



Procesos estocásticos irregularmente espaciados

Wilmar Sepulveda Herrera

Universidad del valle
Facultad de Ingeniería, Escuela de Estadística
Cali, Valle del Cauca
2022

Procesos estocásticos irregularmente espaciados

by

Wilmar Sepulveda Herrera

Trabajo de investigación para optar por el título de:

Magister en

En

Estadística

Aceptamos esta tesis
Conforme a los requerimientos de la norma

Universidad del valle
Facultad de Ingeniería, Escuela de Estadística
Cali, Valle del Cauca
2022

Dedicatoria

A mi familia.

Agradecimientos

Estoy agradecido con...

Abstract

El resumen es una presentación abreviada y precisa. Se debe usar una extensión máxima de 12 renglones. Se recomienda que este resumen sea analítico, es decir, que sea completo, con información cuantitativa y cualitativa, generalmente incluyendo los siguientes aspectos: objetivos, diseño, lugar y circunstancias (u objetivo del estudio), principales resultados, y conclusiones. Al final del resumen se deben usar palabras claves tomadas del texto (mínimo 3 y máximo 7 palabras), las cuales permiten la recuperación de la información.

Palabras clave: palabras clave en español (máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado)

Resumen

Es el mismo resumen pero traducido al inglés. Se debe usar una extensión máxima de 12 renglones. Al final del Abstract se deben traducir las anteriores palabras clave tomadas del texto (mínimo 3 y máximo 7 palabras), llamadas keywords. Es posible incluir el resumen en otro idioma diferente al español o al inglés, si se considera como importante dentro del tema tratado en la investigación, por ejemplo: un trabajo dedicado a problemas lingüísticos del mandarín seguramente estaría mejor con un resumen en mandarín.

Keywords: palabras clave en inglés (máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado)

Contents

Abstract	vii
List of figures	xi
List of tables	xii
1 Introducción	1
1.1 Objetivos	2
1.1.1 Objetivo general	2
1.1.2 Objetivos específicos	2
2 Marco teórico	3
2.1 Procesos estocásticos irregularmente espaciados	3
2.2 Formas de describir un proceso estocástico	4
2.3 El proceso de medias móviles irregular de primer orden (IMA)	4
2.4 El proceso Autorregresivo de primer orden irregular (IAR)	6
2.4.1 Estimación	7
2.5 Estimación Bootstrap en Procesos Irregulares	8
3 Modelo autorregresivo de primer orden irregularmente espaciado	9
3.1 Construcción del modelo	9
3.2 Implementando propuesta de ϕ_n	11
3.3 El modelo autorregresivo irregularmente espaciado de primer orden	11
3.4 Propiedades del proceso	12
3.5 Predicción	12
3.6 Estimación por máxima verosimilitud	12
3.7 Estimación Bootstrap	13
3.8 Estudio de Simulación Monte Carlo	14

4 Conclusion	19
A Appendix	20
References	21

List of Figures

3-1	simulación del modelo IAR con coeficientes positivos	14
3-2	simulación del modelo IAR con coeficientes negativos	15
3-3	Esquema general del estudio de Monte Carlo.	16
3-4	Esquema general del estudio de Monte Carlo.	17
3-5	Esquema general del estudio de Monte Carlo.	17

List of Tables

3-1 Resultados del estudio de Monte Carlo 18

Chapter 1

Introducción

Los procesos estocásticos, particularmente el análisis de series de tiempo, hacen parte de las metodologías estadísticas más utilizadas en la actualidad, ya sea para automatizar procesos de Machine Learning enfocados al pronóstico de valores futuros o para conocer las características propias de un fenómeno y su estructura de autocorrelación. Dicha función de autocorrelación puede ser positiva, con importancia en muchos contextos como la demografía (Novoseltseva et al. 2019, p.ej.) o la explicación de fenómenos ecológicos (Yang et al. 2019, p.ej.), entre otras. Las autocorrelaciones negativas también tienen mucha importancia práctica, por ejemplo; en el sector financiero para conocer el comportamiento de los rendimientos de acciones (Kuttu & Bokpin 2017, p.ej.) y en la biología (Rindorf et al. 2020, p.ej.). Bajo este contexto, los procesos estocásticos son importantes en diversos campos del conocimiento para darnos a conocer las características de la dependencia temporal de un evento de interés y con ella tomar decisiones importantes.

La teoría de series temporales ha proporcionado diversos métodos para modelar la autocorrelación de una secuencia de interés, entre ellos, los modelos autorregresivos integrados de medias móviles (ARIMA), sin embargo, estos métodos asumen que las observaciones son tomadas regularmente en el tiempo, supuesto que en muchos casos no es válido, pues la presencia de valores faltantes o la misma naturaleza de las observaciones, pueden inducir a que la serie de tiempo sea irregularmente espaciada (Elorrieta et al. 2019). Es común encontrar series de tiempo irregularmente espaciadas en diversas aplicaciones, para diferentes áreas del conocimiento: sistemas caóticos, medicina, imágenes satelitales y calidad del aire; ver por ejemplo Shamsan et al. (2020), Liu et al. (2019), Ghaderpour & Vujadinovic (2020) y Dilmaghani et al. (2007), entre otros.

En la actualidad, la literatura proporciona varios métodos para modelar series de tiempo

irregularmente espaciadas, algunos autores como Adorf (1995) transforman las series de tiempo irregulares en regulares vía interpolación, método que puede generar sesgos; Robinson (1977) aborda el problema tratando la serie de tiempo como una realización discreta de un proceso estocástico de tiempo continuo, aproximación que también presenta problemas. Recientemente, Eyheramendy et al. (2018) proponen el Modelo Autorregresivo Irregular (*IAR*) de primer orden, el cual tiene la limitación no de aceptar valores de negativos en su función de autocorrelación (ACF); este problema fue resuelto por los autores en un estudio posterior, en el cual proponen el modelo Autorregresivo Complejo Irregular (*CIAR*). (Elorrieta et al. 2019)

En la misma línea ojedá:inpress-a proponen el modelo de medias móviles irregular (*IMA*), el cual bajo normalidad, es estrictamente estacionario, mientras que, teniendo en cuenta otros supuestos distribucionales, se comporta cómo un proceso débilmente estacionario; en ambos casos, los autores evalúan las propiedades estadísticas de los modelos y proponen métodos de estimación, encontrando muy buenos resultados. Sin embargo, el modelo IMA tiene una desventaja, en términos de que su ACF es únicamente positiva, siendo incapaz de detectar autocorrelaciones negativas y dejándolo muy limitado debido a la importancia práctica que estas suponen; problema que no tienen los modelos de medias móviles clásicos. El propósito de este trabajo es modificar el modelo IMA de modo que permita modelar estructuras de autocorrelación positivas y negativas sin necesidad de irse al plano de los números complejos.

1.1 Objetivos

1.1.1 Objetivo general

Proponer un modelo autorregresivo irregular de primer orden que permita modelar estructuras de autocorrelación positivas y negativas subyacentes de un proceso estocástico.

1.1.2 Objetivos específicos

- Encontrar estimadores para el modelo propuesto.
- Evaluar las propiedades del modelo propuesto y sus estimaciones vía simulación.
- Ajustar el modelo propuesto a un conjunto de datos para valorar su comportamiento en ejercicios de aplicación.

Chapter 2

Marco teórico

Para el tratamiento de procesos estocásticos y series de tiempo, ha sido mucha la teoría estadística implementada y se ha convertido en una herramienta importante para tratar observaciones dependientes y entender la naturaleza de dicha dependencia, la mayoría de estos métodos, asumen que las series de tiempo son regularmente espaciadas, sin embargo este escenario no siempre se da y ahí es donde cobran importancia los procesos estocásticos irregularmente espaciados.

Este capítulo tiene la siguiente estructura, en la sección 2.1 presentamos la definición matemática de un proceso estocástico irregularmente espaciado. En la sección 2.2 se presenta un proceso estocástico autorregresivo irregularmente espaciado. En la sección 2.3 se presenta un proceso estocástico irregularmente espaciado de medias móviles y en la sección 2.3 se presenta un proceso estocástico autorregresivo de medias móviles irregularmente espaciado.

2.1 Procesos estocásticos irregularmente espaciados

Sea (Ω, \mathcal{B}, P) un espacio de probabilidad. Si se define un proceso estocástico como una medida que mapea $x : \Omega \mapsto \mathbb{R}^{\mathbb{T}}$, donde

$$x(\omega) = \{X_\tau(\omega), \tau \in \mathbb{T}\}$$

\mathbb{T} es llamado conjunto de índices y la variable aleatoria $X_\tau(\omega)$ es llamada coordenada del proceso o trayectoria. Consideremos ahora $\mathbb{T}' = \{t_1, t_2, t_3, \dots\}$, con $\Delta_{n+1} = t_{n+1} - t_n$, para $n \geq 1$

$$x' = \{X_\tau(\omega), \tau \in \mathbb{T}'\}$$

x' es un proceso estocástico irregularmente espaciado. Una serie de tiempo irregularmente espaciada es una realización finita de un proceso estocástico irregularmente espaciado. Note que si $\Delta_{n+1} = t_{n+1} - t_n = 1$ para $n \geq 1$, el proceso x' es un proceso regularmente espaciado, por tanto esta definición es más general.

2.2 Formas de describir un proceso estocástico

Una forma de describir un proceso estocástico x , es especificar la función de distribución conjunta de $\{X_{\tau_1}, X_{\tau_2}, \dots, X_{\tau_n}\}$ para todo n , este es llamado *punto de vista distribucional*. Por otra parte, se puede describir el proceso proporcionando una formula para el valor X_τ para cada punto τ en términos de una familia de variables aleatorias con comportamiento probabilístico conocido, esto hace que podamos ver el proceso como una función de otros procesos (o cómo familias de procesos iid); esta forma es llamada *punto de vista constructorista*.

2.3 El proceso de medias móviles irregular de primer orden (IMA)

Teniendo en cuenta el conjunto $\mathbb{T}' = \{t_1, t_2, t_3, \dots\}$ propuesto anteriormente, tal que sus diferencias Δ_{n+1} para $n \geq 1$ están acotados uniformemente lejos de cero. Ahora, sea $m : \mathbb{T}' \mapsto \mathbb{R}$ una función tal que $m(t_n) = 0$, para cualquier $t_n \in \mathbb{T}'$. A continuación, sea $\Gamma : \mathbb{T}' \times \mathbb{T}' \mapsto \mathbb{R}$ una función tal que para cualquier pareja $t_n, t_s \in \mathbb{T}'$,

$$\Gamma(t_n, t_s) = \begin{cases} \gamma_0 & |n - s| = 0 \\ \gamma_1, \Delta_{\max(n,s)} & |n - s| = 1 \\ 0 & |n - s| \geq 1 \end{cases}$$

Note que Γ puede ser representada cómo una matriz diagonal así

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1, \Delta_2 & 0 & 0 & \dots \\ \gamma_1, \Delta_2 & \gamma_0 & \gamma_1, \Delta_3 & 0 & \\ 0 & \gamma_1, \Delta_3 & \gamma_0 & \gamma_1, \Delta_4 & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \cdot & & & & & \cdot \end{bmatrix} \quad (2-1)$$

2.3. EL PROCESO DE MEDIAS MÓVILES IRREGULAR DE PRIMER ORDEN (IMA) 5

Ser Γ_n la truncación $n \times n$ de Γ y asumiendo $\gamma_1, \Delta_j \neq 0$, para $j = 2, \dots, n$, Γ_n es definida positiva si γ_0 y $(\frac{\gamma_1, \Delta_{n+1}}{\gamma_0})^2 \leq 1/4$, para $j = 2, \dots, n$

Esto implica que existe un proceso Gaussiano estacionario $\{X_{t_n}, t_n, \tau \in \mathbb{T}\}$, único hasta la equivalencia, con media 0 y covarianza Γ . Este proceso es llamado, proceso de medias móviles de primer orden irregularmente espaciado de forma general. A continuación se darán las expresiones particulares de este proceso desde los dos puntos de vista antes definidos.

El punto de vista distribucional

En (2-1), γ_0 y γ_1, Δ_{n+1} , para $n \geq 1$, representa la varianzas y las covarianzas de primer orden respectivamente. Definimos la varianzas como $\gamma_0 = \sigma^2(1 + \theta^2)$ y las covarianzas de primer orden como $\gamma_1, \Delta_{n+1} = \sigma^2 \theta^{\Delta_{n+1}}$, $\sigma^2 > 0$ y $0 < \theta < 1$. Por tanto, obtenemos el proceso estocástico irregularmente espaciado de primer orden con matriz de covarianzas

$$\Gamma = \begin{bmatrix} 1 + \theta^2 & \theta^{\Delta_2} & 0 & 0 & \dots \\ \theta^{\Delta_2} & 1 + \theta^2 & \theta^{\Delta_3} & 0 & \\ 0 & \theta^{\Delta_3} & 1 + \theta^2 & \theta^{\Delta_4} & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \end{bmatrix} \quad (2-2)$$

el cual contiene el modelo de medias móviles convencional como caso especial. Este es llamado proceso Gaussiano irregular de medias móviles de primer orden.

El punto de vista constructor

Ahora, como es usual, especificaremos el proceso IMA como función de otros procesos estocásticos. Sea $\{\epsilon_{t_n}\}_{n \geq 1}$ variables aleatorias independientes que siguen una distribución normal $N(0, \sigma^2 c_n(\theta))$ con $\sigma^2 > 0, 0 < \theta < 1, c_1(\theta) = 1 + \theta^2$ y

$$c_n(\theta) = 1 + \theta^2 - \frac{\theta^{2\Delta_n}}{c_{n-1}(\theta)} \text{ para } n \geq 2$$

donde $\Delta_n = t_n - t_{n-1}$. El proceso $\{X_{t_n}, t_n \in \mathbb{T}'\}$, es decirse tiene un proceso IMA si $X_{t_1} = \epsilon_{t_1}$ y para $n \geq 2$

$$X_{t_n} = \epsilon_{t_n} + \frac{\theta^{\Delta_n}}{c_{n-1}(\theta)} \epsilon_{t_{n-1}} \quad (2-3)$$

Decimos que $\{X_{t_n}, t_n \in \mathbb{T}'\}$ es un proceso IMA con media μ si $\{X_{t_n} - \mu, t_n \in \mathbb{T}'\}$ es un proceso IMA

2.4 El proceso Autorregresivo de primer orden irregular (IAR)

Se denota X_{t_n} a una observación medida en el tiempo t_n y consideramos que $\{t_n\}$ para $n = 1, \dots, n$, se define el proceso IAR de primer orden así:

$$X_{t_n} = \phi^{t_n - t_{n-1}} X_{t_{n-1}} + \sigma \sqrt{1 - \phi^{2(t_n - t_{n-1})}} \varepsilon_{t_n} \quad (2-4)$$

Donde ε_{t_n} son variables aleatorias independientes con media cero y varianza unitaria, observe que $E(X_{t_n}) = 0$ y $Var(X_{t_n}) = \sigma^2$ para todo X_{t_n} .

Así, la covarianza entre X_{t_k} y X_{t_j} es $E(X_{t_k}, X_{t_j}) = \sigma^2 \phi^{t_k - t_j}$ para $k \geq j$. La matriz de varianzas y autocovarianzas se muestra en 2-5.

$$\begin{bmatrix} 1 & \phi^{t_2 - t_1} & \phi^{t_3 - t_1} & \cdot & \cdot & \cdot & \cdot & \phi^{t_n - t_1} \\ & 1 & \phi^{t_3 - t_2} & \phi^{t_4 - t_2} & \cdot & \cdot & \cdot & \phi^{t_n - t_2} \\ & & 1 & \phi^{t_4 - t_3} & \phi^{t_5 - t_3} & \cdot & \cdot & \phi^{t_n - t_3} \\ & & & 1 & \cdot & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot & \cdot \\ & & & & & \cdot & \cdot & \cdot \\ & & & & & & \phi^{t_n - t_{n-1}} & \\ & & & & & & & 1 \end{bmatrix} \quad (2-5)$$

Así, para cualquier par de tiempos observacionales $s < t$, podemos definir la función autocovarianza cómo:

$$\gamma(t - s) = E(X_t, X_s) = \sigma^2 \phi^{t-s}$$

También, la función de autocorrelación (ACF) está dada por $\rho(t - s) = \frac{\gamma(t-s)}{\gamma(0)} = \phi^{t-s}$

La secuencia X_{t_n} corresponde a un proceso debilmente estacionario de segundo orden, también se puede demostrar que bajo algunas condiciones el proceso es estacionario y

ergódico. Ver (Eyheramendy et al. 2018).

Teorema 2.4.1. *Considere el proceso definido por la ecuación 2-4 y supongamos que el ruido de entrada es una secuencia iid de variables aleatorias con media cero y varianza unitaria. Además, suponga que $t_j - t_{j-n} \geq C \log n$ como $n \rightarrow \infty$, $0 \leq \phi \leq 1$, donde C es una constante positiva que satisface $C \log \phi^2 \geq -1$. Luego, existe una solución al proceso definido por la ecuación 2-4 y la secuencia $\{X_{t_j}\}$ es estacionaria y ergódica.*

Note también que si $t_n - t_{n-1} = 1$ para todo n , la ecuación 2-4 se convierte en:

$$X_{t_n} = \phi X_{t_{n-1}} + \sigma \sqrt{1 - \phi^2} \varepsilon_{t_n} \quad (2-6)$$

La ecuación 2-6 corresponde al modelo autorregresivo de primer orden para datos regularmente espaciados AR(1). Por lo tanto, el modelo IAR es una extensión del modelo autorregresivo regular

2.4.1 Estimación

La verosimilitud de una muestra aleatoria $X_{t_1} \dots X_{t_n}$ puede ser expresada como

$$f(X_{t_1} \dots X_{t_n}; \theta) = f(X_{t_1}; \theta) * f(X_{t_2} | X_{t_1}; \theta) * \dots * f(X_{t_n} | X_{t_{n-1}}; \theta)$$

donde $\theta = (\sigma^2, \phi)$ es el vector de parámetros del modelo, para poder describir el proceso de estimación, se asume que las distribuciones condicionales son Gaussianas, es decir, se asume que

$$f(X_{t_1}) \sim N(0, \sigma^2) \text{ y } f(X_{t_n} | X_{t_{n-1}}; \theta) \sim N(\phi^{t_j - t_{j-1}} X_{t_{j-1}}, \sigma^2(1 - \phi^{2(t_j - t_{j-1})})) \text{ para } j = 1, 2, 3, \dots, n$$

basado en la ecuación 2-4, la log-verosimilitud del proceso puede ser escrita como

$$l(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{j=1}^n \log v_{t_j} + \frac{1}{2} \sum_{j=1}^n \frac{e_{t_j}}{v_{t_j}} \quad (2-7)$$

se define $e_{t_1} = X_{t_1}$, $e_{t_j} = X_{t_j} - \phi^{t_j - t_{j-1}} X_{t_{j-1}}$ para $j > 1$ y sus varianzas $v_{t_1} = \text{Var}(e_{t_1})$.

Observe que el predictor pasado de tiempo finito del proceso en el tiempo t_j es dado por $\hat{X}_{t_1} = 0$, y $\hat{X}_{t_j} = \phi^{t_j - t_{j-1}} X_{t_{j-1}}$, para $j = 2, \dots, n$

Por tanto, $e_{t_j} = X_{t_j} - \hat{X}_{t_j}$ es el error de predicción con varianza $v_{t_1} = \text{Var}(e_{t_1} = \sigma^2)$ y $v_{t_j} = \text{var}(e_{t_j}) = \sigma^2[1 - \phi^{2(t_j - t_{j-1})}]$, para $j = 2, \dots, n$

Por maximización directa de la función de log-verosimilitud (2-7), se puede obtener el estimador de máxima verosimilitud de σ^2

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n \frac{(X_{t_j} - \hat{X}_{t_j})}{\tau_{t_j}}, \text{ donde } \tau_{t_j} = v_{t_j}/\sigma^2. \quad (2-8)$$

Pero no es posible encontrar $\hat{\phi}$, el estimador por máxima verosimilitud de ϕ , por maximización directa de la verosimilitud, es necesario recurrir a metodos iterativos.

2.5 Estimación Bootstrap en Procesos Irregulares

Para llevar a cabo la inferencia estadística es necesario poder derivar las distribuciones de los estadísticos utilizados para la estimación de los parámetros a partir de los datos. Si N es pequeño, o si los parámetros están cerca de los límites, las aproximaciones asintóticas pueden ser bastante pobres (Shumway et al. 2017). Además, en el caso de tiempo irregularmente espaciado, las aproximaciones asintóticas necesitan establecer condiciones fuertes que son difíciles de cumplir. Para superar estas dificultades y poder obtener aproximaciones de las distribuciones de muestras finitas, podríamos utilizar el método bootstrap. La idea en series temporales es ajustar un modelo adecuado a los datos, construir los residuos del modelo ajustado y, a continuación, generar nuevas series incorporando muestras aleatorias de los residuos al modelo ajustado. Los residuos suelen ser recientes para que tengan la misma media que las innovaciones del modelo.

Siguiendo a Bose (1990), para $n \geq 2$, define las innovaciones estimadas o residuales de la predicción un paso adelante, como

$$e_{\tau_n} = \sum_{j=0}^{n-1} (-1)^j \frac{\prod_{k=n-j+1}^n \theta^{\Delta_k}}{\prod_{l=n-j}^{n-1} c_l(\theta)} X_{\tau_{n-j}}$$

y $e_{\tau_1} = X_{\tau_1}$. Utilizando la estructura del proceso y asumiendo que el modelo ajustado es, de hecho, el modelo verdadero para los datos tenemos:

$$e_{\tau_n} = \varepsilon_{\tau_n} + (-1)^{n-1} \frac{\prod_{k=1}^n \theta^{\Delta_k}}{\prod_{l=0}^{n-1} c_l(\theta)} \varepsilon_{t_0}.$$

Por lo tanto, e_{τ_n} y ε_{τ_n} son cercanos para todo n grande si $|\theta| < 1$, lo que demuestra que el remuestreo es adecuado en esta situación.

Chapter 3

Modelo autorregresivo de primer orden irregularmente espaciado

En este capítulo, proponemos un modelo basado en la propuesta de Eyheramendy et al. (2018), la cual permite modelar unicamente estructuras de autocorrelacion positivas, el llamado $IAR(1)$. Nosotros generalizamos esta propuesta para que se permitan estructuras tanto positivas como negativas, teniendo en cuenta que nuestra novedosa propuesta, debe tener como caso particular el modelo $IAR(1)$ y el clasico modelo $AR(1)$

3.1 Construcción del modelo

En esta sección, construiremos un proceso estocástico estacionario que tiene una estructura autorregresiva y que además considera espacios de tiempo irregulares. Suponemos que el comportamiento irregularmente espaciado es independiente de las propiedades estocásticas del proceso.

Sea $\{\varepsilon_\tau, \tau \in \mathbb{T}\}$ una secuencia iid de variables aleatorias cada una con distribución $N(0, 1)$ y defino:

$$\begin{aligned} X_{\tau_1} &= W_1^{1/2} \varepsilon_{\tau_1} \\ X_{\tau_{n+1}} &= \phi_{n+1} X_{\tau_n} + W_{n+1}^{1/2} \varepsilon_{\tau_{n+1}} \end{aligned} \tag{3-1}$$

Donde, $0 \leq \phi \leq 1$, $E(X_{\tau_n}) = 0$ y $\{W_n\}_{n \geq 1}$ es una secuencia que caracteriza los momentos del proceso, ϕ_n es una funcion de Δ_n y del parametro ϕ del modelo autorregresivo que definiremos mas adelante. así, para $n \geq 1$, tenemos $E(X_{\tau_n}) = 0$,

$$V(X_{\tau_1}) = W_1$$

$$V(X_{\tau_{n+1}}) = \phi_{n+1}^2 V(X_{\tau_n}) + W_{n+1}$$

Ahora, bajo este modelo, calculemos $Cov(X_{\tau_n}, X_{\tau_{n+k}})$, esto puede hacerse mirando casos particulares

- si $k=1$

$$Cov(X_{\tau_n}, X_{\tau_{n+1}}) = Cov(X_{\tau_n}, \phi_{n+1} X_{\tau_n} + W_{n+1}^{1/2}) = \phi_{n+1} Cov(X_{\tau_n}, X_{\tau_n}) = \phi_{n+1} Var(X_{\tau_n})$$

- si $k=2$

$$Cov(X_{\tau_n}, X_{\tau_{n+2}}) = Cov(X_{\tau_n}, \phi_{n+2} X_{\tau_{n+1}} + W_{n+2}^{1/2}) = \phi_{n+2} Cov(X_{\tau_n}, X_{\tau_{n+1}}) = \phi_{n+2} \phi_{n+1} Var(X_{\tau_n})$$

- si $k=3$

$$Cov(X_{\tau_n}, X_{\tau_{n+3}}) = Cov(X_{\tau_n}, \phi_{n+3} X_{\tau_{n+2}} + W_{n+3}^{1/2}) = \phi_{n+3} Cov(X_{\tau_n}, X_{\tau_{n+2}}) = \phi_{n+3} \phi_{n+2} \phi_{n+1} Var(X_{\tau_n})$$

- Reemplazando recursivamente hasta $k = n$ tenemos:

$$Cov(X_{\tau_n}, X_{\tau_{n+k}}) = \prod_{i=1}^k \phi_{n+i} Var(X_{\tau_n}) \quad (3-2)$$

Para que este proceso sea estacionario, debemos demostrar que la varianza es constante en cualquier momento τ , por tanto, debemos mostrar que, para $n \geq 1$; $Var(X_{\tau_{n+1}}) = Var(X_{\tau_1}) = \gamma_0$, es decir:

$$\phi_{n+1}^2 \gamma_0 + W_{n+1} = W_1 = \gamma_0 \quad (3-3)$$

despejando W_{n+1} de 3-3 tenemos:

$$W_{n+1} = \gamma_0(1 - \phi_{n+1}^2) \quad (3-4)$$

Finalmente, para que nuestro modelo coincida con el modelo $AR(1)$ en el momento en que $\Delta_{n+1} = 1$ para todo n , hacemos $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$

3.2 Implementando propuesta de ϕ_n

El foco de la investigación está en generar una propuesta estable y teóricamente coherente para la secuencia ϕ_{n+1} , después de mucho estudiarlo, llegamos a esta expresión.

$$\phi_{n+1} = \text{sign}(\phi)|\phi|^{\Delta_{n+1}} \quad (3-5)$$

Si reemplazamos 3-5 en 3-4 obtenemos:

$$W_{n+1} = \frac{\sigma^2}{1-\phi^2} \left(1 - \text{sign}(\phi)^2 |\phi|^{2\Delta_{n+1}} \right) \quad (3-6)$$

El modelo final, reemplazando 3-5 y 3-6 en 3-1 tiene la expresión:

$$X_{\tau_{n+1}} = \text{sign}(\phi)|\phi|^{\Delta_{n+1}} X_{\tau_n} + \left[\frac{\sigma^2}{1-\phi^2} \left(1 - \text{sign}(\phi)^2 |\phi|^{2\Delta_{n+1}} \right) \right]^{1/2} \varepsilon_{\tau_{n+1}} \quad (3-7)$$

por definición la varianza de la expresión 3-7 es:

$$V(X_{\tau_{n+1}}) = \frac{\sigma^2}{1-\phi^2} \text{ y } \text{Cov}(X_{\tau_n}, X_{\tau_{n+k}}) = \left(\frac{\sigma^2}{1-\phi^2} \right) \text{sign}(\phi)^k |\phi|^{\sum_{i=1}^k \Delta_{n+i}}$$

3.3 El modelo autorregresivo irregularmente espaciado de primer orden

Basado en los calculos y definiciones anteriores, esta nueva propuesta de modelo autorregresivo irregularmente espaciado IAR es realizada desde un punto de vista constructorista. Más adelante en este documento presentamos las propiedades del proceso.

Definición 3.3.1 (Proceso IAR - Autocorrelaciones positivas y negativas). Sea $\{\varepsilon_{\tau_n}\}_{n \geq 1}$ variables aleatorias, cada una con distribución $N(0,1)$, siendo $\sigma^2 > 0$ y $-1 < \phi < 1$, se dice es un proceso IAR Si $X_1 = \left(\frac{\sigma^2}{1-\phi^2} \right)^{1/2} \varepsilon_{\tau_1}$ y para $n \geq 2$.

$$X_{\tau_n} = \text{sign}(\phi)|\phi|^{\Delta_n} X_{\tau_{n-1}} + \left[\frac{\sigma^2}{1-\phi^2} \left(1 - \text{sign}(\phi)^2 |\phi|^{2\Delta_n} \right) \right]^{1/2} \varepsilon_{\tau_n} \quad (3-8)$$

Donde $\Delta_n = \tau_n - \tau_{n-1}$

3.4 Propiedades del proceso

If $X_n = [X_{\tau_1}, \dots, X_{\tau_n}]'$ es un vector aleatorio de un proceso IAR, luego X_n es un vector aleatorio Gaussiano con $m_n = 0$ y matriz de varianzas y covarianza dada por

$$\Gamma_n = \begin{bmatrix} \gamma_0 & & & & & \\ \phi_2 \gamma_0 & \gamma_0 & & & & \\ \phi_2 \phi_3 \gamma_0 & \phi_3 \gamma_0 & \ddots & & & \\ \vdots & \vdots & & \gamma_0 & & \\ \prod_{i=2}^{n-1} \phi_i \gamma_0 & \prod_{i=3}^{n-1} \phi_i \gamma_0 & \cdots & \phi_{n-1} \gamma_0 & \gamma_0 & \\ \prod_{i=2}^n \phi_i \gamma_0 & \prod_{i=3}^n \phi_i \gamma_0 & \cdots & \phi_{n-1} \phi_n \gamma_0 & \phi_n \gamma_0 & \gamma_0 \end{bmatrix},$$

donde $\phi_n = \text{sign}(\phi)|\phi|^{\Delta_n}$ y $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$, para $n \geq 2$. Así, el proceso IAR es, por definicion un proceso Gaussiano debilmente estacionario y por tanto estrictamente estacionario. Además, note que cuando $\Delta_n = 1$, para $n \geq 2$, obtenemos el proceso AR convencional.

3.5 Predicción

3.6 Estimación por máxima verosimilitud

2.3 Estimation The likelihood of the data $\{y_{t_1}, \dots, y_{t_n}\}$ can be expressed as $f(y_{t_1}, \dots, y_{t_n}; \theta) = f(y_{t_1}; \theta) f(y_{t_2} | y_{t_1}; \theta) \times \dots \times f(y_{t_n} | y_{t_{n-1}}; \theta)$, (9) where $\theta = (\sigma^2, \phi)$ is the parameter vector of the model. To describe clearly the estimation process, we assume here that the marginal and conditional distributions of the time series are Gaussian. Note that this assumption is not necessary to obtain the statistical properties stated in Theorem 1. In Section 5, we show an example where the conditional distribution is assumed to be Gamma, and in Section 6, we show an example where the conditional distribution is assumed to be a Student's t -distribution. Assume that $f(y_{t_1}; \sigma^2, \phi) \sim N(0, \sigma^2)$ and

$$f(y_{t_j} | y_{t_{j-1}}; \sigma^2, \phi) \sim N(\phi^{t_j - t_{j-1}} y_{t_{j-1}}, \sigma^2 (1 - \phi^{2(t_j - t_{j-1})}))$$

for $j = 2, \dots, n$. Based on equation (5), minus the log-likelihood of this process can be written as

$$\ell(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{j=1}^n \log v_{t_j} + \frac{1}{2} \sum_{j=1}^n \frac{e_{t_j}^2}{v_{t_j}},$$

where we define $e_{t_1} = y_{t_1}$, $e_{t_j} = y_{t_j} - \phi^{t_j - t_{j-1}} y_{t_{j-1}}$ for $j > 1$ and their variances as $v_{t_j} = \text{Var}(e_{t_j})$. Observe that the finite past predictor of the process at time t_j is given by $\widehat{y}_{t_1} = 0$, and $\widehat{y}_{t_j} = \phi^{t_j - t_{j-1}} y_{t_{j-1}}$, for $j = 2, \dots, n$. Therefore, $e_{t_j} = y_{t_j} - \widehat{y}_{t_j}$ is the prediction error with variance $v_{t_1} = \text{Var}(e_{t_1}) = \sigma^2$, $v_{t_j} = P \text{Var}(e_{t_j}) = \sigma^2 \left[1 - \phi^{2(t_j - t_{j-1})} \right]$, for $j = 2, \dots, n$. By direct maximization of the log-likelihood (12), we can obtain the maximum likelihood estimator of σ^2 , $\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n \frac{(y_{t_j} - \widehat{y}_{t_j})^2}{\tau_{t_j}}$, where $\tau_{t_j} = v_{t_j}/\sigma^2$. But it is not possible to find $\widehat{\phi}$, the maximum likelihood estimator of ϕ , by direct maximization of the likelihood, but iterative methods

3.7 Estimación Bootstrap

Ahora, aplicamos el método bootstrap descrito en 2.5 para estimar ϕ_0 en el proceso IMA con $\sigma_0^2 = 1$. Sea X_τ observado en los puntos τ_1, \dots, τ_N y considere $\hat{\phi}_N$ como la estimación MLE respectiva. Las innovaciones estandarizadas estimadas son

$$e_{\tau_n}^s = \frac{X_{\tau_n} - \hat{X}_{\phi_n}(\hat{\phi}_N)}{\sqrt{W_n(\hat{\phi}_N)}},$$

para $n = 2, \dots, N$. El llamado remuestreo basado en el modelo podría proceder por muestreo equiprobable con reemplazo de los residuos centrados $e_{t_2}^s - \bar{e}, \dots, e_{t_N}^s - \bar{e}$, donde $\bar{e} = \sum_{n=2}^N e_{t_n}^s / (N-1)$, para obtener innovaciones simuladas $\zeta_{t_1}^*, \dots, \zeta_{t_N}^*$, y luego establecer:

$$X_{\tau_1}^* = \sqrt{W(\hat{\phi}_N)} \zeta_{\tau_1}^*, \quad (3-9)$$

$$X_{\tau_n}^* = \hat{\phi}_n X_{\tau_{n-1}}^* + \sqrt{W(\hat{\phi}_N)} \zeta_{\tau_n}^* \quad (3-10)$$

A continuación, estimamos los parámetros a través de MLE suponiendo que los datos son $X_{t_n}^*$. Por lo tanto, podemos repetir este proceso un gran número, B , de veces generando una colección de estimaciones de parámetros bootstrap. A continuación, podemos aproximar la distribución de muestra finita del estimador, $\hat{\phi}_N$, a partir de los valores de los parámetros bootstrap.

3.8 Estudio de Simulación Monte Carlo

Para probar el modelo 3-8 se genera un escenario de simulación, donde se muestran sus propiedades y las de los estimadores por máxima verosimilitud y bootstrap, la simulación se realiza considerando $\sigma_0 = 1$, $\phi_0 \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$ y $n \in \{20, 50, 100\}$ para que pueda observarse el comportamiento irregularmente, los tiempos τ_1, \dots, τ_n pueden ser considerados tanto regulares como irregulares con este modelo, en este caso, consideramos $\Delta_n = \tau_n - \tau_{n-1} \sim 1 + \text{poisson}(\lambda = 2)$ (para asegurar que no haya 0). La figura 3-1 muestra el comportamiento de una trayectoria simulada para cada una de las combinaciones de los parametros establecidos en el caso de $\phi > 0$, las marcas rojas en la parte de arriba del gráfico muestran mas claramente el comportamiento irregular. La figura 3-2 muestra el mismo comportamiento en el caso de $\phi < 0$. Para efectos de la simulación montecarlo simulamos $M = 1000$ trayectorias, $\{S_m\}_{m=1}^M$ y estimamos ϕ_0 para cada configuración.

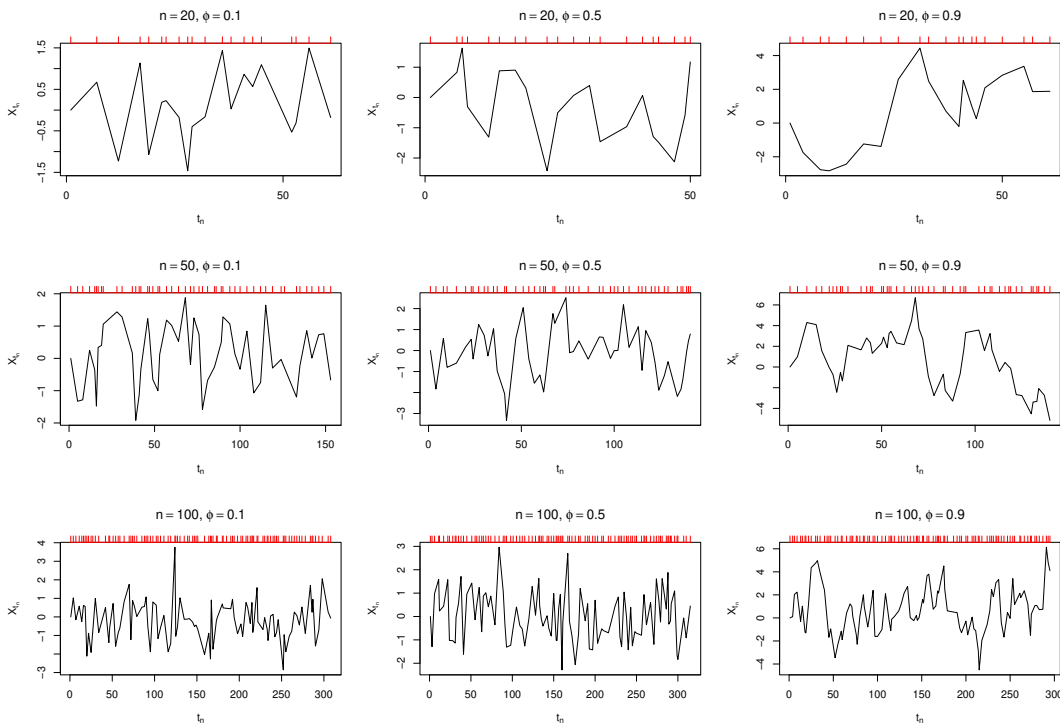


Figure 3-1: simulación del modelo IAR con coeficientes positivos

Sea $\hat{\phi}_m^{\text{MLE}}$ la estimación ML y $\widehat{\text{se}}(\hat{\phi}_m^{\text{MLE}})$ el error estandar estimado para la m -ésima trayectoria. El error estandar se estima por la curvatura de la superficie de verosimilitud en

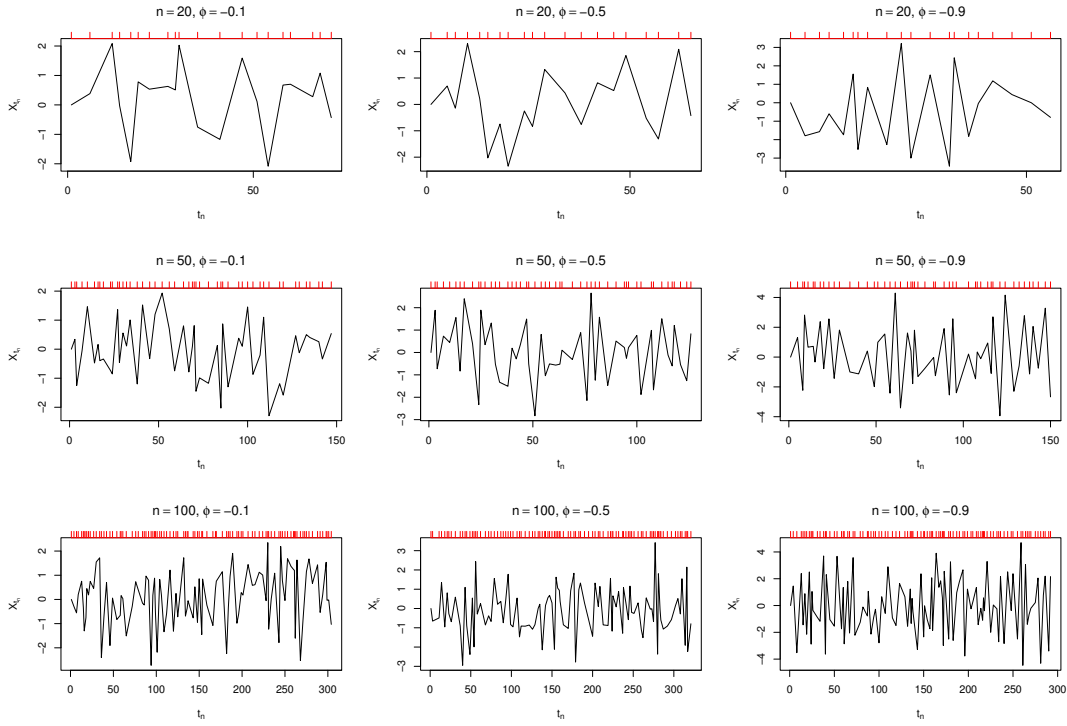


Figure 3-2: simulación del modelo IAR con coeficientes negativos

$\hat{\phi}_m^{\text{MLE}}$. Resumimos las estimaciones de máxima verosimilitud M mediante

$$\hat{\phi}^{\text{MLE}} = \frac{1}{M} \sum_{m=1}^M \hat{\phi}_m^{\text{MLE}} \quad \text{y} \quad \widehat{\text{se}}(\hat{\phi}^{\text{MLE}}) = \frac{1}{M} \sum_{m=1}^M \widehat{\text{se}}(\hat{\phi}_m^{\text{MLE}}).$$

Por otra parte, para cada trayectoria, hemos simulado $B = 500$ trayectorias bootstrap representadas por $\{\{S_{m,b}\}_{b=1}^B\}_{m=1}^M$. Entonces, obtuvimos $\{\{\hat{\phi}_{m,b}^b\}_{b=1}^B\}_{m=1}^M$ (las estimaciones ML). La estimación bootstrap y su error estándar estimado se definen como

$$\hat{\phi}_m^b = \frac{1}{B} \sum_{b=1}^B \hat{\phi}_{m,b}^b \quad \text{and} \quad \widehat{\text{se}}^2(\hat{\phi}_m^b) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\phi}_{m,b}^b - \hat{\phi}_m^b)^2,$$

para $m = 1, \dots, M$. Por último, resumimos las M estimaciones bootstrap mediante

$$\hat{\phi}^b = \frac{1}{M} \sum_{m=1}^M \hat{\phi}_m^b \quad \text{and} \quad \widehat{\text{se}}(\hat{\phi}^b) = \frac{1}{M} \sum_{m=1}^M \widehat{\text{se}}(\hat{\phi}_m^b).$$

Además, como medida del rendimiento del estimador, utilizamos el error cuadrático medio (RMSE), y el coeficiente de variación (CV) estimado definido por

$$\widehat{\text{RMSE}}_{\hat{\phi}^{\text{MLE}}} = \left(\widehat{\text{se}}(\hat{\phi}^{\text{MLE}})^2 + \widehat{\text{bias}}_{\hat{\phi}^{\text{MLE}}}^2 \right)^{1/2}, \text{ and}$$

$$\widehat{\text{CV}}_{\hat{\phi}^{\text{MLE}}} = \frac{\widehat{\text{se}}(\hat{\phi}^{\text{MLE}})}{|\hat{\phi}^{\text{MLE}}|},$$

donde $\widehat{\text{bias}}_{\hat{\phi}^{\text{MLE}}} = \hat{\phi}^{\text{MLE}} - \phi$. Por último, estimamos la varianza del estimador mediante

$$\widetilde{\text{se}}^2(\hat{\phi}^{\text{MLE}}) = \frac{1}{M-1} \sum_{m=1}^M (\hat{\phi}_m^{\text{MLE}} - \hat{\phi}^{\text{MLE}})^2$$

De la misma manera, definimos $\widehat{\text{RMSE}}_{\hat{\phi}^{\text{b}}}$, $\widehat{\text{CV}}_{\hat{\phi}^{\text{b}}}$, $\widehat{\text{bias}}_{\hat{\phi}^{\text{b}}}$ y $\widetilde{\text{se}}^2(\hat{\phi}^{\text{b}})$ para el caso Bootstrap. La figura 3-3 muestra flujo de trabajo de este estudio de simulación

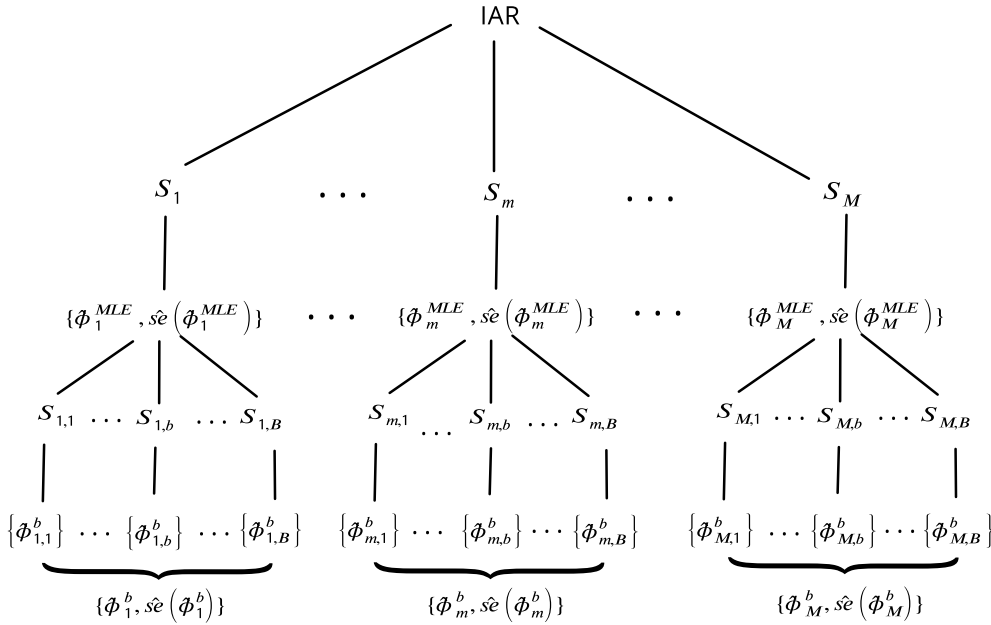


Figure 3-3: Esquema general del estudio de Monte Carlo.

En este estudio, se evaluaron los resultados del modelo en el caso de tiempo irregularmente espaciado, aunque también se comprobó que el modelo es válido para tiempos regulares con $\Delta_n = 1$ para todos los valores de n .

Para analizar el comportamiento de los estimadores de máxima verosimilitud y bootstrap, se realizaron simulaciones de distribuciones de muestras finitas. Los resultados de estas simulaciones se muestran en las figuras 3-4 y 3-5. Se puede observar que ambos estimadores presentan un comportamiento satisfactorio, parecen ser insesgados y consistentes en sus resultados.

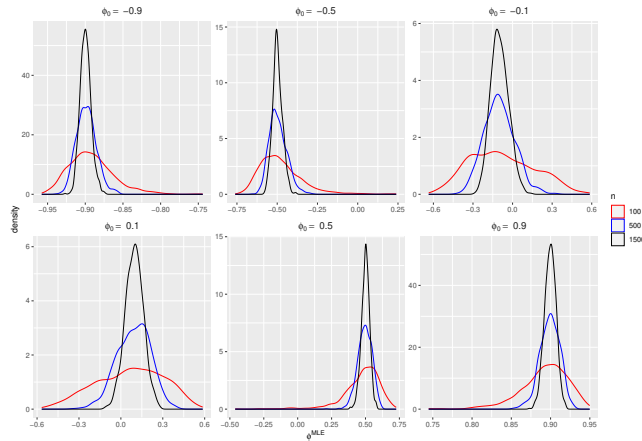


Figure 3-4: Esquema general del estudio de Monte Carlo.

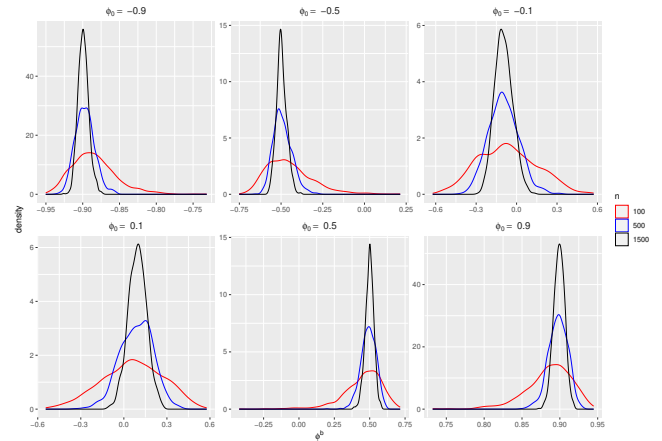


Figure 3-5: Esquema general del estudio de Monte Carlo.

Note que para valores pequeños de n y para valores cercanos a cero de ϕ , los estimadores tienden a tener una mayor variabilidad. Este hallazgo es importante para el análisis de los resultados, ya que sugiere que la precisión de los estimadores puede verse afectada por factores como la cantidad de datos disponibles y la magnitud del parámetro de interés. En segundo lugar, para ambos estimadores, presentamos las medidas de rendimiento para ambos métodos de estimación las cuales se ven reflejadas en la tabla 3-1

Table 3-1: Resultados del estudio de Monte Carlo

n	Máxima Verosimilitud						Bootstrap				
	ϕ	$\hat{\phi}^{MLE}$	$\hat{se}(\phi^{MLE})$	$\widehat{bias}_{\hat{\phi}^{MLE}}$	$\widehat{RMSE}_{\hat{\phi}^{MLE}}$	$\widehat{CV}_{\hat{\phi}^{MLE}}$	$\hat{\phi}^b$	$\hat{se}(\phi^b)$	$\widehat{bias}_{\hat{\phi}^b}$	$\widehat{RMSE}_{\hat{\phi}^b}$	$\widehat{CV}_{\hat{\phi}^b}$
100	-0.9	-0.892	0.029	-0.008	0.031	0.033	-0.885	0.032	-0.015	0.035	0.036
	-0.5	-0.479	0.126	-0.021	0.128	0.263	-0.450	0.153	-0.050	0.161	0.340
	-0.1	-0.081	0.210	-0.019	0.211	2.586	-0.068	0.222	-0.032	0.224	3.267
	0.1	0.062	0.211	0.038	0.214	3.388	0.060	0.222	0.040	0.226	3.726
	0.5	0.473	0.126	0.027	0.129	0.267	0.450	0.145	0.050	0.154	0.323
	0.9	0.892	0.029	0.008	0.030	0.033	0.886	0.032	0.014	0.035	0.036
500	-0.9	-0.898	0.013	-0.002	0.013	0.014	-0.897	0.013	-0.003	0.013	0.014
	-0.5	-0.495	0.052	-0.005	0.053	0.106	-0.491	0.054	-0.009	0.055	0.110
	-0.1	-0.101	0.110	0.001	0.110	1.095	-0.097	0.111	-0.003	0.111	1.153
	0.1	0.091	0.111	0.009	0.112	1.227	0.087	0.111	0.013	0.112	1.274
	0.5	0.495	0.052	0.005	0.052	0.105	0.491	0.054	0.009	0.054	0.109
	0.9	0.899	0.013	0.001	0.013	0.014	0.898	0.013	0.002	0.013	0.014
1500	-0.9	-0.900	0.007	0.000	0.007	0.008	-0.899	0.007	-0.001	0.007	0.008
	-0.5	-0.498	0.030	-0.002	0.030	0.060	-0.497	0.030	-0.003	0.030	0.060
	-0.1	-0.098	0.066	-0.002	0.066	0.673	-0.096	0.066	-0.004	0.066	0.688
	0.1	0.096	0.067	0.004	0.067	0.695	0.094	0.066	0.006	0.067	0.705
	0.5	0.500	0.030	0.000	0.030	0.059	0.498	0.030	0.002	0.030	0.060
	0.9	0.900	0.007	0.000	0.007	0.008	0.899	0.007	0.001	0.007	0.008

Chapter 4

Conclusion

Las conclusiones constituyen un capítulo independiente y presentan, en forma lógica, los resultados del trabajo de grado. Las conclusiones deben ser la respuesta a los objetivos o propósitos planteados. Se deben titular con la palabra conclusiones en el mismo formato de los títulos de los capítulos anteriores (Títulos primer nivel), precedida por el numeral correspondiente (según la presente plantilla).

Appendix A

Appendix

Los Anexos son documentos o elementos que complementan el cuerpo de la tesis o trabajo de investigación y que se relacionan, directa o indirectamente, con el trabajo de grado, tales como acetatos, cd, normas, etc.

References

- Adorf, H.-M. (1995), Interpolation of irregularly sampled data series—a survey, in ‘Astronomical Data Analysis Software and Systems IV’, Vol. 77, p. 460.
- Bose, A. (1990), ‘Bootstrap in moving average models’, *Annals of the Institute of Statistical Mathematics* **42**(4), 753–768.
- Dilmaghani, S., Henry, I. C., Soonthornnonda, P., Christensen, E. R. & Henry, R. C. (2007), ‘Harmonic analysis of environmental time series with missing data or irregular sample spacing’, *Environmental science & technology* **41**(20), 7030–7038.
- Elorrieta, F., Eyheramendy, S. & Palma, W. (2019), ‘Discrete-time autoregressive model for unequally spaced time-series observations’, *arXiv preprint arXiv:1906.11158*.
- Eyheramendy, S., Elorrieta, F. & Palma, W. (2018), ‘An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves’, *Monthly Notices of the Royal Astronomical Society* **481**(4), 4311–4322.
- Ghaderpour, E. & Vujadinovic, T. (2020), ‘Change detection within remotely sensed satellite image time series via spectral analysis’, *Remote Sensing* **12**(23), 4001.
- Kuttu, S. & Bokpin, G. A. (2017), ‘Feedback trading and autocorrelation patterns in sub-saharan african equity markets’, *Emerging Markets Finance and Trade* **53**(1), 213–225.
- Liu, M., Stella, F., Hommersom, A., Lucas, P. J., Boer, L. & Bischoff, E. (2019), ‘A comparison between discrete and continuous time bayesian networks in learning from clinical time series data with irregularity’, *Artificial intelligence in medicine* **95**, 104–117.
- Novoseltseva, M., Gutova, S. & Glinchikov, K. (2019), Spatial and econometric modeling of the demographic situation in the kemerovo region, in ‘IOP Conference Series: Earth and Environmental Science’, Vol. 272, IOP Publishing, p. 032155.

- Rindorf, A., Cadigan, N., Howell, D., Eero, M. & Gislason, H. (2020), 'Periodic fluctuations in recruitment success of atlantic cod', *Canadian Journal of Fisheries and Aquatic Sciences* **77**(2), 236–246.
- Robinson, P. (1977), 'Estimation of a time series model from unequally spaced data', *Stochastic Processes and their Applications* **6**(1), 9–24.
- Shamsan, A., Wu, X., Liu, P. & Cheng, C. (2020), 'Intrinsic recurrence quantification analysis of nonlinear and nonstationary short-term time series', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30**(9), 093104.
- Shumway, R. H., Stoffer, D. S. & Stoffer, D. S. (2017), *Time series analysis and its applications with R examples*, Vol. 4, Springer Texts in Statistics.
- Yang, Q., Fowler, M. S., Jackson, A. L. & Donohue, I. (2019), 'The predictability of ecological stability in a noisy world', *Nature ecology & evolution* **3**(2), 251–259.