Statistical modelling section

# Statistical modelling

Multilevel generalized linear models (GLMs) provide a powerful framework for modelling structured data (Brown and Prescott 2014),(Vonesh 2006), and are a natural fit for analysing the current data given the many important sources of heterogeneity like mouse, vessel type, vessel, treatment etc.

Bayesian multilevel GLMs allow information about latent parameters to be encoded using prior distributions, thereby conferring a number of advantages over non-Bayesian approaches including regularization, computational tractability and model identification among others (Fong, Rue, and Wakefield 2010). Bayesian multilevel GLMs have successfully been applied to many similar problems (Cobigo et al. 2022),(Sorensen and Vasishth 2016),(Wang et al. 2017).

This analysis presents a range of Bayesian multilevel GLMs, which are structurally similar but differ in their measurement distributions, parameter dependencies and prior distributions. These differences arose organically: in each case we started with a simple, naive model of the target measurement type, then iteratively added and removed components as described in (Gelman, Vehtari, et al. 2020). Our aim was to achieve the best possible quantitative and qualitative description of the underlying data generating process while avoiding computational issues.

Following standard practice for Bayesian statistics (Gelman, Carlin, et al. 2020, Ch. 1) we based all our model evaluations and conclusions on integrals over our models’ posterior distributions, which we estimated using adaptive Hamiltonian Monte Carlo via Stan (Carpenter et al. 2017).

To assess how well our models described their target data generating processes, we evaluated their out of sample predictive performance using expected leave-one-observation-out log predictive density (Vehtari, Gelman, and Gabry 2017). We complemented this quantitative evaluation with a qualitative assessment of agreement between our models’ posterior predictive distributions and the observed measurements based on graphical checks.

When we were satisfied with a model, we extracted conclusions from it by specifying a meaningful test statistic in terms of model parameters and examining the marginal posterior distribution of that statistic. For example, to evaluate the impact of a treatment on measurements of a vessel’s diameter, we could choose as a test statistic the difference in expected diameter between two otherwise similar mice, one treated and the other untreated. If the marginal posterior distribution of this statistic concentrates above zero, we conclude that our model indicates a positive effect. If the marginal posterior distribution concentrates around a certain value, we conclude that the model indicates an effect of around that size. While we generally present these results directly, for summaries we sometimes interpreted thresholds of 95% or 99% posterior mass as signifying qualitative certainty in a quantity being above or below zero.

Note that our method is different from the approach of identifying effects based on the results of null hypothesis significance testing. See (Kruschke, John K. 2015, Ch. 11) for a detailed description of the differences between these two methods. In particular, our approach does not involve null models or hypothetical unrealized datasets: the primary questions are simply whether each model adequately describes the actually realized data and, if so, what the model says.

Brown, Helen, and Robin Prescott. 2014. *Applied Mixed Models in Medicine*. John Wiley & Sons.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76 (1): 1–32. <https://doi.org/10.18637/jss.v076.i01>.

Cobigo, Yann, Matthew S. Goh, Amy Wolf, Adam M. Staffaroni, John Kornak, Bruce L. Miller, Gil D. Rabinovici, et al. 2022. “Detection of Emerging Neurodegeneration Using Bayesian Linear Mixed-Effect Modeling.” *NeuroImage: Clinical* 36 (January): 103144. <https://doi.org/10.1016/j.nicl.2022.103144>.

Fong, Youyi, Håvard Rue, and Jon Wakefield. 2010. “Bayesian Inference for Generalized Linear Mixed Models.” *Biostatistics (Oxford, England)* 11 (3): 397–412. <https://doi.org/10.1093/biostatistics/kxp053>.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2020. “Bayesian Data Analysis, Third Edition,” 656.

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. “Bayesian Workflow.” *arXiv:2011.01808 [Stat]*, November. <https://arxiv.org/abs/2011.01808>.

Kruschke, John K. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS , and Stan*. 2nd ed. Elsevier. <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>.

Sorensen, Tanner, and Shravan Vasishth. 2016. “Bayesian Linear Mixed Models Using Stan: A Tutorial for Psychologists, Linguists, and Cognitive Scientists.” *The Quantitative Methods for Psychology* 12 (3): 175–200. <https://doi.org/10.20982/tqmp.12.3.p175>.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing* 27 (5): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.

Vonesh, Edward F. 2006. “Mixed Models: Theory and Applications.” Taylor & Francis.

Wang, Ming, Zheng Li, Eun Young Lee, Mechelle M. Lewis, Lijun Zhang, Nicholas W. Sterling, Daymond Wagner, Paul Eslinger, Guangwei Du, and Xuemei Huang. 2017. “Predicting the Multi-Domain Progression of Parkinson’s Disease: A Bayesian Multivariate Generalized Linear Mixed-Effect Model.” *BMC Medical Research Methodology* 17 (1): 147. <https://doi.org/10.1186/s12874-017-0415-4>.