CAPÍTULO DE LIBRO - XV

Métodos inferenciales sobre cadenas de Markov en tiempo discreto con espacio de estados finito

Inferential methods on discrete-time Markov chains with finite state space

Métodos inferenciais sobre cadeias de Markov em tempo discreto com espaço de estados finito

Jhonier Rangel
UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ – BOGOTÁ, COLOMBIA
jrangelg@ecci.edu.co (correspondencia)
https://orcid.org/0000-0002-6849-5551

DOI: https://doi.org/10.35622/inudi.c.02.15

Resumen

El presente ensayo explora algoritmos computacionales para la estimación de parámetros, ampliando así el entendimiento y la implementación práctica de estos métodos en la investigación y la resolución de problemas en distintos ámbitos. La simulación estocástica se presenta como una herramienta indispensable para explorar la dinámica temporal de las cadenas de Markov. Su capacidad para generar múltiples trayectorias y evaluar diversos escenarios proporciona una visión rica y detallada de los procesos subyacentes. Además, los estimadores por momentos, basados en propiedades estadísticas de las observaciones, ofrecen una perspectiva analítica sólida para la estimación de parámetros mediante el cálculo de momentos empíricos. La máxima verosimilitud, como enfoque estadístico riguroso, busca los valores de los parámetros que maximizan la probabilidad de observar datos reales, brindando estimaciones precisas y eficientes. Este conjunto de métodos, convergiendo en un enfoque integral, facilita la inferencia en cadenas de Markov, proporcionando una comprensión más profunda de su comportamiento y su aplicabilidad en diversos contextos científicos y tecnológicos.

Palabras clave: Markov, estocástico, sistema, estimación, simulación.

Abstract

The present essay explores computational algorithms for parameter estimation, thus expanding the understanding and practical implementation of these methods in research and problem-solving across different domains. Stochastic simulation emerges as an indispensable tool for exploring the temporal dynamics of Markov chains. Its ability to generate multiple trajectories and evaluate various scenarios provides a rich and detailed insight into underlying processes. Additionally, moment estimators, grounded in statistical properties of observations, offer a solid analytical perspective for parameter estimation through the calculation of empirical moments. Maximum likelihood, as a rigorous statistical approach, seeks parameter values that maximize the probability of observing real data, providing precise and efficient estimations. This collection of methods, converging into a comprehensive approach, facilitates inference in Markov chains, offering a deeper understanding of their behavior and applicability in various scientific and technological contexts.

Keywords: Markov, stochastic, system, estimation, simulation**.**

Resumo

O presente ensaio explora algoritmos computacionais para a estimação de parâmetros, ampliando assim o entendimento e a implementação prática desses métodos na pesquisa e na resolução de problemas em diferentes áreas. A simulação estocástica surge como uma ferramenta indispensável para explorar a dinâmica temporal das cadeias de Markov. Sua capacidade de gerar múltiplas trajetórias e avaliar diversos cenários proporciona uma visão rica e detalhada dos processos subjacentes. Além disso, os estimadores de momentos, baseados em propriedades estatísticas das observações, oferecem uma perspectiva analítica sólida para a estimação de parâmetros por meio do cálculo de momentos empíricos. A máxima verossimilhança, como uma abordagem estatística rigorosa, busca os valores dos parâmetros que maximizam a probabilidade de observar dados reais, fornecendo estimativas precisas e eficientes. Este conjunto de métodos, convergindo para uma abordagem abrangente, facilita a inferência em cadeias de Markov, oferecendo uma compreensão mais profunda de seu comportamento e aplicabilidade em diversos contextos científicos e tecnológicos.

Palavras chave: Markov, estocástico, sistema, estimativa, simulação.

INTRODUCCIÓN

Las cadenas de Markov en tiempo discreto constituyen un marco matemático esencial en la teoría de probabilidad, utilizadas históricamente para modelar sistemas que evolucionan en pasos discretos a través de estados definidos. Estos modelos tienen diversas aplicaciones; por ejemplo, en marketing, Styan (1964) describe cómo aplicar cadenas de Markov en estrategias de marketing. En medicina, Pegels (1970) presenta una aplicación sobre el estado de la sangre en pacientes. También se presentan aplicaciones actuales, como la de Lin (2021), que evalúa el estado de la velocidad de un motor. Este enfoque encuentra aplicaciones significativas en una variedad de disciplinas, desde ingeniería y finanzas hasta biología y aprendizaje automático. Sin embargo, en dichas aplicaciones no se ha explotado completamente el potencial inferencial que poseen estos modelos y que aporta significativamente a la investigación.

En ingeniería de comunicaciones, las cadenas de Markov se emplean para analizar la transmisión de datos en canales ruidosos y protocolos de red. En el ámbito financiero, se utilizan para modelar la dinámica de precios en mercados y gestionar riesgos. En biología, estas cadenas son fundamentales para entender la evolución de poblaciones y la propagación de enfermedades. Además, en aprendizaje de máquina, los modelos ocultos de Markov son esenciales en reconocimiento de patrones y procesamiento del lenguaje natural. Por ejemplo, Mo (2022) describe una interesante aplicación de las cadenas de Markov en sistemas de comunicación.

En la investigación de operaciones, las cadenas de Márkov se aplican en la optimización de sistemas y la toma de decisiones, por ejemplo "Describimos nuestro algoritmo para identificar sitios de unión de factores de transcripción (TFBS) a través de Optimización de la cadena de Markov. La clave es un algoritmo para ordenar a la cadena de Markov que capture la dependencia más significativa entre posiciones en el enlace del objetivo secuencia." (Yang, 2012, p. 4), también se usa para abordar problemas como la planificación de inventarios y la gestión de recursos. Es así que, estas cadenas proporcionan un marco robusto para modelar y entender procesos estocásticos en distintos contextos, ofreciendo una herramienta valiosa para analizar, simular y optimizar sistemas complejos que evolucionan en pasos discretos en el tiempo.

Como se ha indicado anterior mente las cadenas de Markov en tiempo discreto son un marco teórico esencial en la teoría de probabilidad y la modelización estocástica. Estas cadenas son utilizadas para describir sistemas que evolucionan a lo largo de una secuencia de instantes discretos en el tiempo, donde la transición de un estado a otro está determinada por probabilidades condicionales. Un aspecto clave de las cadenas de Markov es su propiedad de "sin memoria", lo que significa que la probabilidad de moverse a un estado futuro depende únicamente del estado actual y no de la secuencia de eventos previos.

La representación matemática de las cadenas de Markov según Saghafian (2018) se realiza a través de la matriz de cambio de estados, comúnmente denotada como P, que describe las probabilidades de transición entre los distintos estados en un solo paso temporal. Esta matriz es fundamental para entender la dinámica a corto plazo de la cadena. Además, las cadenas de Markov exhiben un comportamiento a largo plazo que se caracteriza por la distribución límite, π, que representa la distribución estacionaria a medida que la cadena evoluciona indefinidamente.

En conjunto, las cadenas de Markov proporcionan un marco versátil y poderoso para modelar y analizar sistemas dinámicos y estocásticos en una amplia gama de disciplinas, desde ingeniería y finanzas hasta biología y aprendizaje automático. Zucchini (2009) explica que su capacidad para capturar la evolución temporal de un sistema mediante probabilidades de transición y describir su comportamiento estacionario las convierte en una herramienta valiosa para abordar problemas complejos en la modelización y simulación de sistemas dinámicos.

Una característica importante intrínseca a un modelo basado en una cadena de Markov es la distribución estacionaria, "Suponga que tenemos N estados en nuestro sistema. La matriz de cambio de estados, comúnmente denotada como P, es una matriz cuadrada N×N donde cada elemento Pij representa la probabilidad de transición del estado i al estado j. Es importante destacar que cada fila de la matriz P suma a 1, ya que las probabilidades de transición desde un estado dado deben sumar 1" (Meyer, 1994, p. 1). El autor explica que es necesario tener la matriz de transición antes de proceder al cálculo de la distribución estacionaria haciendo uso del contexto. Sin embargo, en la vida real en muchas ocasiones nos encontramos con una observación parcial del proceso estocástico y con estos datos se quiere realizar inferencia sobre los parámetros.

Por otro lado, teniendo la matriz de transición se hace necesario el cálculo de la distribución límite (Meyn, 1993) resalta que este es un concepto clave que describe el comportamiento a largo plazo de la cadena de Markov. Para hacer cálculo de dicha distribución se puede seguir el siguiente proceso:

> a largo plazo la distribución de probabilidad sobre los estados deja de cambiar en su evolución temporal. Denotemos esta distribución límite como π, y es única para una cadena de Markov irreducible y aperiódica. La distribución límite π se puede encontrar resolviendo la ecuación π P= π, donde P es la matriz de cambio de estados. Esta ecuación establece que la distribución límite es invariante cuando la cadena se encuentra en equilibrio (Castañeda, 2012, p. 369).

El propósito de este estudio consiste en llevar a cabo un análisis comparativo entre los métodos inferenciales clásicos de máximo verosimilitud y el método de los momentos aplicados a cadenas de Markov con espacio de estados finito y tiempo discreto. Durante este análisis, se destacarán algunas limitaciones inherentes a estos enfoques tradicionales, lo cual servirá como base para la exploración y justificación de la aplicación de inferencias bayesianas o no paramétricas en estos modelos. Este abordaje busca ofrecer una perspectiva más amplia y robusta, superando las posibles desventajas identificadas en los métodos clásicos, y promoviendo así un entendimiento más profundo de la dinámica subyacente en las cadenas de Markov consideradas.

DESARROLLO

La aplicación de la inferencia clásica a las cadenas de Markov requiere una sólida revisión de conceptos fundamentales de la estadística matemática. Entre estos conceptos, los momentos poblacionales desempeñan un papel esencial al proporcionar medidas descriptivas de la forma y tendencia central de una distribución, resultandos fundamentales para comprender la dinámica de las transiciones entre estados en las cadenas de Markov.

Además, la familiaridad con las distribuciones multinomial y binomial es crucial, ya que estas modelan los eventos en el contexto de múltiples categorías o resultados binarios, respectivamente, aspectos que pueden ser inherentes a las cadenas de Markov. La aplicación de métodos de estimación, tales como el método de momentos y el método de máximo verosimilitud, se revela como un paso indispensable para determinar los parámetros que caracterizan la evolución temporal de la cadena. Complementariamente, el método de Montecarlo, mediante la generación de muestras aleatorias, proporciona un enfoque valioso para simular el comportamiento de la cadena y obtener resultados estadísticos robustos. En conjunto, este conjunto de herramientas estadísticas provee la base necesaria para explorar y comprender a fondo las cadenas de Markov, facilitando así una inferencia rigurosa sobre su comportamiento en diversas situaciones.

En Zhang (2020) se explica que el primer momento es la esperanza matemática defina como sigue:

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx \\ \sum_{x} xP(X = x) \end{cases} \qquad (1)$$

El primer momento describe la centralidad de la distribución, también en (Zhang, 2020) se explica la necesidad del segundo momento como fuente de variabilidad de la variable aleatoria, dicho momento centrado en la esperanza se define como:

ISBN: 978-612-5130-24-2

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx \\ \sum_{x} (x - E(X))^2 P(X = x) \end{cases} \tag{2}$$

Estos dos momentos juegan un papel crucial al realizar inferencia, y son fundamentales para llevar a cabo la estimación. En el proceso de inferencia, se inicia recordando el método de momentos como una herramienta valiosa para realizar estimaciones. Andrews (2010) destaca la robustez de este método bajo la condición de que se cumplan los supuestos distribucionales necesarios. Es importante señalar que el método de momentos opera mediante la igualación de momentos teóricos y empíricos. En este contexto particular, nos limitamos a considerar la estimación de la esperanza y la varianza. Este enfoque, al centrarse en estos momentos específicos, simplifica el proceso de inferencia al abordar directamente las características centrales de la distribución en cuestión. Así, al aplicar el método de momentos a la estimación de la esperanza y la varianza, se establece un marco robusto para la inferencia en el contexto de cadenas de Markov con espacio de estados finito y tiempo discreto.

$$\hat{\mu} = \frac{\sum_i xi}{n} \tag{3}$$

$$\widehat{\sigma^2} = \frac{\sum_i (xi - \bar{x})^2}{n - 1} \tag{4}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_i (xi - \bar{x})^2}{n - 1}} \tag{5}$$

En las cadenas de Markov la inferencia propuesta bajo el método de los momentos consiste en: hallar las frecuencias de transición entre los estados de la cadena de Markov. Estas frecuencias representan el número de veces que la cadena pasa de un estado a otro durante un conjunto de observaciones, luego se definimos los momentos empíricos en términos de estas frecuencias de pasos. Por ejemplo, el primer momento empírico puede representar la media de las transiciones entre estados, y el segundo momento empírico puede corresponder a la varianza de estas transiciones.

Otro método importante a la hora de realizar estimación es el método de máximo verosimilitud, según Li (2020, p. 4) el método de máximo verosimilitud se define como:

$$\hat{\theta} := argmáx_{\theta\epsilon\Theta}\{L(\theta|X_1, \dots, X_n)\} \qquad (6)$$

Suponga que la muestra aleatoria es una familia de variables aleatorias independientes e idénticamente distribuidas.

$$\hat{\theta} := argmáx_{\theta\epsilon\Theta}\{\prod_i f(x_i|\theta)\} \qquad (7)$$

Se invita al lector demostrar que bajo la distribución binomial el estimador por el método de los momentos y por máximo verosimilitud coincide y es:

$$\widehat{\pi_j} = \frac{n_j}{n} \qquad (8)$$

A su vez bajo un estado en el tiempo t, la estimación de pasar a otro estado por el método de los momentos es:

$$\widehat{p_{ij}} = \frac{n_{ij}}{n_{i.}} \qquad (9)$$

Una pregunta muy interesante que nos planteamos es: ¿ p_{ij} *es estadísticamente significativo* ?, luego planteamos el siguiente sistema de hipótesis:

$$\begin{cases} H_0: p_{ij} = 0 \\ H_a: p_{ij} > 0 \end{cases} \qquad (10)$$

Este contraste de hipótesis puede ser sometido bajo diversos métodos, el que será considerado en este caso es la prueba bajo teorema de límite central expuesto en Wasserman (2020).

ISBN: 978-612-5130-24-2

$$z = \frac{\widehat{p_{ij}}}{\sqrt{\frac{\widehat{p_{ij}}(1-\widehat{p_{ij}})}{n_{i.}}}} \qquad (11)$$

Si z es mayor al cuantil de la distribución normal de probabilidad correspondiente a 1 - ὰ, se rechaza la hipótesis nula. Esta prueba presenta algunos problemas de consistencia, los cuales se explican en Bonkhoff (2020) o en Linde (2023). Aunque este sea un problema, requiere un estudio minucioso y se deja como un tema abierto a discusión.

La necesidad de realizar inferencia sobre los parámetros de una cadena de Markov radica en obtener un modelo más parsimonioso. Recordemos que el criterio de parsimonia sugiere que un modelo es mejor cuando maximiza la información retenida al minimizar el número de parámetros. Al considerar la cantidad de parámetros de una cadena de Markov, observamos que se deben estimar k*k parámetros, lo cual es una cantidad bastante significativa. Si utilizamos métodos inferenciales para evaluar qué parámetros son significativos, reduciremos el número de parámetros a estimar. Para profundizar en el criterio filosófico de parsimonia, se puede consultar (Sober, 1981).

Para la simulación se debe tener cuidado ya que globalmente la simulación que haremos no es sobre solo una distribución dada. En la siguiente tabla se expone el algoritmo en seudocódigo para realizar inferencia sobre una cadena de Márkov. Si desea profundizar las reglas generales del seudocódigo pue ver (Oda, 2015). Para simular la cadena se sigue el siguiente código:

Tabla 1

Pseudocódigo para realizar la simulación de una Cadena de Markov

Algoritmo: Simulación de Cadena de Markov
Entradas: - Matriz de Transición (matriz_transicion) - Estado Inicial (estado_actual) - Número de Pasos (num_pasos)
Salida: - Secuencia de Estados Simulados
1. Definir la Matriz de Transición: - matriz_transicion: matriz NxN con probabilidades de transición.
2. Definir el Estado Inicial: - estado_actual: elegir un estado inicial.
3. Número de Pasos de Simulación: - num_pasos: determinar la cantidad de pasos a simular.
4. Simulación: - Para cada paso desde 1 hasta num_pasos: a. Imprimir o almacenar el estado actual. b. Elegir el próximo estado basado en la matriz de transición y actualizar estado actual.

En esta tabla se expone el pseudocódigo para realizar la simulación de una cadena de Markov. Este pseudocódigo puede adaptarse a cualquier lenguaje de programación, como R o Python. Para ver la implementación en R puede ver (Rangel, 2023) y en Python ver (Rangel, 2023).

Observemos que la generación de aleatorios pide un estado inicial, esto se debe a que por definición, los procesos markovianos se definen bajo la distribución condicional, note que dicha distribución condicional si se comporta como un vector aleatorio aleatoria multinomial del tamaño del conjunto de estados de la cadena, esta relación la puede ver en (Zhou, 2009).

Otra perspectiva más ajustada a la realidad y de interés es la exploración del método bayesiano, este trabajo solo considera inferencia clásica sin embargo se puede adaptar la metodología propuesta por (Kelter, 2020). Se destaca la naturaleza secuencial de la estadística bayesiana en el aprendizaje estadístico, según se discute en (Waldmann, 2022). Este enfoque se plantea como un trabajo futuro de gran interés, ya que integra tres aspectos fundamentales de la estadística: procesos estocásticos, estadística matemática para la estimación de parámetros y pruebas de hipótesis, así como la estadística bayesiana, que permite incorporar experiencia pasada para realizar predicciones con los datos. En ámbitos financieros, como el modelo de Markowitz, ya se han implementado mecanismos bayesianos, como se evidencia en (Chen, 2020). Como perspectiva para futuros estudios, se propone realizar un análisis comparativo entre el estadístico de significancia propuesto en este trabajo y el intervalo de credibilidad propuesto por la estadística bayesiana.

CONCLUSIONES

En este trabajo, se propone el uso de métodos de estadística clásica para la estimación y simulación de una cadena de Markov en tiempo discreto con espacio de estados finito. Además, se abordan algunas dificultades asociadas a los métodos clásicos, lo cual subraya la necesidad de recurrir a la estadística bayesiana para la estimación. Dos motivos principales respaldan esta elección: en primer lugar, la naturaleza secuencial de los procesos estocásticos y el enfoque de aprendizaje que ofrece el método bayesiano.

Adicionalmente, se propone la realización de pruebas de significancia clásica con el objetivo de reducir el número de parámetros, cumpliendo así el criterio de parsimonia. Este principio sugiere que un modelo es considerado mejor si logra explicar de manera óptima los datos con la menor cantidad de parámetros posible. La aplicación de pruebas de significancia clásica permitirá evaluar la relevancia estadística de cada parámetro, contribuyendo a la simplificación del modelo y facilitando su interpretación. Esta estrategia no solo busca mejorar la eficiencia del modelo, sino también proporcionar una base más sólida para la toma de decisiones y la generalización de los resultados en diferentes contextos.

REFERENCIAS BIBLIOGRÁFICAS

Andrews, D. W., & Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica, 78*(1), 119-157. https://doi.org/10.3982/ECTA7502

Bonkhoff, A. K., Hope, T., Bzdok, D., Guggisberg, A. G., Hawe, R. L., Dukelow, S. P., ... & Bowman, H. (2020). Bringing proportional recovery into proportion: Bayesian modelling of post-stroke motor impairment. *Brain, 143*(7), 2189-2206. https://doi.org/10.1093/brain/awaa146

Castañeda, L. B., Arunachalam, V., & Dharmaraja, S. (2012). *Introduction to probability and stochastic processes with applications.* John Wiley & Sons. https://doi.org/10.1002/9781118344972

Chen, S. D., & Lim, A. E. (2020). A Generalized black–litterman model. *Operations research, 68*(2), 381-410. https://doi.org/10.1287/opre.2019.1893

Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology, 20*(1), 1-12. https://doi.org/10.1186/s12874-020-00980-6

Li, M., & Liu, X. (2020). Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. *International Journal of Control, Automation and Systems*, *18*(6), 1581-1592. https://doi.org/10.1007/s12555-019-0191-5

Lin, X., Zhang, G., & Wei, S. (2021). *Velocity prediction using Markov Chain combined with driving pattern recognition and applied to Dual-Motor Electric Vehicle energy consumption evaluation*. Applied Soft Computing, 101, 106998. https://doi.org/10.1109/eSmarTA52612.2021.9515733

Linde, M., Tendeiro, J. N., Selker, R., Wagenmakers, E. J., & van Ravenzwaaij, D. (2023). Decisions about equivalence: A comparison of TOST, HDI-ROPE, and the Bayes factor. *Psychological Methods*, *28*(3), 740. https://psycnet.apa.org/doi/10.1037/met0000402

Meyer, C. D. (1994). Sensitivity of the stationary distribution of a Markov chain. *SIAM Journal on Matrix Analysis and Applications, 15*(3), 715-728. https://doi.org/10.1137/S0895479892228900

Mo, J. (2022). *Performance modeling of communication networks with Markov chains*. Springer Nature. https://doi.org/10.1109/TCOMM.2003.809787

Meyn, S., & Tweedie, R. L. (1993). *Markov Chains and stochastic stability.* Springer eBooks. https://doi.org/10.1007/978-1-4471-3267-7

Oda, Y., Fudaba, H., Neubig, G., Hata, H., Sakti, S., Toda, T., & Nakamura, S. (2015). Learning to generate pseudo-code from source code using statistical machine translation. In *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 574-584). IEEE. https://doi.org/10.1109/ASE.2015.36

Pegels, C. C., & Jelmert, A. E. (1970). An evaluation of blood-inventory policies: A Markov chain application. *Operations Research*, *18*(6), 1087-1098. https://doi.org/10.1287/opre.18.6.1087

Rangel, J. (2023). *Estimación de los parámetros de una cadena de Markov*. Recuperado el 10 de enero de 2023 de https://rpubs.com/JhonierRangel/1127722

Rangel, J. (2023). *Simulación cadena de Markov en tiempo discreto*. Recuperado el 10 de enero de 2023 de https://cutt.ly/pwN3jY78

Saghafian, S. (2018). Ambiguous partially observable Markov decision processes: Structural results and applications. *Journal of Economic Theory*, 178, 1-35. https://doi.org/10.1016/j.jet.2018.08.006

Sober, E. (1981). The principle of parsimony. *The British Journal for the Philosophy of Science*, *32*(2), 145-156. https://doi.org/10.2307/2413239

Styan, G. P., & Smith Jr, H. (1964). Markov chains applied to marketing. *Journal of Marketing Research*, *1*(1), 50-55. https://doi.org/10.1177/002224376400100109

Waldmann, M. R., & Martignon, L. (2022). A Bayesian network model of causal learning. *Proceedings of the twentieth annual conference of the Cognitive Science Society* (pp. 1102-1107). Routledge. https://doi.org/10.4324/9781315782416-198

Wasserman, L., Ramdas, A., & Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, *117*(29), 16880-16890. https://doi.org/10.1073/pnas.1922664117

Yang, X., Zheng, X. Q., & Lv, L. N. (2012). A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, 233, 11-19. https://doi.org/10.1016/j.ecolmodel.2012.03.011

Zhang, X., Low, Y. M., & Koh, C. G. (2020). Maximum entropy distribution with fractional moments for reliability analysis. *Structural Safety*, 83, 101904. https://doi.org/10.1016/j.strusafe.2019.101904

Zhou, H., & Lange, K. (2009). Composition Markov chains of multinomial type. *Advances in Applied Probability*, 41(1), 270-291. https://doi.org/10.1239/aap/1240319585

Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov models for time series*. Chapman and Hall/CRC eBooks. https://doi.org/10.1201/9781420010893