

Bayesian Hierarchical Modeling: An Introduction and Reassessment

Myrthe Veenman¹, Angelika M. Stefan², & Julia M. Haaf³

¹ Leiden University

² Universität der Bundeswehr München

³ University of Amsterdam

Abstract

With the recent development of easy-to-use tools for Bayesian analysis, psychologists have started to embrace Bayesian hierarchical modeling. Bayesian hierarchical models provide an intuitive account of inter- and intraindividual variability and are particularly suited for the evaluation of repeated-measures designs. Here, we provide guidance for model specification and interpretation in Bayesian hierarchical modeling and describe common pitfalls that can arise in the process of model fitting and evaluation. Our introduction gives particular emphasis to prior specification and prior sensitivity, as well as to the calculation of Bayes factors for model comparisons. We illustrate the use of state-of-the-art software programs **Stan** and **brms**. The result is an overview over best practices in Bayesian hierarchical modeling that, as we hope, will help psychologists in making the best use of Bayesian hierarchical modeling.

Keywords: Multilevel, Tutorial, rstan, brms, Bayes factor

Section in the Intro about our tutorial: Our tutorial is directed at researchers who have a basic [conceptual](#) understanding of Bayesian inference and parameter estimation and are planning to perform more complex Bayesian analyses [of psychological phenomena](#). We hope to achieve a balance of detail and overview—while we want to provide guidance for a full Bayesian workflow (similar to Gelman et al. and Schad et al.) we don't shy away from in-depth explanations for the most curious scientists. To ensure that a broad audience is able to follow the tutorial, we limit and carefully explain equations. However, rudimentary knowledge of calculus is advantageous to fully understand Bayesian analysis.

This paper is written in R Markdown. The R Markdown file, other R code, and additional information on this study can be found in the Online Supplement: <https://github.com/MyrtheV/Bayesian-Hierarchical-Modelling-An-Introduction-and-Reassessment>.

Correspondence concerning this article should be addressed to Myrthe Veenman, Wassenaarseweg 52, Leiden. E-mail: myrthe.veenman@gmail.com

Section about prior propriety. Another important issue is that wide priors are often improper probability distributions (Hobert & Casella, 1996). A prior is proper if two conditions hold: 1. All values of the probability distribution are positive definite, and 2. the probability distribution sums to one for discrete data or integrates to one for continuous data. An example of a proper and improper probability distribution are provided in Figure XXX. The figure shows two hypothetical priors for parameter θ . The first condition can be assessed by checking the y-values—both priors have only positive definite values as function output (0.1, 0.2, and 0.4). To assess the second condition, we need to sum up the function values. Panel A shows a proper distribution where the values sum to 1; panel B shows an improper distribution where the values sum to 1.2.

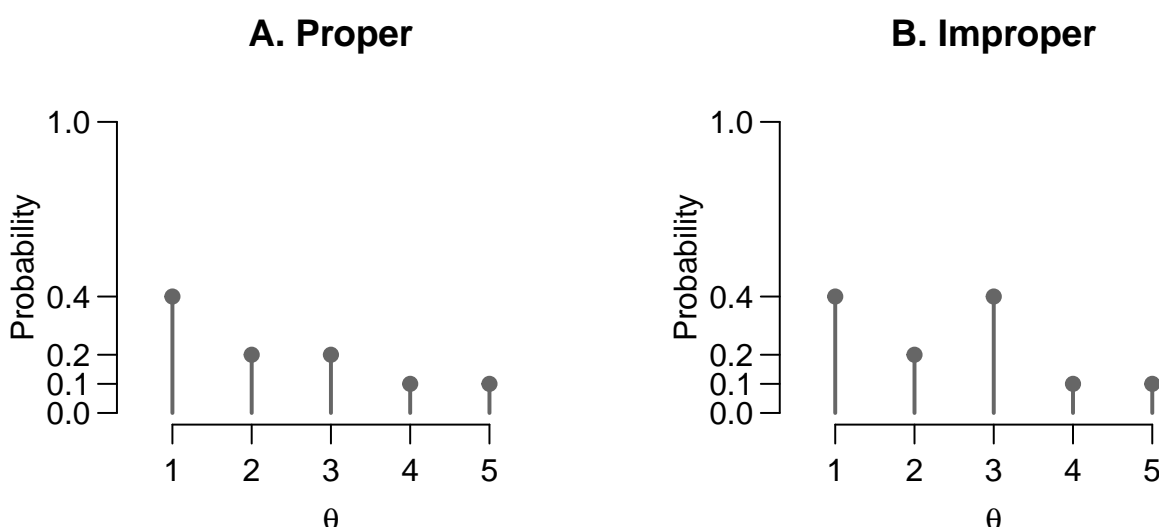


Figure 1. Two hypothetical priors for a discrete parameter θ . A. A proper probability distribution that sums to one. B. An improper probability distribution that sums to 1.2.

A more common example of an improper prior is a uniform prior distribution ranging from minus infinity to infinity. In case of improper priors, it is often impossible to obtain correct estimation of general and individual effects in hierarchical models because the resulting posteriors will again be improper probability distributions. However, even proper priors can result in improper posteriors (Hobert & Casella, 1996). Although estimation should not be possible in this situation, statistical software might still provide results without notifying the user that the posterior distribution does not exist. If users are not aware of the impropriety problem, this can result in misleading conclusions. Therefore, it is important to check if (a) the priors are proper and (b) the priors result in proper posteriors. Unfortunately, there is no fail-safe method that guarantees proper posteriors for hierarchical models. The best approach is to choose well-reasoned informative priors for the model, for instance, by using the proposed four-step procedure below. Priors that are proper and not too diffuse will in most cases lead to proper posteriors.

Section on marginal likelihood STILL WORK IN PROGRESS.

Model Comparison

The symbolic distance hypothesis postulates that response times (RTs) are slower when digits are closer to 5 (digit effects), and that RTs may be influenced by whether the target digit is smaller or greater than 5 (side effect). Solely using Bayesian estimation, we are not able to directly test these hypotheses. Credible intervals often contain zero, but they always also contain lots of other values that have a chance of being the “true” parameter value, making posterior-based intervals unsuitable to test hypotheses. Therefore, our goal for the following section is to set up a Bayesian hypothesis test for the parameters of interest in our model. Specifically, we will achieve this by comparing the predictive accuracy of a null model where we set one or some parameters to zero, and effects models where these parameters are unconstrained. [This approach is called Bayes factor model comparison.](#)

Bayes factor

[Rather than submerging us in technical details or providing a formal justification of the Bayesian approach, we provide an intuition of Bayesian model comparison. A formal justification of Bayes factors is provided by @jeffreys1961theory, @kass1995bayes and @rouder2018teaching.](#) The Bayes factor (BF) is a measure of how well one model predicts the data compared to another.

Bayesian analysis allows for predictions of data from any model because models are fully specified with a probability distribution for the data and priors for all parameters. Let the vector of all parameters of a model be $\boldsymbol{\theta}$. With the prior distributions these parameters, $f(\boldsymbol{\theta})$, the prediction for data from the model can be computed by integrating over all parameters:

$$p(\mathbf{Y}|\mathcal{M}) = \int_{\boldsymbol{\Theta}} p(\mathbf{Y}|\mathcal{M}, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (1)$$

Note that $\int_{\boldsymbol{\Theta}}$ refers to a multidimensional integral over the entire parameter space. As our models consist of many parameters computing these predictions for any one model becomes somewhat difficult. Once the integral is computed, the resulting probability distribution of the data, $p(\mathbf{Y}|\mathcal{M})$ is not a function of the model parameters anymore. This marginal probability distribution is therefore also referred to as the marginal likelihood. Before observing the data, the marginal likelihood serves as prediction for possible data; once the data are observed, the function tells us how well these specific observations were predicted by the model. The Bayes factor then is the ratio of marginal likelihoods of two models, or the relative predictive accuracy of \mathcal{M}_1 over \mathcal{M}_2 :

$$\text{BF}_{1,2} = \frac{P(\mathbf{Y}|\mathcal{M}_1)}{P(\mathbf{Y}|\mathcal{M}_2)}. \quad (2)$$

There is a second popular interpretation of the Bayes factor as the relative evidence of two competing models. One of the key insights from Bayes’ rule is that these two things, the relative predictive accuracy of two models for the data, and the relative evidence from the data for the two models, are one and the same [Rouder:Morey:2018].

Bayes factors have a few very convenient characteristics. One of them is that Bayes factors are transitive, that is, evidence for the full model over the null model can be obtained

by flipping numerator and denominator in equation 13. For a full overview of the advantages of Bayesian inference with Bayes factor, see Wagenmakers and colleagues (2018).

References to include:

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ... & Modrák, M. (2020). Bayesian workflow. arXiv preprint arXiv:2011.01808.

Schad, D. J., Betancourt, M., & Vasisht, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological methods*, 26(1), 103.

Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436), 1461–1473. doi:10.2307/2291572

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al.others. (2018). Bayesian inference for psychology. Part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. doi:10.3758/s13423-017-1343-3