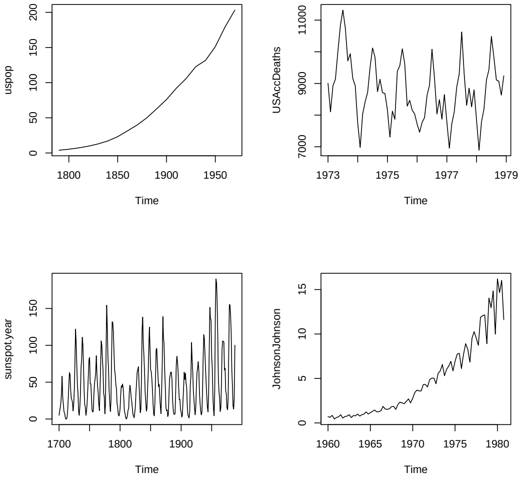
Series de tiempo

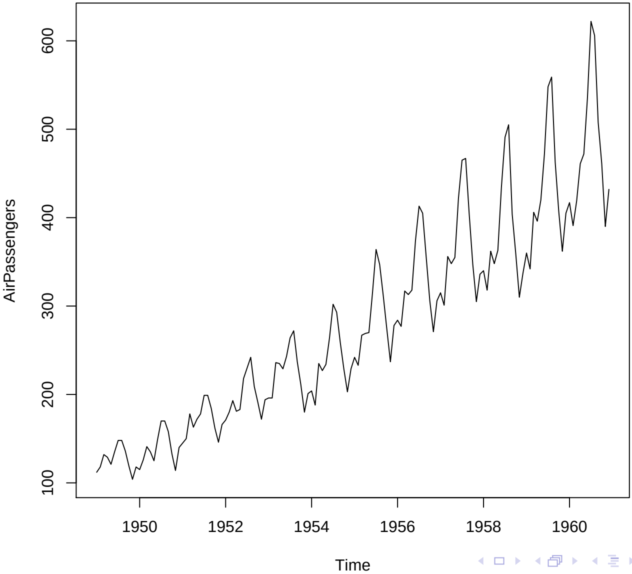
Estadística
Miguel Ángel Chong R.
miguel@sigma.iimas.unam.mx

5 de febrero del 2013

El primer paso en el análisis de series de tiempo, consiste en graficar la serie. A continuación graficaremos las siguientes series



1. uspop Esta serie de tiempo muestra los censos de la población de los E.U. realizados cada 10 años, de 1790 a 1970. Esta serie parece tener un tener un tendencia al menos cuadrática. No parece tener una parte estacional (cíclica).
2. USAccDeaths Representa el total de muertes accidentales mensuales entre los años 1973 a 1978. Esta serie parece tener un comportamiento estacional cada año (un periodo de 12)
3. sunspot.year Muestra el número anual de manchas solares de 1700 a 1988. Esta serie muestra un comportamiento estacional, pero aquí no claro el periodo en el cual se repide el cíclo como en la serie anterior.
4. JohnsonJohnson son las ganancias trimestrales de la compañía Johnson & Johnson de 1960-80. Esta serie tiene tanto una parte cíclica (con periodo 3) y como una tendencia no lineal.
5. AirPassengers captura el total mensual de los pasajeros de líneas aéreas internacionales de 1949 a 1960. Como en la serie anterior podemos ver que esta serie tiene tendencia no lineal y una parte cíclica de periodo 12.


AirPassengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Time

La inspeción gráfica puede sugerir la posibilidad de representar los datos como una realización de un proceso que puede tener todas o alguna de las siguientes componentes:

$$X_t = f(m_t, s_t, Y_t) = \begin{cases} m_t + s_t + Y_t & \text{modelo aditivo} \\ m_t \cdot s_t \cdot Y_t & \text{modelo multiplicativo} \end{cases},$$

donde

- m_t es la componente de tendencia,
- s_t es el componente estacional de periodo d, donde $s_t = s_{t+d} = s_{t+2d} = \ldots$
- Y_t es el componente aleatorio.

Es claro que un modelo multiplicativo $X_t = m_t \cdot s_t \cdot Y_t$ lo podemos llevar a un modelo multiplicativo, siempre que $X_t > 0$, al tomar logaritmo, $X'_t = \log(X_t) = \log(m_t \cdot s_t \cdot Y_t) = \log(m_t) + \log(s_t) + \log(Y_t) = m'_t + s'_t + Y'_t$. Este tipo de trasnformaciones son utiles para linealizar los datos y reducir la varianza de la serie de tiempo. Un ejemplo de esto son las series, JohnsonJohnson y AirPassengers.

Primero veamos como podemos leer datos y ponerlos en formato serie de tiempo usando R.

Primero hay que leer datos de algún archivos (preferentemente .txt)
datos = read.table("archivo", sep=" ", encabezado)
y darle formato de serie de tiempo con el comando
serie = ts(datos, start, frequency)
plot(serie)

A continuación veremos algunos métodos que se han propuesto para identificar y describir las componentes de tendencia m_t y la parte estacional o cíclica s_t de la serie de tiempo.

Primero estudiaremos tres métodos para estimar la tendencia de una serie de tiempo de la forma

$$X_t = m_t + Y_t, \text{ o } X_t = m_t + s_t + Y_t$$

Método 1

Si tenemos una serie que tiene solo una parte de tendencia $X_t = m_t + Y_t$, como por ejemplo la serie uspop. Una manera de estimar de m_t sería vía mínimos cuadrádos. Es decir, procederemos a estimar la tendencia de entre una familia de funciones de la forma

$$m_t = a_0 + a_1 t + a_2 t^2,$$

y escogeremos los a_0, a_1 y a_2 que minimizen $\sum_t (X_t - m_t)^2$.
Y así obtener

$$\hat{m}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2.$$

Una vez estima da la tendencia podemos estimar la parte aleatoria de la siguiente manera

$$\hat{Y}_t = X_t - \hat{m}_t,\ t \in \{1, \ldots, n\}.$$

Método 2

Suavizamineto de la media vía un promedios moviles.

Sea q un entero no negativo y consideremos el promedio movil de dos lados como

$$\hat{m}_t = \sum_{j=-q}^{q} a_j X_{t+j}, \quad \text{para } t \in \{q+1, q+2, \ldots, n-q\},$$

donde $\sum_{j=-q}^{q} a_j = 1$.

Un caso particular de este tipo de ajuste de la tendencia es si suponemos que $a_j = \frac{1}{2q+1}$, es decir

$$\hat{m}_t = \frac{1}{2q+1} \sum_{-q}^{q} X_{t+j}, \ t \in \{q+1, q+2, \ldots, n-q\}.$$

Antes de describir el último método introduzcamos los operadores retraso y diferencia.

El **operador retraso**, denotado por B, actua sobre el tiempo de la siguiente manera

$$\begin{aligned} BX_t &= X_{t-1} \\ B^2X_t &= B\left(BX_t\right) = BX_{t-1} = X_{t-2} \\ &\vdots \\ B^jX_t &= X_{t-j}. \end{aligned}$$

Definamos a $B^0X_t = X_t$.

A partir del operador retraso definimos el **operador diferencia** como

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

Notemos que el operador diferencia lo podemos manipular como si fuera un polinomio común y corriente es decir

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

Y por lo tanto si deseamos diferenciar X_t la serie j veces tenemos que

$$\begin{aligned}\nabla^j X_t &= (1-B)^j X_t \\ &= \sum_{k=0}^{j} \binom{j}{k} (-1)^k B^k X_t \\ &= \sum_{k=0}^{n} \binom{j}{k} (-1)^k X_{t-k}.\end{aligned}$$

Método 3

Cuando la serie de tiempo X_t tiene una tendencia lineal, $m_t = at + b$, entonces al aplicar un el operador diferencia ∇ obtenemos $\nabla m_t = a$.

De la misma forma, si nosotros tenemos una serie de tiempo

$X_t = m_t + Y_t$ donde la tendenca es polinomial de grado k, $m_t = \sum_{j=0}^{k} a_j t^j$, entonces al aplicar el operador diferencia a la tendencia tenemos que $\nabla^k m_t = k! a_k$ y por lo tanto tenemos que

$$\nabla^k X_t = k! a_k + \nabla^k Y_t.$$

Lo que pretendemos ahora es quitar la tendencia y la parte estacional en un modelo general

$$X_t = m_t + s_t + Y_t$$

donde $\mathbb{E}[Y_t] = 0$, $s_t = s_{t+d}$ y $\sum_{j=1}^{d} s_j = 0$

Método S1

Para ilustrar este método usemos la serie de muertes accidentales USAccDeaths, y cambiemos la notación de la serie de tiempo $\{x_1, x_2, \ldots, x_n\}$ por los sub índices $x_{j,k}$, donde $j \in \{1, 2, \ldots, 6\}$ representa los años y $k = \{1, \ldots, 12\}$ representa los meses y claramente el periodo es $d = 12$

Para este método calculemos la tendencia **anual** que cambia suavemente como

$$\hat{m}_j = \frac{1}{12} \sum_{k=1}^{12} x_{j,k}.$$

Mientras que la parte estacional la calcularemos como

$$\hat{s}_k = \frac{1}{6} \sum_{j=1}^{6} \left(x_{j,k} - \hat{m}_j\right).$$

Por último, calculamos la parte aleatoria como

$$\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k,$$

para $j \in \{1, 2, \ldots, 6\}$ y $k = \{1, \ldots, 12\}$

Lo que pretendemos ahora es quitar la tendenciay la parte estacional

$$X_t = m_t + s_t + Y_t$$

donde $\mathbb{E}[Y_t] = 0$, $s_t = s_{t+d}$ y $\sum_{j=1}^{d} s_j = 0$