Proyecto 2: Series de Tiempo

Clemente Ferrer, Rodrigo Pizarro

1. Considere un modelo de la forma

$$Y_t = \beta_0 + \sum_{i=1}^{6} \beta_i \cos(2\pi t/T_i) + \epsilon_t,$$

donde el proceso ϵ_t es un ruido blanco con varianza σ^2 y T_i son los periodos de la serie.

a. Escriba este modelo en la forma $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$

Solución. Obseve que para $t_1, \ldots, t_n$ el modelo está determinado por el sistema

$$\begin{cases} Y_{t_1} = \beta_0 + \sum_{i=1}^{6} \beta_i \cos(2\pi t_1/T_i) + \epsilon_{t_1} \\ \vdots \\ Y_{t_n} = \beta_0 + \sum_{i=1}^{6} \beta_i \cos(2\pi t_n/T_i) + \epsilon_{t_n} \end{cases}$$

Luego, podemos definir los vectores $\boldsymbol{Y}$, $\boldsymbol{\beta}$, $\boldsymbol{e}$ y la matriz $\boldsymbol{X}$ como

$$\boldsymbol{Y} = (Y_{t_1}, \ldots, Y_{t_n})^\top \in \mathbb{R}^n,$$

$$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_6)^\top \in \mathbb{R}^7,$$

$$\boldsymbol{e} = (\epsilon_{t_1}, \ldots, \epsilon_{t_n})^\top \in \mathbb{R}^n,$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & \cos(2\pi t_1/T_1)) & \cdots & \cos(2\pi t_1/T_6)) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(2\pi t_n/T_1)) & \cdots & \cos(2\pi t_n/T_6)) \end{pmatrix} \in \mathbb{R}^{n\times 7}.$$

Así, el modelo queda escrito como $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$. □

b. Explique cómo obtener estimaciones de $\beta_0, \ldots, \beta_6$ y σ^2.

Solución. Dado que el modelo queda escrito de la forma $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, podemos aplicar mínimos cuadrados para estimar los coeficients β_i y σ^2. En particular por lo visto en MAT266, sabemos que las estimaciones vienen dadas por

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^7}{\arg\min} \left\{ (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right\} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y},$$

$$\hat{\sigma^2} = \frac{\boldsymbol{e}^\top \boldsymbol{e}}{n-k} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-k},$$

donde k corresponde al número de coeficientes (incluido el intercerpto) a estimar, en este caso $k = 6$. Junto a ello, observe que $\boldsymbol{X}^\top \boldsymbol{X}$ es no singular y por ende invertible. □

c. ¿Qué consideraciones hay que establecer para que el modelo (1) incluya una tendencia cuadrática?

Solución. Para que el modelo incluya una tendencia cuadrática basta con agregar los términos $\beta_7 t$ y $\beta_8 t^2$, es decir,

$$Y_t = \beta_0 + \sum_{i=1}^{6} \beta_i \cos(2\pi t/T_i) + \beta_7 t + \beta_8 t^2 + \epsilon_t.$$

Luego el vector $\boldsymbol{\beta}$ y la matriz $\boldsymbol{X}$ quedarían redefinidos como

$$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_8)^\top \in \mathbb{R}^9 \quad \text{y} \quad \boldsymbol{X} = \begin{pmatrix} 1 & \cos(2\pi t_1/T_1)) & \cdots & \cos(2\pi t_1/T_6)) & t & t^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(2\pi t_n/T_1)) & \cdots & \cos(2\pi t_n/T_6)) & t & t^2 \end{pmatrix} \in \mathbb{R}^{n\times 9}.$$

□

2. Sea $\{X_t : t \in T\}$ un proceso estacionario normal con función de media μ_Xy función de autocovarianza $C_X(\cdot)$. Definamos la serie no lineal

$$Y_t = \exp\left(X_t\right), t \in T.$$

a. Exprese la media del proceso Y_t en términos de μ_X y $C(0)$.

Solución. Recordemos que la función generadora de momentos de una variable aleatoria X está dada por $M_X(\lambda) := \mathbb{E}[\exp(\lambda X)]$. En particular, como X_t es un proceso estacionario normal,

$$M_X(\lambda) = \exp\left(\mu_X \lambda + \frac{C_X(0)\lambda^2}{2}\right).$$

Luego la media de Y_t está dada por

$$\mu_Y := \mathbb{E}[Y_t] = \mathbb{E}[\exp(X_t)] = M_X(1) = \exp\left(\mu_X + \frac{C_X(0)}{2}\right).$$

□

b. Determine la función de autocovarianza de Y_t.
Solución. De manera directa

$$\begin{aligned}
C_Y(h) &= \mathrm{Cov}[Y_{t+h}, Y_t] \\
&= \mathbb{E}[(Y_{t+h} - \mathbb{E}[Y_{t+h}])(Y_t - \mathbb{E}[Y_t])] \\
&= \mathbb{E}[(Y_{t+h} - \mu_Y)(Y_t - \mu_Y)] \\
&= \mathbb{E}[Y_{t+h}Y_t - \mu_Y Y_{t+h} - \mu_Y Y_t + \mu_Y^2] \\
&= \mathbb{E}[Y_{t+h}Y_t] - 2\mu_Y^2 + \mu_Y^2 \\
&= \mathbb{E}[\exp(X_{t+h} + X_t)] - \mu_Y^2
\end{aligned}$$

Ahora bien, como $X_{t+h} \sim N(\mu_X, C_X(0))$ y $X_t \sim N(\mu_X, C_X(0))$ son procesos estacionarios normales, $(X_{t+h} + X_t) \sim N(2\mu_X, C_X(0) + 2\mathrm{Cov}(X_{t+h}, X_t) + C_X(0))$, es decir,

$$(X_{t+h} + X_t) \sim N(2\mu_X, 2C_X(0) + 2C_X(h)).$$

Luego, la función generadora de momentos para $X_{t+h} + X_t$ está dada por

$$M_{X_{t+h}+X_t}(\lambda) = \exp\left(2\mu_X\lambda + (C_X(0) + C_X(h))\lambda^2\right).$$

Considerando lo anterior, deducimos que

$$\begin{aligned}
C_Y(h) &= \mathbb{E}[\exp(X_{t+h} + X_t)] - \mu_Y^2 \\
&= M_{X_{t+h}+X_t}(1) - \mu_Y^2 \\
&= \exp\left(2\mu_X + C_X(0) + C_X(h)\right) - \mu_Y^2 \\
&= \exp(2\mu_X + C_X(0))\left[\exp(C_X(h)) - 1\right]
\end{aligned}$$

□

3. Si $C_j(h)$ son funciones de covarianza de un proceso estacionario débil para todo $j = 1, \ldots, n$. Demuestre que $\sum_{j=1}^n b_j C_j(h)$ también es una función de covarianza si $b_j \geq 0, \forall j$.

Solución. Debemos verificar que la combinación lineal de funciones de covarianza con pesos no negativos de un proceso estacionario débil es una función de covarianza. Para ello debe satisfacerse la paridad y la condición de ser semidefinida positiva. Definamos $C(h) = \sum_{j=1}^n b_j C_j(h)$ y notemos que

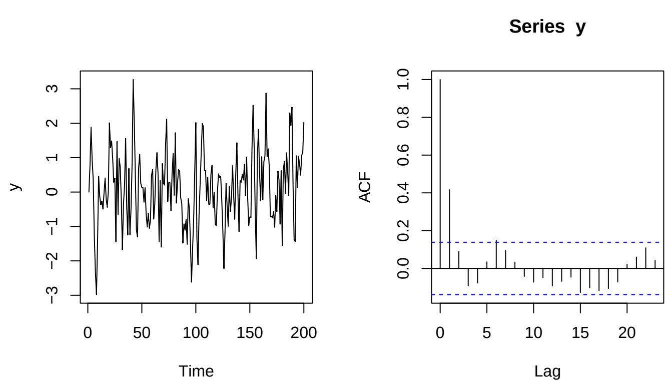
- Dado que $C_j(h)$ son funciones de covarianza para todo $j = 1, \dots, n$, se tiene que para todo $a_1, \dots, a_m \in \mathbb{R}$ se satisface que
$$\sum_{i,k=1}^{m} a_i a_k C_j(h) \geq 0.$$
Ahora bien, observe que para $b_j \geq 0$ obtenemos
$$\sum_{i,k=1}^{m} a_i a_k C(h) = \sum_{i,k=1}^{m} a_i a_k \left(\sum_{j=1}^{n} b_j C_j(h) \right) = \sum_{j=1}^{n} b_j \underbrace{\left(\sum_{i,k}^{m} a_i a_k C_j(h) \right)}_{\geq 0} \geq 0.$$
Así, $C(h)$ es semidefinida positiva.
- Por otro lado, cada C_j es una función de covarianza y por ende $C_j(h) = C_j(-h)$, para todo $j = 1, \dots, n$. Luego
$$C(h) = \sum_{j=1}^{n} b_j C_j(h) = \sum_{j=1}^{n} b_j C_j(-h) = C(-h),$$
lo que comprueba la paridad de $C(h)$.

Por lo anterior, concluimos que $C(h)$ es una función de covarianza. □

4. Describa que hace exactamente la siguiente rutina en R

```
x=rnorm(200,0,1)
y=vector(mode="numeric", length=200)
for (i in 2:200){
y[i]=0.5*y[i-1]+x[i]
}
par(mfrow=c(1,2),pty = "s")
plot.ts(y)
acf(y)
```

Solución. Primeramente, creamos un elemento `x` que mediante la función `rnorm` genera 200 observaciones distribuidas normalmente con media 0 y varianza 1. Enseguida, inicializamos un vector `y` de 200 observaciones, todas rellenadas con el valor 0. Luego, mediante un ciclo `for` generamos un modelo de la forma
$$Y_t = \frac{1}{2} Y_{t-1} + \epsilon_t, \quad t = 2, \dots, 200$$
donde $\epsilon_t \sim N(0,1)$. Finalmente, graficamos la serie temporal creada mediante el comando `plot.ts` y obtenemos un gráfico de autocorrelación de la misma usando `acf`.



Observación: Del gráfico de autocorrelación notamos que los valores de nuestra serie temporal no están lo suficientemente correlacionados. □

5. Considere el proceso $X_t = \delta + X_{t-1} + \epsilon_t$, donde $t = 1, 2, \ldots, \epsilon_t$ es una secuencia de variables aleatorias iid con media cero y varianza σ^2.

a. Escriba la ecuacion del proceso X_t como sigue

$$X_t = \delta t + \sum_{j=1}^{t} \epsilon_j$$

Solución. Observe que, por definición del proceso

$$X_1 = \delta + X_0 + \epsilon_1 = \delta + \epsilon_1$$

y

$$X_2 = \delta + X_1 + \epsilon_2 = \delta + \delta + \epsilon_1 + \epsilon_2 = 2\delta + \sum_{j=1}^{2} \epsilon_j.$$

Por inducción puede generalizarse para todo t. En efecto, asumamos que lo anterior es cierto para X_{t-1} y note que

$$X_t = \delta + X_{t-1} + \epsilon_t = \delta + \delta(t-1) + \sum_{j=1}^{t-1} \epsilon_j + \epsilon_t = \delta t + \sum_{j=1}^{t} \epsilon_j.$$

Lo que comprueba lo pedido. □

b. Calcule $\mu(t) = \mathbb{E}\left[X_t\right]$ y $V(t) = \mathbb{V}\left[X_t\right]$.

Solución. Usando que X_t puede escribirse como el item anterior plantea se obtiene que

$$\mu(t) = \mathbb{E}\left[X_t\right] = \mathbb{E}\left[\delta t + \sum_{j=1}^{t} \epsilon_j\right] = \mathbb{E}[\delta t] + \mathbb{E}\left[\sum_{j=1}^{t} \epsilon_j\right] = \delta t + \sum_{j=1}^{t} \underbrace{\mathbb{E}[\epsilon_j]}_{=0} = \delta t.$$

y de manera análoga

$$V(t) = \mathbb{V}\left[X_t\right] = \mathbb{V}\left[\delta t + \sum_{j=1}^{t} \epsilon_j\right] = \mathbb{V}[\delta t] + \mathbb{V}\left[\sum_{j=1}^{t} \epsilon_j\right] = \sum_{j=1}^{t} \underbrace{\mathbb{V}[\epsilon_j]}_{=\sigma^2} = t\sigma^2.$$

□

c. ¿Es el proceso X_t débilmente estacionario?

Solución. El proceso X_t **no** es débilmente estacionario, pues su función de media y varianza dependen explicitamente del tiempo como se pudo apreciar en el item anterior. □

6. Sea X_t un proceso intrínsecamente estacionario. El semivariograma de X_t se define como

$$\gamma_X(h) = \frac{1}{2}\mathbb{E}\left[(X_{t+h} - X_t)^2\right].$$

a. Si X_t es un ruido blanco con varianza σ^2, calcule $\gamma_X(h)$.

Solución. Si X_t es un ruido blanco, por definición $\mathbb{E}[X_t] = 0$. Luego,

$$\mathbb{V}[X_t] = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 = \mathbb{E}[X_t^2] = \sigma^2.$$

Considerando lo anterior

$$\begin{aligned}\gamma_X(h) &= \frac{1}{2}\mathbb{E}\left[(X_{t+h}-X_t)^2\right]\\ &= \frac{1}{2}\mathbb{E}\left[X_{t+h}^2 - 2X_{t+h}X_t + X_t^2\right]\\ &= \frac{1}{2}\left(\mathbb{E}[X_{t+h}^2] - 2\mathbb{E}[X_{t+h}X_t] + \mathbb{E}[X_t^2]\right)\\ &= \frac{1}{2}\left(\sigma^2 - 2\mathrm{Cov}[X_{t+h},X_t] + \sigma^2\right),\end{aligned}$$

donde lo anterior se justifica pues,

$$\mathrm{Cov}[X_{t+h},X_t] = \mathbb{E}[(X_{t+h}-\mathbb{E}[X_{t+h}])(X_t-\mathbb{E}[X_t])] = \mathbb{E}[X_{t+h}X_t].$$

Finalmente, dado que $\mathrm{Cov}[X_t,X_t] = \mathbb{V}[X_t]$ y, en particular para un ruido blanco $\mathrm{Cov}[X_{t+h},X_t]=0$, deducimos que

$$\gamma_X(h) = \sigma^2 - \mathrm{Cov}[X_{t+h},X_t] = \begin{cases} 0, & \text{si } h=0,\\ \sigma^2, & \text{si } h\neq 0.\end{cases}$$

□

b. Si $X_t = \beta_0 + \beta_1 t + \epsilon_t$, donde ϵ_t es un ruido blanco con varianza σ^2, calcule $\gamma_X(h)$.

Solución. Usando el resultado anterior, se obtiene que

$$\begin{aligned}\gamma_X(h) &= \frac{1}{2}\mathbb{E}\left[(X_{t+h}-X_t)^2\right]\\ &= \frac{1}{2}\mathbb{E}\left[(\beta_0+\beta_1(t+h)+\epsilon_{t+h} - (\beta_0+\beta_1 t+\epsilon_t))^2\right]\\ &= \frac{1}{2}\mathbb{E}[(\beta_1 h + \epsilon_{t+h}-\epsilon_t)^2]\\ &= \frac{1}{2}\mathbb{E}[\beta_1^2h^2 + 2\beta_1 h(\epsilon_{t+h}-\epsilon_t) + (\epsilon_{t+h}-\epsilon_t)^2]\\ &= \frac{1}{2}\left(\mathbb{E}[\beta_1^2h^2] + 2\beta_1 h\mathbb{E}[\epsilon_{t+h}] - 2\beta_1 h\mathbb{E}[\epsilon_t] + \mathbb{E}[(\epsilon_{t+h}-\epsilon_t)^2]\right)\\ &= \frac{1}{2}\beta_1^2h^2 + \frac{1}{2}\mathbb{E}[(\epsilon_{t+h}-\epsilon_t)^2]\\ &= \begin{cases} 0, & \text{si } h=0,\\ \frac{1}{2}\beta_1^2h^2+\sigma^2, & \text{si } h\neq 0.\end{cases}\end{aligned}$$

□

7. Sea $C_X(\cdot)$ la función de covarianza asociada a un proceso de media nula. Si

$$C_X(T) = C_X(0)$$

para algún $T>0$, Demuestre que $C_X(\cdot)$ es periódica.

Solución. Probaremos que, para todo h se tiene que $C_X(h) = C_X(h+T)$, donde T es el periodo del proceso.

Como $\mathbb{E}[X_t] = 0$, para todo t, observe que

$$\begin{aligned}
0 \leq (C_X(h) - C_X(h+T))^2 &= (\mathrm{Cov}[X_{t+h}, X_t] - \mathrm{Cov}[X_{t+h+T}, X_t])^2 \\
&= (\mathbb{E}[(X_{t+h} - \mathbb{E}[X_{t+h}])(X_t - \mathbb{E}[X_t])] - \mathbb{E}[(X_{t+h+T} - \mathbb{E}[X_{t+h+T}])(X_t - \mathbb{E}[X_t])])^2 \\
&= (\mathbb{E}[X_{t+h}X_t] - \mathbb{E}[X_{t+h+T}X_t])^2 \\
&= (\mathbb{E}[X_t(X_{t+h} - X_{t+h+T})])^2 \\
&\leq \mathbb{E}[X_t^2]\mathbb{E}[(X_{t+h} - X_{t+h+T})^2] && \text{(Cauchy-Schwarz)} \\
&= C_X(0)\left(\mathbb{E}[X_{t+h}^2] - 2\mathbb{E}[X_{t+h+T}X_{t+h}] + \mathbb{E}[X_{t+h+T}^2]\right) \\
&= C_X(0)\,(C_X(0) - 2C_X(T) + C_X(0)) \\
&= C_X(0)\,(2C_X(0) - 2C_X(0)) && (C_X(0) = C_X(T)) \\
&= 0
\end{aligned}$$

Así, concluimos que $C_X(h) - C_X(h+T) = 0$ y por ende, $C_X(h) = C_X(h+T)$, es decir, $C_X(\cdot)$ es periódica de periodo T. $\square$