The Journal of Open Source Education

Course Materials for an Introduction to Bayesian Modeling in Spanish

Osvaldo A Martin [1¶]

1 IMASL-CONICET. Universidad Nacional de San Luis. San Luis, Argentina ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Submitted: 19 December 2023
Published: unpublished

Summary

Modelado Bayesiano: Una introducción al modelado probabilista is a course aimed at advanced undergraduate students or graduate students of STEM with little to no prior knowledge of statistics. The course consists of 8 units and is taught in Spanish. The material has been used under several modalities, including intense 5-day courses, semester-long courses, and both in-person and remote learning. Besides the exercises at the end of each chapter, students are encouraged to explore the examples by changing the values of variables, using their datasets, collaborating with classmates, etc. The material is hosted online using Quarto (Allaire et al., 2022).

Statement of need

Bayesian statistics has emerged as a highly flexible and powerful technique, witnessing increased adoption in recent years. In contrast, there is a lack of widespread courses at the university level specifically addressing Bayesian methods, and when they do they tend to be very heavy on the theoretical side, and/or computational tools are absent or they tend to be outdated. Furthermore, the challenge intensifies when considering the limited availability of resources in Spanish. While English-language resources and courses exist, they are often geared toward first-world audiences and can be prohibitively expensive for students and professionals in Latin America. This linguistic and economic disparity exacerbates the exclusion of a significant portion of the scientific and academic community that could greatly benefit from Bayesian statistics. We believe this course contributes to the democratization of Bayesian statistics by providing a free and accessible resource in Spanish.

Target Audience

The course is aimed at graduate students or advanced undergrad students of STEM careers, with working knowledge of Python or a similar language. In our experience, students with no prior knowledge of Python, but knowledge of high-level languages like R or Julia, can also follow the course with little extra guidance.

Content

The main objective of the course is to provide students with a practical understanding of Bayesian modeling, allowing them to formulate models that address specific scientific questions. It is intended as the first Bayesian course. It seeks to familiarize students with key computational tools, especially PyMC (Abril-Pla et al., 2023), ArviZ (Kumar

JOSE
The Journal of Open Source Education

et al., 2019), Bambi Capretto et al. (2022), and Preliz (Icazatti et al., 2023), to perform Bayesian data analysis effectively. However, the focus is on practical statistical ideas and hence the material could be easily adapted to use a different set of tools even from other programming languages like R or Julia. Additionally, the course content encompasses basic concepts of probability, Bayesian inference, probabilistic programming, generalized linear models, hierarchical models, sampling diagnostics, variable selection, and model selection. All these concepts are discussed under the framework of the Bayesian workflow (Gelman et al., 2020) and notions of exploratory analysis of Bayesian models. In summary, the course is designed to equip students with the skills and knowledge necessary to effectively and critically apply Bayesian statistics.

The full course includes both lectures and practical lab sessions, emphasizing a project-based approach. Encouragement is given for students to utilize their datasets during the course. To complete the course, students are required to work on a final project. Collaborative work in pairs is actively encouraged, but individual projects are also allowed. Students must apply the acquired knowledge to their datasets or find openly available datasets (with guidance from tutors if needed). They must develop at least two alternative models, conduct thorough analyses, perform sampling diagnostics, and articulate and defend their conclusions. Whenever feasible, students are expected to present their results to their peers. This format helps students to apply the concepts to their domain problems and encourages discussion of modeling decisions, conclusions, and motivations for the analysis.

Student feedback has been very positive and many of the examples and problems originally designed for this course have been further developed into the books *Bayesian with Python* (O. Martin, 2018, 2024) and *Bayesian Modeling and Computation in Python* (O. A. Martin et al., 2021), usually the examples are first tested and evaluated with students and then included in the books.

Conclusions

Altogether, this material offers a practical and conceptual introduction to Bayesian modeling. By providing a free and comprehensive resource in Spanish, the course contributes significantly to the democratization of Bayesian statistics, fostering a diverse and inclusive community of learners well-equipped to apply these advanced statistical methods in their scientific endeavors.

Acknowledgments

We would like to thank Pablo Garay for his help proofreading the material and all the students who have taken this course and provided valuable feedback. We also thank the many contributors to the open-source software packages used in this course. We received the support of:

- the National Agency of Scientific and Technological Promotion (ANPCyT), Grant PICT-02212.
- the National Scientific and Technical Research Council (CONICET), Grant PIP-0087.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: A modern and comprehensive probabilistic programming

framework in python. *PeerJ Computer Science*, *9*, e1516. https://doi.org/10.7717/peerj-cs.1516

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2022). *Quarto* (Version 1.2). https://doi.org/10.5281/zenodo.5960048

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). Bambi: A simple interface for fitting bayesian linear models in python. *Journal of Statistical Software*, *103*(15), 1–29. https://doi.org/10.18637/jss.v103.i15

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). *Bayesian workflow*. https://arxiv.org/abs/2011.01808

Icazatti, A., Abril-Pla, O., Klami, A., & Martin, O. A. (2023). PreliZ: A tool-box for prior elicitation. *Journal of Open Source Software*, *8*(89), 5499. https://doi.org/10.21105/joss.05499

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Martin, O. (2018). *Bayesian Analysis with Python: Introduction to Statistical Modeling and Probabilistic Programming Using PyMC3 and ArviZ* (2 edition). Packt Publishing.

Martin, O. (2024). *Bayesian Analysis with Python: A Practical Guide to probabilistic modeling* (3rd Edition). Packt Publishing.

Martin, O. A., Kumar, R., & Lao, J. (2021). *Bayesian Modeling and Computation in Python*. https://doi.org/10.1201/9781003019169