Proposal for recognition of a Specialist Course

Ivan Grahek, Ladislas Nalborczyk, & Antonio Schettino

Introduction to Bayesian statistical modelling

The most common statistical procedure used in psychology and other scientific literature is Null Hypothesis Statistical Testing (*NHST*). This framework requires formalizing a statistical model of what the data would look like given that chance or random processes alone were responsible for the results (Stockburger, 2007). The typical end product of NHST procedures is the p-value, which indicates the probability of obtaining results at least as extreme as those observed assuming that the null hypothesis is true (Wasserstein & Lazar, 2016). Unfortunately, the majority of researchers routinely misinterpret the information provided by p-values (Gigerenzer, 2004). One common misinterpretation is that $p < .05$ warrants the rejection of the null hypothesis, consequently supporting the acceptance of the alternative hypothesis (Cohen, 1994). Another popular statistical framework makes use of *Bayesian statistics*. In a Bayesian analysis, every unknown quantity is given a *prior distribution* which embodies the state of knowledge before observing the data. Once data have been observed, Bayes theoreom is used to update the prior distribution into the *posterior distribution*, which represents the probability of a statistical model (i.e., the formalization of a hypothesis) knowing the data at hand and prior knowledge. Posterior distributions are the end-product of Bayesian analysis, as they contain all the information needed for inference and can be interpreted in an intuitive way (Kruschke, 2015; Lee & Wagenmakers, 2015; McElreath, 2016). Bayesian inference is extremely flexible: it allows researchers to update prior knowledge according to available data, quantify relative evidence for each model of interest (including the null hypothesis), and monitor evidence until the result is sufficiently compelling or available resources have been depleted (Wagenmakers et al., 2017).

The advantages of Bayesian over NHST approaches briefly outlined above allow researchers to obtain much more information from their data without unnecessarily wasting resources (Wagenmakers, Morey, & Lee, 2016). Therefore, a practical understanding of Bayesian parameter estimation can greatly improve statistical inferences and is regarded as a strategically advantageous skill to acquire during one's academic preparation. However, as to date, this specific type of data analysis is not adequately covered in the statistical courses for Masters in Psychology, nor do courses for PhD students in Psychology exist at Ghent University. For this purpose, we propose a three-day course focusing on conceptual aspects as well as its wide application potential.

In order to allow as many researchers as possible to enjoy the aforementioned advantages that Bayesian parameter estimation has to offer, it is pivotal that these procedures can be executed via free, open-source, cross-platform, and user-friendly software. One ideal candidate is *R*, a popular programming language and software environment for statistical computing (R Core Team, 2017). Within the *R* environment, one package that stands out for its flexibility and ease of use is `brms` (Bürkner, 2017), which allows the implementation of a wide range of statistical models (e.g., from simple linear regressions to complex survival models) using simple and intuitive *R* syntax and to estimate posterior distributions of the parameters of interest via Markov-Chain Monte-Carlo sampling procedures.

The program of the proposed three-day Doctoral School course will start, on **day 1**, with a general overview of the model comparison approach to data analysis (as implemented in *R*), to guide participants from a *"hypothesis testing"* to a *"parameter estimation"* mindset. Notions of model comparisons will subsequently be outlined, in order to provide practical ways to implement model selection methods in everyday research. The Bayesian data analysis framework will then be introduced, with particular emphasis on how formalized prior knowledge and observed data are integrated to generate posterior distributions.

Day 2 will be devoted to the introduction of the `brms` package by its creator, Dr. Paul-Christian Bürkner. Practical examples will follow, focussing on data analysis problems typically encountered in everyday research: modelling of continuous data via linear regression, generalized linear models for categorical (count and percentage) data, and ordinal regression to adequately analyse data from Likert scales.

During **day 3**, Dr. Bürkner will introduce multilevel models and its straightforward implementation in `brms`. Examples of models built during day 2 will be extended to estimate uncertainty arising from subject-specific variability. Finally, Dr. Bürkner will show how to conduct meta-analyses within a hierarchical Bayesian framework.

Participants will be technically supported by the course organizers (Grahek, Nalborczyk, & Schettino) throughout the whole duration of the course, and will additionally be encouraged to ask questions regarding their own data and research plans.

Organising committee

- Ivan Grahek
  Department of Experimental-Clinical and Health Psychology
  Email: Ivan.Grahek@UGent.be
- Ladislas Nalborczyk
  Department of Experimental-Clinical and Health Psychology
  Email: Ladislas.Nalborczyk@UGent.be
- Antonio Schettino
  Department of Experimental-Clinical and Health Psychology
  Email: Antonio.Schettino@UGent.be
- Prof. Dr. Ernst H. W. Koster, PhD
  Department of Experimental-Clinical and Health Psychology
  Email: Ernst.Koster@UGent.be

Tentative dates: 13/06/2018 – 15/06/2018

Contact hours: 21h (3 x 7h)

Location: Room 1.3 (PC-Lokaal 1), Faculty of Psychology and Educational Sciences, Henri-Dunantlaan 2, Ghent University, 9000 Ghent

Registration fee: There will be no registration fee.

Number of participants: A maximum of 20 participants will be allowed to take the course, ensuring intensive feedback during the practical sessions. Based on a first informal poll, early-career researchers (PhD students and postdocs) of the Faculty of Psychology and Educational Sciences confirmed their interest in taking this course on the proposed dates.

Instructors

Paul-Christian Bürkner
Research Associate
Department of Psychology, University of Muenster WWU
Fliednerstr. 21
48149 Muenster, Germany
Tel: 0251 83-39418
paul.buerkner@uni-muenster.de
https://www.uni-muenster.de/PsyIFP/AEHolling/personen/buerkner.html

Paul-Christian Bürkner will be the main lecturer of this course. Paul is a research associate in the Psychology department at the University of Muenster. His PhD dissertation, completed in 2017, was awarded the Gustav A. Lienert award for the best methodological dissertation. During his PhD, he developed the `brms` package (Bürkner, 2017), currently one of he most used R packages for Bayesian data analysis. Over the past years, Paul has published several peer-reviewed manuscripts regarding Bayesian data analysis and meta-analysis. Furthermore, Paul has been invited to give workshops on Bayesian data analysis at several universities. He also has extensive experience as a lecturer in statistic methods, for which he received this year the award of the best lecture of the Institute of Psychology in Muenster. Given his expertise on the topic and his innovative work, Paul will lead the lectures during day 2 and day 3.

Tentative program

Course objectives

Students start by acquiring basic skills in order to be able to understand and conduct Bayesian data analysis in the R software environment. This allows students to understand the analyses and complete the hands-on exercises during the second and third day of the course. At the end of this course the students should:

- Be familiar with the R environment and the available packages for Bayesian data analysis
- Know the elementary commands and statements used in R and Bayesian data analysis packages
- Understand the theoretical underpinnings of Bayesian statistical models and their potential applications
- Be able to report output of Bayesian data analysis

Study material

The instructors will provide the students of a syllabus and will give theoretical presentations which will be alternated with practical demonstrations and hands-on computer exercises in the R environment.

Used Software

The course will be taught using R (free statistical open-source software) and the RStudio GUI, and packages relevant for Bayesian data analysis (mostly the `brms` package).

Course organisation

Theoretical lectures will introduce basic concepts, followed by demonstrations of analytical procedures and practical sessions to get hands-on experience. During the second and third day two lecturers will be present to guide the practical exercises.

Course schedule

Day 1: Statistical modelling, model comparison and Bayesian statistics (Ivan Grahek, Ladislas Nalborczyk, & Antonio Schettino)

9:00 Registration
9:30 Introducing the model comparison approach to data analysis
12:00 Break
13:00 Notions of model comparison (information criteria, Bayes Factors)
15:00 Introduction to Bayesian data analysis
17:30 End of day

Day 2: Introduction to Bayesian statistical modelling (Paul-Christian Bürkner)

9:00 Introducing the brms package, a simple linear regression case study
12:00 Break
13:00 Toward generalised linear models, modelling categorical data
15:00 Ordinal regression with brms, modelling Likert-type data
17:00 End of day

Day 3: Advanced Bayesian statistical modelling with brms (Paul-Christian Bürkner)

9:00 Introducing multilevel models with brms
12:00 Break
13:00 Toward generalised multilevel models
15:00 Meta-analysis with brms
17:00 End of day

Budget

	Travel	Hotel	Honorarium	Total
Paul-Christian Bürkner	Transport to train station + retour trip Munster-Ghent 200 EUR	300 EUR (3 nights in Europa Hotel inc. breakfast & city tax)	1400 EUR	1900 EUR

This Specialist course is organised for graduate students as part of their Doctoral Schools course. As a result, no registration fee will be asked (no profit). Of course, postdoc researchers are also welcome to join the Specialist course.

References

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606.

Kruschke, J. K. (2015). *Doing bayesian data analysis: A tutorial with r, jags, and stan* (2nd ed.). Amsterdam: Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2015). *Bayesian cognitive modeling: A practical course*. Thousand Oaks, CA: Cambridge University Press.

McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in r and stan*. Boca Raton: CRC Press/Taylor & Francis Group.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Stockburger, D. W. (2007). Hypothesis and hypothesis testing. In S. N. J. (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 446–449). Thousand Oaks, CA: SAGE Publications Ltd.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2017). Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*.

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176.

Wasserstein, R L, & Lazar, N. A. (2016). The asa's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.