

Inferencia estadística y Aplicaciones de Procesos Estocásticos: Series de tiempo y Procesos de Markov. Parte I

1. Introducción.

Los procesos estocásticos son modelos que se utilizan para estudiar y comprender el comportamiento de fenómenos que evolucionan a través del tiempo. Estos modelos permiten incorporar la dependencia temporal presente en estos fenómenos. La aplicación de los Procesos Estocásticos, así como las áreas de la investigación donde están presentes han ido incrementándose en los últimos años de manera notable, y prueba de esto es que la revista *Statistical Inference for Stochastic Processes* ya cumplió 20 años contribuyendo en la difusión de los resultados y aplicaciones en esta área.

Dos tipos de procesos estocásticos muy utilizados son los Modelos para Series de Tiempo y los Modelos de Markov.

Sobre los Modelos para Series de Tiempo, los más famosos por sus aplicaciones son los modelos ARMA, GARCH y sus distintas variantes. Trabajos previos del grupo de Estadística sobre este tema se encuentran publicados en los artículos Argaez, et.al (2014), Chan Cab, et.al (2017), en los cuales se presentan aplicaciones de estos modelos.

Como generalizaciones de estos modelos, han surgido los modelos VAR (Vector Autoregressive) y el concepto de Cointegración, los cuales se abordan en este proyecto.

A pesar de que los modelos de Markov han sido utilizados en muchos contextos, aun se requieren métodos de estimación de los parámetros y análisis exploratorio de las muestras aleatorias. Por lo que en este proyecto se espera desarrollar metodología que complemente o mejore la conocida hasta el momento. Modelos basados en cadenas de Markov, como el proceso de Markov escondido, serán estudiados en este proyecto.

2. Marco Teórico

Los modelos basados en Procesos Estocásticos son útiles en problemas donde los datos se recolectan de acuerdo a un orden en el tiempo y presentan una estructura de dependencia entre ellos.

Un proceso estocástico o aleatorio es una colección de variables aleatorias indexadas por un conjunto T al que se le denomina tiempo, es decir, $X = \{X_t: t \in T\}$ con cada $X_t: \Omega \rightarrow R$, es una variable aleatoria. Si T es discreto se llama proceso estocástico en tiempo discreto o Cadena. Si T es continuo se denomina proceso estocástico en tiempo continuo

Series de Tiempo y cointegración.

Las series de tiempo son una herramienta estadística que permite modelar procesos en los cuales la dependencia temporal es su ingrediente principal. Su aplicación resuelve dos problemáticas:

1. Modelan adecuadamente el comportamiento de un conjunto de datos.
2. Estimar pronósticos con una precisión adecuada

Las series de tiempo se encuentran fuertemente sustentadas por la teoría de los procesos estocásticos. De hecho, una serie de tiempo se define como una trayectoria u observación de un proceso estocástico.

En la década de los 30's, el trabajo de George Yule y Gilbert Walker provocó una revolución en el análisis de las series de tiempo, pues se introdujo el proceso autorregresivo. Posteriormente surgió el modelo de medias móviles.

George E.P. Box y Gwilym Jenkins consolidaron la aplicación de las series de tiempo al introducir la metodología ARMA en la década de los 80's, la cual impulsó a su vez el desarrollo de nuevos modelos, como es el caso del modelo GARCH para las series de tiempo financieras.

Sin embargo, en el caso de las series de tiempo económicas, el ajuste de modelos ARMA o ARIMA presenta deficiencias. Los primeros en señalar que la metodología que se había estado implementando en las series económicas no era del todo correcta fueron Clive Granger y Paul Newbold, quienes en 1974 introdujeron el concepto de regresión espuria, para aquellas series económicas que parecían tener una relación directa, pero que en realidad era un factor externo el que provocaba la correlación.

Más adelante, Engle y Granger, presentaron en 1987 su artículo "*Co-integration and error correction: Representation, estimation and testing*", con el cual, introdujeron el concepto de *cointegración* como una posible solución a la modelación de parejas de series de tiempo económicas.

En términos muy generales, se dice que dos series de tiempo no estacionarias están cointegradas si existe una combinación lineal de ellas que genera una serie de tiempo estacionaria. Para determinar si una serie es estacionaria, es necesario aplicar pruebas de raíces unitarias, por lo que las pruebas de Dickey-Fuller son una herramienta indispensable en esta metodología.

El caso más sencillo es en el que la relación de cointegración es entre dos series, pero el concepto se puede generalizar a un número mayor de series, dando pie al concepto de multicointegración.

En 1991, Johansen introdujo una nueva manera de abordar la cointegración, a través de la reparametrización de un modelo VAR y del Vector Error Correction Model. En su método, Johansen presenta un modelo que permite considerar más de una relación de cointegración, lo cual es un avance significativo de la técnica.

En este proyecto se propondrán modelos basados en cointegración para series de tiempo económicas mexicanas.

Series de Tiempo Multivariadas: el VAR.

El modelo Vectorial Autorregresivos de orden p , denotado por VAR(p), representa una generalización para el modelo AR(p). En este contexto se parte de la existencia de k variables de interés y se supone que el comportamiento conjunto de todas las variables se puede explicar por medio de una combinación lineal de las n observaciones registradas en los períodos previos:

$$Y_t = v + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + U_t,$$

donde cada Y_t es un vector k -dimensional que contiene las observaciones de las k variables del proceso, los elementos A_i son matrices de coeficientes tomadas en $R[k \times k]$ y $\{U_t\}$ un proceso estocástico que satisface $E[U_t] = 0$, $E[U_t U_t'] = \Sigma_u$ y $E[U_t U_s'] = 0$ si $t \neq s$.

Los retos al utilizar este modelo radican en la determinación del orden p , así como en la estimación y validación de los supuestos.

Se espera como parte de este proyecto, la aplicación de este modelo para comprender la dinámica de la productividad del Manglar, en sitios localizados en la zona de Mahahual.

Procesos de Markov.

Una cadena de Markov en tiempo discreto se define como una sucesión $\{X_n \mid n = 0, 1, 2, \dots\}$ de variables discretas, que satisfacen la propiedad de Markov, es decir, dado X_0, \dots, X_n la variable aleatoria X_{n+1} depende solamente de X_n , es decir,

$$P[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = P[X_{n+1} = j \mid X_n = i]$$

Una cadena de Markov en tiempo continuo es un proceso estocástico, $X = \{X_t: t \geq 0\}$ tal que $\forall t \geq 0, s \geq 0$,

$$P[X_{s+t} = j \mid X_s = i; X_u : 0 \leq u < s] = P[X_{s+t} = j \mid X_s = i].$$

En este proyecto se pretende realizar una revisión del estado del arte, acerca de la inferencia para Procesos de Markov, y el proceso de Markov escondido para luego aplicarlo a dos distintas áreas:

- Tráfico Web
- Patrones en actividades de la vida diaria en adultos mayores

Trafico Web.

El proceso de Poisson Modulado o MMPP (Markov Modulated Poisson Process) es un proceso de Poisson cuya tasa está gobernada por una cadena de Markov a tiempo continuo, irreducible y con espacio de estados finito.

El MMPP ha sido utilizado para modelar el comportamiento del tráfico web, en el cual se estudia la llegada de paquetes a un enrutador, proporcionando una herramienta para la planeación de redes de comunicación (Muscariello, Mellia, Meo, Marsan, y Cigno, 2005). Debido a esto, es de gran interés poder proporcionar estimaciones de los parámetros que caracterizan a este proceso.

Patrones en actividades de la vida diaria en adultos mayores

En el contexto del hogar del adulto mayor independiente, es particularmente importante contar con plataformas que faciliten o adapten el ambiente a las necesidades con el objetivo de mejorar su calidad de vida. Es así como surge, la Vida Asistida por el Entorno (Ambient Assisted Living AAL) que es un área de estudio que surgió con el propósito de ayudar a los usuarios en su vida diaria (De Paola 2017)

Para lograr un hogar o ambiente que se adapte a las necesidades de las personas de la tercera edad, se requiere detectar o reconocer las Actividades de la Vida Diaria (Activities of Daily Living ADL). Algunos ejemplos de las ADL son dormir, preparar el desayuno, ir al baño, entre otros, (Lotfi 2015).

El interés por modelar el comportamiento del ser humano involucra el reconocimiento de patrones, en este caso, de las actividades de la vida diaria humana. El reconocimiento de patrones tiene el objetivo de detectar un patrón común dentro de un número de conjunto de datos. El principal motivo de utilizar el reconocimiento de patrones en el modelado del comportamiento humano es por la precisión. Los algoritmos con mayor utilidad en el reconocimiento de patrones son las redes neuronales artificiales y los modelos ocultos de Markov (Ma'arif 2017), Wei (2011).

Hasta el momento, los trabajos realizados en estos tres problemas, carecen de cierta formalidad en algunos aspectos de su aplicación, por lo que serán un reto a superar en este proyecto. Específicamente, se pretende atender las siguientes debilidades:

- Determinación del número de estados en la cadena de Markov subyacente en el caso del MMPP y sus aplicaciones.
- Metodología para descomponer el día en intervalos de tiempo, que presenten un comportamiento de Cadena de Markov homogénea dentro de cada intervalo, para así describir de manera adecuada los patrones de actividades de la vida diaria en adultos mayores.
- Proponer metodología para la comparación de cadenas de Markov

3. Justificación

Los trabajos que se realizarán bajo este proyecto están enfocados en el desarrollo y aplicación de la estadística en problemas relacionados con:

- Estudios del patrón de actividades del adulto mayor en la vida asistida por su entorno
- Modelado de series de tiempo bajo cointegración o en forma multivariada, que permita generar mejores predicciones en los valores de las series de tiempo estudiadas.

Por tanto, la realización del proyecto proporcionará soluciones bien fundamentadas a varios problemas de actualidad y que requieren de mayor trabajo estadístico.

4. Objetivo

Objetivo General: Desarrollar y/o aplicar métodos estadísticos relacionados con series de tiempo y Procesos de Markov.

Objetivos específicos:

1. Aplicar los modelos basados en Procesos de Markov, considerando las distintas formas en que se presenta la muestra aleatoria.
2. Revisión bibliográfica de la metodología de la teoría inferencial reciente de los modelos de Markov escondidos y sus diferentes aplicaciones. Elaboración de un compendio de lo revisado.
3. Aplicar modelos de series de tiempo multivariados (VAR).
4. Aplicar las técnicas de cointegración al modelado de series de tiempo.

5. Métodos.

Para alcanzar los objetivos se presente realizar lo siguiente:

- Revisiones bibliográficas para conocer el estado del arte de los problemas
- Involucrar a estudiantes de Licenciatura y Posgrado mediante trabajos de Tesis
- Realizar ponencias con los avances obtenidos
- Desarrollar la teoría inferencial requerida, basándose en la función de verosimilitud y otras metodologías
- Implementación de la Metodología desarrollada utilizando el lenguaje R.

6. Metas.

Al terminar el primer año de este proyecto se esperan los siguientes productos:

- Artículo de Aplicación del modelo VAR en la productividad del Manglar en la Zona de Mahahual
- Artículo de divulgación sobre la aplicación del Proceso de Poisson Modulado al tráfico web
- Artículo de divulgación sobre el estado del arte en la Inferencia Estadística en Cadenas de Markov
- Tesis de Maestría sobre el modelado de los patrones de actividades del adulto mayor utilizando cadenas de Markov
- Tesis de Licenciatura sobre el tema de Cointegración.

7. Integrantes:

- José Luis Batún Cutz (FMAT-UADY)
- Ernesto Antonio Guerrero Lara (FMAT-UADY)
- María Diódora Kantún Chim (FMAT-UADY)
- Henry Gaspar Pantí Trejo (FMAT-UADY)
- Jorge Armando Argáez Sosa (FMAT-UADY)
- Ricardo Torres Lara (DCI-UQROO)

Financiamiento.

Este proyecto está parcialmente financiado por el proyecto de Ciencia Básica Conacyt CB-2015-01/252996, MODELOS CON ESTRUCTURAS DE DEPENDENCIA Y SUS APLICACIONES II

Bibliografía:

- Argaez, J. Batún, J., Guerrero, E., Kantun, D., Medina, S., Pantí, H. (2014) Un paseo por el modelo GARCH y sus variantes. *Abstraction & Application* 10 (2014) 35 – 50
- Bhat, U. N. (1972). *Elements of Applied Stochastic Processes*, 3ra ed., New York: J. Wiley (Clásico)
- Basawa, I.V., Prakasa Rao, B.L.S (1980) *Statistical Inference for Stochastic Processes*, London: Academic Press Inc.
- Bosq, Q., Nguyen, H. (1996) *A course in Stochastic Process: Stochastic Models and Statistical Inference*, Springer
- Bolivar C. Addy, Notario Carmen y Pérez Aroldo. (2016). Modelo de Markov para la

- trayectoria académica de estudiantes de la UJAT. *Miscelánea Matemática*, vol.62, pp.29-43
- Box, G., Jenkins, G. (2016). *Time series analysis: Forecasting and Control*. Wiley. Fifth edition.
 - Chan, R., Medina, S., Kantun, D., Colorado, L. (2017) Comparación y evaluación de modelos GARCH con errores de distribución normal en el modelado del IPyC. *Abstraction & Application* 18 (2017) 1 – 17
 - De Paola, A., Ferraro, P., Gaglio, S., Re, G., Morana, M., Ortolani, M., and Peri, D. (2017) An ambient intelligence system for assisted living. In 2017 AEIT International Annual Conference, pages 1-6.
 - Engle, R., Granger, C. (1987). Co-Integration and error correction: Representation, estimation and testing. *Econometrica: Journal of the Econometric Society*. 59. 1551-1580.
 - Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society*. 55. 251-276.
 - Ma'arif, M. (2017) Revealing daily human activity pattern using process mining approach. In 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pages 1-5.
 - Moody, V., Ducloux, K. (2014) Application of Markov Chains to Analyze and Predict the Mathematical Achievement Gap between African American and White American Students. *J Appl Computat Math* 3: 161 doi:10.4172/2168-9679.1000161
 - Muscariello, L., Mellia, M., Meo, M., Marsan, M. A., y Cigno, R. L. (2005). Markov models of internet traffic and a new hierarchical MMPP model. *Computer communications*, 28 (16), 1835-1851.
 - Lotfi, A, Langensiepen, C., and Elbayoudi, A. (2015) . Interpretation of behavior evolution in activities of daily living. In *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pages 84:1-84:2, New York, NY, USA.
 - Rajarshi, M. B. (2013) *Statistical Inference for Discret Time Stochastic Processes*, New Delhi: Springer
 - Rao, M. M. (2014) *Stochastic Processes-Inference Theory*, 2^a Ed., Springer
 - Ross, S. M. (2014). *Introduction to probability models*. (11 Ed.). New York: Academic Press
 - Rincón, L. (2012). *Introducción a los procesos estocásticos*. México: Las Prensas de Ciencias, UNAM.
 - Wei, H., He, J., Tan, J. (2011) Layered hidden Markov models for real-time daily activity monitoring using body sensor networks. *Knowl Inf Syst* (2011) 29:479–494
 - Yin, J., Zhang, Q., Karunanithi, M. (2015) Unsupervised Daily Routine and Activity Discovery in Smart Homes. In *EMBC 2015; Aug 25 ~ Aug 29; Milan, Italy. IEEE; 2015*. 72.

Febrero de 2019

Elaboró: Dr José Luis Batún Cutz