Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего профессионального образования "Национальный исследовательский университет

"Высшая школа экономики"

Факультет Социологии

Программа дисциплины

Bayesian Statistics
(Байесовская статистика)

для направления 040100.68 «Социология» подготовки магистра
для магистерской программы «Сравнительные социальные исследования»

Авторы программы: Б.С. Соколов, bssokolov@gmail.com

Одобрена на заседании совета магистерской программы «Сравнительные социальные исследования» «___»_____ 20 г

Руководитель магистерской программы: К.С. Сводер

Рекомендована секцией УМС: Профессиональной коллегией по направлению «Социология»
«___»_____ 20 г
Председатель Е.Р. Ярская –Смирнова

Утверждена УС факультета Социологии «___»_____20 г.
Ученый секретарь _____

Москва, 2014

SUMMARY

Bayesian Data Analysis is a rapidly developing field of statistics, which has many useful applications in various areas of comparative social research. The goal of this course is to provide a brief introduction to the theory and application of Bayesian statistical methods. The course begins with basic concepts of Bayesian statistics. Then we consider the general approach to the estimation and assessment of Bayesian models. Because of the focus of the master program on comparative studies, we then will discuss applications of Bayesian modelling to specific tasks arising in cross-cultural research, including such topics as multilevel/hierarchical analysis, Bayesian structural equation modelling (BSEM), multilevel BSEM, and Bayesian approximate measurement invariance. In the end of the course, several advanced applications of Bayesian analysis are highlighted, such as informative hypothesis testing, multiple imputation and simulation-based approach to model interpretation.

Students are assumed to have basic knowledge of statistics and be familiar with several conventional statistical methods, including regression analysis and factor analysis. Knowledge of advanced topics, such as multilevel analysis, structural equation modelling (SEM), or maximum-likelihood estimation, is helpful, but not critical.

GRADING COMPONENTS:

- home assignments (cumulative grade - 60%)
- final project presentation (40%)
- participation/attendance: If unexcused absences are greater than two, then final grade = (final project grade) + (cumulative grade) x (attended weeks / total weeks)
- Late assignments will be graded down.
- If you plagiarize, you will fail. You may not recycle papers used in other classes.

THEMATIC PLAN OF THE COURSE

Lesson №	Theme	Total hours in theme	Hours in classroom		Independent work
			Lecture	Seminar	
1	Introduction. Basic Concepts of Bayesian analysis	14	2	2	10
2	General Principles of Bayesian Inference. Priors and Likelihood	16	2	2	12
3	Estimating Bayesian Model: choice of sampling algorithm, assessment of model fit, model comparisons	16	2	2	12
4	Bayesian Hierarchical Analysis	16	2	2	12

5	Bayesian Structural Equation Modelling	16	2	2	12
6	Bayesian Approximate Measurement Invariance	14	2	2	10
7	Informative hypothesis testing	14	2	2	10
8	Multiple Imputation. Simulation	14	2	2	10
	Total	120	16	16	88

Lesson 1: Introduction. Basic Concepts of Bayesian analysis:

Required Readings:

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. CRC press, 2013. Chapter 1, p. 3-28

Supplementary Readings:

http://www.bayesian-inference.com/index

Joyce, J., "Bayes' Theorem", The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/

Lesson 2: General Principles of Bayesian Inference. Priors and Likelihood.

Required Readings:

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. CRC press, 2013. Chapters 2-3, p. 29-83.

Lesson 3: Estimating Bayesian Model: choice of the sampling algorithm, assessment of model fit, model comparisons.

Required Readings:

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. CRC press, 2013. Chapters 6-7, p. 141-196

Supplementary Readings:

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. CRC press, 2013. Chapters 10-13: 259-350

Lesson 4: Bayesian Hierarchical Analysis

Required Readings:

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian data analysis. CRC press, 2013. Chapter 15

Lesson 5 Bayesian Structural Equation Modelling: applications to factor analysis, multilevel factor analysis, and latent class analysis.

Required Readings:

Muthén, B., and T. Asparouhov. "Bayesian structural equation modeling: a more flexible representation of substantive theory." *Psychological methods* 17, no. 3 (2012): 313.

Supplementary Readings:

Asparouhov, T., and B. Muthén. "Bayesian analysis of latent variable models using Mplus." Unpublished manuscript. www. statmodel. com/download/BayesAdvantages18. pdf (2010).

Hoijtink, H. "Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics." *Multivariate Behavioral Research* 36, no. 4 (2001): 563-588.

Davidov, E., Schmidt, P., & Billiet, J. (2011). Cross-cultural analysis. Methods and applications. New York: Routledge.

Lesson 6. Bayesian Approximate Measurement Invariance

Required Readings:

Van De Schoot, R., A. Kluytmans, L. Tummers, P. Lugtig, J. Hox, and B. Muthén. "Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance." *Frontiers in psychology* 4 (2013): 770.

Muthén, B., and T. Asparouhov. "BSEM measurement invariance analysis." *Mplus Web Notes* 17 (2013): 1-48.

Supplementary Readings:

Steenkamp, J.-B., and H. Baumgartner. "Assessing measurement invariance in cross-national consumer research." *Journal of consumer research* 25, no. 1 (1998): 78-107.

Muthén, B., and T. Asparouhov. New methods for the study of measurement invariance with many groups. *Technical report*. http://www. statmodel. com, 2013.

Lesson 7: Informative Hypothesis testing

Required Readings:

Van de Schoot, R., Dekovic, M., and Hoijtink, H. (2010). Testing inequality constrained hypotheses in SEM models. Structural Equation Modeling, 17, 443–463.

Supplementary Readings:

Van de Schoot, R., Hoijtink, H., Hallquist, M. N., & Boelen, P.A. (2012). Bayesian Evaluation of inequality-constrained Hypotheses in SEM Models using Mplus. Structural Equation Modeling,19:1–17, 2012

Lesson 8: Multiple Imputation. Simulation-based approach to interpretation.

Required Readings:

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. Journal of Statistical Software, 45(7), 1-47.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*, 44: 347-361.

http://datascience.iq.harvard.edu/files/datascience/files/making.pdf

Supplementary Readings:

Imai K., King G., Lau O. "Zelig: Everyone's Statistical Software", 2006. http://r.iq.harvard.edu/docs/zelig.pdf

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons.

ASSIGNMENTS (Components of Final Grade)

In-Class Participation and Attendance:

Participation is required and expected. Come prepared, having read the relevant texts, and prepared to discuss. For students who miss more than two lessons (seminars or lectures) without a valid doctor's excuse:

final grade = 0.4*project grade + 0.6*(cumulative grade)* (attended weeks / total weeks)

For example, if you missed 4 of 8 weeks with no excuse, and your cumulative grade was an 8 and your research project grade is 7, then your final grade will be 0.4*7 + 0.6*8*(4/8) = 5.2 (*instead of 7.6 without attendance penalty*). Your attendance penalty will also apply to your re-examination grade. Furthermore, only your "research project" can be re-examined. There is no possibility to make up your attendance or any late or missed assignments. If you have only missed 2 lessons unexcused, there will be no grading penalty.

Research project:

All course participants must write a short research paper (15-20 pages) in which they will try to apply Bayesian methods to the topic in cross-cultural social research that they are interested in. The most important aspects of the paper to be graded are the creativity of the research idea, the operationalization and proper statement of hypotheses, and the appropriate use of Bayesian methods. Final project paper must be written alone, independent of other student projects.

Software: we will use R package *laplace's demon* as a main software, so students are expected to have a basic knowledge of programming in R. In some applications, we will also use MPLUS software. However, no prior knowledge of MPLUS modelling language is assumed.

Useful links:

http://www.bayesian-inference.com/index

http://www.statmodel.com/

http://www.r-project.org/

Please send any questions and course-related exchanges to my email at bssokolov@gmail.com with “Bayesian data analysis” in the subject line.