Bayesian Statistics II

Università Bicocca (MILANO)
PhD in Economics, Statistics and Data Science
Instructor: **Raffaele Argiento**

Contact Information

e-mail: raffaele.argiento@unibg.it
Office hours: send me an email.
Web site: www.raffaeleargiento.com/

Objective of the course

The aim of this course is to give an overview of the large class of mixture models in the Bayesian setting. Finite and infinite mixtures will be treated using an unique approach.

The course focuses on mixture models as statistical tools for density estimation and (model based) clustering. Moreover, the computational approaches to perform Bayesian inference under the this class of models will be deeply discussed.

By the end of the course students will know how to – and when – set-up a Bayesian mixture model and how to design a Markov Chain Monte Carlo algorithm to approximate posterior functionals of inferential interest.

A successful completion of the course should also make students able to independently analyse a dataset providing an estimation the unknown density as well as a of the parameter "cluster". They will also be able to implement the algorithms discussed within the course using the R statistical software.

Prerequisites

The minimal prerequisites for this course are the familiarity with probability theory, the knowledge of frequentist parametric statistics as well as the bases of Bayesian statistics (see Bayesian Statistics I of this Ph.D program). Moreover, the student should be also familiar with the family of Markov Chain Monte Carlo algorithms (See Computational Statistics II of this Ph.D program). A minimal knowledge of the R software (www.r-project.org) is required.

Contents

- Finite and infinite mixture models, the basic concepts of kernel, mixing measure and components of a mixture.
- The latent component allocation variables and the clustering it induces on the data by a mixture model. Difference between component and cluster.

- The prior on the parameter "cluster" induced by the mixing measure: the exchangeable product partition function (eppf) and its properties. The prior on the number of components and the prior on the number of clusters.
- Linear and non-linear functionals of the posterior distributions and how to approximate them via Markov Chain Monte Carlo (MCMC) algorithms.
- Marginal and conditional algorithms for mixture model, the Chinese restaurant process and its generalization.

Reading List

A list of papers that inspired this course are:

- Argiento, R., & De Iorio, M. (2019). "Is infinity that far? A Bayesian nonparametric perspective of finite mixture models". arXiv preprint arXiv:1904.09733.
- Früwirth-Schnatter, S. Celux, G., and Robert, C. P. (2019). "Handbook of mixture analysis". Chapman and Hall/CRC.
- Früwirth-Schnatter, S. & Malsiner-Walli, G. (2018). "From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering". Adv. Data Anal. Classif.
- Pitman, J. (1995). "Exchangeable and partially exchangeable random partitions." Probab. Theory and Relat. Fields 102, 145–158.
- Pitman, J. (1996). "Some developments of the Blackwell–MacQueen urn scheme". In Statistics, Probability and Game Theory (T.S. Ferguson, L. S. Shapley and J.B. MacQueen, eds.). IMS Lecture Notes Monogr. Ser. 30, Inst. Math. Statist., Hayward, CA
- Regazzini, E., Lijoi, A. and Prünster, I. (2003). "Distributional results for means of random measures with independent increments". Ann. Statist. 31, 560–585.
- Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 689–710.

Evaluation and Grading

The final grade for this course will be given jointly with the course "Bayesian Statistics I" thought by Prof. B. Nipoti and the course "Computational Statistics II" thought by Prof. T. Rigon.

In particular

- At the end of the course each student will be assigned with a project (which includes, for example, the study of one or more articles on the topics discussed during the three courses, the implementation of a method, a data analysis or a possible extensions of the studied method);
- Each student should prepare a short paper summarizing the activities of the project;
- Finally, there will be an oral test in which the student presents his short paper and answers our questions.