

MÉTODOS DE SERIES DE TIEMPO DE INTERVALO

Adriana Mallea¹, Cecilia Martínez²,
Andrea Salas¹

¹Departamento de Matemática, FFHA, Universidad Nacional de San Juan

²Departamento de Informática, FCFN, Universidad Nacional de San Juan

lamallea@ffha.unsj.edu.ar

RESUMEN

El objetivo principal de esta línea de investigación es mostrar el desarrollo de métodos de predicción para series temporales de intervalo (ITS) y su aplicación en ejemplos reales para demostrar su efectividad.

Este plan surge de la confluencia de dos disciplinas: el análisis de datos simbólicos (área que se ocupa del análisis estadístico de datos de intervalo, en este caso) y la predicción de series temporales. De acuerdo a esto y para lograr el objetivo general, se proponen los siguientes objetivos particulares:

1. Estudio de métodos de predicción de series temporales simbólicas para datos de intervalo. Proponemos estudiar métodos propuestos en la literatura, siguiendo dos enfoques:

- Predecir la serie mediante métodos de predicción clásicos: consiste en tratar la serie temporal de intervalos como un par de series temporales clásicas que pueden ser la serie de los mínimos y la de los máximos o, alternativamente, la de los centros y la de los radios.
- Predecir la serie mediante métodos de predicción que consideren el intervalo como tal, todos ellos basados en la

aritmética de intervalos (Moore, 1966)

2. Estudio de modelos vectoriales autorregresivos de promedios móviles para series de intervalo.

Recientemente, en la literatura de pronóstico de ITS, se ha propuesto el modelo Vectorial Autorregresivo de Promedios Móviles (IVARMA) para capturar la dinámica de cross-dependencia presente en este tipo de series.

3. Aplicación de los métodos de predicción estudiados a series provenientes de diferentes ámbitos.

Se ilustrará la metodología estudiada en series clásicas provenientes de diversas áreas de la actividad humana, sobre todo en contextos econométricos. Para ello se procederá a la agregación temporal, en cada caso, de los datos univariados a fin de lograr el intervalo de valores en instantes específicos de tiempo, se aplicarán los distintos métodos para pronóstico de ITS estudiados y se evaluará su performance con el error definido como la raíz cuadrada del RECM (error cuadrático medio escalado).

Palabras clave: Series, Tiempo, intervalo.

CONTEXTO

Esta línea de investigación se inserta en el proyecto *Análisis de Datos Simbólicos para Data Science*, cuyo tipo de actividad de I+D es investigación básica, que inició en el año 2020 y finaliza en diciembre de 2022, financia la UNSJ. Tiene como unidad ejecutora el Departamento de Matemática de la FFHA y sus integrantes desarrollan sus tareas de docencia e investigación en las áreas de matemática e informática.

1. INTRODUCCIÓN

Los datos simbólicos son un paradigma de representación de la información que surge a finales de los ochenta (Diday, 1987) bajo la premisa de que las variables clásicas, i.e., aquellas que a cada individuo le asignan un único valor, no son capaces de representar con fidelidad algunas situaciones. Los datos simbólicos, a diferencia de los clásicos, permiten representar conceptos de una manera sintética y descriptiva.

La característica fundamental de los datos simbólicos es que permiten la descripción de elementos o fenómenos donde exista una variabilidad interna.

Los conceptos implican variabilidad ya que las distintas realizaciones de ese concepto pueden ser algo diferentes entre sí. La variabilidad surge de manera natural al agregar observaciones. Por agregación se entiende la recopilación de observaciones que satisfacen un requisito que les permite ser agrupadas.

La agregación puede ser:

- Contemporánea, si se recopilan observaciones recogidas en un mismo instante temporal o cuando el instante temporal no es relevante.
- Temporal, si el criterio de agregación es el tiempo y se

recopilan observaciones ocurridas a lo largo de una unidad de tiempo, e.g. un día.

Al tener una estructura distinta que la de los datos clásicos, las técnicas de análisis del paradigma clásico no son válidas para analizar los datos simbólicos. Por ello, es necesario desarrollar un nuevo catálogo de métodos que sean capaces de extraer el conocimiento de este nuevo tipo de datos. Éste es el propósito del análisis de datos simbólicos. Este tipo de análisis es un área novedosa dentro de la estadística y de la minería de datos. Su objetivo es el de extraer información sobre datos expresados mediante variables simbólicas, como por ejemplo variables de intervalo. Estas variables son más complejas que las variables clásicas que describen los elementos mediante un único valor, pero describen de manera más fiel la realidad al poder representar variabilidad.

A lo largo de las últimas tres décadas las técnicas del Análisis de Datos Simbólicos (ADS) se han incrementado notablemente. En esta línea de investigación, en particular, interesa mostrar el avance de contribuciones en el análisis de series temporales de intervalo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación se enmarca dentro de Data Science y Data Mining aplicados a los datos estándar.

Los primeros trabajos para el pronóstico de ITS que se proponen en la literatura son un modelo ARMA y un modelo híbrido que combina el modelo ARMA con una red neuronal artificial (RNA). Teles y Brito (2005) proponen una extensión de los modelos ARMA para trabajar con series temporales de intervalos. Ellos asumen que los procesos que generan los mínimos y los máximos

de los intervalos tienen los mismos parámetros pero diferentes medias, siendo la media de la ecuación de los máximos mayor que la media de la ecuación de los mínimos. En Maia et al. (2006a) y Maia, de Carvalho y Ludermir (2006b), los autores proponen una extensión de los modelos ARMA y de los modelos híbridos para predecir series temporales de intervalo. La extensión de los modelos ARMA es distinta a la propuesta por Teles y Brito (2005) ya que modelan las series temporales de los centros y de los radios de los intervalos, en lugar de los mínimos y los máximos. El modelo híbrido que proponen se basa en el modelo desarrollado por Zhang (2003) para series temporales clásicas. Dicho modelo combina un modelo ARMA y un perceptrón multicapa para obtener las predicciones.

En los últimos años, han surgido varios estudios que han documentado métodos referentes a la modelización y pronóstico de ITS (Arroyo & Maté, 2006; Arroyo et al., 2011; Arroyo, Roque, Maté, & Sarabia, 2007; González-Rivera & Lin, 2013; Hanet et al., 2015; He & Hu, 2009; Maia & Carvalho, 2011; Maia et al., 2008; Neto & Carvalho, 2008, 2010). Arroyo et al. (2007) propusieron el método de suavizado exponencial y lo compararon con otros enfoques para el modelado de series. Consideraron que la forma más directa de transformar una serie simbólica de intervalo es por un par de series de tiempo clásica. González-Rivera and Lin (2013) propusieron un modelo de regresión restringido para los bordes de la ITS que preserva el orden natural del intervalo en todos los casos. La idea fundamental de estos enfoques antes mencionados para modelar una ITS univariada es representarla como un vector, esto es, el punto medio y el rango, o ambos bordes de la serie y luego hacer la predicción de estas series, basados en

herramientas estadísticas clásicas. Han et al. (2015) propusieron una nueva estrategia via una clase de modelos intervalo condicional autorregresivo (ACI) para una ITS univariada. A diferencia de otros, el enfoque ACI propone un modelo parsimonioso que considera una observación intervalar como una unidad inseparable. Yang and Han (2015) exploraron la potencia del pronóstico de un modelo ACI para los datos bursátiles diarios de EE. UU. Y mostraron que estos modelos producen pronósticos de volatilidad más precisos que los métodos basados en modelos de series de tiempo puntuales. Han et al. (2016) proponen un modelo parsimonioso IVARMA para investigar la dinámica de un sistema de vectores ITS y para pronosticar un vector ITS. El modelo IVARMA generaliza el modelo vectorial autorregresivo de promedios móviles clásico para una serie de tiempo vectorial a la configuración de intervalo. Esto se logra extendiendo el trabajo de Han et al. (2015) para una ITS univariada al caso de un sistema ITS multivariado. Para estimar los parámetros de un modelo IVARMA usando datos de intervalo, se propone un método de estimación de mínima distancia y se establece la teoría asintótica de los estimadores propuestos.

3. RESULTADOS OBTENIDOS/ESPERADOS

4.

En el contexto del mencionado proyecto, en Octubre del año 2021, se presentó la comunicación *Pronóstico de la tasa de natalidad haciendo uso del Análisis de Datos Simbólicos*, en el XIV Congreso Latinoamericano de Sociedades Estadísticas. En este trabajo, considerando como fuente el Instituto Nacional de Estadísticas y Censos (INDEC) y la Dirección de Estadísticas e Información de Salud (DEIS), se realizó

una predicción de la tasa de natalidad en Argentina, utilizando una adaptación del método de Alisado Exponencial Simple. Para esto se emplearon paquetes del Software R y algoritmos desarrollados en Octave (versión libre de Matlab).

Por otra parte, la Lic. Cecilia Martínez trabajará en su tesis de maestría pronósticos de ITS simulados y con datos reales, empleando el método IVARMA y métodos de suavizado exponencial, propuestos en la literatura.

5. FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación del proyecto mencionado en la introducción está formado por docentes investigadores de dos facultades de la UNSJ, algunos de ellos son jóvenes investigadores. En particular, la Lic. Cecilia Martínez y Lic. Andrea Salas han investigado durante los dos primeros años acerca de los métodos desarrollados, en el contexto de ADS, para el pronóstico de ITS. La primera desarrolla su tesis de Maestría en Matemática, de la Universidad Nacional de San Luis, en esta línea. La segunda es alumna del Doctorado en Estadística de la Universidad Nacional de Rosario.

6. BIBLIOGRAFÍA

Arroyo, J., Espínola, R., & Maté, C. (2011). *Different approaches to forecast interval time series: A comparison in finance*. Computational Economics, 37(2), 169–191.

Froelich, W., & Salmeron, J. L. (2014). *Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series*. International Journal of Approximate Reasoning, 55(6), 1319–1335.

García-Ascanio, C., & Maté, C. (2010). *Electric power demand forecasting using*

interval time series: A comparison between VAR and iMLP. Energy Policy, 38(2), 715–725.

Golan, A., & Ullah, A. (2015). *Interval estimation: An infometrics approach*. Working Paper.

González-Rivera, G., & Lin, W. (2013). *Constrained regression for interval-valued data*. Journal of Business & Economic Statistics, 4, 473–490.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton university press.

Han, A., Hong, Y., & Wang, S. (2016). *A Vector Autoregressive Moving Average Model for Interval-Valued Time Series Data*.

Published online: 23 Jun 2016; 417-460.

Han, A., Hong, Y., & Wang, S. (2015). *Autoregressive conditional models for interval-valued time series data*, Working Paper.

Han, A., Lai, K. K., Wang, S., & Xu, S. (2012). *An interval method for studying the relationship between the Australian dollar exchange rate and the gold price*. Journal of Systems Science and Complexity, 25(1), 121–132.

Maia, A. L. S., & Carvalho, F.A.T. (2011). *Holt's exponential smoothing and neural network models for forecasting interval-valued time series*. International Journal of Forecasting, 27(3), 740–759.

Nakajima, J. (2011). *Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications*. Institute for Monetary and Economic Studies, Bank of Japan.

Teles, P. y Brito, M. P. (2005) *Modelling interval time series data*. En Proceedings of the 3rd IASC World Conference on Computational Statistics & Data Analysis. Limassol, Cyprus.

Yang, W., & Han, A. (2015). *A new approach for forecasting the price range with financial intervalvalued time series*

data. ASCE-ASME Journal of Risk
Uncertainty Engineering Systems, Part B:
Mechanical Engineering, 1(2), 021004,
1–8. doi:10.1115/1.4028751