

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	хi
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	Metodología	7
	3.1 Citas bibliográficas	7
	3.2 Presentación y citación de figuras	8
	3.3 Presentación y citación de tablas	10
	ciones	11
4	Conclusion	12
A	Appendix	13
	References	14

List of Figures

3-1	Histograma	de costos t	totales de	garantía.	Política PRW.		9
-----	------------	-------------	------------	-----------	---------------	--	---

List of Tables

3-1	Prueba Chi-cuadrado de igualdad distribucional entre aproximación caja	
	negra y aproximación física. Política FRW	10

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 cómo 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 autorregrevos 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 timpo son regulamente espaciadas, sin embargo este escenario no siempre se da y ahi es donde cobran importancia los procesos estocásticos irregularmente espaciados.

Este capitulo tiene la siguiente estructura, en la seccción 2.1 presentamos la definición matematica de un proceso estocastico 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 moviles y en la sección 2.3 se presenta un proceso estocástico atorregresivo de medias moviles irregulamente 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 $x : \Omega \to \mathbb{R}^{\mathbb{T}}$, 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 $\mathbb{T}'=\{t_1,t_2,t_3,...\}$, 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 $n \ge 1$, 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 $\{X_{\tau_1}, X_{\tau_2}, ..., X_{\tau_n}\}$ 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 construccionista.

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

Teniendo en cuenta el conjunto $\mathbb{T}'=\{t_1,t_2,t_3,...\}$ 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 $\Gamma:\mathbb{T}'x\mathbb{T}'\mapsto\mathbb{R}$ una función tal que para cualquier pareja $t_n,t_s\in\mathbb{T}'$,

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

Note que Γ puede ser representada cómo una matriz diagonal así

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1, \Delta_2 & 0 & 0 & \dots \\ \gamma_1, \Delta_2 & \gamma_0 & \gamma_1, \Delta_3 & 0 & \\ 0 & \gamma_1, \Delta_3 & \gamma_0 & \gamma_1, \Delta_4 & \\ \vdots & & & \ddots & \\ \vdots & & & & \ddots \end{bmatrix}$$

$$(2-1)$$

Ser Γ_n la truncacion nxn de Γ y asumiendo $\gamma_1, \Delta_j \neq 0$, para j=2,...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,...,nEsto 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 $\gamma_{1,\Delta_{n+1}}$, para $n \ge 1$, representa la varianza y las covarianzas de primer orden respectivamente. Definimos la varianza cómo $\gamma_0 = \sigma^2(1+\theta^2)$ y las covarianzas de primer orden cómo $\gamma_{1,\Delta_{n+1}} = \sigma^2\theta^{\Delta_{n+1}}$, $\sigma^2 > 0$ y $0 < \theta < 1$. Por tanto, obtenemos el proceso estocástico irregularmente espaciado de primer orden con matriz de covarianzas

$$\Gamma = \begin{bmatrix} 1 + \theta^2 & \theta^{\Delta_2} & 0 & 0 & \dots \\ \theta^{\Delta_2} & 1 + \theta^2 & \theta^{\Delta_3} & 0 \\ 0 & \theta^{\Delta_3} & 1 + \theta^2 & \theta^{\Delta_4} \\ \vdots & & & \ddots \\ \vdots & & & & \ddots \end{bmatrix}$$

$$(2-2)$$

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

El punto de vista construccionista

Ahora, cómo es usual, especificaremos el proceso IMA cómo función de otros procesos estocásticos. Sea $\{\epsilon_{t_n}\}n \ge 1$ 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

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

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 $n \ge 2$

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

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

Cambiando el marco teorico en rama, hola mamamamam ssjsjsjsj nuevamente cambio el marco jajajja

estoy subiendo esto al remoto

Chapter 3

Metodología

La Norma Técnica Colombiana NTC 1486 fija la reglamentación nacional para la "presentación de tesis, trabajos de grado y otros trabajos de investigación" (?). Se recomienda a los autores revisar cuidadosamente la Norma para familiarizarse con su contenido antes de empezar a escribir su propio documento. Pero además, se recomienda mantener el documento a mano como fuente de referencia permanente.

Por otra parte, la Norma Técnica Colombiana NTC 5613 se refiere a las referencia bibliográficas, su contenido, forma y estructura (?). En esta plantilla se recogen las ideas generales de la Norma NTC 5613, para lo cual se utiliza el estilo de citas *agsm* (tipo autoraño), tomado de la familia de estilos bibliográficos *Harvard* (?), pero usando el paquete de referencias y citas para las ciencias naturales *Natbib* (??).

Las figuras, tablas y ecuaciones se enumeran consecutivamente dentro de cada capítulo.

Las siguientes secciones de este capítulo describen cómo citar las referencias bibliográficas y como referenciar en el texto las figuras, las tablas y las ecuaciones.

3.1 Citas bibliográficas

En primer lugar, se debe crear un archivo con todas las referencias bibliográficas. En esta plantilla este archivo tiene la extensión *bib* y se llama *BiblioPAE*. Los archivos tipo *bib* tienen entradas que el programa BibTeX interpreta e incorpora en el documento final.

Por ejemplo, la cita del trabajo? se introduce en el archivo BiblioPAE.bib de la siguiente

manera:

```
@Manual{Harvard:2008,
title = {The {H}arvard {F}amily of {B}ibliography {S}tyles},
author = {Peter Williams and Thorsten Schnier},
month = {Noviembre},
year = {2008},
note = {ftp://ftp.tex.ac.uk/tex-archive/macros/latex/contrib/harvard/harvard.pdf},
}
```

Dependiendo de la forma en que se introduzca la cita, ésta lucirá de manera diferente. Por ejemplo, si se usa \citet{Harvard:2008}, la referencia será adecuada para una expresión de la forma: "... de acuerdo con ?..."

Pero si se usa \citep{Harvard: 2008}, la referencia será más adecuada para la expresión: "... si se usa el estilo *Harvard* (?)..."

Para otras variaciones de la forma de citar, consulte el manual de Natbib (?).

Los siguientes ejemplos pueden resultar útiles para ilustrar la forma de citar en el documento:

- El sistema de procesamiento estadístico de libre distribución R (?) se utiliza para ejecutar el procesamiento de los datos en este Trabajo de Grado.
- En este Trabajo de Grado se utiliza la librería Lattice de *R* (ver ?).
- Según ?, la suavización spline es una de las técnicas más comunes para estimar la función de regresión en los métodos de regresión no paramétrica.
- El modelo de regresión lineal podría considerarse el método estadístico más utilizado en el análisis de datos (?????).
- ?, pág. 12 presenta la idea básica del concepto de *suavización*.

3.2 Presentación y citación de figuras

Las ilustraciones forman parte del contenido de los capítulos. Se deben colocar en la misma página en que se mencionan o en la siguiente (deben siempre mencionarse en el

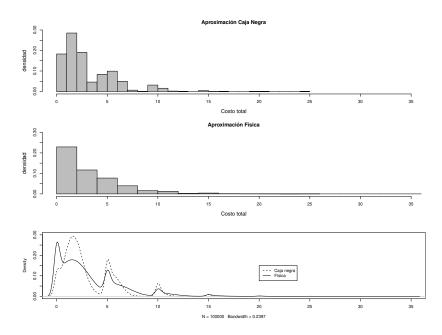


Figure 3-1: Histograma de costos totales de garantía. Política PRW.

texto).

Un ejemplo para la presentación y citación de figuras, se presenta a continuación (citación directa):

La Figura **3-1** presenta un histograma del costo total bajo la politica PRW, los histogramas bajo las dos aproximaciones sugieren una mezcla de sub-poblaciones.

Acerca de las notas al pie

En general se evitará el uso de llamadas fuera del texto principal para introducir aclaraciones o complementos. Pero en la situación eventual de que esto se necesite (algo altamente improbable) estas llamadas deben hacerse con un nota al pie¹.

La fuente documental se debe escribir al final de la figura con los elementos de la referencia (de acuerdo con las normas seleccionadas) y no como pie de página.

¹Solo en caso de necesidad extrema, las notas se usarían como "notas al pie". Se utilizan para explicar, comentar o hacer referencia al texto de un documento, así como para introducir comentarios detallados. Las notas al pie no se usan en este informe final de Trabajo de Grado para introducir referencias bibliográficas.

3.3 Presentación y citación de tablas

Para la edición de tablas, cada columna debe llevar su título; la primera palabra se debe escribir con mayúscula inicial y preferiblemente sin abreviaturas. En las tablas y cuadros, los títulos y datos se deben ubicar entre líneas horizontales y verticales cerradas (como se realiza en esta plantilla).

La numeración de las tablas se realiza de la misma manera que las figuras o ilustraciones, a lo largo de todo el texto. Deben llevar un título breve, que concreta el contenido de la tabla; éste se debe escribir en la parte superior de la misma. Para la presentación de cuadros, se deben seguir las indicaciones dadas para las tablas.

Un ejemplo para la presentación y citación de tablas (citación indirecta), se presenta a continuación:

Los resultados de la prueba chi-cuadrado sugieren fuerte evidencia para rechazar la hipótesis nula de igualdad distribucional. Ver Tabla 3-1

Table **3-1**: Prueba Chi-cuadrado de igualdad distribucional entre aproximación caja negra y aproximación física. Política FRW.

	χ^2	df	valor-p
costo falla I	12379.12	6	$< 2.2 \times 10^{-16}$
costo falla II	5893.437	17	$< 2.2 \times 10^{-16}$
costos total	10700.75	26	$< 2.2 \times 10^{-16}$

NOTA: en el caso en que el contenido de la tabla sea muy extenso, se puede cambiar el tamaño de la letra, siempre y cuando ésta sea visible por el lector.

11

3.3.1 Consideraciones adicionales para el manejo de figuras, tablas y ecuaciones

Cuando una tabla, cuadro o figura ocupa más de una página, se debe repetir su identificación numérica, seguida por la palabra continuación.

Adicionalmente los encabezados de las columnas se deben repetir en todas las páginas después de la primera.

Los anteriores lineamientos se contemplan en la presente plantilla.

• Presentación y citación de ecuaciones.

La citación de ecuaciones, en caso que se presenten, debe hacerse como lo sugiere esta plantilla. Todas las ecuaciones deben estar numeradas y citadas detro del texto.

Para el manejo de cifras se debe seleccionar la norma según el área de conocimiento del trabajo de grado.

$$Z_w^{FRW} = \sum_{j=1}^{\eta} \left(c^I N_{T_j^{II}}^I + c^{II} \right) - \left(c^I N_{T_{\eta}^{II}}^I + c^{II} \right) + c^I N_w^I$$
 (3-1)

y el valor esperado del proceso de costo (3-1) se encuentra dado por,

$$E(Z_w^{FRW}) = E\left[\sum_{j=1}^{\eta} \left(c^I N_{T_j^{II}}^I + c^{II}\right) - \left(c^I N_{T_{\eta}^{II}}^I + c^{II}\right)\right] + E\left[c^I N_w^I\right]$$
(3-2)

Para obtener la expresión del valor esperado del proceso de costo Z_w^{FRW} en (3-2) se realiza el cálculo de la esperanza para cada uno de los términos que componen la expresión.

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.
- 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 subsaharan 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.

REFERENCES 15

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.