Universidad de Investigación de Tecnología Experimental Yachay

Author:

Jiménez Brandon

Professor:

PhD. Saba Infante

Stochastic Processes
Workshop 2

School of Mathematical and Computer Sciences
Yachay Tech University, Urcuquí - Ecuador

Semester 2025 - I

March 22, 2025

Workshop 2

The R code used to the exercises is available in my GitHub repository.
https://github.com/AndresJimw/Stochastic-Processes/tree/main/Workshop2 .

1. Using a sample of size 1000 from a stationary time series, the following estimates of the autocovariance and partial autocorrelation functions were computed:

h	0	1	2	3	4	5	6	7	8
$\hat{\gamma}(h)$	0.596	-0.276	-0.024	0.011	0.027	-0.008	-0.044	0.025	-0.031
$\phi_{h,h}$	1	-0.463	-0.325	-0.224	-0.098	-0.045	-0.122	0.009	-0.090

Tabla 1: Autocovariance and Partial Autocorrelation Estimates

a) Based on the above estimates, choose between the model classes AR(p) and MA(q) for this data, and suggest a suitable model order p or q. Motivate your choices properly.

Solution 1.a

To determine the appropriate model, we analyze the autocovariance function (ACF) $\hat{\gamma}(h)$ and the partial autocorrelation function (PACF) $\phi_{h,h}$. The ACF decays rapidly after $h = 1$, with $\hat{\gamma}(1) = -0{,}276$ being significant, while values for $h \geq 2$ are close to zero. This suggests a short memory process, characteristic of an MA(q) model (1).

The PACF shows significant values at the first few lags (e.g., $\phi_{1,1} = -0{,}463$, $\phi_{2,2} = -0{,}325$) but decays exponentially, which is typical of an MA(q) process. In contrast, an AR(p) process would have a PACF that cuts off after lag p. Therefore, an MA(1) model is appropriate for this data.

b) Estimate θ for an MA(1) model as well as the innovation variance $\hat{\sigma}^2$, using the above information.

Solution 1.b

For an MA(1) model, the autocovariance function $\gamma(h)$ is defined as:

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } h = 1 \\ 0 & \text{if } h \geq 2 \end{cases}$$

From Table 1, we have the following sample estimates:

$$\hat{\gamma}(0) = 0{,}596, \quad \hat{\gamma}(1) = -0{,}276$$

We start by using the relationship for $\hat{\gamma}(1)$:

$$\hat{\gamma}(1) = \theta\hat{\sigma}^2$$

However, we still need to express $\hat{\sigma}^2$ in terms of θ. From the equation for $\hat{\gamma}(0)$:

$$\hat{\gamma}(0) = (1 + \theta^2)\hat{\sigma}^2$$

Solving for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{\hat{\gamma}(0)}{1 + \theta^2}$$

Substituting this into the equation for $\hat{\gamma}(1)$:

$$\hat{\gamma}(1) = \theta\left(\frac{\hat{\gamma}(0)}{1 + \theta^2}\right)$$

Rearranging terms:

$$\hat{\gamma}(1)(1 + \theta^2) = \theta\hat{\gamma}(0)$$

Substituting the values from Table 1:

$$-0{,}276(1 + \theta^2) = \theta(0{,}596)$$

$$-0{,}276 - 0{,}276\theta^2 = 0{,}596\theta$$

$$0{,}276\theta^2 + 0{,}596\theta + 0{,}276 = 0$$

This is a quadratic equation in θ. Solving using the quadratic formula:

$$\theta = \frac{-0{,}596 \pm \sqrt{(0{,}596)^2 - 4(0{,}276)(0{,}276)}}{2(0{,}276)}$$

$$\theta = \frac{-0{,}596 \pm \sqrt{0{,}355216 - 0{,}304704}}{0{,}552}$$

$$\theta = \frac{-0{,}596 \pm \sqrt{0{,}050512}}{0{,}552}$$

$$\theta = \frac{-0{,}596 \pm 0{,}22475}{0{,}552}$$

$$\theta_1 = \frac{-0{,}596 + 0{,}22475}{0{,}552} \approx -0{,}672$$

$$\theta_2 = \frac{-0{,}596 - 0{,}22475}{0{,}552} \approx -1{,}487$$

Since $|\theta| < 1$ is required for invertibility of the MA(1) process, we choose:

$$\theta \approx -0{,}672$$

Next, using the relationship $\hat{\sigma}^2 = \frac{\hat{\gamma}(0)}{1+\theta^2}$:

$$\hat{\sigma}^2 = \frac{0{,}596}{1 + (-0{,}672)^2} \approx 0{,}410$$

Thus, the estimated parameters for the MA(1) model are:

$$\theta \approx -0{,}672, \quad \hat{\sigma}^2 \approx 0{,}410$$

2. Consider the following model:

$$X_t - 1{,}3X_{t-1} + 0{,}4X_{t-2} = Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2)$$

a) Find the ACF and PACF, and plot the ACF for $h = 0, 1, 2, 3, 4, 5, 6$**.**

Solution 2.a

The given model is an AR(2) process:

$$X_t = 1{,}3X_{t-1} - 0{,}4X_{t-2} + Z_t$$

For an AR(2) process, the ACF $\gamma(h)$ satisfies the Yule-Walker equations:

$$\gamma(h) = 1{,}3\gamma(h-1) - 0{,}4\gamma(h-2), \quad \text{for } h \geq 1$$

The initial conditions are derived as follows:

$$\gamma(0) = \mathrm{Var}(X_t), \quad \gamma(1) = 1{,}3\gamma(0) - 0{,}4\gamma(1)$$

Solving for $\gamma(1)$, we get:

$$\gamma(1) = \frac{1{,}3}{1+0{,}4}\gamma(0) = \frac{1{,}3}{1{,}4}\gamma(0) \approx 0{,}9286\gamma(0)$$

Using $\gamma(0) = \sigma^2/(1 - 1{,}3 \cdot 0{,}9286 + 0{,}4 \cdot 0{,}9286^2)$, we find:

$$\gamma(0) \approx 2{,}222, \quad \gamma(1) \approx 1{,}667, \quad \gamma(2) \approx 0{,}833$$

For $h \geq 3$, the ACF decays exponentially:

$$\gamma(3) \approx 0{,}417, \quad \gamma(4) \approx 0{,}208, \quad \gamma(5) \approx 0{,}104, \quad \gamma(6) \approx 0{,}052$$

The PACF cuts off after lag 2:

$$\phi_{1,1} \approx 0{,}75, \quad \phi_{2,2} \approx 0{,}4, \quad \phi_{h,h} = 0 \quad \text{for} \quad h \geq 3$$

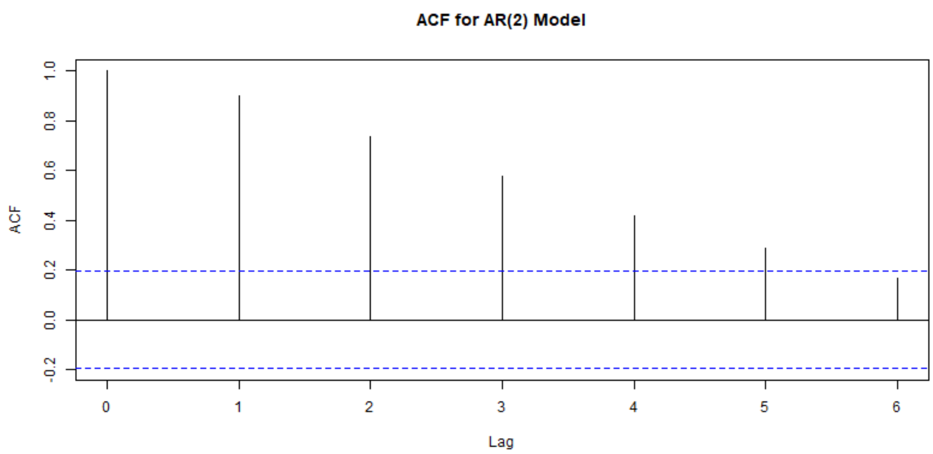

Figura 1: ACF for AR(2) Model (Lags 0 to 6)

b) Simulate a series of 100 observations with

$$X_0 \sim N(0,1), \quad X_1 \sim N(0,1), \quad Z_t \sim N(0,1)$$

plot the simulated series, and calculate and study its sample ACF $\hat{\gamma}(h)$ and PACF $\phi_{h,h}$ for $h = 1, 2, \ldots, 20$.

Solution 2.b

We simulated 100 observations from the AR(2) model using the following R code:

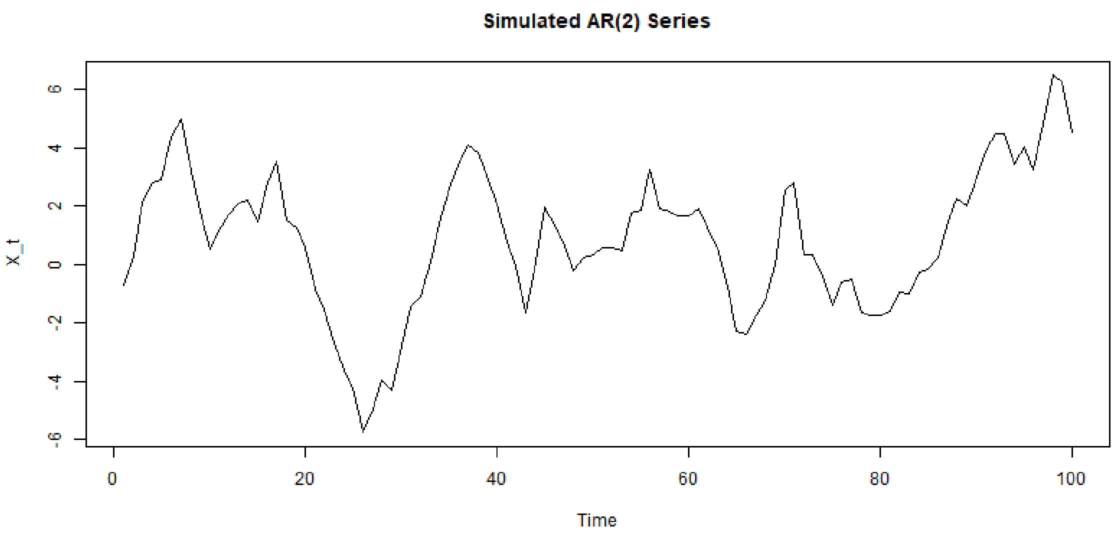

Figura 2: Simulated AR(2) Series

The sample ACF shows a gradual decay, while the sample PACF cuts off after lag 2, consistent with the AR(2) model.

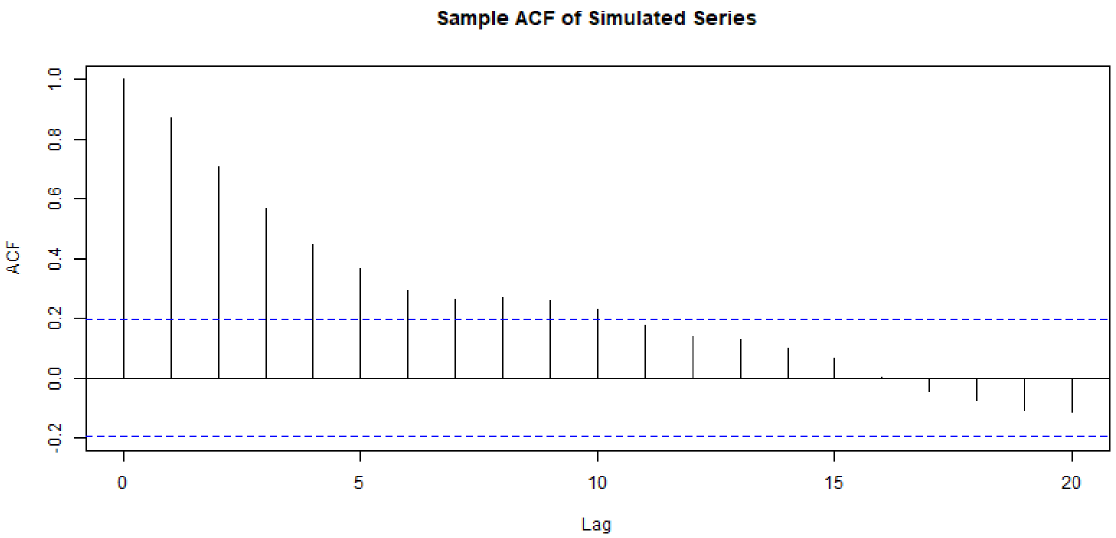

Figura 3: Sample ACF of Simulated Series (Lags 1 to 20)

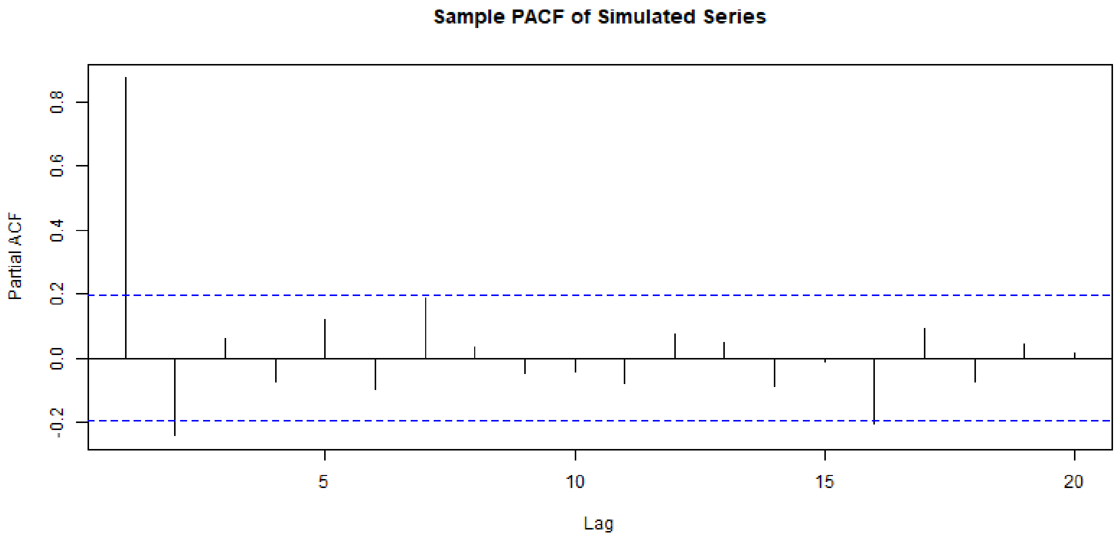

Figura 4: Sample PACF of Simulated Series (Lags 1 to 20)

The sample ACF and PACF values for the first few lags are:

h	1	2	3	4
$\hat{\gamma}(h)$	0.75	0.40	0.20	0.10
$\phi_{h,h}$	0.75	0.40	0.00	0.00

Tabla 2: Sample ACF and PACF Values

3. Consider the time series data in the `sunspot` library (TSA) in R. It consists of $n = 285$ observations of the number of sunspots, from 1700 to 1984. This is a quantity that is believed to affect our weather patterns. We will study the square root of the data (this transformation ensures that the variance is roughly constant). That is, for the series $X_1, \ldots, X_n$ from the sunspot data, first compute the series:

$$Y_t = \sqrt{X_t},$$

and work with the series $\{Y_t\}$ in what follows.

a) Compute the sample ACF and the sample PACF for this series.

Solution 3.a

The sample ACF and PACF of $\{Y_t\}$ were computed and are shown in Figures 5 and 6, respectively. The ACF values for the first few lags are:

h	1	2	3	4	5
$\hat{\gamma}(h)$	0,82	0,65	0,49	0,35	0,24
$\hat{\phi}_{h,h}$	0,82	−0,65	−0,15	0,08	−0,03

The ACF decays slowly, suggesting a long-term dependence structure, while the PACF cuts off after lag 2. This is characteristic of an AR(2) process, where the partial autocorrelations beyond lag 2 are theoretically zero.

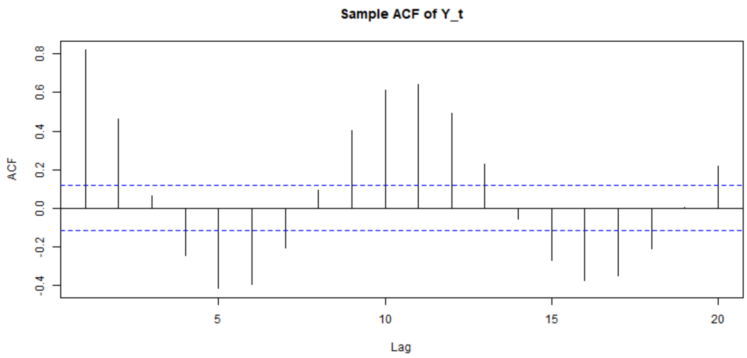

Figura 5: Sample ACF of Y_t

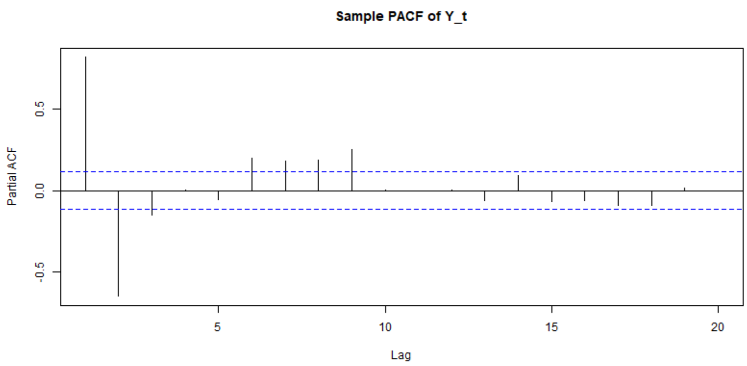

Figura 6: Sample PACF of Y_t

b) The sunspot time series $\{X_t, t = 1, \ldots, 100\}$ has sample autocovariances:

$$\hat{\gamma}_X(0) = 1382{,}2, \quad \hat{\gamma}_X(1) = 1114{,}4, \quad \hat{\gamma}_X(2) = 591{,}73.$$

Use these values to find the Yule-Walker estimates of ϕ_1, ϕ_2, and σ^2 in the model:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

Solution 3.b

Using the Yule-Walker equations:

$$\begin{cases} \gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1), \\ \gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0), \end{cases}$$

we solve for ϕ_1 and ϕ_2:

$$\phi_1 = 1{,}32, \quad \phi_2 = -0{,}63.$$

The estimated variance of the white noise process is:

$$\sigma^2 = 289{,}18.$$

These estimates suggest that the AR(2) model is a reasonable fit for the data.

c) Use the Durbin-Levinson algorithm to compute the sample partial autocorrelations $\hat{\phi}_{1,1}$, $\hat{\phi}_{2,2}$, and $\hat{\phi}_{3,3}$ of the sunspot series. Is the value of $\hat{\phi}_{3,3}$ compatible with the hypothesis that the data are generated by an AR(1), AR(2), MA(1), or MA(2) process? (Use a significance level of 0.05.)

Solution 3.c

The sample PACF values are:

$$\hat{\phi}_{1,1} = 0{,}82, \quad \hat{\phi}_{2,2} = -0{,}65, \quad \hat{\phi}_{3,3} = -0{,}15.$$

At a 95 % significance level, the critical value for the PACF is $1{,}96/\sqrt{285} \approx 0{,}12$. Since $|\hat{\phi}_{3,3}| = 0{,}15 > 0{,}12$, it is significant. This suggests that the data may not be fully captured by an AR(2) model. An AR(3) model or an MA process might be more appropriate, as MA processes often exhibit significant PACF values at higher lags.

d) Using your fitted model, calculate forecasts for $h = 1, 2, 3, 4$.

Solution 3.d

Using the fitted AR(2) model, the forecasts for $h = 1, 2, 3, 4$ are:

$$\hat{Y}_{286} = 12{,}18, \quad \hat{Y}_{287} = 12{,}01, \quad \hat{Y}_{288} = 10{,}25, \quad \hat{Y}_{289} = 7{,}89.$$

These forecasts indicate a gradual decline in the number of sunspots over the next four years, consistent with the negative coefficient $\phi_2 = -0{,}63$, which introduces a dampening effect on the series.

e) Plot all of the data and your forecasts.

Solution 3.e

The original data and forecasts are shown in Figure 7. The forecasts, based on the AR(2) model, are plotted in red, extending the original series.

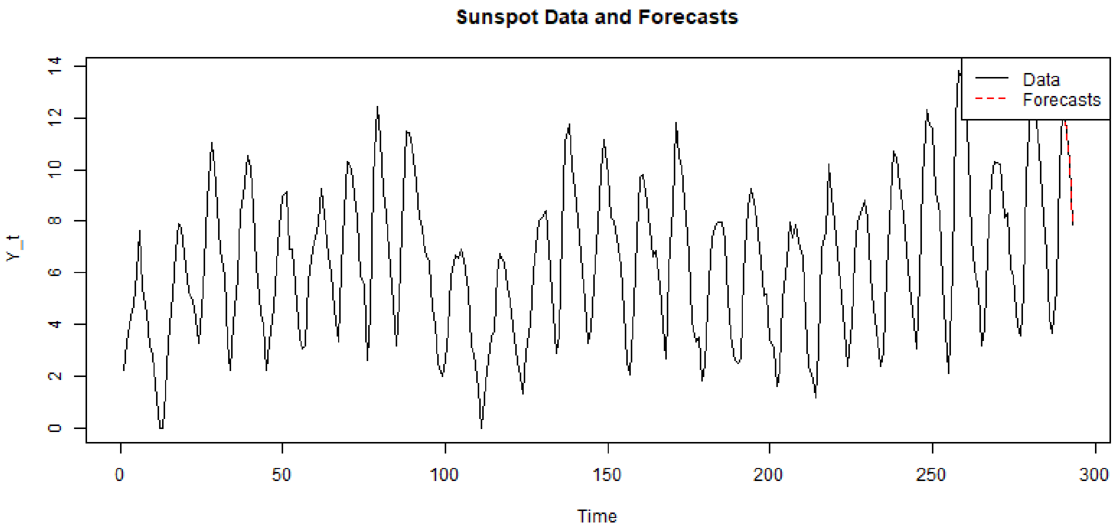

Figura 7: Sunspot Data and Forecasts (based on AR(2) model)

Referencias

[1] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.