Department of Psychology, Stockholm University
Doctoral program in Psychology, spring semester 2024

Statistics 2.5: Bayesian Multi-level Modelling the Rethinking Way (3 hp)

This course (Stat2.5) is a follow up course to Statistics 2 (Stat2). Stat2 follows Richard McElreath's book "Statistical Rethinking" (2020, 2nd ed.) up to but not including the last chapters on Bayesian multilevel modeling (BMLM). In this course, we will work through some of the examples of BMLM in chapters 13 and 14, using the function `ulam()` of the `rethinking` package. In addition, we will work through some examples from other sources, including a simpler way of doing BMLM, using the R-package `rstanarm`, function `stan_glmer()`, and a more complicated, but more general, way involving `Stan` code and the R-package `rstan`.

Prior knowledge

The course assumes prior knowledge corresponding to "Stat2: Bayesian data analysis", that is, McElreath's book Chs. 1-12. But if you have not taken Stat2, it is still possible join after obtaining basic understanding of the R library `rethinking` and in a particular the function `ulam()`. There will also be an introductory seminar reviewing the fundamentals of Stat2 needed to follow Stat2.5.

Learning outcomes

- Understanding of the main ideas of BMLM and its strengths and weaknesses in relation to conventional approaches to data analysis.
- The easy way: Skills to conduct BMLM using the standardized method implemented in the R-package `rstanarm`, function `stan_glmer()` or the R-package `brms`, function `brm()`.
- The middle way: Skills to conduct BMLM using the method implemented in the R-package `rethinking`, function `ulam()`.
- The hard way: Skills implementing BMLM in `Stan` code, using the R-package `rstan`

Course content

- Key concepts: No, partial and complete pooling, shrinkage, varying intercepts, varying slopes.
- Simulating hierarchical data from specified data generating model.
- BLMM the easy way: R-package `rstanarm`, function `stan_glmer()` or R-package `brms`, function `brm()`
- BLMM the middle way: R-package `rethinking`, function `ulam()`.
- BLMM the hard way: `Stan` coding implemented using R-package `rstan`.
- R programming

Activities

A series of 5 seminars. The seminars will cover theoretical discussions of topics covered by sections of the literature, followed by practical work on solving a set of exercises including analysis of specific data sets. Much of the seminar discussions will concern how to address problems and illustrate phenomena using R. It is therefore a good idea to bring a laptop with R and R-studio installed to each seminar.

Examination

The course is graded *Pass* or *Fail*. *Pass* requires passing both of the two examination parts described below.

1. Solving a set of exercises, some from McElreath (2020), that we worked on in the practical part of the course. We will help each other out, but you have to compile and submit your own solutions. It is OK to get help from others, but you need to write your own code. Solutions should be delivered no later than 2 months after the last seminar. If revision is needed, the revision should be delivered no later than 3 months after the last seminar.
2. A report of analyses of data of the student's own choice (real or simulated data). The analyses should follow some of the analytic approaches to multi-level modelling discussed in the

course. The student will present an outline of the planned analyses at the last seminar, and should submit a report no later than 2 months after the last seminar. If revision is needed, the revision should be delivered no later than 3 months after the last seminar.

Literature

Abbr vtn	R-library	Reference
SR	`rethinking`	McElreath, R. (2020). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan (2nd ed)*. New York: CRC Press. **Chapters 13-14** [Available in electronic format from Stockholm University Library.]
BR	`rstanarm`	Johnson, A. A., Ott, M. Q., & Dogucu, M. (2023). *Bayes Rules! An Introduction to Applied Bayesian Modeling*. London: CRC Press. **Chapters 15-19** [Available in electronic format at https://www.bayesrulesbook.com/]
STV	`rstan`	Sorensen, T., Hohenstein, S. & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology, 12*, 175-200.
S	`brms`	Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, 26(1), 103.

Note: STV and S use the same data for illustration.

Schedule

<u>Time</u>: 10.00–12.00 (Theory), 13.00–16.00 (Practice).
<u>Place</u>: Seminar room at Department of Psychology, SU

	Date	Topic	Literature	Exercises	Data sets
0	April 1	Review Bayes the rethinking way	SR:1-12	MN0	`BR:running` `SR:reedfrogs`
1	April 2	Basic idea: Partial pooling and shrinkage	SR:13, BR:15	BR:15.3-4, 15.6, 15.8 MN1	`lme4:sleepstudy` `SR:Simulated tadpoles`
2	April 4	Varying intercepts	SR:13	SR:13M1-4; 13H4 MN2	`SR:reedfrogs` `lme4:sleepstudy`
3	April 8	Varying slopes	SR:14.1	SR:14M1-2 MN3	`SR:Simulated cafés`
4	April 11	Varying slopes continued	SR:14.2	MN4	`BR:running` `lme4:sleepstudy`
5	April 15	Rstan	SR:14, STV	MN5	`STV:gibsonwu` `S:gibsonwu`

<u>Exercises</u>:

BR: Exercises from BR

SR: Exercises from SR

MN0: (a). Import the running data set (see BR, p. 376), and fit a model of running time as a function of age, ignoring id (complete pooling) and separately for each id (no pooling). Visualize model fits. Do this exercise using both `rstanarm::stan_glm` and `rethinking::ulam`. (b). Import the reedfrogs data from the rethinking package, and fit m13.1 using `rethinking::ulam`, visualize the model fit in a figure like 13.1 (see Ch. 13). Repeated with wider priors, what happens, and why?

MN1: Do the tadpole simulation from the book (section 13.2), but for another scenario (other true parameters of your choice), and display result as Fig. 13.1 <u>and</u> Fig 13.3. Briefly compare results of your simulation with McElreath's simulation.

MN2: (a) Fit a varying intercept model to the running data, using `rethinking::ulam` or `rstanarm::stan_glmer` or `brms::brm`. Visualize the results. (b) Do the same for the sleepstudy data.

MN3: Repeat the café simulation from the book (section 14.1), but for scenarios with less and more visits to each café (N_visits = 10 in the book), display result for each simulation in way that visualizes the difference (shrinking) in MLM compared to a no-pooling model. At what sample size (N_visits) do partial and no pooling lead to the same results?

MN4: (a) Fit a varying intercept & slope model to the running data, using `rethinking::ulam` or `rstanarm::stan_glmer` or `brms::brm`. Visualize the results and briefly compare to results from MN 2. (b) Do the same for the sleepstudy data.

MN5: Import data used in STV and run the four models discussed in the paper using Rstan (see Stan code in the paper). Run the same models using `rstanarm::stan_glmer` or `brms::brm`. Discuss any differences between results from the two programs?