Series de Tiempo 2018

Maestría en Estadística Aplicada, UNR
Unidad 3

Luis Damiano
damiano.luis@gmail.com

2018-04-17

Contenido

- Procesos autorregresivos.
- Procesos de media móvil.
- Procesos ARMA.

Procesos autorregresivos

Proceso autorregresivo de primer orden

Proceso autorregresivo de primer orden $AR(1)$

$$Z_t = \phi_1 Z_{t-1} + a_t,\ \ t = 0, \pm 1, \ldots$$

con $a_t \sim \mathrm{WN}(0, \sigma^2)$, $|\phi_1| < 1$, y a_t no está correlacionado con Z_s para todo $s < t$ (Brockwell and Davis 2016, 15).

Discusión en clases

¿Cómo simularían una muestra del proceso, dado el valor de los parámetros?

Simulación

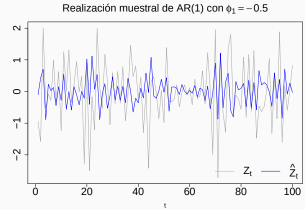
$$Z_t - \phi_1 Z_{t-1} = a_t, \; t = 0, \pm 1, \ldots, \; a_t \sim \mathcal{N}(0, \sigma^2)$$

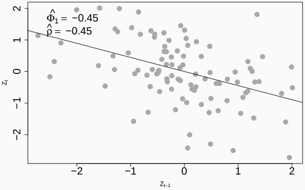
```
simAR1 <- function(phi1, sigma, Z0, T) {
  TT <- 2 * T

  # Ruido
  at <- rnorm(TT, 0, sigma)

  # Observaciones
  Zt <- vector("numeric", TT)
  Zt[1] <- Z0
  for (t in 2:TT) {
    Zt[t] <- phi1 * Zt[t - 1] + at[t]
  }

  # Descartamos la primera mitad para eliminar la influencia del valor inicial
  tail(Zt, T)
}
```

Ejemplo

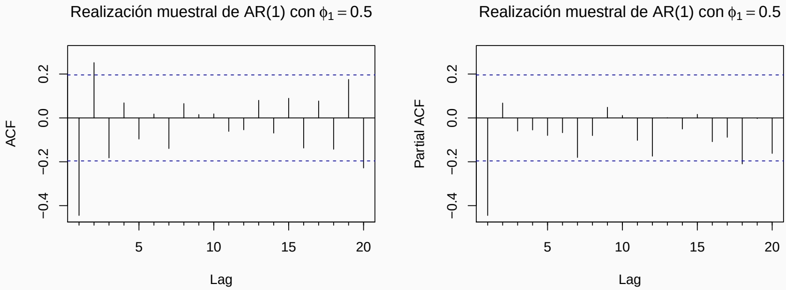
$$Z_t - 0.5Z_{t-1} = a_t,\ \ t = 0, \pm 1, \ldots,\ \ a_t \sim \mathcal{N}(0, 1)$$

```r
set.seed(9000)
z <- simAR1(-0.5, 1, 0, 100)

library(forecast)
fit   <- Arima(
  z,
  order = c(1, 0, 0),
  include.mean = FALSE
  )

lmfit <- lm(z[-1] ~ z[-length(z)] - 1)
```





Estimaciones muestrales



Ejercicio en clases

Analíticamente, encontrar la función de autocorrelación y la función de autocorrelación parcial.

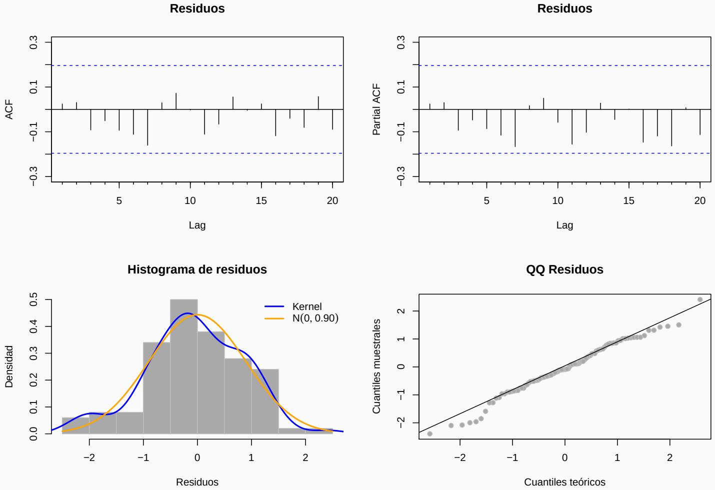
Modelo ajustado

```
## Series: z
## ARIMA(1,0,0) with zero mean
##
## Coefficients:
##           ar1
##       -0.4457
## s.e.   0.0896
##
## sigma^2 estimated as 0.8121:  log likelihood=-131.1
## AIC=266.19   AICc=266.32   BIC=271.4
```

Discusión en clases

¿Cuáles son todos los supuestos del modelo? ¿Cómo validarían cada uno de ellos?

Diagnóstico de residuos

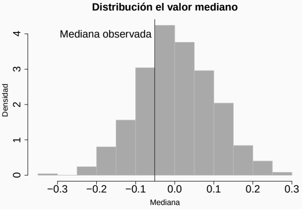


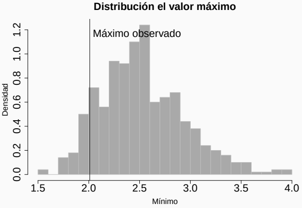
¿Qué supuestos estamos diagnosticando? Hay un supuesto implícito que no estamos probando. . .

Diagnóstico ¿predictivo?

Una forma diferente de diagnosticar el modelo: simular datos bajo el modelo ajustado y comparar estadísticos de resumen contra la muestra observada. ¿Qué decisiones hubiesen tomado con una muestra diferente?[1]

```r
phiHat   <- coef(fit)
sigmaHat <- sd(residuals(fit))
genN     <- 500 # Generar genN muestras según modelo
genMin   <- vector("numeric", genN)
genMed   <- vector("numeric", genN)
genMax   <- vector("numeric", genN)
for (n in 1:genN) {
  gen <- simAR1(phiHat, sigmaHat, 0, length(z))
  genMin[n] <- min(gen)
  genMed[n] <- median(gen)
  genMax[n] <- max(gen)
}
gen      <- simAR1(phiHat, sigmaHat, 0, length(z))
```





[1] If the model fits, then replicated data generated under the model should look similar to observed data. [. . .] Any systematic differences between the simulations and the data indicate potential failings of the model (Gelman et al. 2014, p 143).

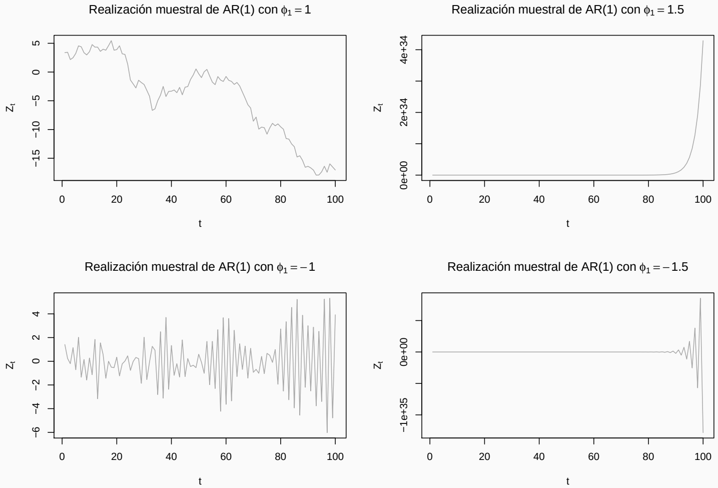
Qué sucede si. . .

Discusión en clases

¿Cuáles son las restricciones que se imponen sobre el coeficiente autorregresivo? ¿Qué imaginan que suceda si no se cumplen?

Qué sucede si. . .

AR(1) siempre es invertible. Para que sea estacionario, se requiere que $|\phi_1| < 1$.



Proceso autorregresivo de segundo orden

El siguiente es un proceso $AR(2)$:

$$Z_t - 1.3Z_{t-1} + 0.4Z_{t-2} = a_t,\ t = 0, \pm 1, \ldots,\ a_t \sim \mathcal{N}(0, 1).$$

Para divertirse en casa:

- Escribir código para simular un conjunto de datos.
- Emplear `Arima` para estimar el valor de los parámetros y corroborar contra los valores prefijados.
- Emplear `Acf` para estimar las funciones ACF y PACF y corroborar con los resultados analíticos.
- Prueben de romper el modelo desafiando las condiciones de estacionariedad :)

Procesos de medias móviles

Proceso medias móviles de primer orden

Proceso de medias móviles de primer orden $MA(1)$

$$Z_t = a_t - \theta_1 a_{t-1},\ t = 0, \pm 1, \ldots$$

con $a_t \sim \mathrm{WN}(0, \sigma^2)$, $\theta \in \mathbb{R}$ (Brockwell and Davis 2016, 15).

Discusión en clases

¿Cómo simularían una muestra del proceso, dado el valor de los parámetros?

Simulación

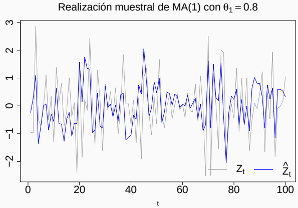
$$Z_t = a_t - \theta_1 a_{t-1},\ \ t = 0, \pm 1, \ldots,\ \ a_t \sim \mathcal{N}(0, \sigma^2)$$

```
simMA1 <- function(theta1, sigma, Z0, T) {
  TT <- 2 * T

  # Ruido
  at <- rnorm(TT, 0, sigma)

  # Observaciones
  Zt <- vector("numeric", TT)
  Zt[1] <- Z0
  for (t in 2:TT) {
    Zt[t] <- at[t] - theta1 * at[t - 1]
  }

  # Descartamos el primer 50%
  tail(Zt, T)
}
```

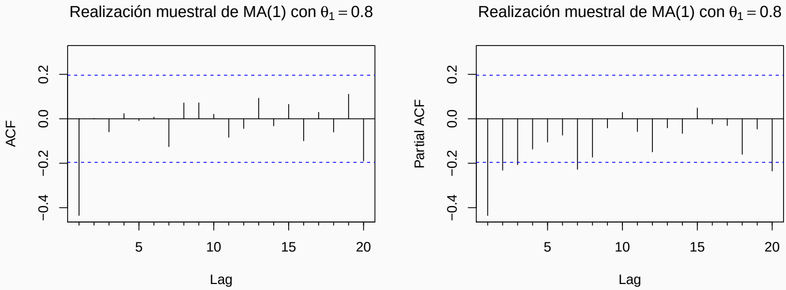
Ejemplo

$$Z_t = a_t - 0.8a_{t-1},\ \ t = 0, \pm 1, \ldots,\ \ a_t \sim \mathcal{N}(0,1)$$

```r
set.seed(9000)
z <- simMA1(0.8, 1, 0, 100)

library(forecast)
fit   <- Arima(
  z,
  order = c(0, 0, 1),
  include.mean = FALSE
  )
```



Estimaciones muestrales



Ejercicio en clases

Analíticamente, encontrar la función de autocorrelación y la función de autocorrelación parcial.

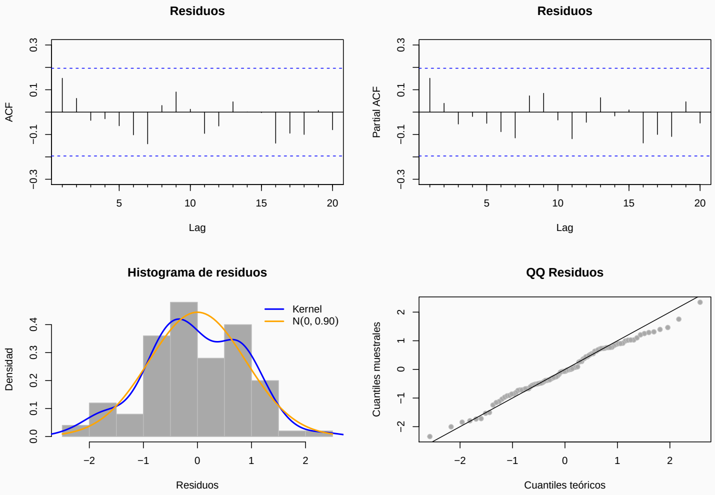
Modelo ajustado

```
## Series: z
## ARIMA(0,0,1) with zero mean
##
## Coefficients:
##           ma1
##       -0.8774
## s.e.   0.0805
##
## sigma^2 estimated as 0.8072:  log likelihood=-131.42
## AIC=266.83   AICc=266.96   BIC=272.04
```

Discusión en clases

¿Cuáles son todos los supuestos del modelo? ¿Cómo validarían cada uno de ellos?

Diagnóstico de residuos



Proceso promedio móvil de segundo orden

El siguiente es un porceso $MA(2)$:

$$Z_t = a_t - 1.2a_{t-1} - 0.5a_{t-2},\ t = 0, \pm 1, \dots,\ a_t \sim \mathcal{N}(0, 1)$$

Para divertirse en casa:

- Escribir código para simular un conjunto de datos.
- Emplear `Arima` para estimar el valor de los parámetros y corroborar contra los valores prefijados.
- Emplear `Acf` para estimar las funciones ACF y PACF y corroborar con los resultados analíticos.

Procesos ARMA

Proceso $ARMA(1,1)$

Proceso ARMA de orden $1,1$

$$Z_t = \phi_1 Z_{t-1} + \theta_1 a_{t-1} + a_t,\ \ t = 0, \pm 1, \ldots$$

con $a_t \sim \mathrm{WN}(0, \sigma^2)$ y $\phi_1 + \theta_1 \neq 0$ (Brockwell and Davis 2016, 48). El proceso es estacionario si y sólo si $\phi_1 \neq \pm 1$.

Alternativamente,

$$\begin{aligned} Z_t - \phi_1 Z_{t-1} &= \theta_1 a_{t-1} + a_t \\ \phi(B) Z_t &= \theta(B) a_t. \end{aligned}$$

Proceso $ARMA(p, q)$

Proceso ARMA de orden p, q

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + a_t,\ t = 0, \pm 1, \ldots$$

con $a_t \sim \mathrm{WN}(0, \sigma^2)$, y los polinomios característicos no tienen factores en común (Brockwell and Davis 2016, 74). El proceso es estacionario si y sólo si $1 - \phi_1 Z - \cdots - \phi_p Z^p \neq 0$.

Alternativamente,

$$\begin{aligned} Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} &= \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + a_t \\ \phi(B) Z_t &= \theta(B) a_t. \end{aligned}$$

Proceso autorregresivo $ARMA(2,1)$

El siguiente es un proceso $ARMA(2,1)$:

$$(1-1.4B+0.6B^2)Z_t = (1-0.8B)a_t,\ t=0,\pm 1,\ldots,\ a_t \sim \mathcal{N}(0,1).$$

Para divertirse en casa:

- Emplear `arima.sim` para simular un conjunto de datos.
- Emplear `Arima` para estimar el valor de los parámetros y corroborar contra los valores prefijados.
- Emplear `Acf` para estimar las funciones ACF y PACF y corroborar con los resultados analíticos.

En resumen

Teoría

Para los modelos AR, MA, y ARMA:

- Especificación.
- Restricciones de los parámetros.
- Funciones de autocovariancia, autocorrelación, y autocorrelación parcial.
- Relación dual AR y MA.

Tareas

Para los procesos AR(2), MA(2), y ARMA(2):

- Plantear el modelo.
- Explicitar (TODOS) los supuestos.
- Simular un conjunto de datos.
- Estimar estas cantidades en R.
- Encontrar analíticamente funciones de autocovariancia y autocorrelación.

Discusión en clases

¿Preguntas, dudas, inquietudes, ansiedades, sugerencias?

Anexo

Algunas propiedades útiles

Sean X e Y variables aleatorias; $k, a, b \in \mathbb{R}$ escalares constantes y finitos.

Esperanza.

$$\mu = \mathrm{E}\langle k \rangle = k \quad \mathrm{E}\langle kX \rangle = k\,\mathrm{E}\langle X \rangle \quad \mathrm{E}\langle X + Y \rangle = \mathrm{E}\langle X \rangle + \mathrm{E}\langle Y \rangle \quad \mathrm{E}\langle XY \rangle \neq \mathrm{E}\langle X \rangle\,\mathrm{E}\langle Y \rangle$$

Varianza.

$$\sigma = \mathrm{V}\langle X \rangle = \mathrm{E}\left\langle (X - \mathrm{E}\langle X \rangle)^2 \right\rangle = \mathrm{E}\left\langle X^2 \right\rangle - \mathrm{E}\langle X \rangle^2 \quad \mathrm{V}\langle X \rangle \geq 0$$

$$\mathrm{V}\langle X + k \rangle = \mathrm{V}\langle X \rangle \quad \mathrm{V}\langle kX \rangle = k^2\,\mathrm{V}\langle X \rangle \quad \mathrm{V}\langle aX \pm bY \rangle = a^2\,\mathrm{V}\langle X \rangle + b^2\,\mathrm{V}\langle Y \rangle \pm 2ab\,\mathrm{Cov}\langle X, Y \rangle$$

Covarianza. Como propiedad general, el operador de valor esperado no cumple con la propiedad multiplicativa. La diferencia está dada por la covariancia.

$$\gamma = \mathrm{Cov}\langle X, Y \rangle = \mathrm{E}\langle (X - \mathrm{E}\langle X \rangle)(Y - \mathrm{E}\langle Y \rangle) \rangle = \mathrm{E}\langle XY \rangle - \mathrm{E}\langle X \rangle\,\mathrm{E}\langle Y \rangle \quad \mathrm{Cov}\langle X, k \rangle = 0 \quad \mathrm{Cov}\langle X, X \rangle = \mathrm{V}\langle X \rangle$$

$$\mathrm{Cov}\langle X, Y \rangle = \mathrm{Cov}\langle Y, X \rangle \quad \mathrm{Cov}\langle aX + bY \rangle = ab\,\mathrm{Cov}\langle X, Y \rangle \quad \mathrm{Cov}\langle X + a, Y + b \rangle = \mathrm{Cov}\langle X, Y \rangle$$

Correlación.

$$\rho = \mathrm{Corr}\langle X, Y \rangle = \frac{\mathrm{Cov}\langle X, Y \rangle}{\sqrt{\mathrm{V}\langle X \rangle\,\mathrm{V}\langle Y \rangle}} \text{con } \mathrm{V}\langle X \rangle \neq 0 \wedge \mathrm{V}\langle Y \rangle \neq 0$$

Referencias

Brockwell, Peter J., and Richard A. Davis. 2016. *Introduction to Time Series and Forecasting*. Springer International Publishing. doi:10.1007/978-3-319-29854-2.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2014. *Bayesian Data Analysis*. Vol. 2. CRC press Boca Raton, FL.