UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Inferenza Bayesiana

2021-2-F8203B011-F8203B012M

Learning objectives

The course allows the student to learn:

- the Bayes’ rule and the use of probability to update beliefs from the data;
- Bayesian inferential methods for a Binomial proportion and the normal mean;
- To use subjective priors, computation of the likelihood and posterior distributions;
- Monte Carlo methods to simulate the posterior distributions;
- calculus of the predictive distribution for features predictions and model checking;
- Markov Chain Monte Carlo algorithms: Metropolis-Hastings and Gibbs sampler;
- Bayesian inference and prediction for the multiple linear and logistic regression models;
- latent Markov models for longitudinal data.

Contents

Bayes’rule.

Model specification and choice of the prior assignment.

Coniugacy: Gaussian, Poisson-Gamma, Beta-Binomial and multinomial-Dirichelet distributions.

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.

Introduction to the latent Markov models for panel data with covariates.

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

R environment and RStudio interface with the RMarkdown to integrate code and output within the knitr library. The main R libraries are the following: probBayes, learnBayes, LMest. SAS software with proc MCMC.

Detailed program

The Bayesian paradigm is introduced and compared with the frequentist approach along with the Bayes ‘rule and the total probability rule. A short introduction to the Bayesian non-parametric methods is provided, the notion of exchangeability and De Finetti’s theorem is explanied. The Bayes’billard example is presented to introduce the Beta-Binomial model. Choice and specification of the prior distribution. Conjugate families: Gaussian, Poisson-Gamma, Beta-Bionomial and Multinomial-Dirichelet distributions and non-informative priors. Methods to draw conclusions from the posterior distribution: Bayesian interval estimation, credible intervals and intervals with the highest posterior density. The prediction context is also considered along with the empirical Bayes estimation.

An introduction to the stochastic processes within the Markov random field is proposed. Properties and features of the Markov chains are illustrated and explained with the use of simulations. A Random Walk is also showed.

Markov chain models for longitudinal data are explained and the Latent Markov models for panel data with covariates are introduced from a theoretical and applied perspective.

Markov Chain Monte Carlo (MCMC) algorithms are provided with a focus on Metropolis-Hastings and Gibbs sampling algorithms. Diagnostic evaluations of the convergence are considered.

Some amount of time is devoted to explain the theory by imparting flavor of the applications on real data collected from different fields arising in epidemiology, pharmacoepidemiology, medicine and biology as well as ecology and environmental sciences. They are developed within the statistical environment R, RStudio, RMarkdown to make reproducible documents. The SAS software is proposed to perform the analyses to estimate Bayesian linear and logistic models with PROC MCMC.

Prerequisites

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

Teaching methods

The lectures are held in the lab since the theoretical part is placed side by side with the applications carried out with the computer using R and SAS software. During the lectures, many practical examples based on real and simulated data referred to different contexts are proposed to the students to be solved with R through the RMarkdown interface and SAS software. The student is also encouraged to develop the cooperative learning in order to interact each other and finalize the required steps of the analysis. Exercises are carried out to report in a written form the results by adding critical comments and create reproducible documents.

During the Covid-19 emergency period the lessons will take place in the online asynchronous mode (videotaped lessons) with scheduled videoconferences meetings and some real meetings according to the availability suggested by the University.

Assessment methods

The following assessment methods are valid also for students not attending lectures. The written examination is performed in the lab where the student has to solve the exercises by showing that she/he is able to apply simulations, Bayesian models and advanced models for longitudinal data to real data in the field of biostatistics. The exercises are planned to evaluate the analytical skills of the students and his/her ability to solve the problem with R and the RMarkdown interface as well as SAS software and provide a reproducible document.

The results of the written examination are published in the e-learning page. With a positive score (from 18/30 and above) the student has to sustain an oral exam where she/he is explaining the theoretical features raised in the written part and the theory in the program of the course. In this way it is possible to evaluate the comprehension of the theoretical models. The written and the oral part are compulsory and should be carried out during the same examination term. Intermediate assessments are not planned.

During the Covid-19 emergency the exam will be same but it will be carried out in the lab or in videoconference *through Webex*

Textbooks and Reading Materials

The teaching material is made by the lecture notes concerning the theory and the applicative examples. Slides, R scripts, datasets and exercises with solutions are available after each lecture. The material is downloadable from the web page of the e-learning platform of the university.

The main texts are illustrated in the lecture notes. Some of them are the following:

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. (2020). *Dispensa di Inferenza Bayesiana -parte di 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 (2020). 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 2020 – January 2021.

Teaching language

The teaching language is Italian. Erasmus students can meet the professor to define proper English textbooks and require to carry out the exam in English.