

Bayesian Modelling

Xavier Fernández i Marín
Ludwig-Maximilians-Universität (LMU Munich)

<http://xavier-fim.net>

July 2018

Contents

1 Objectives	1
2 Prerequisites	2
2.1 Install R, JAGS (and RStudio)	2
2.2 Bring a dataset	2
2.3 Be able to import the dataset into R	2
2.4 Understand some basic functions and operations	2
3 Organization	3
3.1 Sessions	3
3.2 Structure	3
4 Contents	3
4.1 Session 1	3
4.2 Session 2	3
4.3 Session 3	3
4.4 Session 4	3
4.5 Session 5	4
4.6 Session 6	4
5 Instructor	4

1 Objectives

The purpose of the course is to provide the participant with the tools to be able to do data analysis through Bayesian inference. At the end of the course, the participant must be able to do data analysis on a dataset of his/her own using Bayesian inference, with the following steps:

- Clean, prepare and arrange data before any analysis.
- Go from an equation specifying a concrete model to its representation in computer code.
- Obtain estimates of the parameters of interest.
- Graphically represent the results of the models and interpret the results.

2 Prerequisites

Prior to the start of the course, participants must ensure the following: have R (and RStudio if it is the first time they are using R) and JAGS installed; have a dataset to be ready to be analyzed, and loaded into R; have some familiarity with basic R functions.

2.1 Install R, JAGS (and RStudio)

Participants can use computers provided by the organization, but they are encouraged to bring their own laptops. In the latter case, the laptops have to have R and JAGS (RStudio) installed (in this order).

- Install R from <http://cran.r-project.org/>
- Install JAGS from <http://mcmc-jags.sourceforge.net/>
- Install RStudio (Desktop) from <http://www.rstudio.com/products/rstudio/download/>
- Installing R and RStudio, explained: <http://socserv.mcmaster.ca/jfox/Courses/R/ICPSR/R-install-instructions.html>

2.2 Bring a dataset

The course includes access to prepared datasets, but in order to get the most of the course it is important that each participant brings at least one dataset of his/her own. The outcome / dependent variable has to be continuous, in order to follow most of the course, although other types of outcomes (binary, counts) will be covered in the later stages of the course.

2.3 Be able to import the dataset into R

The participant has to ensure that he/she is able to import a dataset from R. In general terms, the *R Data Import/Export* manual is useful: <http://cran.r-project.org/doc/manuals/r-release/R-data.html>.

2.4 Understand some basic functions and operations

The following functions have to be understood prior to start the course: `c()`, `seq()`, `rep()`, `gl()`, `names()`.

It is important also to be familiar with R's own logic of object selection using `[]` and conditionals.

The following materials can be helpful for covering the aforementioned prerequisites:

- Tutorial inside R itself <http://swirlstats.com/>. Course "R Programming".
- Very pedagogical guided course with exercises <https://www.datacamp.com/courses/introduction-to-r>.
- Online tutorial <http://www.r-tutor.com/r-introduction>, part "R introduction".
- Online tutorial <https://www.codeschool.com/courses/try-r>.
- Hands-on guide <http://www.computerworld.com/article/2497143/business-intelligence-beginner-s-guide-to-r-introduction.html>.

- Official documentation <http://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>, chapter 2 and Appendix A.

3 Organization

The course will be structured as follows:

3.1 Sessions

The course has *6 sessions* of *2 hours* divided in *3 days*, with a total of 12 hours.

1. July 2, 09:00-11:00
2. July 2, 11:00-13:00
3. July 3, 09:00-11:00
4. July 3, 11:00-13:00
5. July 4, 09:00-11:00
6. July 4, 11:00-13:00

3.2 Structure

Each session will be divided between parts of lecture and parts of autonomous work solving a set of exercises, with the supervision of the instructor.

The last part sessions will be devoted to solve specific individual problems that the participants have, either concerning data preparation or data analysis.

4 Contents

The expected contents are the following, but please notice that the specific content of each day may vary according to the pace of the course.

4.1 Session 1

- Introduction to Bayesian modelling.
- Inference: from the linear model to a simple two-level hierarchical model.
- Tools setup.

4.2 Session 2

- Why being Bayesian when working with survey data?
- Foundations of Bayesian inference: convergence and diagnostics.
- Practical tips.

4.3 Session 3

- Interpretation.
- Model fit.

4.4 Session 4

- Hierarchical / Multilevel models

4.5 Session 5

- Heteroskedasticity.
- Robust regression.
- Other types of outcomes: counts (Poisson and Negative Binomial), binary (logit).

4.6 Session 6

- Measurement models
- Missing data.

5 Instructor

Xavier Fernández-i-Marín is lecturer at the Geschwister-School-Institute for Political Science at the Ludwig-Maximilians-Universität (LMU) in Munich, working on comparative public policy. He has been previously senior researcher at ESADE working on the entrepreneurial roles of end users to shape a green EU economy, and assistant professor at the University of Konstanz in 2014 working on the diffusion of social and environmental policies. Before he has been research fellow at ESADEgeo (Center for Global Economy and Geopolitics) working on the determinants of the cooperation amongst countries and Global Governance, and "Juan de la Cierva" post-doc research fellow at IBEI (Institut Barcelona d'Estudis Internacionals) from 2009 until 2011. He presented his PhD on Technology and Public Policy: An Evaluation of Internet and e-Government policies in Spain in 2008 at the Political and Social Sciences Department of the Universitat Pompeu Fabra.

He has been trained in methodology at the University of Essex, obtaining a postgraduate degree, and teaching in the summer school for several years. He has large experience with hierarchical/multilevel models and Bayesian inference, and also in cluster analysis, principal component analysis, factor analysis, survival (event history) analysis, spatial models and quantitative methods in general. This has allowed me to collaborate with several disciplines, bringing scientific, methodological and systematic value to the different teams.



References

- Congdon, P. *Applied Bayesian modelling*. Wiley series in probability and statistics. Wiley, 2003.
- Gelman, Andrew and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, Dec. 2006.
- Kabacoff, R. *R in Action: Data Analysis and Graphics with R*. Manning Publications Company, 2014.