

# Research on Bayesian Statistics Teaching with R

Yanmin Wang<sup>1, \*</sup>, Jing Jin<sup>2</sup>

<sup>1</sup>School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, China

<sup>2</sup>School of Accountancy, Anhui University of Finance and Economics, Bengbu, 233030, China

\*Corresponding author: Yanmin Wang (Email: 120081542@aufe.edu.cn)

---

**Abstract:** Bayesian statistics uses the mathematical rules of probability to combine data with prior information to yield inferences which are always more precise than would be obtained by either source of information alone. The current Bayesian education has not yet attracted enough attention from the education community in China. The rise of MCMC and probabilistic programming languages has profoundly reshaped Bayesian statistics. Compared to precise mathematical analysis, the simulation-based teaching can help students to avoid tedious statistical calculations and cultivate Bayesian thinking.

**Keywords:** Bayesian Statistics; Precise Mathematical Analysis; Simulation-based Teaching.

---

## 1. Introduction

Bayesian statistics is a branch of mathematical statistics named after its founder, Thomas Bayesian (1702-1761). It provides us with mathematical tools to update our beliefs about random events as we see new data or evidence about those events.

The statistics textbooks in Chinese colleges base on the Classical statistical theory -- hypothesis testing, estimation, and the design of experiments and sample surveys -- is mainly the creation of two men: Ronald A. Fisher (1890-1962) and Jerzy Neyman (1894-1981). This theory only utilizes population and sample information. Compared to the Classical statistical theory, Bayesian statistics provides us with a different and perhaps more intuitive perspective: the prior distribution (experience) is updated using current observations to obtain the posterior distribution (new knowledge); the updated knowledge is in turn used as prior information to initiate the next round of learning, and thus. Consequently, Bayesian statistics is considered a statistical learning method. The machine learning applied Bayesian ideas in a variety of tools. The more successful applications are a variety of Bayesian classification models and Bayesian networks and so on.

In the curriculum of statistics majors in China, Bayesian statistics is an edge course. The first National Standard for Teaching Quality of Undergraduate Programs in Colleges, formulated by the Teaching Guidance Committee of Higher Education of the Ministry of Education, does not specify Bayesian Statistics as required courses for statistics majors, but only as an elective course for applied statistics major. However, Bayesian statistics is widely used in both scientific research and industrial application. To promote Bayesian education in China, it is necessary to explore the more efficient method for Bayesian teaching.

## 2. Teaching Guidelines

In Bayesian statistics, integral is always used to obtain the posterior distribution of a parameter or its function. As we know, integral is tedious, especially for more than two parameters. To avoid integral, traditional teaching of Bayesian statistics spends a great deal of time obtaining the exact form of the posterior distribution of the parameters, for

example, by means of conjugate or uninformative prior based on the common distributions. This process is complicated and requires proficiency in calculus, linear algebra, probability theory, and mathematical statistics, as well as a keen statistical intuition for students.

Except for simple problems such as low-dimensional models, it is often difficult, even impossible, to obtain explicit form of the posterior distributions of the parameters. In most cases, we must resort to Markov Chain Monte Carlo (MCMC), which allows one to characterize a distribution without knowing all the distribution's mathematical properties by randomly sampling values out of the distribution. MCMC is useful not only for the incredibly complicated statistical models but also for quite simple statistical models whose Bayesian inference is still analytically intractable. The rise of MCMC in the 1990s reshaped Bayesian statistics. Since then, Bayesian computation has rarely resorted to exact mathematical analysis, but rather to stochastic simulations.

In general, the traditional teaching of Bayesian statistics was mathematics-centered, and the statistical property of the estimators is clear. Modern Bayesian teaching is simulation-centered, where statistical inferences are based on simulation results, and is regarded as simulation-based teaching.

## 3. Curriculum Orientation

My experience on Bayesian teaching suggests that the traditional teaching of Bayesian statistics is not favored for students. The reason is that Bayesian statistics requires students to have a solid foundation in mathematics and statistics. This prevents most students from learning it, even students for statistics majors. The students for non-statistics majors are able to understand the models and their application, however, they have no interest in the details of the derivation behind the Bayesian method. Simulation-based teaching of Bayesian statistics has greatly reduced the difficulty of Bayesian learning by introducing probabilistic programming. As a result, the students only master basic Bayesian and programming knowledge. It can be taught like R or Python with a variety of cases. So, Bayesian statistics should be a methodology course for all natural and social sciences, rather than being restricted to statistics students.

According to Dogucu & Hu (2022), 46 of the top 152 high-ranking institutions in the US have Bayesian courses. Of the

45 courses offered at universities, 60% are cross-listed between undergraduate and graduate programs, meaning that undergraduate students take these courses along with graduate-level peers. And the Bayesian courses are offered as part of a variety of majors, including majors in statistics, mathematics, biology, economics, psychology, and political science, etc. America is on the front rank in Bayesian education.

## 4. Teaching Content

The most popular textbook for Bayesian statistics in China is wrote by Shisong Mao and Yincai Tang. The content of this book have three parts: Bayesian Inference (3 chapters), Bayesian Decision (3 chapters) and Bayesian Computation (1 chapter). The structure deviates from that of popular textbooks in the world. for example, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, Second Edition (by John K. Kruschke), *Bayesian Data Analysis* (by Andrew Gelman et al.), and *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (by Richard McElreath). These books have greatly reduced the content of Bayesian decision and increased the content of Bayesian data analysis. The recently published textbook, *Bayesian Statistics* (Second Edition) by Laisheng Wei and Weiping Zhang, includes only one chapter on Bayesian decision. However, the probabilistic programming language used in that textbook is WinBUGS, which had not been updated since 2007.

To meet the requirement of simulation-based teaching, it is necessary to add the content of probabilistic programming. In recent years, the most popular software for Bayesian inference is JAGS and Stan. Due to Andrew Gelman, Stan seems to have a better future. Therefore, Stan should be recommended to Bayesian teaching. RStan is the R interface to Stan, and there is a "rstan" package as well as other application packages. As a result, one chapter should be added to introduce the programming language and case studies, and another chapter on R is also needed. Accordingly, the class time needs to be added. In my university, Bayesian statistics is scheduled for 36 lessons. There is a need to add 18 lessons hours for laboratory instruction. In addition, examples of real statistical analyses should be added to tell students to perform data mining, inference, and prediction.

## 5. Epilogue and Discussion

Bayesian statistics is an important approach to applying probability to statistical problems and has been a great success in both theory and practice. The key to Bayesian

science is classroom education. Bayesian statistics should be oriented to students for most majors, besides statistics students. Bayesian education has not been paid enough attention in China, and the existing textbooks and traditional teaching has no help to the promotion of Bayesian education.

The traditional teaching focusing on accurate mathematical analysis is only applicable to low-dimensional models. The analysis of high-dimensional models must turn to MCMC. From the application point of view, the center of classroom education should shift from mathematics-based teaching to simulation-based teaching. The latter is much less mathematically demanding, and students only need to master the elements of regression and are not required to master algorithm.

The probabilistic programming languages provide tools for Bayesian modeling, making it possible for students to model with only a basic knowledge of Bayes. It is also helpful for students to develop statistical thinking while freeing them from tedious statistical calculations. Therefore, promoting the teaching of Bayesian statistics based on simulation can promote the popularization of Bayesian statistics in China, which is a more urgent teaching task.

## Acknowledgment

This work is supported by 2020 provincial quality project program offline course "Bayesian statistics" (No. 2020kfk014).

## References

- [1] Dogucu, M., & Hu, J. (2022). The current state of undergraduate Bayesian education and recommendations for the future. *The American Statistician*, 76(4), 405-413.
- [2] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Aki, V., & Rubin, D. B. (2014). *Bayesian data analysis*(Third Edition). Chapman and Hall/CRC.
- [3] Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- [4] Mao Shisong, Tang Yincai. *Bayesian statistics* (second edition). Beijing: China Statistics Press, 2012.
- [5] McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
- [6] Ministry of Education Higher Teaching Guidance Committee. (2018) *National Standards for Teaching Quality of Undergraduate Professional Categories in General Colleges and Universities*. Beijing: Higher Education Press,2018.
- [7] Wei Laisheng, Zhang Weiping. *Bayesian statistics* (second edition). Beijing: China Statistics Press, 2012.