

# Introducción al análisis de datos Bayesiano

Héctor Nájera y Curtis Huffman

2023-01-31

## Descripción del curso

El curso provee una introducción coherente a los métodos y conceptos bayesianos con énfasis en el análisis de datos. Cubre adecuadamente las bases de la materia que consideramos esenciales para la investigación empírica en economía y ciencias sociales en general.

Discutiremos los fundamentos de la estadística bayesiana y cómo hacer inferencia de problemas reales y relevantes para la economía basada en simulaciones (aproximaciones numéricas a integrales complejas) usando técnicas de Monte Carlo vía cadenas de Markov. Usaremos esas simulaciones para analizar varios modelos estadísticos, discutir técnicas para aquilatar modelos y su comparación, escoger distribuciones a priori (priors), análisis de sensibilidad y el concepto de modelos jerárquicos.

El curso pretende cubrir los temas previstos de manera profunda, pero intuitiva y relativamente poco técnica frente a la complejidad de algunas soluciones analíticas de las integrales involucradas. En ese sentido el curso está más cargado al material relativo al cómputo y la interpretación de los resultados. Con estas metas en mente, hemos seleccionado un conjunto de temas de fundamental importancia para una comprensión de las técnicas que permitirán a los alumnos continuar por cuenta propia en el estudio de temas más avanzados.

Ajustar modelos bayesianos requiere un ambiente computacional estadístico. Después de las sesiones introductorias, el curso está pensado para utilizar, rutinariamente, el ambiente R (R-project), paquetería que está ganando popularidad y es usada ampliamente en el modelaje bayesiano. Desde luego, los estudiantes tendrán la libertad de escoger el programa estadístico de su preferencia para dar seguimiento a los ejercicios del curso (JAGS/Stan). En último término, nuestra esperanza es que los alumnos desarrollen un entendimiento sólido e intuitivo de la importancia, significado y utilidad del análisis bayesiano en la investigación social, así como prepararlos para el estudio de los temas de frontera en el campo.

## Objetivos

- El objetivo del curso es introducir conceptos básicos, métodos y paquetería estadística de punta para llevar a cabo inferencia bayesiana.
- Preparar al estudiante con experiencia práctica y nociones metodológicas para aplicar los métodos bayesianos a problemas reales.

- Introducir a los estudiantes al abordaje bayesiano de planteamiento de problemas y toma de decisiones bajo incertidumbre.
- Actualizar el conocimiento de los estudiantes respecto a la construcción de modelos para hacer inferencia probabilística.
- Aumentar la empleabilidad de los estudiantes mediante el desarrollo de habilidades para el uso de paquetería estadística y datos nacionales.
- Desarrollar la capacidad analítica y crítica de los estudiantes en torno a las prácticas actuales de inferencia estadística.

## Prerequisitos y formato

Nuestra intención es ofrecer un curso para cualquiera con cursos básicos cubiertos en probabilidad y estadística, en un formato que incluirá exposiciones frente a pizarrón o a distancia según lo permitan las circunstancias, discusiones en clase, talleres-laboratorios y proyectos. El número de sesiones y el volumen del material están pensados para un curso semestral de 16 semanas (3 horas por semana).

## Bibliografía básica

McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press.

## Calificaciones

- **50%** asistencia y participación.
- **30%** ejercicio 2
- **30%** ejercicio 1.

## Contenidos y lecturas de cada clase

Los estudiantes deberán leer una lectura esencial antes de cada clase. Las lecturas son en español y en inglés.

### Semana 01: Presentación del curso.

*Responsables: Héctor Nájera y Curtis Huffman*

- Tendencias recientes y retos actuales en inferencia estadística
- ¿Por qué estudiar análisis Bayesiano?

### Semana 02: Credibilidad, modelos y posibilidades

*Responsable: Curtis Huffman*

- Capítulo 2: Doing Bayesian analysis. John K. Kruschke

### Semana 03: ¿Qué es probabilidad?

*Responsable: Curtis Huffman*

- Capítulo 4: Doing Bayesian analysis. John K. Kruschke

### Semana 04: Teorema de Bayes: Priors, verosimilitud y posterior

*Responsable: Héctor Nájera*

- Capítulo 5: Doing Bayesian analysis. John K. Kruschke
- Gelman et al. (2017) [The prior can generally only be understood in the context of the likelihood](#)

### Semana 05: Hipótesis a la Bayes v estadística clásica 1.

*Responsable: Curtis Huffman*

- Capítulo 11: Los peligros de los p-values en NHST. Doing Bayesian analysis. John K. Kruschke

### Semana 06: Hipótesis a la Bayes v estadística clásica 2.

*Responsable: Héctor Nájera*

- Capítulo 12: Aproximaciones Bayesianas a la NHST. Doing Bayesian analysis. John K. Kruschke
- Estimación bayesiana le gana al t-test. Kruschke, J. K. Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, American Psychological Association, 2013, 142, 573

## **Semana 07: Muestreo de las distribuciones posteriores 1: MCMC. Metropolis y Gibbs**

*Responsable: Héctor Nájera y Curtis Huffman*

- Capítulo 3: Muestreando el imaginario. Statistical rethinking: A Bayesian course with examples in R and Stan Chapman and Hall/CRC, 2020
- Capítulo 7: Markov Chain Monte Carlo. Doing Bayesian analysis. John K. Kruschke

## **Semana 08: Muestreo de las distribuciones posteriores 2: MCMC, Hamiltonian**

*Responsable: Curtis Huffman*

- Capítulo 14: STAN. Doing Bayesian analysis. John K. Kruschke
- Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo arXiv preprint arXiv:1701.02434, 2017

## **Semana 09: Convergencia de cómputo Bayesiano. Suficiencia y evaluación. GIBBS y HMC**

*Responsable: Curtis Huffman*

- Capítulo 8: MCMC. Statistical rethinking: A Bayesian course with examples in R and Stan Chapman and Hall/CRC, 2020
- Vethari et al. (2021) [Rank-Normalization, Folding, and Localization: An Improved  \$\hat{R}\$  for Assessing Convergence of MCMC \(with Discussion\)](#)

## **Semana 10: Modelos jerárquicos 1. Especificación y escritura de modelos a la Bayes**

*Responsable: Héctor Nájera*

- Chapter 9: Modelación jerárquica. Parámetros
- Estimación bayesiana le gana al t-test

## **Semana 11: Evaluación de modelos**

- Capítulo 6: Sobreajuste, regularización y criterios de información. Statistical rethinking: A Bayesian course with examples in R and Stan Chapman and Hall/CRC, 2020
- Evaluación práctica de modelos usando LOO y WAIC. Vehtari, A.; Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC Statistics and computing, Springer, 2017, 27, 1413-1432
- Capítulo 10: Comparación de modelos y modelación jerárquica. Doing Bayesian analysis. John K. Kruschke

*Responsable: Curtis Huffman*

## Semana 12: Clase práctica. Instalación y estimación con Rstan

*Responsable: Héctor Nájera*

## Semana 13: Clase práctica. Instalación y estimación con Rstan

*Responsable: Héctor Nájera*

## Semana 14: Modelos jerárquicos 2.

*Responsable: Curtis Huffman*

- Capítulo 12: Modelos multinivel. Statistical rethinking: A Bayesian course with examples in R and Stan Chapman and Hall/CRC, 2020
- Capítulo 13: Aventuras en la covarianza. Statistical rethinking: A Bayesian course with examples in R and Stan Chapman and Hall/CRC, 2020

## Semana 15: Flujo de trabajo Bayesiano

*Responsable: Curtis Huffman y Héctor Nájera*

Flujo de trabajo bayesiano. Gelman, A.; Vehtari, A.; Simpson, D.; Margossian, C. C.; Carpenter, B.; Yao, Y.; Kennedy, L.; Gabry, J.; Bürkner, P.-C. & Modrák, M. Bayesian workflow

## Semana 16: Llevando Bayes al siguiente nivel

- Modelos avanzados
- Cómputo
- Datos perdidos

*Responsable: Curtis Huffman y Héctor Nájera*

## Referencias

Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo arXiv preprint arXiv:1701.02434, 2017

Gelman, A.; Vehtari, A.; Simpson, D.; Margossian, C. C.; Carpenter, B.; Yao, Y.; Kennedy, L.; Gabry, J.; Bürkner, P.-C. & Modrák, M. Bayesian workflow arXiv preprint arXiv:2011.01808, 2020

Kruschke, J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press, 2014

Kruschke, J. K. Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, American Psychological Association, 2013, 142, 573

McElreath, R. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2020

Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: An improved Rhat for assessing convergence of MCMC (with discussion) Bayesian analysis, International Society for Bayesian Analysis, 2021, 16, 667-718

Vehtari, A.; Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC Statistics and computing, Springer, 2017, 27, 1413-1432