UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Inferenza Bayesiana

2627-2-F8205B011-F8205B011-2

Learning objectives

The course belongs to the learning areas of statistical sciences, computer science, and social sciences, and enables students to acquire analytical and inferential procedures in the field of Bayesian inference. Bayesian reasoning is presented in an integrated way with the classical approach to statistical inference. The course enables students to acquire solid foundations in Bayesian theory and to develop applications through a problem-solving approach using real and simulated data, with reference to applied problems in biostatistics. Students also acquire skills in written scientific communication, as they are required to write texts accompanying the results obtained from the analyses carried out.

Knowledge and understanding

Students are introduced to the main Bayesian statistical models for the analysis of data with different types of response variables, to the assumptions underlying these models, and to models for the analysis of longitudinal data. They also acquire an understanding of Markov Chain Monte Carlo (MCMC) procedures and related estimation algorithms, as well as the ability to evaluate their effectiveness. Students are also introduced to the R programming language, within the RMarkdown environment, which makes it possible to create reproducible documents containing code, results, comments, and specific procedures for Bayesian analysis using the SAS software. The applied examples concern real and simulated data from different fields relevant to the degree programme. Students also learn how to provide a written description of the results obtained in light of the research questions, together with a critical assessment of the limitations of the analyses performed. In this way, students develop independent judgement and strengthen their communication skills, both in presenting and justifying the methodological choices adopted and in interpreting and communicating the results in relation to the research questions.

Applying knowledge and understanding

The course provides skills in the use of Bayesian models with conjugate distributions, in the choice of prior distributions, and in the use of estimation algorithms for complex models. Through R, RStudio, and RMarkdown, students learn how to organise statistical reasoning in a coherent way by analysing data and writing reports that

illustrate the code, the analyses, and the results. Through the use of the SAS software, students learn how to estimate complex Bayesian models using MCMC simulation algorithms and how to correctly set up the inputs required by the estimation algorithms.
The course enables students to acquire solid theoretical foundations and the ability to apply the statistical models presented to real data. Students will be able to evaluate the most appropriate model according to the available data and the research questions. They also learn how to write and comment on the code used to generate the results, adopting an approach oriented towards open science that ensures the reproducibility and replicability of the analyses.

By the end of the course, thanks to the materials provided, the instructor’s lecture notes, accompanied by an extensive bibliography, the code for the R and SAS software, and the RMarkdown interface, students will be able to continue studying this discipline independently.

The course is essential for the subsequent university path, as it provides the key concepts for the development of Bayesian methods, through solid theoretical and applied foundations for the relevant professional contexts of the students enrolled in the Master’s degree programme in Biostatistics, including biostatistical, statistical, demographic, and related fields.

Contents

Introduction to Bayesian inference and Bayes’ rule.

Methods for specifying the model and prior distributions.

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

Nonparametric Bayesian inference.

Methods for summarising the posterior distribution, credibility regions, and highest posterior density intervals.

Introduction to Markov stochastic processes and properties of Markov chains.

Random walk model. Transition model for longitudinal data.

Latent Markov models for longitudinal data and extensions with covariates in both the observed model and the latent model.

Markov Chain Monte Carlo (MCMC) methods: Metropolis-Hastings algorithm and Gibbs sampling.

Diagnostic tests for convergence.

Practical sessions related to specific applied problems using R, RStudio with the RMarkdown text editor, and the SAS software.

Detailed program

During the course, Bayes’ rule and the law of total probability are reviewed. The course develops the aspects concerning the specification of prior distributions, the exact estimation of posterior distributions, and the interpretation of models. The beta-binomial model and other conjugate families are introduced: Gaussian, Poisson-gamma, multinomial-Dirichlet, with particular emphasis also placed on the predictive distribution. Bayesian point inference is compared with classical inference. The characteristics involved in choosing and determining the prior distribution, both informative and non-informative, are illustrated. The notion of exchangeability is presented through De Finetti's representation theorem. The posterior distribution is summarised through credibility regions and highest posterior density intervals.

Markov stochastic processes are introduced by presenting the properties and characteristics of Markov chains. The random walk is illustrated through simulations of trajectories for stochastic matrices of different dimensions. The transition model for longitudinal data and the latent Markov model are introduced. The estimation algorithms used within the MCMC method to approximate the posterior distribution are also illustrated from a computational point of view: the Metropolis-Hastings algorithm and the Gibbs sampling algorithm. Several measures are discussed, referring both to graphical analyses and to statistical tests that allow for the diagnostic evaluation of convergence.

The theory is accompanied by numerous examples of applications of Bayesian models in the field of biostatistics, using real and simulated data concerning epidemiology, pharmacoepidemiology, medicine, biology, ecology, and environmental sciences. The course also aims to facilitate the development of knowledge of the R environment and the SAS software. The examples are carried out in RStudio with the support of RMarkdown. During practical sessions, students are encouraged, also through cooperative learning, to develop reproducible documents that include critical comments on the results of the analyses. The following packages are used: probBayes, learnBayes, stan, LMest, LaplaceDemon, Nimble, and RMarkdown through the knitr package, in order to integrate code, analysis results, and comments. The analysis with the SAS software is carried out using the PROC MCMC procedure.

Prerequisites

Students are advised to have acquired and to review the notions taught in the following courses: Statistics, Probability and Statistical Inference, and Statistical Models II.

Teaching methods

The course includes face-to-face lectures delivered in person, concerning the theoretical aspects of the course, and accompanied by practical sessions. The classes take place in the computer laboratory. During the teaching activities, with the support of R, RStudio, the RMarkdown environment, and the SAS software, students learn how to analyse data, estimate Bayesian models, and accompany the analyses with comments and interpretations, producing documents that make it possible to reproduce and replicate the analyses carried out.

Weekly review exercises are assigned, based on real or simulated data, through which students are encouraged to address applied problems consistent with the theoretical topics presented in class, also fostering the development of cooperative learning. The course consists of 35 hours of lectures, devoted to the presentation of theoretical and methodological aspects, and 12 hours of interactive teaching, devoted to guided exercises, applications using R/RStudio/RMarkdown and SAS, discussion of results, and problem-solving activities. The asynchronous video recordings made available on the e-learning platform constitute supplementary study support material and do not replace in-person teaching activities.

Assessment methods

The following assessment methods apply both to attending and non-attending students. The exam consists of a written test, with open-ended questions and applied exercises, which may be followed, at the request of either the student or the instructor, by an optional oral interview. No mid-term tests are scheduled. The written test lasts a maximum of two hours and takes place in the computer laboratory. During the exam, students must answer open-ended theoretical questions and solve applied exercises related to the topics covered during the lectures and practical sessions.
The theoretical questions are aimed at assessing the understanding of the fundamental concepts of Bayesian statistical inference and related advanced methods. The applied exercises, carried out using R, RStudio, RMarkdown, and the SAS software, make it possible to assess the student's ability to apply Bayesian statistical models to real or simulated data, as well as to produce reproducible reports containing code, analyses, and interpretation of the results.
The written examination is graded on a scale out of 30, according to a scoring rubric communicated to students before the exam. The assessment takes into account: the theoretical correctness of the answers; the appropriateness of the model choice; the correctness of the computational implementation; the critical interpretation of the results; and the clarity and completeness of the written communication. The optional oral examination, if requested by the student or by the instructor, consists of a discussion of the written examination and of the theoretical topics covered in the course, and may confirm, supplement, or modify the final grade. During the examination, students are allowed to use the teaching materials and the materials provided by the instructor, including the R and SAS code developed during the course activities. The exam is considered passed with a minimum grade of 18/30.

Textbooks and Reading Materials

The main teaching material consists of lecture notes prepared by the instructor, covering the theoretical topics of the course, the applications developed with the R software, the exercises, and the corresponding solutions. The lecture notes are made available on the course page on the University e-learning platform.
At the end of each lecture, the slides used, the computing programs, the datasets used during the practical sessions, and the support materials are made available. Exercises are assigned weekly, together with the corresponding solutions. Examples of exam papers from previous years are also available on the same page of the e-learning platform.
The main bibliographic references are reported in the course lecture notes. The following fundamental reference texts, available at the University library, including in e-book format, are listed below:

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. (2026). Dispensa di Inferenza Bayesiana: Teoria e applicazioni con R e SAS. Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca.

Reich, B. J. (2026). Bayesian Statistical Methods: With Applications to Machine Learning. CRC Press.

Robert, C., Casella, G. (2004). Monte Carlo Statistical Methods (2nd ed.). Springer–Verlag, New York.

SAS Institute. (2012). SAS/STAT® 12.1 User's Guide: The MCMC Procedure (PROC MCMC).

R Core Team (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical

Computing, Vienna, Austria. https://www.R-project.org/

Semester

1st semester, Cycle II, November 2026 – January 2027.

Teaching language

The course is delivered in Italian. Erasmus students may use the teaching material available in English, provided by the instructor upon request. They may also request to take the exam in English.

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING