Statistique Bayésienne: projet

Anna Simoni

November 4, 2020

To validate the course of Statistique Bayésienne you have to return two homeworks (Devoir Maisons) and this project. The final grade will be obtained from the grades of the two homeworks and the project.

Here are the instructions for the project:

(1) The project must be done by groups of two students. (2) The project length has to be about 5 pages (plus code), with reasonable font size and margins. (3) It is an applied project. Each group has to choose a statistical application with real or simulated data and then analyse it with Bayesian methods related to the material seen in class (or extensions of it). You are free to choose to work either with a real data set or with simulated data. You are free to consider a statistical problem that you have already analysed with frequentist procedure in another class or during an internship. Alternatively, you can consider a statistical problem related to an academic article (below I propose a partial list or you can propose an article). (4) **Before starting to work on a project it must be approved by me. So, please email me (anna.simoni@ensae.fr): the names of the participants of the group and a short description of the project so that I can validate it.** The project has to be chosen and validated by me before December 23, 2020 so that you will have enough time to work on the project. (5) You can use any language among Python, R or Matlab for the programming. (6) The due date of the project is January 29, 2021. By the due date please send

- your report as a pdf,
- a zipped folder containing your code and a detailed readme file with instructions to run the code

to anna.simoni@ensae.fr.

A possible structure of the project is: (1) definition of the problem under study and explanation of why it is interesting, (2) choice of the appropriate Bayesian technique (you should explain the methodology used and the motivation why you have chosen it), (3) description of the computational method used and difficulties encountered, (3) explanation and interpretation of the results. You can also compare the results you find with the results that you find with a frequentist approach.
If you choose to develop a project based on an academic article you can, for instance, either replicate and extend the simulations in the paper, or find a real data set and apply the method of the paper to it, or develop a simulation study if the paper does not contain it. An article can be chosen only by one group on the first-in-first-out basis. So, if you want to choose an article among the ones proposed below please email me the titles of three articles in order of preference and I will assign you the first article in your list that is still available.

Proposed academic articles

High-dimensional models:

1. Hauzenberger, M., Huber, F. and G. Koop, (2020). Dynamic Shrinkage Priors for Large Time-varying Parameter Regressions using Scalable Markov Chain Monte Carlo Methods. arXiv:2005.03906

2. Jin, R. and A. Tan, (2019). Fast Markov chain Monte Carlo for high dimensional Bayesian regression models with shrinkage priors. arXiv:1903.06964

3. Mogliani, M. and A. Simoni, (2020). Bayesian MIDAS Penalized Regressions: Estimation, Selection, and Prediction. *Journal of Econometrics*, forthcoming. arXiv: 1903.08025

4. Huber, F., Koop, G. and L. Onorante, (2020). Inducing Sparsity and Shrinkage in Time-Varying Parameter Models. *Journal of Business and Economic Statistics.*

5. Kowal, D.R., Matteson, D.S. and D. Ruppert, (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society*, 81, 781 – 804.

Computations:

6. McCulloch, R. and P.E. Rossi, (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*.

7. Chib, S. and E. Greenberg, (1998). Analysis of multivariate probit models. *Biometrika*, 85, 347-361.

8. Imai, K. and D.A. van Dyk, (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124, 311-334.

9. Basu, S. and S. Chib, (2003). Marginal Likelihood and Bayes Factors for Dirichlet Process Mixture Models. *Journal of the American Statistical Association*, 98, 224 – 235.

10. Deshpande, S.K., V. Rockova, E.I. George, (2017). Simultaneous Variable and Covariance Selection with the Multivariate Spike-and-Slab Lasso. *Journal of Computational and Graphical Statistics*, arXiv:1708.08911.

Partial Identification:

11. Gustafson, P., (2012). On the behaviour of Bayesian credible intervals in partially identified models, *Electronic Journal of Statistics*, 6, 2107 – 2124.

12. Moon, R.H. and F. Schorfheide, (2012). Bayesian and Frequentist Inference in Partially Identified Models, *Econometrica*, 80, 755 – 782.

Clustering:

13. Chib, S. and B.H. Hamilton, (2000). Bayesian analysis of cross-section and clustered data treatment models, *Journal of Econometrics*, 97, 25 – 50.

14. Bakker, B. and T. Heskes, (2003). Task Clustering and Gating for Bayesian Multitask Learning, *Journal of Machine Learning Research*, 83 – 99.

15. Quintana, F.A. (2006). A predictive view of Bayesian clustering, *Journal of Statistical planning and inference*, 136, 2407 – 2429.

16. Kim, S., Tadesse, M.G., and M. Vannucci, (2006). Variable selection in clustering via Dirichlet process mixture models, *Biometrika*, 93, 877 – 893.

17. Kyung, M., J. Gill, and G. Casella, (2010). Estimation in Dirichlet Random Effects Models, *The Annals of Statistics*, 38, 979 – 1009.

18. Kaufmann, S. (2010). Dating and forecasting turning points by Bayesian clustering with dynamic structure: a suggestion with an application to Austrian data, *Journal of Applied Econometrics*, 25, 309 – 344.

19. Shotwell, M.S. and E.H. Slate, (2011). Bayesian Outlier Detection with Dirichlet Process Mixtures, *Bayesian Analysis*, 6, 665 – 690.

20. Zhang, Z., D. Pati, and A. Srivastava, (2015). Bayesian clustering of shapes of curves, *Journal of Statistical Planning and Inference*, 166, 171 – 186.

21. Rigon, T., Herring, A.H. and D. Dunson, (2020). A generalized Bayes framework for probabilistic clustering, ArXiv: 2006.05451v1.

Other topics and applications:

22. Hirano, K., (2002). Semiparametric Bayesian Inference in Autoregressive Panel Data Models, *Econometrica*, 70, 781 – 799.

23. Chipman, H.A., E.I. George, and R. E. Mcculloch (2010). Bart: Bayesian Additive Regression Trees, *The Annals of Applied Statistics*, 10, 266 – 298.

24. Chib, S. and L. Jacobi, (2008). Analysis of treatment response data from eligibility designs, *Journal of Econometrics*, 144, 465 – 478.