UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Inferenza Bayesiana

2526-2-F8203B042-F8203B042M

Learning objectives

The course falls within the areas of learning in statistical sciences, computer science, and social sciences.
It enables students to learn analytical and inferential procedures within the Bayesian framework. Bayesian reasoning is presented in an integrated manner alongside the classical approach to statistical inference.
The course provides students with a solid foundation in Bayesian theory and fosters the development of applications through a problem-solving approach using real and simulated data, particularly in relation to applied problems in biostatistics. Students also acquire written communication skills, as they are required to write reports to accompany the results of the analyses performed.

Knowledge and understanding

Students are introduced to the main Bayesian statistical models for the analysis of data with different types of response variables, the assumptions underlying these models, and models for longitudinal data analysis.
They also develop an understanding of Markov Chain Monte Carlo (MCMC) procedures and the related estimation algorithms, as well as the ability to assess their effectiveness.
Furthermore, students are introduced to the R programming language within the RMarkdown environment, which enables them to create reproducible documents containing code, results, comments, and specific procedures for Bayesian analysis using SAS software.
Applied examples are based on real and simulated data from various fields relevant to the degree program. Students also learn how to write a structured report of the results obtained, considering the research questions, and to critically evaluate the limitations of the analyses carried out.
In this way, they develop independent judgment and refine their communication skills, both in presenting and justifying the analytical procedures adopted and in illustrating the results obtained in relation to the research questions.

Ability to apply knowledge and understanding

The course provides skills in the use of conjugate Bayesian models, in selecting prior distributions, and in employing estimation algorithms for complex models.

Through R and RStudio, students learn to structure statistical reasoning in an organized way, by analyzing data and drafting reports that explain the code, the analyses, and the results.
Using SAS software, students learn to estimate complex Bayesian models through simulations and to correctly configure the input required by estimation algorithms.
The course enables students to acquire a solid theoretical foundation and the ability to apply the statistical models introduced to real data. Students learn to evaluate the most appropriate model based on the available data and research questions.
They also learn to write and comment on the code used to generate results, adopting an open-source approach to ensure the reproducibility and replicability of their analyses.
By the end of the course, with the support of the provided materials—namely the instructor's lecture notes with a comprehensive bibliography, the R and SAS code, and the RMarkdown interface—students are able to independently pursue further study of this discipline.

This teaching is fundamental for the subsequent university course as it provides essential concepts for the development of Bayesian methods in both theoretical and applied fields, relevant to the target job contexts (biostatistics/statistics/demography and related fields) of students in the Biostatistics degree course.

Contents

Introduction to Bayesian inference and Bayes' rule. Methods of model specification and prior distributions.

Determination of the posterior distribution by exact methods.

Conjugate families: Gaussian, Poisson-gamma, beta-binomial, multinomial-Dirichelet.

Introduction to Bayesian non-parametric inference.

Methods to summarize the posterior distribution: credibility intervals and intervals with the highest posterior density.

Introduction to stochastic Markov processes, random walk.

Markov chain models for longitudinal data and latent Markov models with covariates.

Introduction to the Markov Chain Monte Carlo Methods: Metropolis-Hastings algorithm and Gibbs sampler.

Diagnostic tools to assess convergence of the MCMC procedure.

R environment and RStudio interface using, in particular, the following libraries: probBayes, learnBayes, LMest, LaplaceDemon.

RMarkdown will be employed to produce reproducible documents and to integrate code and output within the knitr library. SAS software with proc MCMC.

Detailed program

The Bayesian paradigm is introduced and compared with the frequentist approach, including Bayes'rule, and the total probability rule. A short introduction to Bayesian non-parametric methods is provided, and the notions of exchangeability and De Finetti's theorem are explained. The beta-binomial model and the other conjugate families such as Gaussian, Poisson-gamma, beta-binomial, and multinomial-Dirichlet distributions, are introduced. The

choice of the prior distribution is considered. Inference is compared with the classical approach. Methods to draw conclusions from the posterior distribution include Bayesian interval estimation, credible intervals, and intervals with the highest posterior density. The prediction context is also explored along with the empirical Bayes estimation.
Several examples of the application of Bayesian models in biostatistics, using real and simulated data concerning epidemiology, drug epidemiology, medicine and biology, ecology, and environmental sciences support the theory.

An introduction to stochastic processes within the Markov random field is proposed. The properties and features of the Markov chains are illustrated and explained using simulations. The random walk process is also described. Markov chain models for longitudinal data are explained, and the latent Markov models with covariates are introduced both from a theoretical and applied perspective.
Markov Chain Monte Carlo (MCMC) algorithms are explained with a focus on Metropolis-Hastings and Gibbs sampling algorithms. Diagnostic evaluations of the convergence are considered.

Some time is devoted to explaining the theory by imparting the flavor of the applications using observed data arising from different fields. The examples are developed within the statistical environment R, RStudio, and RMarkdown to create reproducible documents. The SAS software is proposed to perform analyses to estimate Bayesian linear and logistic models using PROC MCMC. During exercises, students are encouraged, also through collaborative learning, to develop reproducible documents concerning critical comments on the results of the analyses.

Prerequisites

The student is encouraged to know the content of the following courses: Statistics, Probability, and Statistical Inference and Statistical Models II.

Teaching methods

The lectures are held in the lab and the theoretical part is developed with applications carried out using R and SAS software. Many practical examples based on real and simulated are proposed, enabling students to learn data analysis and Bayesian modeling using R through the RMarkdown interface and SAS software. They are also encouraged to engage in collaborative learning and interact with their peers and finalize the required steps of the analysis. Weekly summarizing exercises are assigned, which involve applying the proposed models to real or simulated data. During the course, with the help of R in the RStudio environment and the RMarkdown interface, students learn to create reproducible documents. The scheduled hours for lecture-based teaching are 30, while those for interactive teaching are 17. The 3-hour lessons, in particular, are structured so that the second part involves students in an interactive manner.
Exercises are conducted interactively and in person at the computer lab.
Asynchronous video recordings of both the lectures and the exercises are made available on the e-learning platform.

Assessment methods

The following assessment methods apply to both attending and non-attending students. The exam is written, with open-ended questions, and includes an optional oral interview. No midterm exams are scheduled.
The written exam has a maximum duration of two hours and takes place in the computer lab. During the exam, students must answer open-ended theoretical questions and solve applied exercises, based on the theoretical

topics covered and the practical exercises assigned weekly throughout the course.
The theoretical questions aim to assess the understanding of fundamental concepts in Bayesian statistical inference using advanced methods.
Empirical analyses, conducted using the R environment, RStudio, RMarkdown, and SAS software, are used to evaluate the students' ability to apply Bayesian statistical models to real or simulated data, as well as to produce reproducible reports describing the data, procedures, and results.
The exam is designed to assess the student's ability to address the issues under study, use the provided code, and clearly communicate the results achieved, in addition to managing the available time effectively.
During the exam, the use of study materials and resources provided by the instructor is allowed, including R and SAS code developed during the course.
Each question is worth 3 points. The exam is considered passed with a minimum score of 18 out of 30.

Textbooks and Reading Materials

The teaching material consists mainly of handouts prepared by the teacher. These cover theory, applications, exercise and solutions developed with R software. All the files are available on the course page of the university's e-learning platform. In addition, the teacher publishes the following material at the end of each lesson: slides, R and SAS code, exercises, datasets, and solutions to some of the exercises. Previous exam texts are also published on the same page.

The main references are listed in the bibliography of the handouts, some of which are the following and are available in the university library, also in ebook format:

Albert, J. (2009). Bayesian computation with R. Springer Science & Business Media.

Albert, J., Hu, J. (2019). Probability and Bayesian modeling. Chapman and Hall/CRC.

Bartolucci, F., Farcomeni, A., Pennoni, F. (2013). Latent Markov Models for longitudinal data, Chapman and Hall/CRC, Boca Raton.

Migon, H. S., Gamerman, D., Louzada, F. (2014). Statistical inference: an integrated approach. Chapman & Hall.

Pennoni, F. (2024). Dispensa di Inferenza Bayesiana: Teoria e Applicazioni con R e SAS. Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca.

Robert, C., Casella, G. (2004). Monte Carlo Statistical Methods (second edition). Springer–Verlag, New York. Dipak, D. K., Ghosh, S. K., Mallick, B. K. (2000). Generalized linear models: A Bayesian perspective. CRC press.

SAS/STAT PROC MCMC, User’s guide, SAS Institute, 2012.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Semester

Semester I, cycle II, November 2025-January 2026

Teaching language

The course is delivered in Italian. Erasmus students may use the teaching material in English and request the teacher to conduct the examination in English.

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING | QUALITY EDUCATION