



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	x
List of tables	xi
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
3 Conclusion	7
A Appendix	8
References	9

List of Figures

List of Tables

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 ojeda: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 practica 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 irregular de medias móviles 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 varianza y las covarianzas de primer orden respectivamente. Definimos la varianza 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

Chapter 3

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.
- 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.
- 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.