Mixed-effects models for EEG analysis

Exposé for a master thesis in the area of Computational Cognitive Science

LUIS LIPS – UNIVERSITY OF STUTTGART

1. INTRODUCTION

Electroencephalography (EEG) is a non-invasive method for acquiring brain activity data. It is cost-efficient with a very high temporal resolution. In research it is common to further focus on event-related potentials (ERPs) within the recorded EEG data. Event-related potentials are oscillations in the EEG signal that are time-locked to an event (e.g. stimulus). However, event-related potentials are very small compared to other signals inside the EEG. Therefore ERPs are typically created via averaging over multiple trials per subject. In addition, researchers generally only look at electrodes that are of interest to their research. For those selected electrodes, the amplitudes and latencies of the corresponding ERP peaks are extracted. Statistical analysis is then performed on those values to find significant differences between groups and/or conditions. Why has this approach become established? EEG data is 3-dimensional data (electrodes / space, time, subjects). In combination with the size of each dimension the analysis and computation is quickly very time consuming. Because computation is so expensive, hierarchical analyses spanning all trials and subjects were impossible, and the two-stage approach is common.

A. Two-stage approach

In this approach the analysis is separated into the (1) individual subject level and (2) the group level. As mentioned above values on the individual subject level are condensed into a single value by averaging or regression. On the second stage the extracted values are then further analyzed using e.g. a paired t-test. This simplifies the complexity of the analysis. In addition averaging over multiple repetitive trials per subject increases the statistical power and zeroes out noise. Nevertheless this also has its drawbacks and limitations. One major drawback is the loss of information about the uncertainty of the extracted location parameter. A more elaborate modelling-scheme are linear mixed-effects models, which explicitly take into account uncertainties at multiple levels.

B. Linear mixed-effects models

Linear mixed-effects models (LMMs) have a long reaching history. LMMs have been widely used in research in psycholinguistics and other research fields (Clark, 1973; Bagiella et al., 2000; Baayen et al., 2008). Especially where nested experimental data or repeated measurements are encountered. Consequently they come in different shapes and names. Linear mixed-effects models are also called nested data models, multilevel models or hierarchical models. LMMs are an extension of simple linear models. Unlike simple linear models, linear mixed-effects models can describe not only fixed effects (average effect of the independent variable on a dependent measure) but also random effects (random variation from subject to subject or item to item). In general linear mixed-effects models can be defined the following way:

$$
\begin{aligned}
y &= X\beta + Zu + \epsilon \\
u &\sim N(0, G) \\
\epsilon &\sim N(0, R)
\end{aligned}
$$

The vector y represents the observations. In the context of EEG this corresponds to the recorded EEG data. β is a unknown vector of fixed effects. u is a unknown vector of random effects with mean $E(u) = 0$ and variance-covariance matrix $var(u) = G$. The matrices X and Z are design matrices for the respective effects (fixed / random). ϵ is a vector of random errors, with mean $E(\epsilon) = 0$ and variance $var(\epsilon) = R$. This allows the possibility to model parameters on the group level as well as within each individual subject. A common method to compute the parameters is restricted maximum likelihood estimation.

C. Mixed-effects multilevel analysis

Another analysis approach is the mixed-effects multilevel analysis (MEMA) proposed by Chen et al. (2012). The models can be described as a special case of the linear mixed-effects models (Chen et al., 2013). Similar to LMMs the approach aims to model the individual subject level more detailed. The general idea is to incorporate the within-subject variability into the group level model. This is achieved by weighting via the variance of the subject level estimates.

D. Motivation & Related Work

Experimental research is generally goal-driven. The purpose is most often to generalize from a small subset of people to the whole population. This is not always straightforward and in the past cutbacks had to be made because of the high computational cost involved. Especially in EEG research a lot of research results over the last decade are not reproducible (Pavlov et al., 2021; OpenScienceCollaboration, 2015). A major factor are the experimental degrees of freedom (Simmons et al., 2011). This includes the experimental setup as well as the analysis and the statistical methods used. Statistics in this case is often seen as more of a burden than a blessing, since statistical intuition is often lacking or complex to understand. Button et al. (2013) already reported that many studies in the field of neuroscience yield very low statistical power (8% - 30%). Baker et al. (2021)

investigated the effects of the number of trials per subject for multiple experimental paradigms and methodologies. They showed for multiple cases that statistical power could be maintained with smaller subject size but increased trial number per subject. Because of the cost-efficiency, the robustness advantage over other methods and the high temporal resolution, EEG is used in a very broad area. Not always perfect experimental conditions are guaranteed. Heise et al. (2022) inspected in their work the consequences of unbalanced designs on statistical models. The work focuses on EEG research in infants and children where it is difficult to collect / record trials due to the practical problems like e.g. short attention or excessive movement. They compare traditional two-level models to linear mixed models with respect to a unbalanced design. They propose linear mixed-effects model (LMMs) as a solution to mitigate the generated bias of a unbalanced design in the traditional two-level models.

As described above, research reporting statistically significant results (and not spurious effects) can have many pitfalls. With EEG research evolving more and more towards more complicated experiment designs, this leads to problems with established statistical analysis methods and approaches (e.g. two-way approach and summary statistics). In particular standard ways of group analysis like the two-stage approach do not incorporate within-subject uncertainty and have difficulties with unbalanced designs. As a consequence the result can be biased or even false. Linear mixed-effects models and mixed-effects multilevel analysis offer multiple advantages over the established two-level approach. I intend to show the effect of varying number of subjects, number of trials per subject and the between-subject variability on the statistical power with regard to the selected modelling-schemes.

2. MANDATORY AND OPTIONAL GOALS

This thesis is split into three main parts. The list below gives an overview over the separate goals. Mandatory goals are enumerated, optional goals are marked with a star (∗).

(A) **Simulation toolbox**
- (1) Simulating event-related responses
- (2) Varying number of subjects, trials
- (3) Addition of Noise (white, realistic)
- (∗) Addition of varying overlap
- (∗) Addition of different variance per condition

(B) **Implementation of modelling-schemes**
- (1) Two-way approach
- (2) Mixed-effect models
- (∗) Meta-models
- (4) Sanity check

(C) **Comparison of selected modelling-schemes varying...**
- (1) Number of subjects
- (2) Number of trials
- (3) Between-subject variability
- (∗) Unbalanced design
- (∗) Within-subject variability
- (∗) Signal-to-noise ratio
- (∗) Item effects

A. Simulation Toolbox

The first mandatory goal is to develop a toolbox for simulating time-series / EEG data. It is not intended to model the EEG signal based on the real underlying source space (Krol et al., 2018). The implementation aims to simulate event-related responses based on a linear model parameterization. The toolbox will be implemented in the programming language julia. The required functionalities are simulating data for a varying number of subjects, trials per subject as well as varying between-subject variability. Another required option is the addition of noise to the signal. Up to now white noise and real extracted EEG noise is considered. Furthermore, the toolbox should be constructed with the intention to be compatible with Unfold.jl (Ehinger, 2019; Ehinger and Dimigen, 2019). The functionality of the toolbox is kept small on purpose to not lose the focus on the comparison in the latter part. The scope of functionality of the simulation toolbox can be extended arbitrarily. Some interesting possibilities would be e.g. adding varying temporal overlap or different overlap/variance per condition.

B. Implementation of the modelling-schemes

The second mandatory goal is the implementation of the different selected modelling-schemes. This includes the before introduced two-way approach and the mixed-effects models. The implementation of the meta-models is considered at the present time optional. This is due to the fact that at this point in time, we are unable to estimate exactly how much time will be required for implementation. As a conclusion to the implementation of the modelling-schemes, I intend to conduct a sanity check. Therefore the implemented models are used for analysing simulated data without an effect. To verify the functionality the type 1 error (false positive) is checked and should be around 5%.

C. Comparison of selected modelling-schemes

The third mandatory goal is a comparison of the statistical models in the context of EEG. The model types selected for the comparison are two-stage models, mixed-effects models and mixed-effects multilevel models (optional). The comparison will be with regard to their statistical power of detecting an event depending on (1) the number of subjects, (2) the number of trials per subject and (3) the between-subject variability. The comparison can be easily extended if spare time is left. Possible considerations are the following factors: Unbalanced design (varying numbers of trials per subjects), within-subject variability, signal-to-noise ratio or item effects. Some of the those optional factors are also accompanied by changes of the simulation toolbox.

D. Further optional goals

Another optional stretch goal is to further research typical numbers in EEG experiments (e.g. number of subjects / trials). Furthermore the idea is to set the results of the comparison in context to the general common used numbers of subjects and trials in EEG research.

3. APPROACH

As previously described, the thesis is is divided into three main parts. The separate parts will be targeted in the order of presentation above: (A) Simulation toolbox, (B) Implementation of modelling-schemes and (C) comparison.

A. Simulation Toolbox

The goal of the simulation toolbox is to simulate time-series / EEG data. More precisely, the implementation aims to simulate event-related responses. Such ground truth is necessary for a controlled comparison of the different statistical models in the second part of the thesis. Since the simulation toolbox should be compatible with the Unfold.jl toolbox, the language of choice is julia. The underlying ground truth model is a Mixed Model with time-varying parameters. This is suitable because we want to simulate a fixed effect in multiple trials and subjects as well as random effects per subject. Advantageous is that julia provides already an extensive implementation of mixed-effects models in the toolbox MixedModels.jl (Bates et al., 2022; Alday et al., 2021). A possible way to approach the simulation could be the following: (1) Simulate experimental conditions with dummy variables using MixedModelsSim.jl. (2) Define the Mixed Model and parameters and fit the model to the dummy variables. (3) Change the time-varying parameters by iterating over each time point. (4) Add noise to the simulated signal.

B. Implementation modelling-schemes

The implementation of the separate modelling-schemes will be addressed sequentially. The two-way approach will be implemented first, afterwards the implementation of the mixed-effects model will be tackled. Exact realization of the implementation will be developed on the go. It will be an iterative process as problems as well as improvements will occur during the whole implementation process.

C. Comparison of statistical models in EEG

The latter part of the thesis focuses on the comparison of the statistical models on artificial generated ground-truth EEG data. It aims to provide a conclusion about the statistical power of each model under varying experimental conditions. In more detail the comparison will focus on varying number of subjects, varying number of trials and different between-subject variability. As first step the event-related potentials will be generated via the toolbox. This depends heavily on the speed of the implementation of the simulation toolbox. Therefore the detailed look and granularity of the data set will be further defined as soon as the necessary dependencies are finished. Thereafter the event-related potentials will be analysed using the selected approaches. Furthermore for each distinct model power contour plots as described in Baker et al. (2021) will be created. Similar to the power contour plots in Baker et al. (2021) the number of subjects, number of trials or the the between-subject variability will be plotted on the x-axis and the statistical power on the y-axis. At this point in time statistical correction for multiple comparison will be circumvented by averaging first. This could mean running only one single model for example on a average P300. If assessed as practical and implementable, cluster permutation correction is considered as an alternative to not bypass but to tackle multiple comparison.

4. INTENDED OUTCOMES

The purpose of this master thesis is to show the influence of a varying number of subjects, number of trials per subject and the between-subject variability on the statistical power of detecting an event with regard to selected modelling-schemes. Furthermore we state that more sophisticated models like linear mixed-effects models or mixed-effect multilevel models are more flexible and suitable for modern EEG analysis. We expect that the statistical power is greater for those, especially in cases where only a small number of subjects and trials per subject are available.

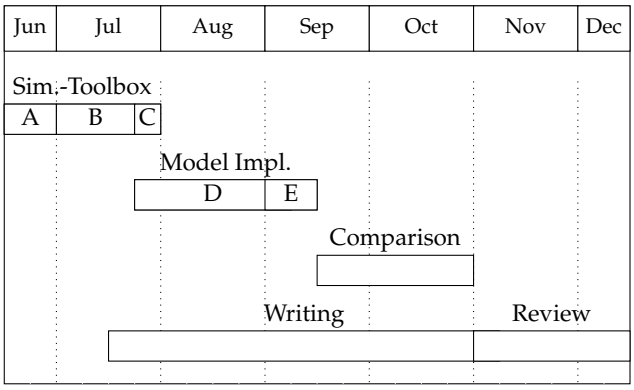

Fig. 1. Schedule

5. SCHEDULE

The project will be conducted over a period of six months. The currently planned start date is mid-June. This would result in a deadline around mid-December. A rough schedule is shown in Figure 1.The first approached task is the implementation of the simulation toolbox. The task is further subdivided into (A) research, (B) implementation and (C) testing. At the end of July the toolbox is expected to be finished and the implementation of the separate models will be addressed (D). This will be concluded by a sanity check on simulated data (E). Around mid-September everything should be set up to generate data for the comparison of the different modelling-schemes. For the comparison itself six weeks are reserved. The written elaboration of the thesis is a continuous process and will be tackled throughout the separate tasks. In addition a time buffer of around six weeks is reserved at the end for reviewing and additional writing. This schedule is not intended to be exact, instead it should give an broad overview over the approximate duration of each task.

REFERENCES

Phillip Alday, Douglas Bates, Lisa DeBruine, PhD José Bayoán Santiago Calderón, and Lisa Schwetlick. Repsychling/mixedmodelssim.jl: v0.2.6, October 2021. URL https://doi.org/10.5281/zenodo.5543934.

R.H. Baayen, D.J. Davidson, and D.M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412, November 2008. ISSN 0749596X. doi: 10.1016/j.jml.2007.12.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0749596X07001398.

Emilia Bagiella, Richard P. Sloan, and Daniel F. Heitjan. Mixed-effects models in psychophysiology. *Psychophysiology*, 37(1): 13–20, January 2000. ISSN 0048-5772, 1469-8986. doi: 10.1111/1469-8986.3710013. URL https://onlinelibrary.wiley.com/doi/10.1111/1469-8986.3710013.

Daniel H. Baker, Greta Vilidaite, Freya A. Lygo, Anika K. Smith, Tessa R. Flack, André D. Gouws, and Timothy J. Andrews. Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26(3):295–314, June 2021. ISSN 1939-1463, 1082-989X.

doi: 10.1037/met0000337. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000337.

Douglas Bates, Phillip Alday, Dave Kleinschmidt, PhD José Bayoán Santiago Calderón, Likan Zhan, Andreas Noack, Alex Arslan, Milan Bouchet-Valat, Tony Kelman, Antoine Baldassari, Benedikt Ehinger, Daniel Karrasch, Elliot Saba, Jacob Quinn, Michael Hatherly, Morten Piibeleht, Patrick Kofod Mogensen, Simon Babayan, and Yakir Luc Gagnon. Juliastats/mixedmodels.jl: v4.6.2, April 2022. URL https://doi.org/10.5281/zenodo.6450229.

Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, May 2013. ISSN 1471-003X, 1471-0048. doi: 10.1038/nrn3475. URL http://www.nature.com/articles/nrn3475.

Gang Chen, Ziad S. Saad, Audrey R. Nath, Michael S. Beauchamp, and Robert W. Cox. FMRI group analysis combining effect estimates and their variances. *NeuroImage*, 60 (1):747–765, March 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.12.060. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811911014625.

Gang Chen, Ziad S. Saad, Jennifer C. Britton, Daniel S. Pine, and Robert W. Cox. Linear mixed-effects modeling approach to FMRI group analysis. *NeuroImage*, 73:176–190, June 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.01.047. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811913000943.

Herbert H. Clark. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4):335–359, August 1973. ISSN 00225371. doi: 10.1016/S0022-5371(73)80014-3. URL https://linkinghub.elsevier.com/retrieve/pii/S0022537173800143.

Benedikt V. Ehinger. Unmixed: Linear Mixed Models combined with Overlap Correction for M/EEG analyses. An Extension to the unfold Toolbox. In *2019 Conference on Cognitive Computational Neuroscience*, Berlin, Germany, 2019. Cognitive Computational Neuroscience. doi: 10.32470/CCN.2019.1102-0. URL https://ccneuro.org/2019/Papers/ViewPapers.asp?PaperNum=1102.

Benedikt V. Ehinger and Olaf Dimigen. Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *peerJ*, 2019. doi: 10.7717/peerj.7838.

Megan J. Heise, Serena K. Mon, and Lindsay C. Bowman. Utility of linear mixed effects models for event-related potential research with infants and children. *Developmental Cognitive Neuroscience*, 54:101070, April 2022. ISSN 18789293. doi: 10.1016/j.dcn.2022.101070. URL https://linkinghub.elsevier.com/retrieve/pii/S1878929322000147.

Laurens R. Krol, Juliane Pawlitzki, Fabien Lotte, Klaus Gramann, and Thorsten O. Zander. SEREEGA: Simulating event-related EEG activity. *Journal of Neuroscience Methods*, 309:13–24, November 2018. ISSN 0165-0270. doi: 10.1016/j.jneumeth.2018.08.001. URL https://www.sciencedirect.com/science/article/pii/S0165027018302395.

OpenScienceCollaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716, 2015. doi: 10.1126/science.aac4716. URL https://www.science.org/doi/abs/10.1126/science.aac4716. _eprint: https://www.science.org/doi/pdf/10.1126/science.aac4716.

Yuri G. Pavlov, Nika Adamian, Stefan Appelhoff, Mahnaz Arvaneh, Christopher S.Y. Benwell, Christian Beste, Amy R. Bland, Daniel E. Bradford, Florian Bublatzky, Niko A. Busch, Peter E. Clayson, Damian Cruse, Artur Czeszumski, Anna Dreber, Guillaume Dumas, Benedikt Ehinger, Giorgio Ganis, Xun He, José A. Hinojosa, Christoph Huber-Huber, Michael Inzlicht, Bradley N. Jack, Magnus Johannesson, Rhiannon Jones, Evgenii Kalenkovich, Laura Kaltwasser, Hamid Karimi-Rouzbahani, Andreas Keil, Peter König, Layla Kouara, Louisa Kulke, Cecile D. Ladouceur, Nicolas Langer, Heinrich R. Liesefeld, David Luque, Annmarie MacNamara, Liad Mudrik, Muthuraman Muthuraman, Lauren B. Neal, Gustav Nilsonne, Guiomar Niso, Sebastian Ocklenburg, Robert Oostenveld, Cyril R. Pernet, Gilles Pourtois, Manuela Ruzzoli, Sarah M. Sass, Alexandre Schaefer, Magdalena Senderecka, Joel S. Snyder, Christian K. Tamnes, Emmanuelle Tognoli, Marieke K. van Vugt, Edelyn Verona, Robin Vloeberghs, Dominik Welke, Jan R. Wessel, Ilya Zakharov, and Faisal Mushtaq. #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex*, 144:213–229, November 2021. ISSN 00109452. doi: 10.1016/j.cortex.2021.03.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0010945221001106.

Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11):1359–1366, November 2011. ISSN 0956-7976, 1467-9280. doi: 10.1177/0956797611417632. URL http://journals.sagepub.com/doi/10.1177/0956797611417632.