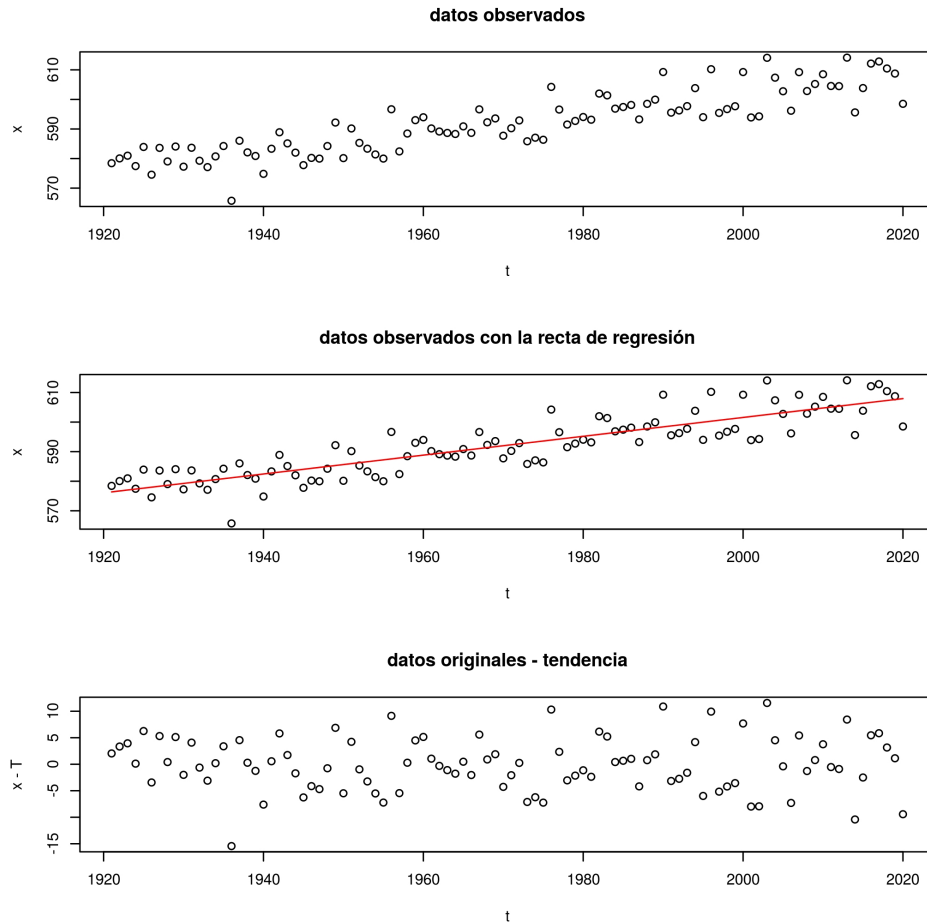


Figura 2: En el gráfico inferior se ve $x_t - \hat{T}_t$, que no muestra ningún tipo de tendencia.

Definición 5 *Proceso estocástico fuertemente estacionario.*

Si X_1, X_2, \dots son infinitas (numerable) variables aleatorias, se dice que el proceso $X = \{X_n\}_{n \in \mathbb{N}}$ es fuertemente estacionario si y sólo si para todos $k, h \in \mathbb{N}$ se cumple que

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h}) \stackrel{d}{=} (X_{t_1}, X_{t_2}, \dots, X_{t_k}) \text{ para todos } t_1, t_2, \dots, t_k \in \mathbb{N}.$$

Observación 6 La definición anterior significa que si elegimos cualquier cantidad de variables digamos $X_{t_1}, X_{t_2}, \dots, X_{t_k}$ trasladamos h instantes de tiempo a los instantes considerados en las misma y miramos ahora a las variables $X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h}$ entonces la distribución conjunta no cambia.

Ejemplo 7 Variables aleatorias X_1, X_2, \dots i.i.d (independientes e idénticamente distribuidas) son un ejemplo de proceso estocástico fuertemente estacionario.

Definición 8 Proceso estocástico débilmente estacionario.

Si X_1, X_2, \dots son infinitas (numerable) variables aleatorias, se dice que el proceso $X = \{X_n\}_{n \in \mathbb{N}}$ es débilmente estacionario si y sólo si se cumplen las siguientes condiciones:

- $\mathbb{E}(X_n^2)$ es finita para todo n (se dice que el proceso está en L^2).
- $\mathbb{E}(X_n) = \mu$ para todo n .
- $\text{COV}(X_{n+h}, X_n) = \gamma(h)$ para todos n, h .

Observación 9 La condición de estar en L^2 asegura la existencia de la esperanza de las variables así como la de su covarianza.

Observación 10 La condición $\text{COV}(X_{n+h}, X_n) = \gamma(h)$ para todos n, h , significa que la covarianza entre dos observaciones de este proceso, sólo depende de la distancia entre los instantes de tiempo de las mediciones, ($n+h$ y n) y no del instante específico en el cual estemos parados, por ejemplo si tenemos 1000 datos entonces $\text{COV}(X_{101}, X_1) = \text{COV}(X_{851}, X_{950})$.

Observación 11 Si un proceso está en L^2 y es fuertemente estacionario, entonces es débilmente estacionario.

Definición 12 Diremos que un proceso es estacionario cuando es débilmente estacionario.

Definición 13 Proceso Gaussiano.

Si X_1, X_2, \dots son infinitas (numerable) variables aleatorias, se dice que el proceso $X = \{X_n\}_{n \in \mathbb{N}}$ es Gaussiano si y sólo si $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ tiene distribución normal multivariada para todos $t_1, t_2, \dots, t_k \in \mathbb{N}$.

Observación 14 Si un proceso es Gaussiano y además débilmente estacionario, entonces es fuertemente estacionario (porque al ser las covarianzas de los distintos elementos del vector $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ dependientes sólo de h y no de t_1, t_2, \dots, t_k , las mismas coincidirán con las del vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$).

Cuando trabajamos con un proceso estocástico indizado en un intervalo $[0, T]$ las definiciones anteriores siguen siendo válidas.

Definición 15 Ruido Blanco. Si X_1, X_2, \dots son variables aleatorias en L^2 no correlacionadas con esperanza cero y varianza σ^2 , le llamaremos a las variables ruido blanco.

Notación: $\{X_t\} \sim WN(0, \sigma^2)$ (white noise).

Observación 16 *El ruido blanco se usará en la definición de los procesos ARMA que veremos en el capítulo próximo.*

2 Pasos a realizar en el modelado de una serie de tiempo

La mayor parte de la teoría de procesos estocásticos consiste en procesos estacionarios, por lo que en particular no contienen tendencia, ni componente estacional ni cíclica.

Para la modelación de una serie de tiempo, dados los datos observados x_t , primero se les aplica un proceso llamado de “desestacionalización” que puede llevarse a cabo de muchas formas, de las cuales veremos algunas, y que consiste en quitarles a los datos la tendencia y las componentes estacional y cíclica, para luego modelar mediante algún proceso estocástico estacionario.

Dado un conjunto de observaciones de una serie de tiempo x_t para distintos instantes de tiempo t , supongamos que estamos bajo un modelo aditivo de la forma $X_t = T_t + E_t + C_t + I_t$, podemos sintetizar lo anterior en los siguientes pasos:

Paso 1: estimo la tendencia (\hat{T}_t) y se la quitamos a los datos (para cada valor de t) con lo que nos queda los nuevos datos $x_t - \hat{T}_t$. Este nuevo conjunto de datos ya no tendría tendencia.

Paso 2: estimo la componente estacional (\hat{E}_t) y se la quitamos a los datos (para cada valor de t) con lo que nos queda los nuevos datos $x_t - \hat{T}_t - \hat{E}_t$.

Paso 3: estimo la componente cíclica (\hat{C}_t) y se las quitamos a los datos (para cada valor de t) obteniendo como nuevos datos $y_t = x_t - \hat{T}_t - \hat{E}_t - \hat{C}_t$.

Paso 4: ajusto algún modelo estacionario al nuevo conjunto de datos y_t (que al haberles quitado la tendencia y las componentes cíclicas y estacionales, son datos “desestacionalizados”).

Los cuatro pasos anteriores se hicieron asumiendo un modelo aditivo, si fuera multiplicativo se procede de manera análoga.

2.1 Modelo aditivo vs modelo multiplicativo

Dada una serie de datos ¿qué modelo corresponde más utilizar? Como es de esperar la respuesta depende del tipo de datos.

Siempre es posible ajustar ambos modelos por separado y luego con las herramientas que se verán después se ve cuál anda mejor. A veces visualizando en un gráfico los datos, se puede intuir a cual modelo corresponde de acuerdo al siguiente sencillo criterio.

Si el modelo es aditivo, entonces podemos asumir que $X_t = T_t + E_t + C_t + I_t$ siendo T_t, E_t, C_t no aleatorios, mientras que I_t sería un proceso estacionario con esperanza μ y varianza σ^2 . Entonces en ese caso tendremos que $\mathbb{V}(X_t) = \mathbb{V}(T_t + E_t + C_t + I_t) = \sigma^2$.

Si el modelo es multiplicativo, entonces $\mathbb{V}(X_t) = \mathbb{V}(T_t E_t C_t I_t) = T_t^2 E_t^2 C_t^2 \sigma^2$, es decir que en el modelo multiplicativo, la varianza depende del instante t , pero en el modelo aditivo la varianza permanece constante para todo valor de t .

A veces, es posible visualizar una mayor (o menor) variación de los datos a medida que el tiempo crece, en esos casos el modelo más adecuado parecería ser el multiplicativo. En la Figura 3, se ve cómo en el gráfico de arriba no se visualiza ningún cambio en la varianza de los datos, mientras que en el gráfico de abajo, la dispersión va aumentando con el paso del tiempo. De todas formas esto suele no ser muy práctico porque rara vez es posible deducirlo a “ojo”.

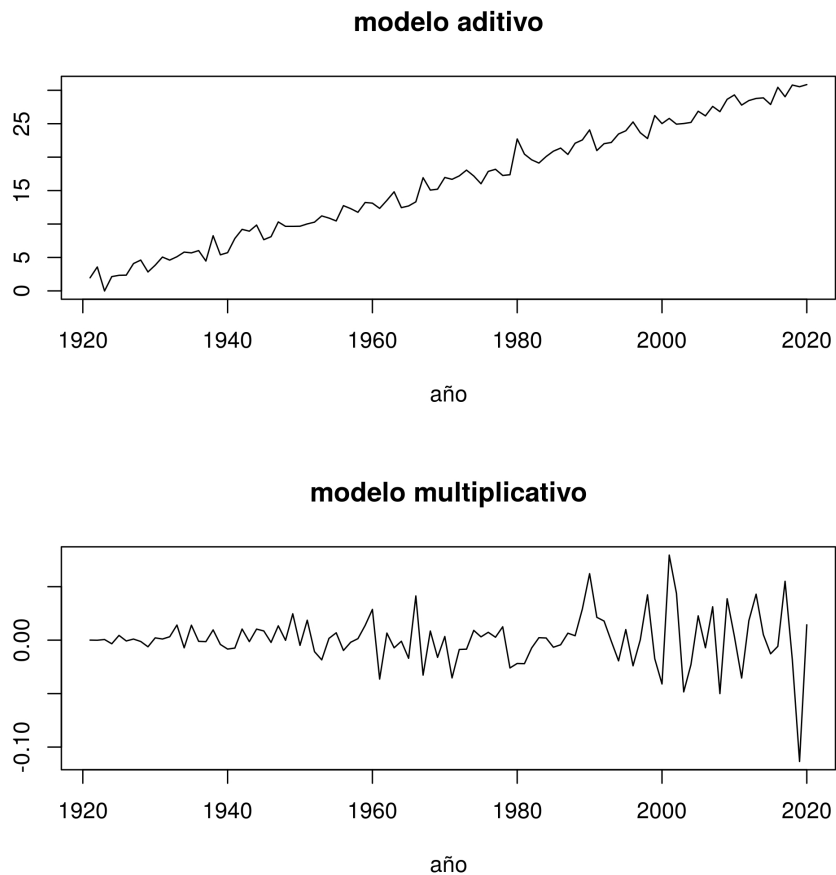


Figura 3: En el gráfico superior no se ven cambios en la variabilidad de los datos cuando pasa el tiempo, mientras que en el inferior se nota que la variabilidad aumenta cuando t crece.

3 Estimación de la tendencia

Hay varias formas de extraer la tendencia, las dos más sencillas y usuales son ajustando a los datos una recta mínimo cuadrática, o calculando los llamados promedios móviles (MA=moving averages en inglés).

3.1 Recta de mínimos cuadrados

Supongamos que tenemos las observaciones $(t_1, x_{t_1}), (t_2, x_{t_2}), \dots, (t_n, x_{t_n})$ (pensemos que nuestras observaciones son $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ y que dibujamos los puntos en el plano, en las abscisas t_1, t_2, \dots, t_n y en las ordenadas $x_{t_1}, x_{t_2}, \dots, x_{t_n}$. Podemos intentar hallar una recta que sea “lo más representativa posible de los puntos”. Si le llamo $T(t) = at + b$ a dicha recta, se obtienen los valores de los parámetros a y b por el método de los mínimos cuadrados que consiste en considerar la función $\varphi(a, b) = \sum_{i=1}^n (x_{t_i} - at_i - b)^2$ y hallar los valores de a y b que minimizan la función φ (notar que estamos midiendo la distancia al cuadrado entre el valor observado al tiempo t_i y el valor que daría la ecuación de la recta al tiempo t_i es decir $at_i + b$. Se puede probar (a partir de argumentos de álgebra lineal sobre proyección a subespacios, o de análisis igualando las derivadas parciales a cero y luego probando convexidad de la función) que los valores que minimizan a la función φ vienen dados por las siguientes fórmulas:

$$\begin{aligned}\hat{a} &= \frac{\sum_{i=1}^n t_i x_{t_i} - n \bar{x} \bar{t}}{S_X^2} \\ \hat{b} &= \bar{x} - \hat{a} \bar{t}\end{aligned}$$

siendo $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$ (promedio de los instantes de tiempo de observación), $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{t_i}$ (promedio del conjunto de datos $x_{t_1}, x_{t_2}, \dots, x_{t_n}$) y $S_X^2 = \frac{1}{n} \sum_{i=1}^n (x_{t_i} - \bar{x})^2$ (varianza muestral del conjunto de datos $x_{t_1}, x_{t_2}, \dots, x_{t_n}$).

Definición 17 *Recta de regresión.*

Dado el conjunto de datos $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ se llama recta de regresión (o también recta mínimo cuadrática) a $T_t = \hat{a}t + \hat{b}$ siendo \hat{a} y \hat{b} calculados con las fórmulas de la ecuación anterior.

Ejemplo 18 *Vamos a generar un ejemplo de 100 datos que se encuentren sobre la recta $x_t = 0.3t + 1 + I_t$ donde $I_t \sim N(0, 25)$ independientes, para los tiempos $t = 1920, 1921, \dots, 2019$. En este ejemplo no tenemos componente estacional y su tendencia es la recta $T_t = 0.3t + 1$ que pasaremos a estimar con el método de los mínimos cuadrados. Para lograrlo, aplicamos los siguientes comandos en R:*

```
t=seq(1921,2020)#genera un vector con los números del 1920 al 2019.  
x=0.3*t+1+rnorm(100,0,5)#aplicamos la fórmula anterior.  
plot(t,x)#dibujamos en el plano t,x los 100 puntos generados.
```

Ahora imaginamos que esos 100 datos del ejemplo anterior, se corresponden con una serie de tiempo de 100 datos anuales, el primero de los cuales corresponde al año 1920. Mirando el gráfico de los puntos se ve claramente que hay

una tendencia. Vamos a hallar la recta y dibujarla encima del mismo gráfico, para ello utilizamos los siguientes comandos:

```
lm(x~ t)$coefficients#por si quiero ver los coeficientes
z=lm(x~ t)
T=z$coefficients[2]*t+z$coefficients[1]
lines(t,T,col="red")
```

Observamos ahora que al restar a los datos originales su tendencia $(x_t - \hat{T}_t)$ nos queda una serie nueva de datos sin tendencia:

```
plot(t,x-T)
```

Lo anterior puede verse en los primeros dos gráficos de la Figura 2.

3.2 Ajustes mínimo cuadráticos para otros tipos de relaciones funcionales

La misma idea de mínimos cuadrados puede ser aplicada para ajustar otro tipo de funciones. Por ejemplo, podemos ajustar por mínimos cuadrados una parábola $y = at^2 + bt + c$ mediante la minimización de la función

$$\varphi(a, b, c) = \sum_{i=1}^n (x_{t_i} - at_i^2 - bt_i - c)^2$$

o ajustar una función de tipo sinusoidal por ejemplo de la forma $y = a \sin(t) + b \cos(t) + c \sin(2t) + d \cos(2t) + e$ mediante la minimización de la función

$$\varphi(a, b, c, d, e) = \sum_{i=1}^n (x_{t_i} - a \sin(t_i) - b \cos(t_i) - c \sin(2t_i) - d \cos(2t_i) - e)^2.$$

En la Figura 4, vemos un conjunto de datos con ajuste lineal y por polinomio de segundo grado. Se ve claramente que el ajuste por polinomio de segundo grado es mucho ms adecuado para describir la tendencia creciente que el ajuste por recta.

En R para obtener este otro tipo de ajustes mínimo cuadráticos debemos incluir las siguientes sentencias:

```
z2=lm(x ~ cbind(t,t^2))
T2=z2$coefficients[1]+z2$coefficients[2]*t+z2$coefficients[3]*t*t
lines(t,T2,col="red",lwd="2")
```

3.3 Promedios móviles

Otra forma muy utilizada de eliminar la tendencia en una serie de tiempo es la de aplicar las medias o promedios móviles.

Si tenemos datos cuatrimestrales, entonces en cada instante de tiempo t calculamos la tendencia a partir de $\hat{T}_t = \frac{x_{t-1} + x_t + x_{t+1}}{3}$ es decir que promediamos 3 observaciones: la observación del momento t con la del momento anterior $t - 1$ y la del momento posterior $t + 1$. Cuando se hace un gráfico con estos datos se

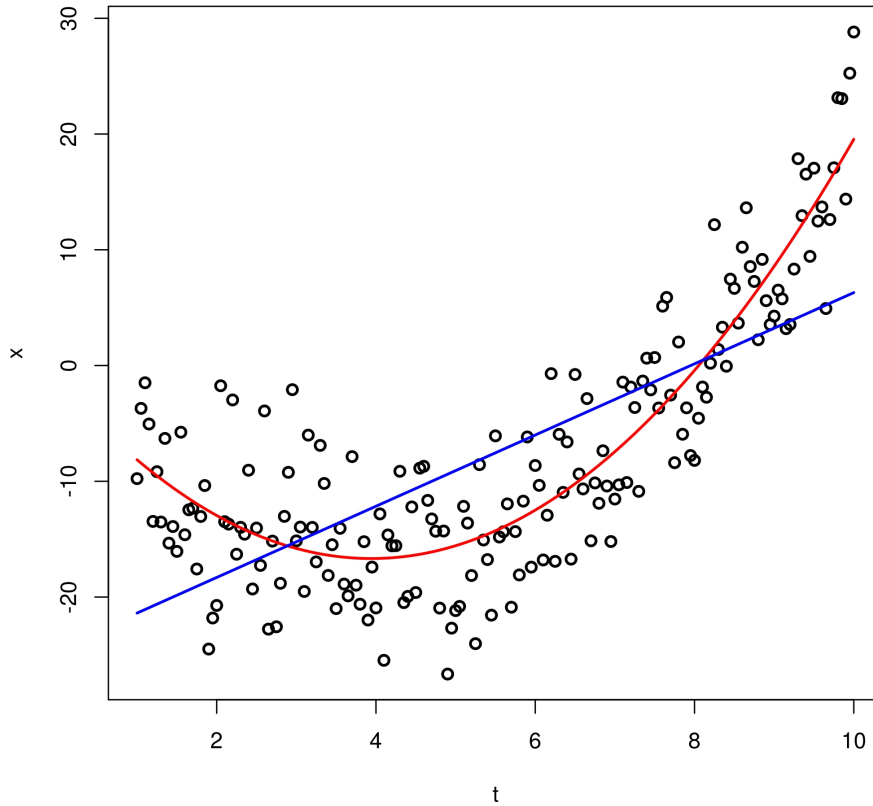


Figura 4: 181 datos en el intervalo $[1,10]$ con su ajuste lineal y parabólico.

logra una curva “más suavizada” que la original, porque si un dato al instante t llegara a ser muy grande y muy distante de sus observaciones cercanas, al promediarla con los dos momentos más cercanos (el del momento $t - 1$ y el del $t + 1$), se disipa la diferencia abrupta entre un dato y otro.

En el caso de datos trimestrales, consideramos la tendencia como $\hat{T}_t = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}$ y de igual manera se razona si los datos son semestrales, semanales, diarios, etc, etc.

Observación 19 *Notar que cuanto más valores tomemos a la hora de realizar promedios móviles más suave quedará la curva de tendencia, en el caso extremo, si tuviéramos n datos diarios y consideramos promedios móviles de n datos, entonces nos quedaría como curva de tendencia, una constante, que sería igual al promedio de todos los datos.*

3.4 Otros métodos de suavizamiento

Existen muchos otros métodos de suavizamiento de la curva observada de datos. El suavizamiento por núcleos, que es una generalización de los promedios móviles promediando los datos cercanos al instante de tiempo t , pero ponderando por una función (el llamado núcleo).

El suavizamiento por regresión local, para predecir un valor al instante t se aplica regresión pero con los datos cercanos a dicho valor.

El suavizamiento de la regresión por Lowess (propuesto por Cleveland en 1979), que es aplicar regresión local ajustando en cada punto con un polinomio de grado bajo.

El suavizamiento por splines, es una generalización del de regresión polinomial. Se divide el intervalo de tiempo en k subintervalos, en cada subintervalo se ajusta un polinomio local, de modo que en los extremos de cada subintervalo los polinomios (del lado izquierdo y del derecho) y sus derivadas coincidan.

4 Componente estacional

Supongamos el modelo aditivo (con el multiplicativo se razona de forma análoga). Ya tenemos calculada la componente estacional por alguno de los dos métodos, de modo que podemos plantear la igualdad $x_t - \hat{T}_t = E_t + I_t$.

La componente estacional se calcula de acuerdo al siguiente razonamiento. Supongamos que los datos que tenemos son trimestrales, en este caso tendríamos 4 componentes estacionales: que les llamamos E_1, E_2, E_3 y E_4 . E_1 es la componente estacional correspondiente a los meses de enero, febrero y marzo y la obtenemos promediando todos los valores $x_t - \hat{T}_t$ cuyos instantes t correspondan a los valores de enero, febrero y marzo. A dicho valor le llamamos \hat{E}_1 (porque es una estimación de E_1). Análogamente construimos \hat{E}_2, \hat{E}_3 y \hat{E}_4 . Notar que para cada instante t su componente estacional toma únicamente cuatro valores posibles de acuerdo a “la estación” en la que se encuentre.

El razonamiento es análogo si nuestros datos fueran cuatrimestrales (3 componentes estacionales), mensuales (12 componentes) o semanales (52 componentes).

Ejemplo 20 *Supongamos que tenemos datos trimestrales, desde el primer trimestre de 1990 hasta el cuarto trimestre de 2019, totalizando de esta forma 120 datos. Los datos se presentan en la Figura 5 donde se observa claramente una tendencia creciente y también se ve una componente estacional, porque los momentos de subidas y de bajadas son siempre en los mismos trimestres. En la Figura 6 se ve que los mínimos se encuentran siempre en el primer semestre de cada año, mientras que los máximos aparecen en el tercer trimestre de cada año.*

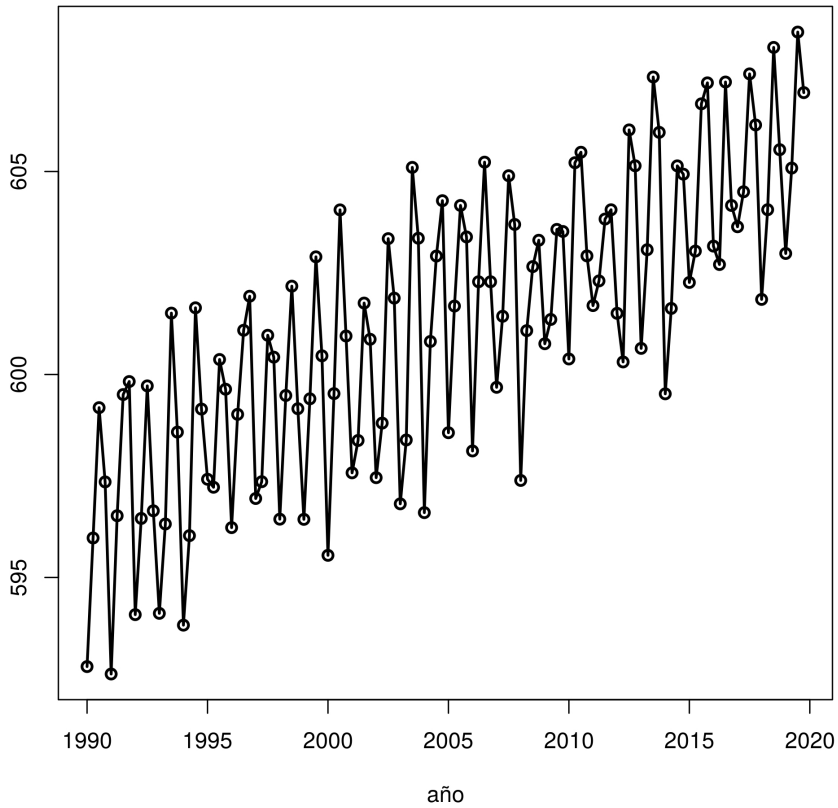


Figura 5: 120 datos simulados trimestrales con tendencia y componente estacional. Hay tendencia creciente en los datos pero dentro de cada año los valores comienzan subiendo y terminan bajando.

5 Función de autocorrelación muestral (ACF)

En series de tiempo una herramienta imprescindible es la llamada función de autocorrelación.

Recordamos que la covarianza entre dos variables X e Y se define mediante la fórmula

$$\text{COV}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

y que si dos variables aleatorias X e Y son independientes, entonces $\text{COV}(X, Y) = 0$.

No vale el recíproco en general, es decir, si $\text{COV}(X, Y) = 0$ (se dice que las

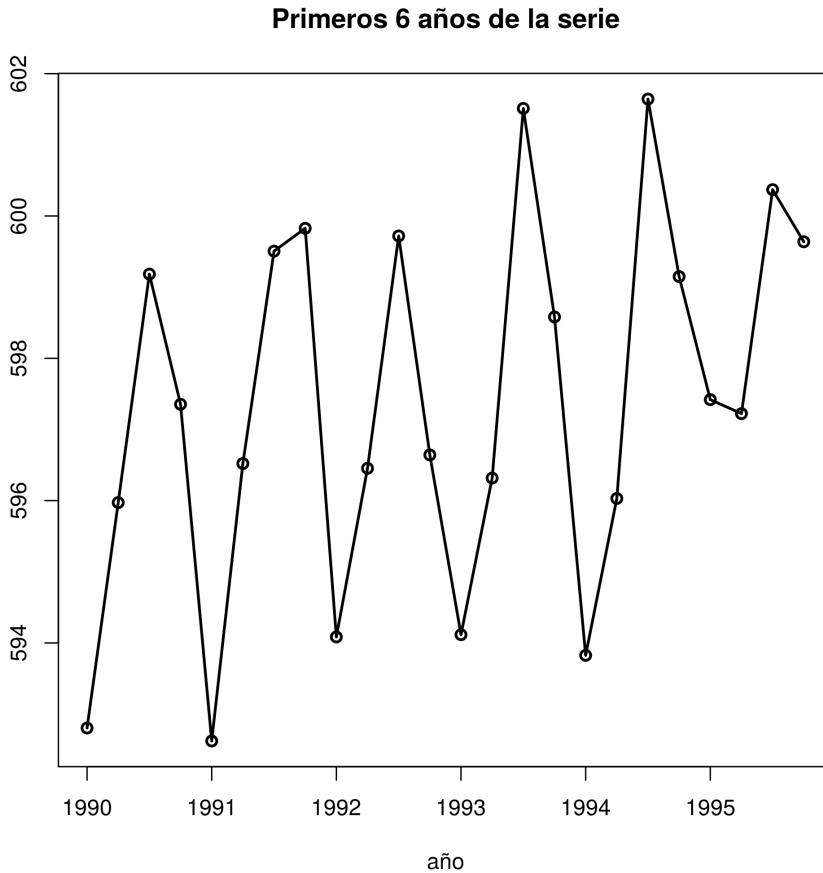


Figura 6: Primeros 6 años de la serie observada. Se observa que los máximos de cada año se obtienen casi siempre en el tercer trimestre y los mínimos en el primer o cuarto trimestre.

variables X e Y no están correlacionadas) no necesariamente son independientes.

Observación 21 *Los procesos gaussianos, sí cumplen el recíproco, porque en todo proceso gaussiano se cumple que todas las posibles distribuciones finito dimensionales tienen distribución conjunta normal multivariada, y en una normal multivariada, las covarianzas cero entre sus variables, implican independencia.*

Recordamos también que si tenemos una muestra $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ del par (X, Y) , el estimador natural de la covarianza es la llamada covarianza

muestral definida mediante

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X}_n \bar{Y}_n$$

siendo

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \text{ las medias muestrales de cada variable.}$$

Cuando tenemos un proceso estocástico $\{X_t\}_{t \in \mathbb{N}}$, tenemos infinitas variables aleatorias X_1, X_2, \dots , por lo tanto tenemos infinitas covarianzas entre todas las elecciones posibles de a dos de las infinitas variables.

A partir de lo anterior, podemos definir de manera intuitiva, estimadores para cada una de las posibles covarianzas, cuando los datos están modelados por un proceso estacionario.

Definición 22 *Función de autocovarianzas de un conjunto de datos X_1, X_2, \dots, X_n .*

Si X_1, X_2, \dots, X_n es un conjunto de datos modelados por un proceso estacionario, definimos

$$\hat{\gamma}(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} X_i X_{i+h} - \bar{X}_n^2.$$

Observación 23 *Con $\hat{\gamma}(h)$ estamos estimando*

$$\text{COV}(X_1, X_{1+h}) = \text{COV}(X_2, X_{2+h}) = \text{COV}(X_3, X_{3+h}) = \dots$$

ya que al ser el proceso estacionario $\mathbb{E}(X_1) = \mathbb{E}(X_{1+h})$ por lo que \bar{X}_n estima tanto $\mathbb{E}(X_1)$ como $\mathbb{E}(X_{1+h})$, mientras que con $\frac{1}{n-h} \sum_{i=1}^{n-h} X_i X_{i+h}$ estamos estimando $\mathbb{E}(X_1 X_{1+h})$.

Observación 24 *Para cada h estamos estimando $\text{COV}(X_1, X_{1+h})$ es decir que estimamos la covarianza entre dos variables distanciadas h unidades de tiempo.*

Observación 25 *En las aplicaciones suele observarse la condición $\hat{\gamma}(h) \rightarrow 0$ cuando $h \rightarrow +\infty$, lo cual en muchos casos, es intuitivo en el sentido de que si dos observaciones están muy distanciadas en el tiempo tienden a ser independientes por lo que la covarianza entre las mismas suele ser muy pequeña.*

Observación 26 *Para estimar la covarianza entre dos variables medidas en tiempos distanciados entre sí de h unidades, debo utilizar $n - h$ datos de modo que con estas fórmulas se pierde precisión en la estimación en la medida que h aumenta.*

Observación 27 *Cuando tomamos $h = 0$, se tiene que $\hat{\gamma}(0) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 = S_X^2$ que es la varianza muestral del conjunto de datos.*

Como la covarianza entre dos variables puede tomar cualquier valor, una forma de relativizarlo y acotar sus valores se obtienen a través del coeficiente de correlación que recordemos que se define como $\rho(X, Y) = \frac{\text{COV}(X, Y)}{\sqrt{\text{V}(X)\text{V}(Y)}}$ y que se demuestra que $-1 \leq \rho(X, Y) \leq 1$, que $\rho(X, Y) = 1$ si y sólo si X e Y están sobre una recta creciente y que $\rho(X, Y) = -1$ si y sólo si X e Y están sobre una recta decreciente. Por lo tanto, el valor del coeficiente de correlación puede tomarse como un indicador que va desde la no correlación ($\rho(X, Y) = 0$) hasta la perfecta dependencia lineal entre las variables (valores del coeficientes cercanos a 1 indican X e Y cercanos a una recta creciente, mientras que valores del coeficientes cercanos a -1 indican X e Y cercanos a una recta decreciente).

A partir de lo anterior, para un proceso estacionario, vemos que

$$\rho(h) = \rho(X_1, X_{1+h}) = \frac{\text{COV}(X_1, X_{1+h})}{\sqrt{\text{V}(X_1)\text{V}(X_{1+h})}} = \frac{\text{COV}(X_1, X_{1+h})}{\text{V}(X_1)}$$

por lo que se puede estimar de manera sencilla la función de autocorrelación como sigue.

Definición 28 *Función de autocorrelación de un conjunto de datos X_1, X_2, \dots, X_n .*

Si X_1, X_2, \dots, X_n es un conjunto de datos modelados por un proceso estacionario, definimos

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

En R, dado un vector de datos x la función de autocorrelación de los mismos viene dado por el comando

`acf(x, lag.max=k)`

el cual me grafica $\hat{\rho}(0), \hat{\rho}(1), \dots, \hat{\rho}(k)$ para determinado valor de k que el usuario desee observar.

Si queremos ver los valores numéricos de $\hat{\rho}(0), \hat{\rho}(1), \dots, \hat{\rho}(k)$ utilizamos

`acf(x, lag.max=k)$acf`

Si queremos la función de autocovarianza simplemente utilizamos `acf(x, lag.max=k, type="covariance")`

Cuando corremos en la consola de R los anteriores comandos, vemos que aparecen líneas punteadas horizontales azules, que representan el intervalo de confianza asintótico al 95% entre las que deberían estar las autocorrelaciones si las variables fueran i.i.d. y estn basadas en el siguiente teorema.

Teorema 29 *Si X_1, X_2, \dots son i.i.d. tales que $\mathbb{E}(X_i^4) < +\infty$ entonces*

$$\sqrt{n}(\hat{\rho}(h) - \rho(h)) \rightarrow N(0, 1)$$

cualquiera sea h .

Observación 30 *El teorema anterior significa que si X_1, X_2, \dots son i.i.d. entonces $\hat{\rho}(h)$ es aproximadamente normal con media $\rho(h)$ y desviación $1/\sqrt{n}$.*

En la Figura 7, se ve cómo luce la función de autocorrelación de 1000 observaciones i.i.d. con distribución $N(0, 1)$.

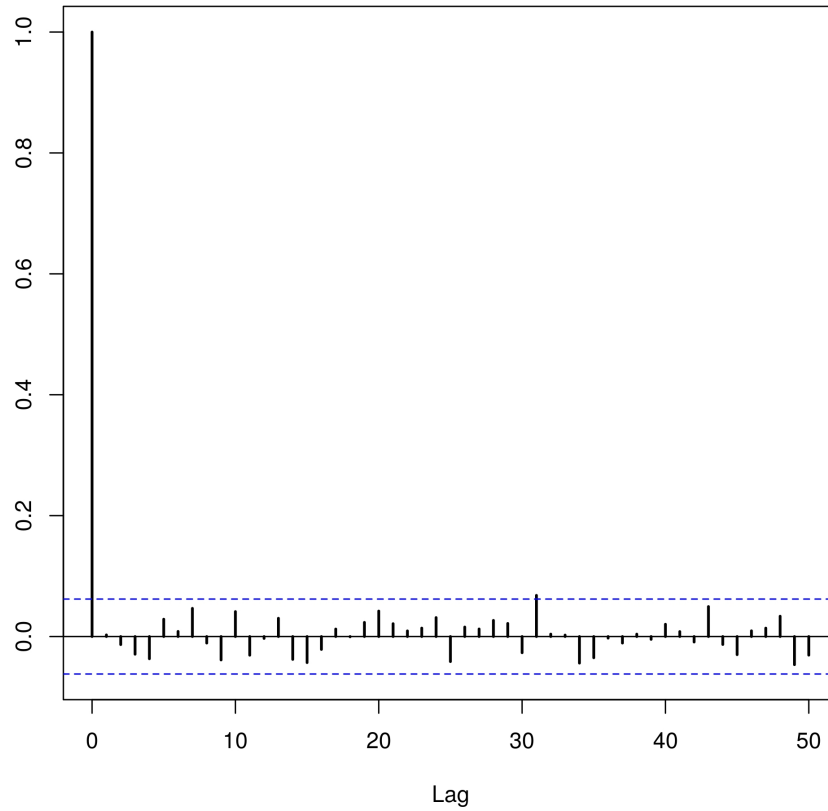


Figura 7: Función de autocorrelación de 1.000 observaciones de variables i.i.d. $N(0,1)$.

6 Algunas funciones en R

La función `ts`, lleva todo vector de datos en un objeto “serie de tiempo”. Por otro lado la función “`decompose`” calcula las componentes de tendencia, estacional e irregular. La componente de tendencia la saca por promedios móviles, para ello es necesario indicarle el tipo de datos sobre si son trimestrales, semestrales, diarios, etc. Luego calcula la componente estacional como se indicó líneas más arriba y luego la componente irregular por diferencia entre cada dato observado a la que se le resta las componentes de tendencia y estacional (si el modelo es aditivo). La función “`decompose`” es capaz de descomponer la serie en sus componentes tanto en el caso aditivo como multiplicativo.

En el ejemplo de la Figura 5, supongamos que el vector x es el de los 120 datos originales. Entonces podemos utilizar las siguientes sentencias para llevar a cabo la descomposición.

```
x.ts=ts(x,start=c(1990,1),end=c(2019,4),fr=4)
```

```
plot(decompose(x,ts))#visualizamos la descomposición.
```

`ts(x,start=c(1990,1),end=c(2019,4),fr=4)` significa que el vector x se lleva a objeto “serie de tiempo”, “fr” significa que son 4 datos por año (es decir, datos trimestrales), y además se indica que el primer dato corresponde al año 1990, primer trimestre y el último dato corresponde al cuarto trimestre del año 2019. La función “decompose” realiza la descomposición de acuerdo al modelo aditivo, si queremos trabajar con el modelo multiplicativo, la sentencia debe ser `decompose(x.ts,type=“multiplicative”)`. Si queremos ver los resultados numéricos de cada componente utilizamos las siguientes sentencias:

```
decompose(x.ts)$trend #nos da la tendencia
```

```
decompose(x.ts)$seasonal #nos da la componente estacional
```

```
decompose(x.ts)$random #nos da la componente aleatoria o irregular
```

En la Figura 8 vemos que la componente estacional se repite (dado que en este ejemplo son sólo 4 valores que se repiten año a año, la curva de tendencia es ms “suave” que la original y para la componente aleatoria o irregular es de esperar que no muestre ningún tipo de patrones, ni de tendencia ni de estacionalidad. Si llegara a mostrarlos, podría indicarnos que existe alguna componente cíclica que habría que detectar y quitar, o simplemente es porque los datos puedan ser ms difíciles de modelar por ejemplo que tenga componentes estacionales o cíclicas con duraciones distintas en el tiempo.

Luego que tengamos la serie descompuesta en sus componentes, deberemos ajustarle algún modelo estacionario a la componente irregular o aleatoria. Antes que nada es importante chequear visualmente en el gráfico de datos de la componente irregular que tenga un comportamiento estacionario en el tiempo. Los gráficos muchas veces permiten visulaizar patrones, pero muchas veces los ocultan. Veremos en lo que sigue algunos test de hipótesis que están disponibles en R para chequear la estacionariedad de la componente aleatoria.

7 Diferenciación como forma de eliminar la tendencia

Si dada una serie de datos, pasamos a una nueva serie considerando las diferencias entre dos términos sucesivos, es posible lograr eliminarla. Para lo que viene y para la teoría, es conveniente suponer que el proceso está definido sobre los enteros (\mathbb{Z}) en lugar de sobre los naturales.

Definición 31 *Definimos el operador B (backward shift) aplicado a un proceso $\{X_t\}_{t \in \mathbb{Z}}$ mediante $BX_t = X_{t-1}$ y el operador ∇ definido como $\nabla X_t = X_t - X_{t-1} = (I - B)X_t$.*

Decomposition of additive time series

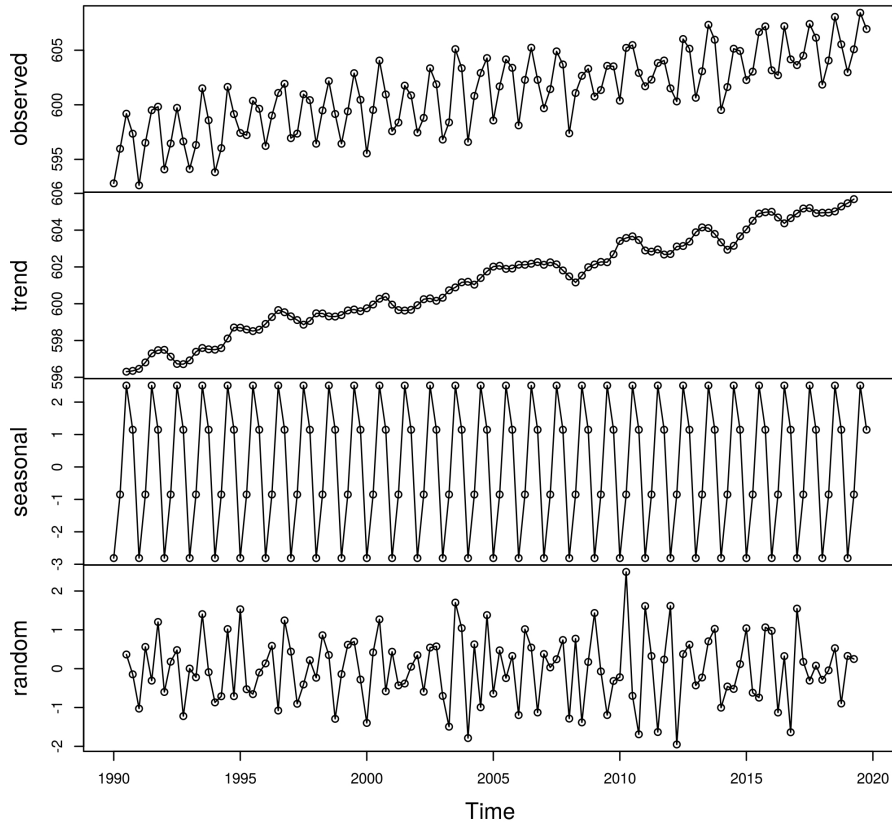


Figura 8: La función decompose nos grafica la serie observada y sus componentes de tendencia y estacionalidad.

Observación 32 Es importante tener en cuenta que BX_t no es un producto, es una notación, ya que B es una función que se la aplica a un proceso $\{X_t\}_{t \in \mathbb{Z}}$ y devuelve otro proceso $\{Y_t\}_{t \in \mathbb{Z}}$ siendo $Y_t = X_{t-1}$. El operador B simplemente significa que a la serie original se le realiza un corrimiento (shift) de una unidad para atrás en el tiempo. De igual forma se interpreta el operador ∇ .

Observación 33 En la igualdad $\nabla X_t = X_t - X_{t-1} = (I - B)X_t$, I es la transformación identidad, que se define como $IX_t = X_t$ para todo t . En muchos casos se escribe también $(1 - B)X_t$ como si fuera un producto.

Observación 34 De forma similar el operador B^2 se interpreta como $B^2X_t = BBX_t = BX_{t-1} = X_{t-2}$ o sea que es un shift de dos pasos para atrás. Simi-

larmente el operador B^k se interpreta como $B^k X_t = X_{t-k}$, o sea, un shift de k pasos para atrás.

Observación 35 Análogamente $\nabla^2 X_t = \nabla \nabla X_t = \nabla (X_t - X_{t-1}) = X_t + X_{t-2} - 2X_{t-1}$ es decir que

$$\nabla^2 X_t = X_t + X_{t-2} - 2X_{t-1}.$$

Veremos a continuación el por qué aplicarle a los datos el operador ∇ en muchos casos puede eliminar la tendencia.

Supongamos que tenemos un proceso estocástico definido mediante $X_t = at + b + \varepsilon_t$ siendo $\{\varepsilon_t\}$ i.i.d. $N(0, \sigma^2)$ para todo $t \in \mathbb{Z}$. Este proceso tiene una componente aleatoria dada por las variables normales i.i.d y una tendencia dada por la recta $y = at + b$. Veremos cómo queda el nuevo proceso $Y_t = \nabla X_t$. Calculamos,

$$Y_t = X_t - X_{t-1} = at + b + \varepsilon_t - (a(t-1) + b + \varepsilon_{t-1}) = a + \varepsilon_t - \varepsilon_{t-1} = a + \varepsilon'_t$$

donde ahora el proceso $\{\varepsilon'_t\}$ es estacionario. ¿Por qué? Es decir que el nuevo proceso $\{Y_t\}$ no tiene tendencia.

Observación 36 Si le pedimos al proceso $\{\varepsilon_t\}$ que sean variables i.i.d. (sin necesidad de pedir normalidad) el resultado anterior sigue siendo válido.

Si la tendencia fuera polinómica de orden k , entonces es posible eliminar la tendencia considerando el nuevo proceso definido por $Y_t = \nabla^k X_t$. Por ejemplo si el proceso fuera de la forma

$X_t = at^2 + bt + c + \varepsilon_t$ siendo $\{\varepsilon_t\}$ i.i.d. entonces (irá dentro de los ejercicios) el nuevo proceso $Y_t = \nabla^2 X_t$ queda de la forma $Y_t = 2a + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2} = 2a + \varepsilon'_t$ siendo $\{\varepsilon'_t\}$ estacionario.

En \mathbb{R} , la función que le aplica el operador ∇ a un vector de datos x es $\text{diff}(x)$.

8 Chequeo de estacionariedad

Teniendo los paquetes “tseries” y “lmtest” podemos realizar algunos test de hipótesis para chequear la estacionariedad de la componente aleatoria. Veremos en lo que sigue cuatro test de hipótesis, los primeros tres son sobre si la serie es no es estacionaria a partir de la existencia de raíces unitarias en el modelo a considerar y dichos tests están basados en modelos autorregresivos, que no tienen por qué ser los modelos que mejor se ajustan a los datos. La definición de raíz unitaria tiene que ver con los modelos autorregresivos que veremos en el próximo capítulo, de modo que por ahora los usaremos simplemente como herramienta. El cuarto test, es para chequear que la varianza de los datos se mantenga constante en el tiempo, y se llama test de heteroscedasticidad. Los cuatro tests de hipótesis son pobres (como ocurre muchas veces en la estadística no paramétrica) en el sentido que en un caso asume modelos autorregresivos para los datos y en el otro caso asume modelo de regresión lineal entre los

datos y el tiempo, cosas que pueden no ser realistas en algunas series de datos. Esto lleva a que los tests podrían indicar que los datos son estacionarios cuando en realidad no lo son. De todas formas, suelen funcionar bastante bien en la práctica de modo que son una herramienta útil de chequeo.

8.1 Test de Dickey–Fuller aumentado (ADF)

Este test fue planteado en su primera versión en 1979, en donde H_0 : el modelo seleccionado para los datos tiene raíz unitaria, o sea que no es estacionario vs H_1 : el modelo es estacionario. Para aplicarlo es necesario tener instalado el paquete tseries.

```
adf.test(x)
```

El argumento x puede ser un vector numérico o una serie de tiempo, lo que quiere decir que puedo aplicar indistintamente `adf.test(x,k)` o `adf.test(x.ts)`.

8.2 Test de Phillips–Perron (PP)

Es todo idéntico al de Dickey Fuller, fue propuesto en 1988 y funciona con el siguiente comando:

```
pp.test(x)
```

8.3 Test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

Este test fue desarrollado en 1992 y plantea al revés de los anteriores H_0 : el modelo seleccionado para ajustar los datos es estacionario, vs H_1 : el modelo es no estacionario. Este test se aplica a través del paquete tseries y funciona con el siguiente comando:

```
kpss.test(x)
```

8.4 Test de Goldfeld–Quandt (GQ)

Este test se usa bajo el paquete lmtest y es para ver que la varianza se mantiene constante. H_0 : la varianza es constante a lo largo de toda la serie, vs H_1 : no H_0 . Se aplica de la siguiente forma: si le llamamos x al vector de datos y t al vector de instantes de tiempo (por ejemplo `t=seq(1:length(x))`), utilizamos el siguiente comando:

```
gqtest(x t,alternative="two.sided")
```

Este test, es un test basado en heteroscedasticidad (cambio en las varianzas) para modelos lineales. En el contexto de series de tiempo debemos tomar como variable regresora a la variable tiempo, eso explica el hecho de ingresar `x t` en la función `gqtest`. La alternativa considerada `two.sided` indica que hay cambio de la varianza en el tiempo. Es posible cambiar `“two.sided”` por `“greater”` (la alternativa es que las varianzas aumentan en el tiempo) o por `“less”` (la alternativa es que las varianzas disminuyen con el tiempo).

El test fue planteado en 1965. Es más antiguo que los tres anteriores debido a que los modelos autorregresivos fueron planteados y comenzados a usarse a

partir de inicios de los años 80, mientras que este test está basado en modelos lineales, tema muchísimo más antiguo en su desarrollo.

Antes de ajustar un modelo estacionario a la componente aleatoria (luego de eliminadas la tendencia y la componente estacional de los datos) se espera que la misma sea estacionaria, o sea debería rechazar H_0 en ADF y PP y no rechazar H_0 en KPSS y GQ.

8.5 Gráfico de la función de autocorrelación

Suele ser útil también una inspección gráfica de la función de autocorrelación de la componente aleatoria. No es de esperar que se comporte como un ruido blanco (Figura 7) porque en la componente aleatoria, generalmente se cumple que el valor en un instante t depende de los anteriores (cuando en el ruido blanco son variables independientes), y de hecho los modelos que se considerarán en el capítulo siguiente incluyen estos casos, pero el gráfico de la función de autocorrelación de los datos originales puede presentar periodicidades que tienen que ver con la componente estacional, que no deberían quedar cuando vemos la función de autocorrelación de la componente aleatoria.

Ejemplo 37 *En el ejemplo de la Figura 9, se tienen datos semanales en m^3/s de las represas “Salto”, “Rincón del Bonete” y “Palmar” desde la primer semana de 1909 hasta la última semana de 2009. Los datos trabajados son un promedio ponderado de las tres represas a los cuales se les llama aportes complejivos. La serie de datos original presenta una clara componente estacional (como era de esperarse de acuerdo a los distintos volúmenes de lluvia que se presentan a lo largo de un año en Uruguay), la cual se ve reflejada claramente en la función de autocorrelación, mientras que esos patrones ya no se ven en el gráfico de la función de autocorrelación de los datos desestacionalizados.*

Nota: en este ejemplo se desestacionalizó la serie mediante la estandarización semana a semana de los datos, es decir que se consideraron 52 componentes estacionales, una por cada semana.

9 Comentarios finales e información adicional

El análisis realizado en este capítulo es llamado análisis clásico. Hay varias variantes para extraer la tendencia y la componente estacional que en muchos casos son mejores. Algunos inconvenientes que pueden presentarse utilizando el método clásico de descomposición son los siguientes.

- La estimación de la tendencia por promedio móviles pierde datos, por ejemplo si estamos trabajando con datos trimestrales y estamos aplicando promedios móviles de frecuencia 4, perdemos los dos primeros datos en la tendencia y los dos últimos. Si trabajáramos con datos mensuales y hacemos promedios móviles de frecuencia 12, en este caso se pierden los 6 primeros y los 6 últimos en la tendencia. No tener datos de tendencia

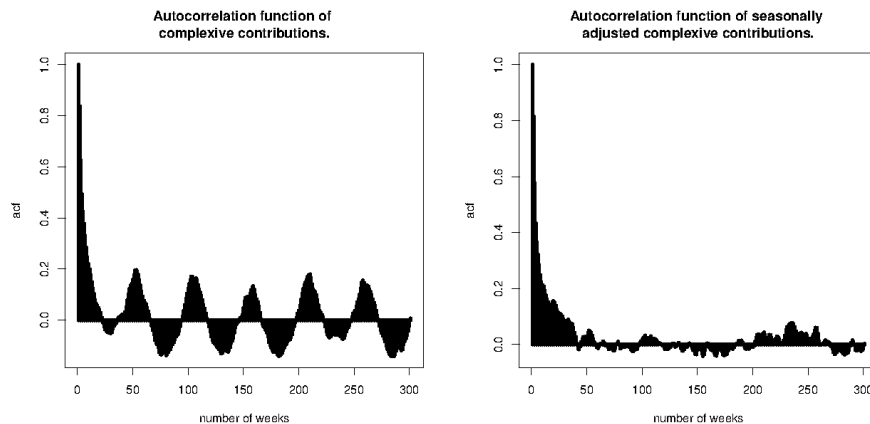


Figura 9: ACF de los datos de aportes complejivos y de su serie desestacionalizada.

justo para los valores finales, que son los que deberían tener más peso a la hora de realizar predicciones no es aconsejable.

- El método de descomposición clásico, asume que la componente estacional es constante. Esto no siempre es realista. Muchas series de tiempo, si son de gran tamaño (incluyen un número grande de años), tienen una componente estacional que no es constante.
- Algunas series de tiempo, tienen en algunos períodos algunos cambios que podrían no ser bien captados por la descomposición clásica.

Como siempre ocurre, es muy importante “el ojo” de quien realice las estadísticas a la hora de hacer un análisis de series de tiempo, todos los problemas anteriores se pueden sortear sin mayores problemas planteando distintas alternativas. Comentamos de manera superficial dos alternativas de manejo de descomposición en series de tiempo más sofisticadas que la clásica.

9.1 Método X-13ARIMA-SEATS

Este método se puede descargar a través del paquete “seasonal” en R. Es un software desarrollado por Unit States Census Bureau. Se basa en la descomposición clásica pero incluye posibilidades de sortear las dificultades planteadas más arriba. Por ejemplo, la tendencia est definida para todos los instantes de tiempo, el método permite variación en la componente estacional, tiene en cuenta valores particulares como las fechas de festividades como fin de año, fin de año chino, etc, y tiende a ser mucho más robusto con respecto a valores extremos que el método clásico. Tiene la desventaja de estar diseñado únicamente para datos mensuales, trimestrales o semestrales. Para mayor información vale

la pena instalar dicho paquete y pedir la ayuda del mismo mediante el comando `?seasonal` en la consola de R.

9.2 Método STL

El método STL (Seasonal and Trend decomposition using Loess) es un método desarrollado por Cleveland, Cleveland, McRae y Terpenning en 1990, es un método que tiene por ventaja el ser robusto a datos atípicos, las componentes de tendencia y estacional pueden variar en el tiempo y es válido para cualquier serie de datos, pueden ser datos diarios, semanales, trimestrales, etc. Permite detectar outliers también. El método está incluido en el paquete básico de estadística de R. La función en R se llama “`stl`” y la información respecto a su funcionamiento junto con ejemplos se obtiene con el comando `?stl` en la consola de R.

10 Ejercicios

1. Simular 360 datos independientes con distribución exponencial de parámetro $\lambda = 2$ y calcular su tendencia mediante aplicación de la recta de regresión. Graficar los datos con la recta de regresión incluida. ¿Cómo se interpreta este resultado?
2. En el ejercicio anterior, dibujar su tendencia aplicando promedios móviles considerando los datos trimestrales. Chequear la estacionariedad de los datos.
3. Simular 180 datos trimestrales que estén formados por una recta de tendencia de $-0.2t$ y que sume 1 en cada primer trimestre, sume 2 en cada segundo trimestre, reste 1 en cada tercer trimestre y reste 5 en cada cuarto trimestre y cuya componente aleatoria tenga distribución $N(1, 4)$. Graficar su componente estacional, tendencia por promedios móviles y componente aleatoria.
4. Con los mismos datos generados en el ejercicio anterior, suponer ahora que los datos son semestrales y que el primero de ellos es del primer semestre de 1905. Dibujar en un único gráfico la recta de tendencia por mínimos cuadrados junto con la tendencia por promedios móviles y que en el eje de las abscisas aparezcan los años. Chequear la estacionariedad de los datos.
5. Dibujar en un gráfico la serie de tiempo del archivo llamada “ejercicio1series.csv” y ajustarle por mínimos cuadrados los siguientes 3 posibles modelos.
Modelo 1: $X_t = a + bt + c \sin(t/5) + \varepsilon_t$.
Modelo 2: $X_t = a + bt + c \cos(t/5) + \varepsilon_t$.
Modelo 3: $X_t = a + bt + c \sin(t/5) + d \cos(t/5) + \varepsilon_t$.
Superponer en el gráfico de los puntos observados las 3 curvas estimadas. ¿Cuál ajusta mejor?

6. Probar que el proceso definido mediante la igualdad $X_t = at^2 + bt + c + \varepsilon_t$ siendo $\{\varepsilon_t\}$ ruido blanco, se transforma en un proceso estacionario si aplicamos el operador ∇^2 . Probar además que el proceso Y_t definido por $Y_t = \nabla^2 X_t$ es estacionario.
7. Idem para el caso de $X_t = at^3 + bt^2 + ct + d + \varepsilon_t$ siendo $\{\varepsilon_t\}$ i.i.d. con el operador ∇^3 . Probar además que el proceso Y_t definido por $Y_t = \nabla^3 X_t$ es estacionario.
8. Para los datos del ejercicio 3, dibujar la curva acf para los datos originales, la acf de la componente de tendencia, la acf de la componente estacional y la acf de la componente aleatoria. Utilizar un lag máximo de 100. ¿Cuál es la explicación intuitiva de los resultados obtenidos?