



UvA-DARE (Digital Academic Repository)

Evaluating Multinomial Order Restrictions With Bridge Sampling

Sarafoglou, A.; Haaf, J.M.; Ly, A.; Gronau, Q.F.; Wagenmakers, E.-J.; Marsman, M.

DOI

[10.1037/met0000411](https://doi.org/10.1037/met0000411)

Publication date

2023

Document Version

Final published version

Published in

Psychological Methods

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care>)

[Link to publication](#)

Citation for published version (APA):

Sarafoglou, A., Haaf, J. M., Ly, A., Gronau, Q. F., Wagenmakers, E.-J., & Marsman, M. (2023). Evaluating Multinomial Order Restrictions With Bridge Sampling. *Psychological Methods*, 28(2), 322-338. <https://doi.org/10.1037/met0000411>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Evaluating Multinomial Order Restrictions With Bridge Sampling

Alexandra Sarafoglou¹, Julia M. Haaf¹, Alexander Ly^{1, 2}, Quentin F. Gronau¹, Eric-Jan Wagenmakers¹,
and Maarten Marsman¹

¹ Department of Psychology, Psychological Methods, University of Amsterdam

² Centrum Wiskunde and Informatica, Machine Learning Group, Amsterdam, the Netherlands



Abstract

Hypotheses concerning the distribution of multinomial proportions typically entail exact equality constraints that can be evaluated using standard tests. Whenever researchers formulate inequality constrained hypotheses, however, they must rely on sampling-based methods that are relatively inefficient and computationally expensive. To address this problem we developed a bridge sampling routine that allows an efficient evaluation of multinomial inequality constraints. An empirical application showcases that bridge sampling outperforms current Bayesian methods, especially when relatively little posterior mass falls in the restricted parameter space. The method is extended to mixtures between equality and inequality constrained hypotheses.

Translational Abstract

When researchers evaluate categorical variables, their hypotheses typically postulate that all category proportions are exactly equal or equal to fixed values. These hypotheses can be evaluated using standard tests. However, whenever researchers postulate hypotheses that entail expectations about increasing or decreasing trends, they must rely on methods that are relatively inefficient and computationally expensive. To address this problem we developed a bridge sampling routine that allows an efficient evaluation of these trends for multinomial variables. An empirical application showcases that bridge sampling outperforms current Bayesian methods. The method is extended so that hypotheses can also include mixtures between equality and inequality constraints.

Keywords: Bayes factors, model selection, inequality constraints, Savage-Dickey density ratio

In many scientific fields the analysis of categorical variables is of major importance. Applications range from the analysis of declared numeric values in forensic accounting, auditing, and fraud detection (Nigrini, 2012; Rauch et al., 2011); the analysis of descriptive measures in survey studies (e.g., Haberman, 1978;

Nuijten et al., 2016; Sedransk et al., 1985; Veldkamp et al., 2014); the analysis of gut microbiome composition (Song et al., 2020) to the validation of model assumptions and axioms in the field of psychometrics (see e.g., Cavagnaro & Davis-Stober, 2014; Davis-Stober, 2009; Guo & Regenwetter, 2014; Myung et al., 2005;

This article was published Online First December 16, 2021.

Alexandra Sarafoglou  <https://orcid.org/0000-0003-0031-685X>

Julia M. Haaf  <https://orcid.org/0000-0001-5122-706X>

Alexander Ly  <https://orcid.org/0000-0003-3925-3833>

Quentin F. Gronau  <https://orcid.org/0000-0001-5510-6943>

Eric-Jan Wagenmakers  <https://orcid.org/0000-0003-1596-1034>

Maarten Marsman  <https://orcid.org/0000-0001-5309-7502>

The method proposed in this article was presented as a talk at the Conference for Psychometric Computing (PsyCoCo) in Prague (February 2019) and at the International Meeting of the Psychometric Society (IMPS) in New York (July 2018).


This research was supported by a Netherlands Organization for Scientific Research (NWO) grant to Alexandra Sarafoglou (406-17-568) and to Quentin F. Gronau (406-16-528), a Veni Grant from the NWO to Maarten

Marsman (451-17-017), as well as a Vici Grant from the NWO to Eric-Jan Wagenmakers (016.Vici.170.083).

Readers can access the data from the empirical example, our online appendix, and the R code all analyses (including the creation of all figures), in our OSF folder at: <https://osf.io/59tce/>. The R package `multibridge` which implements the proposed bridge sampling method can be downloaded from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=multibridge>.

The study was approved by the local ethics board of the University of Amsterdam.

 The data are available at <https://osf.io/59tce/>

 The experiment materials are available at <https://osf.io/59tce/>

Correspondence concerning this article should be addressed to Alexandra Sarafoglou, Department of Psychology, Psychological Methods, University of Amsterdam, P.O. Box 15906, 1001 NK Amsterdam, the Netherlands. Email: alexandra.sarafoglou@gmail.com

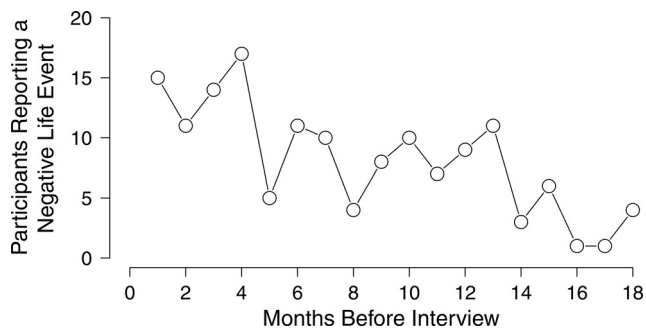
Regenwetter et al., 2011, 2018; Tijnstra et al., 2015). The breadth and depth of these examples underscore the importance of having efficient tools for their analysis readily available.

In each of the examples above, researchers are interested in quantifying evidence for hypotheses that impose certain restrictions on the underlying category proportions. These hypotheses often predict that all category proportions are exactly equal (e.g., the prevalence for a statistical reporting error is equal across different psychological journals; Veldkamp et al., 2014), or that they are fixed and follow a specific pattern (e.g., the digit proportions in nonfraudulent auditing data conform to Benford's law; Benford, 1938; Nigrini, 2012). However, research hypotheses also often stipulate ordinal expectations among category proportions (e.g., students with higher abilities have a higher chance to solve any particular item correctly; Grayson, 1988), or a mix of equality and inequality parameter constraints (for example, according to the recognition heuristic, when laypeople predict which sports team will win a tournament they assign a higher probability of winning to more familiar teams and equal but lower probabilities to unknown teams; Goldstein & Gigerenzer, 2002).

Ordinal expectations about underlying category proportions are a regular occurrence in scientific theories. However, the evaluation of hypotheses that go beyond exact equality constraints is not very popular, particularly among researchers who use frequentist statistics (Iverson, 2006). As a motivating example, consider the study conducted by Uhlenhuth et al. (1974), who surveyed 735 adults to investigate the association between symptoms of mental disorders and experienced life stress. To measure participants' life stress, the authors asked them to indicate, out of a list of negative life events, life stresses, and illnesses, which event they had experienced during the last 18 months prior to the interview. A subset of these data was reanalyzed by Haberman (1978, p. 3). Haberman noted that retrospective surveys tend to fall prey to the fallibility of human memory, causing participants to report primarily those negative events that happened most recently. He therefore investigated the 147 participants who reported only one negative life event over this time span and tested whether the frequency of the reported events was equally distributed over the 18-month period. However, Haberman did not directly test the ordinal pattern implied by his assumption of forgetting, namely that the number of reported negative life events decreases as a function of the time passed. Figure 1 shows the frequency of reported negative life events in Haberman's sample.

Figure 1

Frequency of Reported Negative Life Events Over the Course of the 18 Months Prior to the Interview for Haberman's (1978) Sample of the Data Collected by Uhlenhuth et al., (1974)



To evaluate ordinal multinomial patterns such as the one hypothesized by Haberman (1978), we focus on Bayesian methods. In the Bayesian statistical framework, researchers may quantify the evidence for or against a specific restriction on the model parameters using the Bayes factor (Jeffreys, 1935; Kass & Raftery, 1995). The Bayes factor is defined as the relative predictive performance of the models with and without the restriction, which is reflected in the ratio of their normalizing constants. For the usual scenario of equal or fixed underlying category proportions, the Bayes factor is available analytically. This is not the case, unfortunately, when inequality constraints are in play. In these cases, the Bayes factor can be approximated using the encompassing prior approach which defines the Bayes factor as the ratio of the prior and posterior probabilities that the inequality constraints hold (Klugkist et al., 2005). In the following, we highlight two particularly popular implementations of the encompassing prior approach. The first implementation is the *unconditional encompassing method* which uses straightforward sampling from the unrestricted (i.e., encompassing) distributions to compute the Bayes factor (Haaf et al., 2020; Haaf & Rouder, 2017, 2020; Hoijtink, 2011; Hoijtink et al., 2008; Klugkist et al., 2005; Schnuerch et al., 2020; van der Lans et al., 2020). The second implementation is the *conditional encompassing method* which decomposes the Bayes factor into a product of conditional probabilities (Gu et al., 2014; Laudy, 2006; Mulder, 2014, 2016; Mulder et al., 2009) and is implemented in the R packages *multinomineq* (Heck & Davis-Stober, 2019), *bain* (Gu et al., 2019), *BFpack* (Mulder et al., 2018), and the software program *BIEMS* (Mulder et al., 2012).¹

The main disadvantage of both methods is computational: The approximation of the Bayes factor becomes harder (i.e., more time-consuming and less accurate) as researchers are interested in a smaller part of the parameter space. For the unconditional encompassing method, the problem is that the probability of samples falling within the parameter space of the restricted distribution is very low, making it practically impossible to obtain accurate estimates of the Bayes factor by sampling from the unrestricted distribution.

For instance, in the Haberman example the ordinal restrictions on the 18 categories are associated with a minuscule prior mass of $1/18!$ or 1 over 6,402,373,705,728,000. Consequently, a single posterior draw that obeys the restriction will catapult the Bayes factor to extreme values unless the number of draws is impractically large. For instance, with 5 million draws, a single posterior draw that obeys the restriction will yield an estimated Bayes factor of 1.28×10^9 , massively favoring the inequality-constrained hypothesis. On the other hand, if none of the 5 million draws obeys the restriction, the estimated Bayes factor shows infinite support *against* the inequality-constrained hypothesis. This illustrates that in order to obtain a precise estimate of the Bayes factor, researchers need to draw millions of samples from the posterior. These ratios become increasingly problematic as the models become more complex (Mulder et al., 2009; Sedransk et al., 1985).

By decomposing the Bayes factor into a product of conditional probabilities, the conditional encompassing method is more stable than the unconditional encompassing method. However, the increased stability of the conditional encompassing method is

¹ For more examples in which the conditional encompassing method has been used, see the website of the R software package *bain* (Gu et al., 2019) at <https://informative-hypotheses.sites.uu.nl/software/bain/>.

accompanied by a steep increase in runtime. This increase has three reasons. The first reason follows directly from the sequential evaluation of the individual constraints. To refer again to our motivating example: Because the associated model features 17 constraints, 17 sets of prior and posterior samples must be drawn for the evaluation. The resulting runtime is thus 17 times higher than that of the unconditional encompassing method. The second reason is that even though the method is more stable, there is still the risk that the relative size of the restricted area for each individual restriction is too small to effectively sample from it (but see Gu et al., 2019 who estimate these conditional probabilities more efficiently). The third reason is the implementation of the conditional encompassing method. When evaluating the individual constraints, it is not enough to simply draw samples from the unrestricted distribution; this is only possible for the first constraint. For each additional constraint, samples are drawn from distributions that are conditional on previous constraints, with a new constraint added at each step. Thus, we need to draw samples from restricted distributions using Markov chain Monte Carlo (MCMC), and this is slower than the standard Monte Carlo methods used in the unconditional encompassing method.

To overcome the above limitation, we present a bridge sampling routine (e.g., Gronau et al., 2017; Meng & Wong, 1996) to estimate the Bayes factor for multinomial inequality constraints. The advantage of the bridge sampling routine is that its efficiency does not suffer when the size of the restricted parameter space decreases. The resulting Bayes factor estimates are relatively unbiased and precise. In addition, the bridge sampling approach has a fixed cost in terms of runtime, which makes it appealing for the implementation in standard statistical software packages. The bridge sampling method outlined in this article can be used to evaluate hypotheses that postulate (a) a monotonic increase or decrease for (a subset of) multinomial parameters (e.g., $\theta_1 < \theta_2 < \theta_3$ or $\theta_1 > \theta_2 > \theta_3$); (b) mixtures of inequality constraints and equality constraints (e.g., $\theta_1 < \theta_2 = \theta_3$); (c) mixtures of inequality constraints and free parameters (e.g., $\theta_1 < \theta_2, \theta_3$); and (d) mixtures of the first three cases (e.g., $\theta_1 < (\theta_2 = \theta_3), \theta_4$).

The outline of this article is as follows. First, we introduce the basic theoretical concepts of Bayesian parameter estimation and the computation of Bayes factors for the multinomial model featuring equality constrained hypotheses. We then extend these concepts to inequality constrained hypotheses to a mixture of equality and inequality constrained hypotheses. Third, we show how the bridge sampling approach compares with the established methods such as the unconditional encompassing method and the conditional encompassing method in terms of precision and efficiency by applying the methods to our motivating example. The last section contains a short discussion and the Appendix compares the accuracy of the bridge sampling approach to the established methods.

Bayesian Analysis of Multinomial Variables

This section introduces the theoretical concepts of Bayesian inference for the multinomial model—that is, Bayesian parameter estimation using posterior distributions and Bayesian hypothesis testing using Bayes factors. We denote the number of observations in a category k with x_k , and the total number of observations with $N = \sum_{k=1}^K x_k$. The multinomial distribution

is a generalization of the binomial distribution to variables that can take values in $K \geq 2$ categories, and it assigns the following probabilities to the different ways that N observations distribute across the K categories,

$$p(\mathbf{x} | \boldsymbol{\theta}) = p(x_1, x_2, \dots, x_K | \theta_1, \theta_2, \dots, \theta_K) = \binom{N}{x_1, x_2, \dots, x_K} \prod_{k=1}^K \theta_k^{x_k},$$

where the first factor in the likelihood denotes an extension of the binomial coefficient known as the multinomial coefficient. The parameters of the multinomial model, θ_k , reflect the probability of observing a value in a particular category, and need to sum to one. Note that due to the sum-to-one constraint, the K -th parameter is sometimes expressed as $\theta_K = 1 - \sum_{k=1}^{K-1} \theta_k$.

Prior and Posterior Distribution Without Inequality Constraints

Bayesian parameter estimation concerns the expression of a posterior distribution for model parameters capturing a priori information and information from the data (i.e., the likelihood). For the vector of probability parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)$, we choose a Dirichlet distribution with concentration parameters $(\alpha_1, \alpha_2, \dots, \alpha_K)$, where each element in $\boldsymbol{\alpha}$ is larger than zero:

$$p(\boldsymbol{\theta}) = p(\theta_1, \theta_2, \dots, \theta_K) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1}.$$

The concentration parameters α_k of the Dirichlet distribution have an intuitive interpretation: They may be interpreted as a priori category counts, and their exact values determine both the relative values of category probabilities and their variability. For the problem at hand, the posterior is also a Dirichlet distribution of the form

$$p(\boldsymbol{\theta} | \mathbf{x}) = \frac{\Gamma(N + \sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(x_k + \alpha_k)} \prod_{k=1}^K \theta_k^{x_k + \alpha_k - 1},$$

with the updated concentration parameters $\alpha'_k = x_k + \alpha_k$ (O'Hagan & Forster, 2004). The concentration parameters of the posterior Dirichlet distribution can be interpreted as a posteriori category counts, the sum of the prior and observed category counts.

Bayes Factor Hypothesis Testing Without Inequality Constraints

When stipulating exact equality constraints on the parameters of interest, researchers formulate a point null hypothesis \mathcal{H}_0 that assigns expected values \mathbf{c} to the underlying category proportions $\boldsymbol{\theta}$, that is $\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{c}$. We first consider the Bayes factor

$$\text{BF}_{0e} = \frac{p(\mathbf{x} | \mathcal{H}_0)}{p(\mathbf{x} | \mathcal{H}_e)},$$

which is defined as the ratio of normalizing constants of the null hypothesis and the encompassing hypothesis. Here, the hypothesis

\mathcal{H}_0 stipulates exact values for all of the model parameters, that is, $\mathcal{H}_0 : \boldsymbol{\theta} = \mathbf{c}$. In the standard multinomial test the null hypothesis states that all model parameters are exactly equal. Because the parameters are sum-to-one constrained, it follows that all elements in \mathbf{c} are set equal to $1/K$. We test the null hypothesis against the encompassing hypotheses which states that all category proportions are free to vary without any ordinal restrictions. We call this hypothesis the encompassing hypothesis \mathcal{H}_e , because it encompasses all possible orders of the parameters. The parameter space of the encompassing hypothesis is denoted as \mathcal{R}_e . When stipulating exact equality constraints, it is assumed that there is no prior uncertainty about the model parameters, and the marginal likelihood of the null hypothesis is simply a multinomial distribution. Due to the conjugacy of the Dirichlet distribution to the parameters of the multinomial model, the marginal likelihood for the encompassing hypothesis has a simple, closed-form solution. Thus, if all model parameters of the null hypothesis are a priori specified, the Bayes factor BF_{0e} is equal to

$$\text{BF}_{0e} = \prod_{k=1}^K c_k^{x_k} \times \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^K \alpha_k)} \times \frac{\Gamma(N + \sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k + x_k)},$$

as derived already by Good (1967). There is another way to express the Bayes factor, which relates the Bayes factor to Bayesian parameter estimation. By rearranging Bayes' rule the marginal likelihood of the encompassing hypothesis can be expressed as:

$$\underbrace{p(\mathbf{x} | \mathcal{H}_e)}_{\substack{\text{marginal} \\ \text{likelihood} \\ \text{of } \mathcal{H}_e}} = \underbrace{p(\mathbf{x} | \boldsymbol{\theta}, \mathcal{H}_e)}_{\text{likelihood}} \underbrace{\frac{p(\boldsymbol{\theta} | \mathcal{H}_e)}{p(\boldsymbol{\theta} | \mathbf{x}, \mathcal{H}_e)}}_{\substack{\text{prior} \\ \text{density} \\ \text{posterior} \\ \text{density}}}$$

which is known as Chib's identity (Chib, 1995). Chib's identity allows us to arrive at an alternative characterization of the Bayes factor that only requires the prior and posterior distribution under the alternative hypothesis at \mathbf{c} :

$$\begin{aligned} \text{BF}_{0e} &= \frac{p(\mathbf{x} | \mathcal{H}_0)}{p(\mathbf{x} | \mathcal{H}_e)} = \frac{p(\mathbf{x} | \boldsymbol{\theta} = \mathbf{c}, \mathcal{H}_e)}{\frac{p(\mathbf{x} | \boldsymbol{\theta} = \mathbf{c}, \mathcal{H}_e) p(\boldsymbol{\theta} = \mathbf{c} | \mathcal{H}_e)}{p(\boldsymbol{\theta} = \mathbf{c} | \mathbf{x}, \mathcal{H}_e)}} \\ &= \frac{\text{Height of posterior density of } \mathcal{H}_e \text{ at } \boldsymbol{\theta} = \mathbf{c}}{\frac{p(\boldsymbol{\theta} = \mathbf{c} | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta} = \mathbf{c} | \mathcal{H}_e)}} \\ &= \frac{\text{Height of posterior density of } \mathcal{H}_e \text{ at } \boldsymbol{\theta} = \mathbf{c}}{\text{Height of priordensity of } \mathcal{H}_e \text{ at } \boldsymbol{\theta} = \mathbf{c}} \end{aligned}$$

This expression is known as the Savage-Dickey density ratio (Dickey, 1971; Dickey & Lientz, 1970; O'Hagan & Forster, 2004; Verdinelli & Wasserman, 1995). The underlying principle of the Savage-Dickey density ratio is to compute the Bayes factor by dividing the height of the posterior density under \mathcal{H}_e at the point of interest (i.e., \mathbf{c}) by the height of the prior density under \mathcal{H}_e at the same point.

For concreteness, we will demonstrate the Bayesian multinomial test for exact equality constraints by reanalyzing the research question of Haberman (1978). The null hypothesis entails that the probability of reporting a negative life event is equally distributed over the 18 months prior to the interview. In particular, the expected category proportions under \mathcal{H}_0 are

$$\mathbf{c} : \theta_1, \theta_2, \dots, \theta_K = 1/K.$$

Assuming that every parameter value is equally likely before we see any data, we assign a uniform prior distribution across the parameter vector $\boldsymbol{\theta}$, such that, $p(\boldsymbol{\theta} | \mathcal{H}_e) \sim \text{Dirichlet}(\boldsymbol{\alpha})$ with all concentration parameters set to one. Using the observed frequencies from Haberman (1978), that is,

$$\mathbf{x} = (15, 11, 14, 17, 5, 11, 10, 4, 8, 10, 7, 9, 11, 3, 6, 1, 1, 4)',$$

the Bayes factor comparing the null and encompassing hypotheses is:

$$\text{BF}_{0e} = \frac{p(\boldsymbol{\theta} = \mathbf{c} | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta} = \mathbf{c} | \mathcal{H}_e)} = \frac{\frac{\Gamma(\sum_{k=1}^K \alpha_k + x_k)}{\prod_{k=1}^K \Gamma(\alpha_k + x_k)} \prod_{k=1}^K \theta_k^{x_k + \alpha_k - 1}}{\frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1}} = \frac{1}{27.1}.$$

This result indicated that the data are about 27 times more likely under \mathcal{H}_e (in which the parameters are free to vary) than under \mathcal{H}_0 (in which the parameters are constrained to be equal).

Sometimes it is desirable to compare the null hypothesis \mathcal{H}_0 not against the encompassing hypothesis \mathcal{H}_e , but against a more informed hypothesis \mathcal{H}_1 that makes specific theoretically-motivated predictions. With Bayes factors against the encompassing hypotheses in hand, the desired comparison between the null and the informed alternative can be obtained through transitivity:

$$\text{BF}_{01} = \frac{\text{BF}_{0e}}{\text{BF}_{1e}} = \frac{\frac{p(\mathbf{x} | \mathcal{H}_0)}{p(\mathbf{x} | \mathcal{H}_e)}}{\frac{p(\mathbf{x} | \mathcal{H}_1)}{p(\mathbf{x} | \mathcal{H}_e)}} = \frac{p(\mathbf{x} | \mathcal{H}_0)}{p(\mathbf{x} | \mathcal{H}_1)}.$$

This transitivity property is especially relevant when comparisons are made with restricted hypotheses, because it is more challenging to compute BF_{01} directly. Furthermore, we can use posterior model probabilities to compare the relative plausibility of any number of hypotheses Berger and Molina (2005). Assuming that all hypotheses are equally likely a priori, the posterior model probability for a particular hypothesis \mathcal{H}_1 is defined as:

$$p(\mathcal{H}_1 | \mathbf{x}) = \frac{p(\mathbf{x} | \mathcal{H}_1)}{\sum_{i=1}^I p(\mathbf{x} | \mathcal{H}_i)} = \frac{\text{BF}_{1e}}{\sum_{i=1}^I \text{BF}_{ie}}.$$

Here we have outlined how to express the prior and posterior distribution for the multinomial model using a Dirichlet prior. In addition, we expressed the Bayes factor in terms of the change of belief about the parameter value and outlined how to compare multiple hypotheses by utilizing the transitivity property of the

Bayes factor and posterior model probabilities. A related expression for the Bayes factor can be derived in the case of inequality constraints, to which we turn next.

Prior and Posterior Distribution With Inequality Constraints

When stipulating inequality-constrained hypotheses we can predict, for instance, an increasing trend of the first two categories, $\mathcal{H}_r : \theta_1 < \theta_2$. We refer to such inequality-constrained hypotheses as \mathcal{H}_r . Here, the parameter space, \mathcal{R}_r , is a subset of \mathcal{R}_e by restrictions imposed on θ , that is, $\mathcal{R}_r = \{\theta \in \mathcal{R}_e; \mathcal{H}_r\}$. The prior and posterior distributions of the parameters subject to an inequality-constrained hypothesis \mathcal{H}_r thus take the following form:

$$p(\theta | \mathcal{H}_r) = \frac{p(\theta | \mathcal{H}_e) \mathbb{I}(\theta \in \mathcal{R}_r)}{p(\theta \in \mathcal{R}_r | \mathcal{H}_e)} \quad (1)$$

$$p(\theta | \mathbf{x}, \mathcal{H}_r) = \frac{p(\theta | \mathbf{x}, \mathcal{H}_e) \mathbb{I}(\theta \in \mathcal{R}_r)}{p(\theta \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)}, \quad (2)$$

where $\mathbb{I}(\theta \in \mathcal{R}_r)$ is an indicator function that is one for parameter values θ in the restricted space \mathcal{R}_r and zero otherwise. As apparent from the equations above, the prior and posterior distributions under an inequality-constrained hypothesis are proportional to their unconstrained counterparts. In principle, whenever the concentration parameters in the Dirichlet distribution are natural numbers for every k , thus, $\alpha_k \in \mathbb{N}$, we are able to achieve an exact result for the normalizing constants for the restricted prior and posterior distribution (see the online appendix for a description of the exact procedure).² However, the exact procedure is far more inefficient than sampling-based methods, especially as the number of categories in the model and the number of observations for a fixed K increases. Here, we were only able to obtain exact results for simple cases, involving models with no more than $K = 6$ categories and no more than $N = 63$ observations. For this reason, in the following we limit our descriptions of Bayesian parameter estimation and the computation of Bayes factors to sampling-based procedures.

In general researchers rely on Monte Carlo sampling methods to compute the normalizing constants of the restricted prior and posterior distribution. In the simplest case we can use rejection sampling to simulate values from the unconstrained prior and posterior distributions and only keep those values that conform to the restrictions. The proportion of the retained samples to the total number of samples is then an approximation for the normalizing constant of the restricted distribution. Unfortunately, when many inequality constraints are proposed, the approach outlined above, although intuitive, can be terribly inefficient. For instance, in the Haberman example, when drawing from a uniform prior only 1 in over $18! = 6.4 \times 10^{15}$ samples will obey the restriction. As an alternative, we can use a MCMC approach, that allows us through random variable transformation to simulate the values directly from the restricted distribution. Devroye (1986, p. 594), for instance, shows that one can simulate values from a Dirichlet distribution by first simulating K independent random variables γ_k with a Gamma(α_k , 1) density, for $k = 1, \dots, K$, and then setting

$$\theta_k = \frac{\gamma_k}{\sum_{k=1}^K \gamma_k}.$$

The variables θ_k that are generated in this way follow the desired Dirichlet(α) distribution (see Klugkist et al., 2010 for an application in the context of contingency tables). Note that with the transformation from θ to γ the sum-to-one constraint is conveniently removed. Additionally, this MCMC method is suitable for drawing values from the restricted distribution because the transformation between θ and γ is order-preserving. Thus, an inequality-constrained hypothesis $\mathcal{H}_r : \theta_1 < \theta_2$ on the category probabilities translates into the inequality-constrained hypothesis $\mathcal{H}_r : \gamma_1 < \gamma_2$ on the gamma variables. If we simulate the gamma variables consistent with the order restrictions imposed by \mathcal{H}_r , that is, $p(\gamma | \mathcal{H}_r)$, the transformed gamma variables then generate Dirichlet variables that are consistent with \mathcal{H}_r , that is, $p(\theta | \mathcal{H}_r)$.

To draw gamma variables that obey the order imposed by the inequality-constrained hypotheses we use the Gibbs sampling algorithm proposed by Damien and Walker (2001). Their Gibbs sampling algorithm assumes fixed upper and lower bounds for each parameter. However, the algorithm can easily be generalized to cases where we wish to draw from gamma variables whose upper and lower bounds are not known, but are itself random variables (as it is the case for inequality-constrained hypotheses).

Instead of simulating values directly from the multivariate distribution of gamma variables that are subject to inequality constraints— $p(\gamma | \mathcal{H}_r)$ —the Gibbs sampler operates by iteratively simulating values from the full-conditional posterior distributions, that is, the distribution of one gamma variable given the remaining gamma variables and inequality constraints— $p(\gamma_k | \gamma^{(k)}, \mathcal{H}_r)$, where $\gamma^{(k)}$ refers to the vector of gamma variables with the k th parameter removed. If there is no constraint on a gamma variable γ_k then the full conditional is simply the regular Gamma(α_k , 1) density. However, if γ_k is subject to a constraint, for instance, $\gamma_j < \gamma_k < \gamma_q$, then the gamma variable γ_k has the bounded support $[\gamma_j, \gamma_q]$ instead of $[0, \infty)$. This implies that the full conditional distribution of γ_k subject to an inequality constraint is a truncated gamma distribution:

$$\begin{aligned} p(\gamma_k | \gamma^{(k)}, \mathcal{H}_r) &= p(\gamma_k | \gamma_j < \gamma_k < \gamma_q) \\ &= \frac{\frac{1}{\Gamma(\gamma_k)} \gamma_k^{\alpha_k - 1} e^{-\gamma_k} \mathbb{I}(\gamma_k \in [\gamma_j, \gamma_q])}{p(\gamma_k \in [\gamma_j, \gamma_q])}. \end{aligned}$$

For gamma variables with bounded support $[\gamma_j, \gamma_q]$, the bounds at iteration t are calculated using the current values of the parameters. After the gamma variables have been simulated in this manner, they can be transformed back into category probabilities to yield samples from the Dirichlet distribution. Sampling from the prior and posterior is useful when we wish to estimate the parameters or when draws from restricted distributions are required to

² Whenever the concentration parameters are natural numbers and we express the problem in terms of stick-breaking parameters the computations involve integrating polynomials, which makes the result exact. For general α , however, we do not have polynomials and thus an exact result is not expected.

compute the Bayes factor, as is the case with the conditional encompassing method and the bridge sampling method.

Bayes Factor Hypothesis Testing for Inequality Constraints

We consider the Bayes factor

$$BF_{re} = \frac{p(\mathbf{x} | \mathcal{H}_r)}{p(\mathbf{x} | \mathcal{H}_e)},$$

where the hypothesis \mathcal{H}_r stipulates inequality constraints on the model parameters, for instance,

$$\mathcal{H}_r : \theta_1 < \dots < \theta_K.$$

In order to obtain the marginal likelihood of the inequality-constrained hypothesis we need to integrate over the restricted parameter space \mathcal{R}_r , which makes the Bayes factor BF_{re} difficult to compute:

$$p(\mathbf{x} | \mathcal{H}_r) = \int_{\mathcal{R}_r} p(\mathbf{x} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathcal{H}_r) d\boldsymbol{\theta}.$$

It is nevertheless possible to arrive at an intuitive expression of the Bayes factor. This expression is a generalization of the Savage-Dickey density ratio mentioned above and follows from an alternative characterization of $p(\mathbf{x} | \mathcal{H}_r)$:

$$p(\mathbf{x} | \mathcal{H}_r) = \frac{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e) p(\mathbf{x} | \mathcal{H}_e)}{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)},$$

which was derived in Klugkist et al. (2005). With this characterization the Bayes factor amounts to

$$BF_{re} = \frac{\text{Proportion of posterior parameter space consistent with the restriction}}{\text{Proportion of prior parameter space consistent with the restriction}}. \quad (3)$$

Like the Savage-Dickey density ratio, this presents the Bayes factor as the change of belief that the parameters lie in the restricted parameter space \mathcal{R}_r (see also Wetzels et al., 2010 and Mulder et al., 2009). We discuss two established procedures to approximate the Bayes factor BF_{re} in the next section.

Established Procedures to Estimate the Bayes Factor for Inequality-Constraints

One popular method to estimate the Bayes factor for inequality-constrained hypotheses is the unconditional encompassing method which relies on simple Monte Carlo estimates (Gelfand et al., 1992; Klugkist et al., 2005; Sedransk et al., 1985). This method estimates the Bayes factor in Equation 3 by considering the proportion of the prior and posterior distributions of the unrestricted distribution that are in agreement with the constraints. That is, the

numerator can be estimated by sampling from the encompassing posterior density and then calculating the proportion of draws in accordance with the restrictions imposed by the inequality-constrained hypothesis. Likewise, the denominator can be estimated by sampling from the encompassing prior density and then calculating the proportion of draws in accordance with the restrictions:

$$BF_{re} = \frac{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)} \approx \frac{\frac{1}{S} \sum_{s=1}^S \mathbb{I}(\boldsymbol{\theta}'_s \in \mathcal{R}_r)}{\frac{1}{S} \sum_{s=1}^S \mathbb{I}(\boldsymbol{\theta}^*_s \in \mathcal{R}_r)},$$

where $\boldsymbol{\theta}^*_s$ and $\boldsymbol{\theta}'_s$ denote the s -th sample from the encompassing prior and posterior distribution, respectively, for samples $s = 1, \dots, S$. The simplicity of the method has stimulated application to numerous statistical scenarios, ranging from contingency tables and the analysis of variance and covariance models (Hojtink et al., 2008), to item response theory (Haaf et al., 2020), to meta-analysis models (Haaf & Rouder, 2020), to linear mixed models (Haaf & Rouder, 2017), and to circular mixed effects models (van der Lans et al., 2020). However, it is also widely recognized that this method is not particularly efficient for models with an increasing number of independent constraints (Myung et al., 2008; Sedransk et al., 1985). The same holds true for models with a small number of constraints that are extremely restrictive or models for which the data do not align with the inequality-constrained hypothesis. This is the case because the efficiency of the method relies on the relative size of the restricted area: If prior and posterior samples almost never fall inside the area of interest, a large number of samples is required to estimate the proportions accurately (Gelfand et al., 1992; Hoijtink, 2011).

A method that is more stable for larger models is the conditional encompassing method (Mulder et al., for an application to multinomial models, see Heck & Davis-Stober, 2019). The conditional encompassing method also utilizes the identity in Equation 3. But instead of estimating the normalizing constants of the constrained distribution based on a single set of samples from the encompassing distribution, Mulder et al. (2009) proposed a stepwise approach. For instance, when evaluating a hypotheses concerning $K = 4$ ordered parameters $\mathcal{H}_r : \theta_1 < \theta_2 < \theta_3 < \theta_4$, the proportion of prior parameter space consistent with the restriction can be factored as follows:

$$p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e) = p(\theta_1 < \theta_2 | \mathcal{H}_e) \times p(\theta_2 < \theta_3 | \theta_1 < \theta_2, \mathcal{H}_e) \times p(\theta_3 < \theta_4 | \theta_1 < \theta_2 < \theta_3, \mathcal{H}_e).$$

The proportion of posterior samples consistent with the restriction are estimated in a similar fashion, which yields the Bayes factor:

$$BF_{re} = \frac{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)} = \frac{p(\theta_1 < \theta_2 | \mathbf{x}, \mathcal{H}_e) \times \dots \times p(\theta_3 < \theta_4 | \theta_1 < \theta_2 < \theta_3, \mathbf{x}, \mathcal{H}_e)}{p(\theta_1 < \theta_2 | \mathcal{H}_e) \times \dots \times p(\theta_3 < \theta_4 | \theta_1 < \theta_2 < \theta_3, \mathcal{H}_e)},$$

where each Bayes factor is estimated independently. By evaluating the constraints sequentially, the conditional encompassing method

yields better results for models featuring larger numbers of constraints (Mulder et al., 2009). It is noteworthy that the conditional encompassing method was first used to evaluate almost-equality constraints: using the transitivity property of the Bayes factor, Laudy (2006, p. 115) and Klugkist (2008) proposed to approximate the Bayes factor for almost-equality constraints by evaluating a series of hypotheses of increasing narrowness, such that for each pair of parameters $\theta_1 \approx \theta_2$ the distance between them approaches zero (i.e., $|\theta_1 - \theta_2| \rightarrow 0$).³ However, care must be taken not to set the values for the distance $|\theta_1 - \theta_2|$ too small, or otherwise the method becomes inefficient (Klugkist, 2008).

A Bridge Sampling Routine to Estimate the Bayes Factor

The main limitations of the unconditional encompassing method and the conditional encompassing method—lack of precision, lack of scalability, and long runtimes—come from the effort to estimate the proportion of the encompassing parameter space in accordance with the constraint. In contrast, bridge sampling (Bennett, 1976; Meng & Wong, 1996) estimates the Bayes factor using a different approach. The basic principle of bridge sampling is that the ratio between two normalizing constants operating on the same parameter space can be estimated by the following identity:

$$\text{BF}_{12} = \frac{p(\mathbf{x} | \mathcal{H}_1)}{p(\mathbf{x} | \mathcal{H}_2)} = \frac{\mathbb{E}_{\mathcal{H}_2}(p(\mathbf{x} | \boldsymbol{\theta}, \mathcal{H}_1)p(\boldsymbol{\theta} | \mathcal{H}_1)h(\boldsymbol{\theta}))}{\mathbb{E}_{\mathcal{H}_1}(p(\mathbf{x} | \boldsymbol{\theta}, \mathcal{H}_2)p(\boldsymbol{\theta} | \mathcal{H}_2)h(\boldsymbol{\theta}))}, \quad (4)$$

where the term $h(\boldsymbol{\theta})$ refers to a bridge function that ensures that the denominator is nonzero. In this case we choose the optimal bridge function as proposed by Meng and Wong (1996). Instead of estimating the Bayes factor directly, we use a modified form of the bridge identity proposed by Overstall and Forster (2010) which estimates only a single normalizing constant instead of the Bayes factor to further increase the precision of the estimates (for a tutorial on the bridge sampling method, see Gronau et al., 2017). The modified form of the bridge identity requires that the second distribution is chosen such that it has overlapping support with the target distribution and has a known normalizing constant. In the following, we will refer to this distribution as proposal distribution $g(\boldsymbol{\theta})$. The modified identity then becomes:

$$p(\mathbf{x} | \mathcal{H}_1) = \frac{\mathbb{E}_{g(\boldsymbol{\theta})}(p(\mathbf{x} | \boldsymbol{\theta})p(\boldsymbol{\theta} | \mathcal{H}_1)h(\boldsymbol{\theta}))}{\mathbb{E}_{\mathcal{H}_1}(g(\boldsymbol{\theta})h(\boldsymbol{\theta}))}, \quad (5)$$

where $p(\mathbf{x} | \mathcal{H}_1)$ indicates a normalizing constant we wish to estimate, that is, the normalizing constant of the constrained prior distribution, or the normalizing constant of the constrained posterior distribution, that is, $p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)$ or $p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)$, respectively. Because these normalizing constants are of the form

$$\int_{\mathcal{R}_r} p(\boldsymbol{\theta} | \mathcal{H}_e) d\boldsymbol{\theta} \quad \text{and} \quad \int_{\mathcal{R}_r} p(\boldsymbol{\theta} | \mathbf{x}, \mathcal{H}_e) d\boldsymbol{\theta}$$

the bridge sampler can be used to estimate them, if the support of the proposal distribution $g(\boldsymbol{\theta})$ is \mathcal{R}_r . That is, the restricted distribution and the proposal distribution need to operate on the same parameter space. As we will discuss in the next section, we

will facilitate this overlapping support by applying a series of transformation on the parameters of the restricted distribution.

To arrive at the expression for the bridge sampling identity for the normalizing constant of the constrained prior distribution we now simply replace the terms related to \mathcal{H}_1 . Specifically, because

$$p(\boldsymbol{\theta} | \mathcal{H}_r) = \frac{p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r)}{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)},$$

we can replace the term for the unnormalized density under \mathcal{H}_1 in the numerator of Equation 5 (i.e., $p(\mathbf{x} | \boldsymbol{\theta})p(\boldsymbol{\theta} | \mathcal{H}_1)$) by the corresponding term for the constrained prior distribution, that is, $p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r)$. Thus, the resulting bridge sampling identity can be described as follows:

$$p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e) = \frac{\mathbb{E}_{g(\boldsymbol{\theta})}(p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r)h(\boldsymbol{\theta}))}{\mathbb{E}_{\text{prior}}(g(\boldsymbol{\theta})h(\boldsymbol{\theta}))}. \quad (6)$$

The normalizing constant for the constrained posterior distribution can be described similarly. Based on this identity, we can now define the corresponding estimator. We substitute the expectations by sample averages, using N_1 samples from the constrained prior distribution, that is, $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta} | \mathcal{H}_r)$ and N_2 samples from a suitable proposal distribution, that is $\tilde{\boldsymbol{\theta}} \sim g(\boldsymbol{\theta})$. Then, we can estimate $p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)$ by:

$$\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e) \approx \frac{\frac{1}{N_2} \sum_{m=1}^{N_2} p(\tilde{\boldsymbol{\theta}}_m | \mathcal{H}_e) \mathbb{I}(\tilde{\boldsymbol{\theta}}_m \in \mathcal{R}_r) h(\tilde{\boldsymbol{\theta}}_m)}{\frac{1}{N_1} \sum_{n=1}^{N_1} g(\boldsymbol{\theta}_n^*) h(\boldsymbol{\theta}_n^*)}. \quad (7)$$

There are many possible choices for $h(\boldsymbol{\theta})$. Meng and Wong (1996) suggested the use of a bridge function that has been shown to minimize the relative mean square error of the estimate. However, when following this recommendation, the specific choice for $h(\boldsymbol{\theta})$ depends on the unknown normalization constant:

$$h(\boldsymbol{\theta}) = c \times \frac{1}{s_1 p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r) + s_2 p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e) g(\boldsymbol{\theta})},$$

where $s_1 = \frac{N_1}{N_2 + N_1}$, $s_2 = \frac{N_2}{N_2 + N_1}$ and c is a constant that has no influence on the results. To be able to estimate the normalizing constant of the constrained prior distribution we use the iterative scheme proposed by Meng and Wong (1996). Thus, we yield the following formula for the bridge sampling estimator at iteration $t + 1$:

$$\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)^{(t+1)} \approx \frac{\frac{1}{N_2} \sum_{m=1}^{N_2} \frac{\ell_{2,m}}{s_1 \ell_{2,m} + s_2 p(\tilde{\boldsymbol{\theta}}_m \in \mathcal{R}_r | \mathcal{H}_e)^{(t)}}}{\frac{1}{N_1} \sum_{n=1}^{N_1} \frac{1}{s_1 \ell_{1,n} + s_2 p(\boldsymbol{\theta}_n^* \in \mathcal{R}_r | \mathcal{H}_e)^{(t)}}}, \quad (8)$$

³Wetzels et al. (2010) showed that the proposed almost-equality constrained method approximates the Savage-Dickey density ratio.

where $\ell_{1,n} = \frac{p(\boldsymbol{\theta}_n^* | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta}_n^* \in \mathcal{R}_r)}{g(\boldsymbol{\theta}_n^*)}$ and $\ell_{2,m} = \frac{p(\tilde{\boldsymbol{\theta}}_m | \mathcal{H}_e) \mathbb{I}(\tilde{\boldsymbol{\theta}}_m \in \mathcal{R}_r)}{g(\tilde{\boldsymbol{\theta}}_m)}$.⁴ We then run the iterative scheme until a predefined tolerance criterion is reached. We follow the suggestion by Gronau et al. (2017) to use a tolerance criterion of

$$\frac{|\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)^{(t+1)} - \hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)^{(t)}|}{\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)^{(t+1)}} \leq 10^{-10},$$

while setting $\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)^{(1)} = 0$ as initial guess. To summarize, we can use bridge sampling to separately estimate the normalizing constants for the restricted prior distribution and the restricted posterior distribution. Then, we can use these two estimates to compute the Bayes factor:

$$\text{BF}_{re} \approx \frac{\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)}{\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)}.$$

Transformations to Facilitate Bridge Sampling

Given that the bridge function is defined on the common support of the proposal and target distribution, both distributions have to operate on the same parameter space. In addition, the normalizing constant of the proposal distribution must be known, which means that we cannot choose another constrained Dirichlet distribution. To resolve this problem we move the prior and posterior draws from the probability space to the real line using a probit transformation. This transformation aims to eliminate the constraints inherent to the restricted Dirichlet distribution, namely the sum-to-one constraint and the inequality constraints. Furthermore, the transformation enables us to choose a convenient proposal distribution that is easy to sample from and easy to evaluate, for instance, the multivariate normal distribution (Overstall & Forster, 2010).

The general idea is as follows: $\boldsymbol{\theta}$ is a probability vector; therefore, its elements must sum to one. As a result, the vector is completely determined by its first $K - 1$ elements. For the transformation we therefore only consider the first $K - 1$ elements and transform them to $K - 1$ elements of a new vector $\boldsymbol{\xi}$ with $\boldsymbol{\xi} \in \mathbb{R}^{K-1}$. To retain the inequality constraints imposed on the parameters, we need to account for the lower bound l_k and the upper bound u_k of each θ_k . These bounds can be determined by adapting a stick-breaking approach (Frigyik et al., 2010; Stan Development Team, 2020). The stick-breaking approach represents $\boldsymbol{\theta}$ as a stick of length one which we subsequently break into K elements. Assuming $\theta_{k-1} < \theta_k$, for $k \in \{1, \dots, K\}$, the lower bound for any element in $\boldsymbol{\theta}$ is defined as

$$l_k = \begin{cases} 0 & \text{if } k = 1 \\ \theta_{k-1} & \text{if } 1 < k < K. \end{cases} \quad (9)$$

The upper bound is defined as

$$u_k = \begin{cases} \frac{1}{K} & \text{if } k = 1 \\ \frac{1 - \sum_{j < k} \theta_j}{K + 1 - k} & \text{if } 1 < k < K, \end{cases} \quad (10)$$

where $1 - \sum_{j < k} \theta_j$ represents the length of the remaining stick and $K + 1 - k$ is the number of elements in the remaining stick. Let ϕ

denote the density of a normal variable with a mean of zero and a variance of one, Φ its cumulative density function, and Φ^{-1} its inverse cumulative density function. Then, the transformation of $\boldsymbol{\theta}$ is given by:

$$\begin{aligned} \xi_k &= \Phi^{-1} \left(\frac{\theta_k - l_k}{u_k - l_k} \right). \\ &= \begin{cases} \Phi^{-1} \left(\frac{\theta_k}{1/K} \right) & \text{if } k = 1 \\ \Phi^{-1} \left(\frac{\theta_k - \theta_{k-1}}{1 - \sum_{j < k} \theta_j} \right) & \text{if } 1 < k < K - 1. \end{cases} \end{aligned}$$

The inverse transformation is given by:

$$\begin{aligned} \theta_k &= (u_k - l_k)\Phi(\xi_k) + l_k \\ &= \begin{cases} \frac{1}{K}\Phi(\xi_k) & \text{if } k = 1 \\ \left(\frac{1 - \sum_{j < k} \theta_j}{K + 1 - k} - \theta_{k-1} \right) \Phi(\xi_k) + \theta_{k-1} & \text{if } 1 < k < K. \end{cases} \end{aligned}$$

In the inverse transformation θ_k depends only on the first k elements of $\boldsymbol{\xi}$. Therefore, we know that the Jacobian matrix will be lower triangular, and the determinant of the Jacobian matrix will be the product of the diagonal entries given by:

$$\frac{\partial \theta_k}{\partial \xi_k} = \begin{cases} \frac{1}{K}\phi(\xi_k) & \text{if } k = 1 \\ (u_k - l_k)\phi(\xi_k) & \text{if } 1 < k < K. \end{cases}$$

Therefore, the Jacobian can be computed using the upper and lower bounds for all samples determined in the transformation step:

$$|J| = \frac{1}{K} \phi(\xi_1) \prod_{k=2}^{K-1} ((u_k - l_k)\phi(\xi_k)).$$

Taking this transformation into account the bridge sampling estimator computes $\ell_{1,n}$ and $\ell_{2,m}$ as follows:

$$\begin{aligned} \ell_{1,n} &= \frac{p(\boldsymbol{\theta}_n^* | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta}_n^* \in \mathcal{R}_r)}{g(\boldsymbol{\xi}_n^*)}, \\ \ell_{2,m} &= \frac{p(\tilde{\boldsymbol{\theta}}_m | \mathcal{H}_e) \mathbb{I}(\tilde{\boldsymbol{\theta}}_m \in \mathcal{R}_r)}{g(\tilde{\boldsymbol{\xi}}_m)}, \end{aligned}$$

where $\boldsymbol{\xi}_n^* = \Phi^{-1} \left(\frac{\boldsymbol{\theta}_n^* - \mathbf{l}}{\mathbf{u} - \mathbf{l}} \right)$, and $\tilde{\boldsymbol{\theta}}_m = ((\mathbf{u} - \mathbf{l})\Phi(\tilde{\boldsymbol{\xi}}_m) + \mathbf{l}) | J |$.

⁴ Equation 8 illustrates another advantage of bridge sampling: its robustness to the tail behavior of the proposal distribution. Because both the numerator and denominator are bounded, samples from the tail region of the distributions cannot dominate the bridge sampling estimate. In this sense the bridge sampler improves on other estimation methods—such as the importance sampling estimator or the generalized harmonic mean estimator—whose variance depend on ratios that are potentially unbounded for poorly chosen proposal distributions (Frühwirth-Schnatter, 2004; Gronau et al., 2017).

Taken together, to apply the proposed bridge sampling routine, the following three conditions must be met. First, we need to be able to sample directly from the constrained prior and posterior densities, which can be achieved by using the adapted version of the Gibbs sampling method by [Damien and Walker \(2001\)](#) described above. Second, we need to select a suitable proposal distribution for the bridge sampling algorithm; here we choose a multivariate normal distribution that achieves sufficient overlap with our target distribution by moving the samples from the restricted Dirichlet distribution to the real line. Third, we need to choose a bridge function; here, we have chosen the bridge function proposed in [Meng and Wong \(1996\)](#) which has the favorable property that it minimizes the estimated relative mean-squared error.

Given that bridge sampling only requires draws of the restricted distribution and the proposal distribution, this method is more efficient than the unconditional encompassing method (because fewer samples are typically needed) and the conditional encompassing method (because fewer instances of the Gibbs sampler are needed). In addition, the precision of the bridge sampling estimator depends not on the relative size of the restricted parameter space, but on the overlap between the target and proposal distribution; when the proposal distribution resembles the target distribution more closely, the resulting estimates are more accurate ([Meng & Wong, 1996](#)).

Bayes Factors for Mixed Constraints

In addition to pure equality-constrained and pure inequality-constrained hypotheses, researchers may want to specify hypotheses with some parameters that are exactly equal to each other while others can vary freely and again others are ordered (see, e.g., [Pericchi Guerra et al., 2008](#)). However, it is not intuitively clear how to compute Bayes factors when parametric constraints are mixed. Without loss of generality, we first consider a mixed hypothesis \mathcal{H}_m where the first j category parameters are constrained to be exactly equal and where the remaining $K - j$ parameters are increasing:

$$\mathcal{H}_m : (\theta_1 = \theta_2 = \dots = \theta_j) < \theta_{j+1} < \dots < \theta_K.$$

As shown in [Equation 3](#), the Bayes factor of restricted hypotheses against the encompassing hypothesis can be formulated as

$$BF_{me} = \frac{p(\boldsymbol{\theta} \in \mathcal{R}_m | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta} \in \mathcal{R}_m | \mathcal{H}_e)}.$$

The mixed hypothesis stipulates the following set of constraints

$$\mathcal{R}_m : (\theta_1 = \dots = \theta_j) \cap (\theta_j < \dots < \theta_K) = \mathcal{R}_0 \cap \mathcal{R}_r.$$

The first set of constraints, which we denote with \mathcal{R}_0 , are the equality constraints, and the second set of constraints, which we denote with \mathcal{R}_r , are the inequality constraints. Using this notation, the Bayes factor can be reformulated as

$$BF_{me} = \underbrace{\frac{p(\boldsymbol{\theta}_r \in \mathcal{R}_r | \boldsymbol{\theta}_0 \in \mathcal{R}_0, \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta}_r \in \mathcal{R}_r | \boldsymbol{\theta}_0 \in \mathcal{R}_0, \mathcal{H}_e)}}_{BF_r} \times \underbrace{\frac{p(\boldsymbol{\theta}_0 \in \mathcal{R}_0 | \mathbf{x}, \mathcal{H}_e)}{p(\boldsymbol{\theta}_0 \in \mathcal{R}_0 | \mathcal{H}_e)}}_{BF_0},$$

that is, a conditional Bayes factor for the inequality constraints given the equality constraints and a Bayes factor for

the equality constraints. The latter is similar to the Savage-Dickey ratio that we discussed before, but involves a correction for marginalization.

The probabilities above crucially depend on the marginal probabilities $p(\theta_0 \in \mathcal{R}_0 | \mathcal{H}_e)$ and $p(\theta_0 \in \mathcal{R}_0 | \mathbf{x}, \mathcal{H}_e)$, which are derived from the prior and posterior Dirichlet distributions, respectively. Because the derivations and results are the same for the prior and posterior probabilities, we derive it here for the prior distribution. The prior probability is of the form

$$p(\boldsymbol{\theta}_0 \in \mathcal{R}_0 | \mathcal{H}_e) = \frac{1}{B(\boldsymbol{\alpha})} \int_{\mathcal{R}_e \setminus \mathcal{R}_0} \theta_j^{\alpha_j - 1} \prod_{k=j+1}^{K-1} \theta_k^{\alpha_k - 1} \left(1 - j\theta_j - \sum_{k=j+1}^{K-1} \theta_k \right)^{\alpha_K - 1} d\boldsymbol{\theta}_r,$$

and involves a Dirichlet integral, except that the first j probabilities are now collapsed. Here, we have used $\mathcal{R}_e \setminus \mathcal{R}_0$ to denote the unconstrained parameter space for the parameters $\boldsymbol{\theta}_r = (\theta_j, \dots, \theta_{K-1})^T$. We introduce a change of variable $\lambda_j = j\theta_j$, and $\lambda_k = \theta_k$, for $k = j + 1, \dots, K - 1$, with $|J| = 1/j$, such that

$$\begin{aligned} p(\boldsymbol{\theta}_0 \in \mathcal{R}_0 | \mathcal{H}_e) &= \frac{1}{jB(\boldsymbol{\alpha})} \int_{\mathcal{R}_e \setminus \mathcal{R}_0} \binom{\lambda_j}{j} \sum_{k=1}^j \alpha_k - j \prod_{k=j+1}^{K-1} \theta_k^{\alpha_k - 1} \left(1 - \lambda_j - \sum_{k=j+1}^{K-1} \theta_k \right)^{\alpha_K - 1} d\lambda_r \\ &= \frac{1}{B(\boldsymbol{\alpha})} \binom{1}{j} \sum_{k=1}^j \alpha_k - j + 1 \text{B} \left(\sum_{k=1}^j \alpha_k - j + 1, \alpha_{j+1}, \dots, \alpha_K \right), \end{aligned}$$

which allows us to express the (marginal) Bayes factor for the equality constraints as

$$BF_{e0} = \frac{B(\boldsymbol{\alpha})}{B(\boldsymbol{\alpha} + \mathbf{x})} \binom{1}{j} \sum_{k=1}^j x_k \frac{\text{B} \left(\sum_{k=1}^j (\alpha_k + x_k) - j + 1, \alpha_{j+1} + x_{j+1}, \dots, \alpha_K + x_K \right)}{\text{B} \left(\sum_{k=1}^j \alpha_k - j + 1, \alpha_{j+1}, \dots, \alpha_K \right)},$$

where the latter factor introduces a correction for marginalizing which originates from the marginalization of the remaining free parameters, including the collapsed category parameter. If it is the case that no free parameters are involved, that is, \mathcal{H}_0 assigns expected category proportions to the entire parameter vector $\boldsymbol{\theta}$ (such as in the multinomial test), then the Bayes factor for the equality constraints corresponds to the Savage-Dickey density ratio.⁵ It readily follows that the conditional Bayes factor of inequality constraints given the equality constraints now involves expectations over the conditional Dirichlet distributions

$$p(\boldsymbol{\theta}_r | \boldsymbol{\theta}_0 \in \mathcal{R}_0, \mathcal{H}_e) = \text{Dirichlet} \left(\sum_{k=1}^j \alpha_k - j + 1, \alpha_{j+1}, \dots, \alpha_K \right)$$

and

⁵ When stipulating exact equality constraints on all parameters, it is assumed that there is no prior uncertainty about the model parameters, and the likelihood of the constrained hypothesis marginalized over the parameter space is simply a multinomial distribution. This expression follows from the fact that the prior distribution under \mathcal{H}_0 is

$$p(\boldsymbol{\theta} | \mathcal{H}_0) = \frac{p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} = \mathbf{c})}{\int_{\mathcal{R}_e} p(\boldsymbol{\theta} | \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} = \mathbf{c}) d\boldsymbol{\theta}} = \frac{p(\boldsymbol{\theta} = \mathbf{c} | \mathcal{H}_e)}{p(\boldsymbol{\theta} = \mathbf{c} | \mathcal{H}_e)} = 1,$$

for $\boldsymbol{\theta} = \mathbf{c}$ and 0 otherwise.

$$p(\theta_r | \theta_0 \in \mathcal{R}_0, \mathbf{x}, \mathcal{H}_e) = \text{Dirichlet} \left(\sum_{k=1}^j (\alpha_k + x_k) - j + 1, \alpha_{j+1} + x_{j+1}, \dots, \alpha_K + x_K \right),$$

which can be computed, as before, using bridge sampling. To generalize the above derivations for any set of mixed constraints, we note that the conditional Dirichlet distribution adds the parameters for the collapsed categories and corrects for the change in degrees of freedom by subtracting the degrees of freedom it lost; $j - 1$ degrees of freedom are lost if j categories are collapsed. Thus, for mixed hypotheses of the form

$$\mathcal{H}_m : \theta_1 < \theta_2 = \theta_3 < \theta_4 = \theta_5 = \theta_6,$$

we find the following conditional Dirichlet distribution $p(\theta_r | \theta_0 \in \mathcal{R}_0, \mathcal{H}_e) = \text{Dirichlet}(\alpha_1, \alpha_2 + \alpha_3 - 1, \alpha_4 + \alpha_5 + \alpha_6 - 2)$, which has two sets of collapsed categories, and we lose one degree of freedom for the first, and lose two degrees for the second collapsed category.

The marginal probability has two corrections. First, a uniform probability is stipulated for the collapsed categories, that is, $1/j$ if j categories are collapsed. Its concentration parameter is equal to the sum of the collapsed categories minus the change in degrees of freedom. Second, a multivariate beta function is introduced that incorporates the corrected concentration parameters. For the mixed hypothesis

$$\mathcal{H}_m : \theta_1 < \theta_2 = \theta_3 < \theta_4 = \theta_5 = \theta_6,$$

we readily find the following marginal probability

$$\frac{B(\alpha_1, \alpha_2 + \alpha_3 - 1, \alpha_4 + \alpha_5 + \alpha_6 - 2)}{B(\alpha)} \left(\frac{1}{2}\right)^{\alpha_2 + \alpha_3 - 1} \left(\frac{1}{3}\right)^{\alpha_4 + \alpha_5 + \alpha_6 - 2},$$

and marginal Bayes factor,

$$BF_{e0} = \frac{B(\alpha)}{B(\alpha')} \left(\frac{1}{2}\right)^{x_2 + x_3} \left(\frac{1}{3}\right)^{x_4 + x_5 + x_6} \frac{B(\alpha'_1, \alpha'_2 + \alpha'_3 - 1, \alpha'_4 + \alpha'_5 + \alpha'_6 - 2)}{B(\alpha_1, \alpha_2 + \alpha_3 - 1, \alpha_4 + \alpha_5 + \alpha_6 - 2)}$$

where we have used $\alpha'_k = \alpha_k + x_k$. Note that this result has also been established for a specific case, albeit for a more general set of hypotheses, in Mulder et al. (2020). What the above analysis of the Bayes factor for the mixed hypotheses \mathcal{H}_m shows is that we are, in general, able to factor the hypotheses and associated likelihoods. This factorization is beneficial because it allows us to compute Bayes factors for parametric constraints with the methods described in the main text, even if these constraints are mixed. Intuitively, parameters that vary freely in both hypotheses do not affect the resulting Bayes factor, because the associated part of the marginal likelihood can be split off from both the mixed and encompassing hypotheses.

Empirical Application: Memory of Negative Life Events

In this section we investigate the precision and efficiency of the estimation methods when applied to a real data set published in Uhlenhuth et al. (1974). Specifically, we conduct a Bayesian reanalysis of Haberman’s sample to test whether the reported negative

life events decrease over time as a function of forgetting. We test this inequality-constrained hypothesis against the encompassing hypothesis without constraints:

$$\begin{aligned} \mathcal{H}_r : \theta_1 > \theta_2 > \dots > \theta_{18} \\ \mathcal{H}_e : \theta_1, \theta_2, \dots, \theta_{18}. \end{aligned}$$

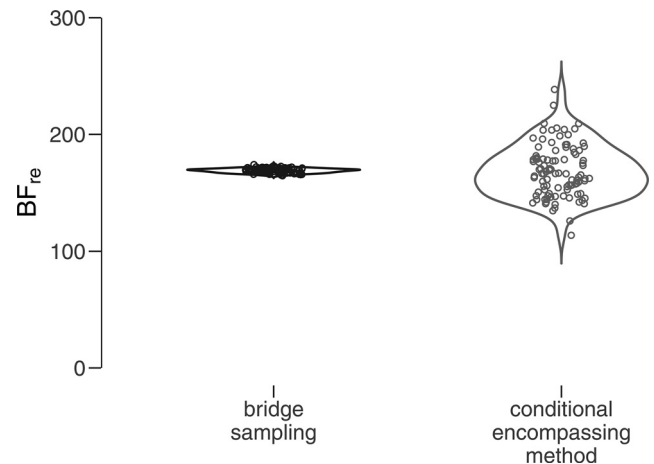
Method

We obtained the Bayes factor using the bridge sampling approach, the conditional encompassing method, and the unconditional encompassing method. To assess the precision and efficiency we computed Bayes factors in favor of \mathcal{H}_r 100 times for the same data set and for each method and recorded the respective values and the runtime to produce a result. We assigned a uniform prior distribution to our parameters of interest, such that we could compute the prior probability of the constraint, $p(\theta \in \mathcal{R} | \mathcal{H}_e)$, analytically. For the bridge sampling method, we drew 20,000 samples from the constrained posterior distribution. For the conditional encompassing method, the marginal probabilities of each constraint holding were estimated using 40,000 draws from the posterior distribution, resulting in a total of $40,000 \times 18$ draws. For the unconditional encompassing method, we drew 5 million samples from the unconstrained posterior distribution.

Results

The estimated Bayes factors BF_{re} are displayed in Figure 2. Bayes factors based on the bridge sampling method and the conditional encompassing method are centered around the same value ($M = 168.88$ and $M = 168.55$, respectively); however, the bridge sampling estimates varied far less ($SD = 1.873$) than the estimates produced by

Figure 2
Bayes Factors for the Bridge Sampling Method (Black), and the Conditional Encompassing Method (Dark Grey), for the Test of an Order-Restriction in Haberman’s (1978) Data on the Reporting of Negative Life Events



Note. Each dot represents one Bayes factor estimate in favor of \mathcal{H}_r obtained by the respective method. The bridge sampling method yields more precise Bayes factor estimates than the conditional encompassing method; the unconditional encompassing method fails to estimate any Bayes factor.

the conditional encompassing method ($SD = 22.23$). To understand the reasons for these differences in variability, we investigated the autocorrelation and the influence of chain length on the Bayes factor estimates, but could not identify a consistent pattern. We suspect that the variability stems from the Monte Carlo error that increases with each sequential evaluation of the individual constraint. If it were possible to estimate the conditional probabilities more efficiently, the variability in the estimates might reduce. Such an improved algorithm has been developed by Gu et al. (2019) for continuous variables but is not yet available for categorical data.

Regarding computational efficiency, the bridge sampling method had the lowest runtimes with a mean of $M = 29.11$ s ($SD = .39$). The conditional encompassing method on the other hand had mean runtimes of $M = 375.84$ s ($SD = 5.04$), which is more than 6 min to estimate one Bayes factor, compared with less than half a minute for the bridge sampling method. In sum, the empirical example demonstrates that the bridge sampling routine outperforms both the conditional encompassing method and the unconditional encompassing method. The bridge sampling estimates are considerably more precise than those of the conditional encompassing method, and are obtained more quickly. The unconditional encompassing method fails to estimate any Bayes factor altogether.

Discussion

In this article we describe a precise, scalable, and efficient bridge sampling routine to estimate Bayes factors for inequality constrained hypotheses on multinomial data. Bridge sampling is a promising alternative to current methods that sample from the unconstrained parameter space and hence may yield imprecise results and long runtimes.

The main reason why the bridge sampling method achieves relatively high precision—even for a model with many categories—is that it does not sample from the unconstrained or increasingly restricted parameter space. Instead, bridge sampling combines the draws from the restricted target distribution with samples from a proposal distribution to estimate the marginal likelihood efficiently. As a result, the precision of the bridge sampling estimate does not depend on the prior probability of the constraint, but rather depends on the similarity between the proposal distribution and the target distribution. Meng and Schilling (2002, p. 584) note that by using more sophisticated methods (e.g., by using warp bridge sampling) to create more overlap between the proposal distribution and the target distribution “[...] we can achieve better and better estimation efficiency based on the same set of draws, and it seems there is no lower bound on the Monte Carlo error.” To achieve sufficient overlap between the two distributions, we applied random variable transformation and used the method of moments to construct a suitable proposal distribution.

Compared with existing methods, the bridge sampling routine requires more effort to implement. As with the conditional encompassing method, researchers who wish to use bridge sampling to evaluate inequality constrained hypotheses need to implement a Gibbs sampling algorithm to draw samples from the constrained prior and posterior distribution. In addition, functions must be implemented to perform the required variable transformations and to apply the bridge sampling algorithm. In order to maximize the accessibility of the proposed method, we developed the R package

multibridge which can be downloaded from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=multibridge>. In the near future we also plan to make the analysis available in the user-friendly statistical software program JASP (JASP Team, 2020). Using JASP does not require any programming experience whatsoever.

At this point we would like to address some discussion points that repeatedly arise in the context of estimating Bayes factors for order restrictions. First, we chose to compare the restricted hypothesis with the encompassing hypothesis in which all parameters are free to vary. This comparison was central in the early work on the encompassing prior approach (e.g., Klugkist et al., 2005). However, this comparison can be critiqued because the encompassing hypothesis overlaps with the restrictive hypothesis (Morey & Rouder, 2011; but see Lee et al., 2019). An alternative comparison pits the restricted hypothesis against its complement (i.e., the part of the encompassing hypothesis that excludes the restricted hypothesis; e.g., Gu et al., 2019; Heck & Davis-Stober, 2019; Hoijtink, 2011; Mulder et al., 2012). Although our implementation is designed to compare against the encompassing hypothesis by default, researchers can easily obtain the comparison to the complement by exploiting the fact that the Bayes factor is transitive (i.e., dividing BF_{re} by $BF_{\neg re}$).

Second, our method involves a Dirichlet prior distribution, which allows researchers to specify values for the concentration parameters. Alternative Bayesian approaches are fully automatic in the sense that the prior distribution is determined by (part of) the sample data (e.g., fractional Bayes factors and adjusted fractional Bayes factors; for details see Böing-Messing & Mulder, 2016; Gu et al., 2019; Mulder, 2014; Mulder et al., 2018; Mulder et al., 2010). These alternative approaches have not yet been applied to multinomial models.

The method proposed here is relatively general and may be extended to problems of higher dimension and increasing sophistication. For instance, the bridge sampling framework could be expanded to multinomial models with complex linear restrictions (e.g., Heck & Davis-Stober, 2019). This would allow researchers to test more complex hypotheses, such as ordinal expectations on the size ratio of the parameters of interest (e.g., $\mathcal{H}_r : \theta_1 > 3 \times \theta_2$), on the differences between category proportions (e.g., $\mathcal{H}_r : (\theta_1 - \theta_2) < (\theta_3 - \theta_4)$), or on odds ratios for data that are summarized in contingency tables (e.g., $\mathcal{H}_r : \frac{\theta_1}{(\theta_1 + \theta_2)} < \frac{\theta_3}{(\theta_3 + \theta_4)}$). Another generalization of the presented methods concerns the application to hierarchical models for cases where participants repeatedly choose a response option and therefore category proportions are nested within participants. The bridge sampling routine may also benefit the field of psychometrics, for instance, when evaluating ordinal rating scales (Schnuerch et al., 2021), or ordinal item-response theory models (possible applications include Haaf et al., 2020; Karabatsos, 2001; Karabatsos & Sheu, 2004; Myung et al., 2005; Tijmstra et al., 2015; Tijmstra & Bolsinova, 2019; see also Meng & Schilling, 1996 for an early use of bridge sampling in the context of IRT modeling). Finally, in the field of cognitive psychology, the bridge sampling routine could be used to test ordinal hypotheses in multinomial processing tree models (e.g., Batchelder & Riefer, 1999; Erdfelder et al., 2009, 2020; Kuhlmann et al., 2019) and in discrete choice models (Johnson, 2007).

Here we used bridge sampling to evaluate order constraints for the multinomial model, which is associated with categorical data. However, bridge sampling might also benefit the evaluation of order constraints in models associated with continuous data, for instance regression models or analysis of variance (ANOVA) models. The key to using bridge sampling for these applications is to bring the integral over the restricted parameter space into a form that works with standard bridge sampling. For the current application to multinomial data, we used a series of transformations to ensure that (a) the support of one parameter does not depend on other parameters, and (b) that the support of the parameters are in a range that matches the one of the multivariate normal proposal distribution. If one can accomplish these goals for any particular problem, then standard bridge sampling immediately applies. As this strategy can be applied to a wide range of problems, we expect bridge sampling to be generalized to other models in the future.

Our results demonstrate that bridge sampling offers considerable improvements in precision and efficiency over existing methods. As our empirical application showed, for multinomial models it is common to have a relatively high number of categories (i.e., $K > 10$) which can easily lead to extreme values of the Bayes factors, if the data either speak for or against the restriction. In other disciplines, such as microbiology, we even find multinomial models with up to $K = 46$ categories, as a study of the relationship between gut microbiome and body mass index showed (Song et al., 2020). In these scenarios we believe that the benefit of the bridge sampling routine is particularly apparent. To conclude, the bridge sampling routine of estimating Bayes factors for inequality constraints in multinomial models constitutes a promising tool to evaluate ordinal expectations reliably and efficiently.

References

- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6(1), 57–86. <https://doi.org/10.3758/BF03210812>
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2), 245–268. [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4)
- Berger, J. O., & Molina, G. (2005). Posterior model probabilities via path-based pairwise priors. *Statistica Neerlandica*, 59(1), 3–15. <https://doi.org/10.1111/j.1467-9574.2005.00275.x>
- Böing-Messing, F., & Mulder, J. (2016). Automatic Bayes factors for testing variances of two independent normal distributions. *Journal of Mathematical Psychology*, 72, 158–170. <https://doi.org/10.1016/j.jmp.2015.08.001>
- Cavagnaro, D. R., & Davis-Stober, C. P. (2014). Transitive in our preferences, but transitive in different ways: An analysis of choice variability. *Decision*, 1(2), 102–122. <https://doi.org/10.1037/dec0000011>
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432), 1313–1321. <https://doi.org/10.1080/01621459.1995.10476635>
- Damien, P., & Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2), 206–215. <https://doi.org/10.1198/10618600152627906>
- Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, 53(1), 1–13. <https://doi.org/10.1016/j.jmp.2008.08.003>
- Devroye, L. (1986). Sample-based non-uniform random variate generation. In J. R. Wilson, J. O. Henriksen, & S. D. Roberts (Eds.), *Proceedings of the 18th Conference on Winter Simulation* (pp. 260–265). ACM.
- Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1), 204–223. <https://doi.org/10.1214/aoms/1177693507>
- Dickey, J. M., & Lientz, B. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, 41(1), 214–226.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Abfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift Für Psychologie/Journal of Psychology*, 217(3), 108–124.
- Erdfelder, E., Hu, X., Rouder, J. N., & Wagenmakers, E.-J. (2020). Cognitive psychometrics: The scientific legacy of William H. Batchelder (1940–2018). *Journal of Mathematical Psychology*, 99, 102468.
- Frigyik, B. A., Kapila, A., & Gupta, M. R. (2010). *Introduction to the Dirichlet distribution and related processes* (Tech. Rep.). Department of Electrical Engineering, University of Washington.
- Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7, 143–167.
- Gelfand, A. E., Smith, A. F., & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87(418), 523–532. <https://doi.org/10.1080/01621459.1992.10475235>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1), 75–90. <https://doi.org/10.1037/0033-295X.109.1.75>
- Good, I. J. (1967). A Bayesian significance test for multinomial distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(3), 399–431. <https://doi.org/10.1111/j.2517-6161.1967.tb00705.x>
- Grayson, D. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika*, 53(3), 383–392. <https://doi.org/10.1007/BF02294219>
- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81, 80–97. <https://doi.org/10.1016/j.jmp.2017.09.005>
- Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: a program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation*, 89(8), 1526–1553. <https://doi.org/10.1080/00949655.2019.1590574>
- Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, 19(4), 511–527. <https://doi.org/10.1037/met0000017>
- Guo, Y., & Regenwetter, M. (2014). Quantitative tests of the perceived relative argument model: Comment on Loomes (2010). *Psychological Review*, 121(4), 696–705. <https://doi.org/10.1037/a0036095>
- Haaf, J. M., Merkle, E. C., & Rouder, J. N. (2020). Do items order? the psychology in IRT models. *Journal of Mathematical Psychology*, 98, 102398. <https://doi.org/10.1016/j.jmp.2020.102398>
- Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in Bayesian mixed models. *Psychological Methods*, 22(4), 779–798. <https://doi.org/10.1037/met0000156>
- Haaf, J. M., & Rouder, J. N. (2020). *Does every study? Implementing ordinal constraint in meta-analysis* [Manuscript submitted for publication]. <https://psyarxiv.com/hf9se/>
- Haberman, S. J. (1978). *Analysis of qualitative data: Introductory topics* (Vol. 1). Academic Press.
- Heck, D. W., & Davis-Stober, C. P. (2019). Multinomial models with linear inequality constraints: Overview and improvements of computational methods for Bayesian inference. *Journal of Mathematical Psychology*, 91, 70–87. <https://doi.org/10.1016/j.jmp.2019.03.004>
- Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Chapman & Hall/CRC.

- Hojtink, H., Klugkist, I., & Boelen, P. (Eds.). (2008). *Bayesian Evaluation of Informative Hypotheses*. Springer-Verlag.
- Iverson, G. J. (2006). An essay on inequalities and order-restricted inference. *Journal of Mathematical Psychology*, 50(3), 215–219. <https://doi.org/10.1016/j.jmp.2006.01.007>
- JASP Team. (2020). JASP (Version 0.13.1.0) [Computer software]. <https://jasp-stats.org/>
- Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31, 203–222. <https://doi.org/10.1017/S030500410001330X>
- Johnson, T. R. (2007). Discrete choice models for ordinal response variables: A generalization of the stereotype model. *Psychometrika*, 72(4), 489–504. <https://doi.org/10.1007/s11336-007-9020-5>
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, 2(4), 389–423.
- Karabatsos, G., & Sheu, C.-F. (2004). Order-constrained Bayes inference for dichotomous models of unidimensional nonparametric IRT. *Applied Psychological Measurement*, 28(2), 110–125. <https://doi.org/10.1177/0146621603260678>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Klugkist, I. (2008). Encompassing prior based model selection for inequality constrained analysis of variance. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 53–83). Springer-Verlag.
- Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, 59(1), 57–69. <https://doi.org/10.1111/j.1467-9574.2005.00279.x>
- Klugkist, I., Laudy, O., & Hoijtink, H. (2010). Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables. *Psychological Methods*, 15(3), 281–299. <https://doi.org/10.1037/a0020137>
- Kuhlmann, B. G., Erdfelder, E., & Moshagen, M. (2019). Testing interactions in multinomial processing tree models. *Frontiers in Psychology*, 10, 2364. <https://doi.org/10.3389/fpsyg.2019.02364>
- Laudy, O. (2006). *Bayesian inequality constrained models for categorical data* [Unpublished doctoral dissertation]. Utrecht University.
- Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, 2(3-4), 141–153. <https://doi.org/10.1007/s42113-019-00029-y>
- Meng, X.-L., & Schilling, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91(435), 1254–1267. <https://doi.org/10.1080/01621459.1996.10476995>
- Meng, X.-L., & Schilling, S. (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics*, 11(3), 552–586. <https://doi.org/10.1198/106186002457>
- Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6(4), 831–860.
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, 16(4), 406–419. <https://doi.org/10.1037/a0024377>
- Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses. *Computational Statistics & Data Analysis*, 71, 448–463. <https://doi.org/10.1016/j.csda.2013.07.017>
- Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, 72, 104–115. <https://doi.org/10.1016/j.jmp.2014.09.004>
- Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., & Hoijtink, H. (2018). *Bfpack: Flexible Bayes factor testing of scientific theories in R*. arXiv. <https://arxiv.org/abs/1911.07728>
- Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140(4), 887–906. <https://doi.org/10.1016/j.jspi.2009.09.022>
- Mulder, J., Hoijtink, H., & Leeuw, C. D. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, 46(2), 1–39. <https://doi.org/10.18637/jss.v046.i02>
- Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, 53(6), 530–546. <https://doi.org/10.1016/j.jmp.2009.09.003>
- Mulder, J., Wagenmakers, E.-J., & Marsman, M. (2020). A generalization of the Savage-Dickey density ratio for testing equality and order constrained hypotheses. *The American Statistician*. Advance online publication. <https://doi.org/10.1080/00031305.2020.1799861>
- Myung, J., Karabatsos, G., & Iverson, G. (2008). A statistician's view on Bayesian evaluation of informative hypotheses. In H. Hoijtink, I. Klugkist, & P. Boelen (Eds.), *Bayesian Evaluation of Informative Hypotheses* (pp. 310–329). Springer-Verlag.
- Myung, J., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, 49(3), 205–225. <https://doi.org/10.1016/j.jmp.2005.02.004>
- Nigrini, M. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). Wiley.
- Nuijten, M. B., Hartgerink, C. H., Assen, M. A. V., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48(4), 1205–1226. <https://doi.org/10.3758/s13428-015-0664-2>
- O'Hagan, A., & Forster, J. (2004). *Kendall's Advanced Theory of Statistics Vol. 2B: Bayesian Inference* (2nd ed.). Arnold.
- Overstall, A. M., & Forster, J. J. (2010). Default Bayesian model determination methods for generalised linear mixed models. *Computational Statistics & Data Analysis*, 54(12), 3269–3288. <https://doi.org/10.1016/j.csda.2010.03.008>
- Pericchi Guerra, L., Liu, G., & Torres, D. (2008). Objective Bayes factors for informative hypotheses: “completing” the informative hypothesis and “splitting” the Bayes factors. In H. Hoijtink, I. Klugkist, & P. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 131–154). Springer-Verlag.
- Rauch, B., Götsche, M., Brähler, G., & Engel, S. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, 12(3), 243–255. <https://doi.org/10.1111/j.1468-0475.2011.00542.x>
- Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision*, 5(2), 63–94. <https://doi.org/10.1037/dec0000069>
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118(1), 42–56. <https://doi.org/10.1037/a0021150>
- Schnuerch, M., Nadarevic, L., & Rouder, J. N. (2020). The truth revisited: Bayesian analysis of individual differences in the truth effect. *Psychonomic Bulletin & Review*, 28, 750–765.
- Schnuerch, M., Haaf, J. M., Sarafoglou, A., & Rouder, J. (2021). *Meaningful comparisons with ordinal-scale items* [Manuscript submitted for publication]. <https://psyarxiv.com/nkydng>
- Sedransk, J., Monahan, J., & Chiu, H. (1985). Bayesian estimation of finite population parameters in categorical data models incorporating order restrictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3), 519–527. <https://doi.org/10.1111/j.2517-6161.1985.tb01382.x>

- Song, Y., Zhao, H., & Wang, T. (2020). An adaptive independence test for microbiome community data. *Biometrics*, 76(2), 414–426. <https://doi.org/10.1111/biom.13154>
- Stan Development Team. (2020). Stan modeling language user's guide and reference manual, version 2.23.0 [Computer software manual]. <http://mc-stan.org/>
- Tijmstra, J., & Bolsinova, M. (2019). Bayes factors for evaluating latent monotonicity in polytomous item response theory models. *Psychometrika*, 84(3), 846–869. <https://doi.org/10.1007/s11336-019-09661-w>
- Tijmstra, J., Hoijtink, H., & Sijtsma, K. (2015). Evaluating manifest monotonicity using Bayes factors. *Psychometrika*, 80(4), 880–896. <https://doi.org/10.1007/s11336-015-9475-8>
- Uhlenhuth, E. H., Lipman, R. S., Balter, M. B., & Stern, M. (1974). Symptom intensity and life stress in the city. *Archives of General Psychiatry*, 31(6), 759–764. <https://doi.org/10.1001/archpsyc.1974.01760180005001>
- van der Lans, R., Cremers, J., Klugkist, I., & Zwart, R. (2020). Teachers' interpersonal relationships and instructional expertise: How are they related? *Studies in Educational Evaluation*, 66, 100902. <https://doi.org/10.1016/j.stueduc.2020.100902>
- Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLoS ONE*, 9(12), e114876. <https://doi.org/10.1371/journal.pone.0114876>
- Verdinelli, I., & Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430), 614–618. <https://doi.org/10.1080/01621459.1995.10476554>
- Wetzels, R., Grasman, R. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio. *Computational Statistics & Data Analysis*, 54(9), 2094–2102. <https://doi.org/10.1016/j.csda.2010.03.016>

Appendix

Simulation Study: Accuracy of Estimation Methods

To illustrate the accuracy of the estimation methods we conducted two simulation studies. The first simulation study features eight different data sets, given in Table A1, for which it is possible to obtain the exact normalizing constants of the restricted prior and posterior distributions (and hence the Bayes factors). These data sets were relatively small with five and six categories. The normalizing constant of the restricted prior distribution was readily available as we assigned a uniform Dirichlet prior on the model parameters. The exact computations for the restricted posterior distributions exploited the fact that integrating an order restriction expressed in the stick-breaking parameterization amounts to integrating a polynomial whenever the Dirichlet parameters are integers (for details see the online appendix). The

exact Bayes factors were then compared to the estimated Bayes factors from the bridge sampling method, the conditional encompassing method, and the unconditional encompassing method.

The second simulation study features five different data sets for which the exact normalizing constants could not be obtained. These data sets featured 18 categories. We compared the variability of the Bayes factor estimates from the bridge sampling method with those from the conditional encompassing method. The comparison did not include the unconditional encompassing method because the prior probability of a sample obeying the restriction is minuscule. As a result, the Bayes factors from this method are liable to be staggeringly overestimated as outlined in the main text.

Models With a Small Number of Categories

Method

The eight data sets and exact results are summarized in Table A1. To quantify accuracy, we estimated the Bayes factors 100 times using the bridge sampling method, the conditional encompassing method, and the unconditional encompassing method. For all data sets, we estimated the Bayes factor in favor of the inequality-constrained hypothesis \mathcal{H}_r that the probabilities of each category are increasing against the encompassing hypothesis \mathcal{H}_e that allows all probabilities to vary freely:

$$\begin{aligned}\mathcal{H}_r &: \theta_1 < \theta_2 < \dots < \theta_K \\ \mathcal{H}_e &: \theta_1, \theta_2, \dots, \theta_K.\end{aligned}$$

For the bridge sampling method, we drew 20,000 samples from the constrained posterior distribution. For the conditional encompassing method the marginal probabilities of each constraint holding were estimated using 40,000 draws from the

Table A1

Data Sets, Exact Normalizing Constant of the Restricted Posterior Distribution, and Corresponding Bayes Factors (Rounded to Two Decimals) in Favor of and Against the Inequality-Constrained Hypotheses That the Parameters Are Increasing

Observations	$p(\boldsymbol{\theta} \in \mathcal{R}_r \mathbf{x}, \mathcal{H}_e)$	BF_{er}	BF_{re}
$\mathbf{x}_1 = (0, 0, 0, 0, 0)'$	0.0083	1	1
$\mathbf{x}_2 = (18, 15, 12, 9, 6, 3)'$	3.07023×10^{-9}	452,373	2.21×10^{-6}
$\mathbf{x}_3 = (3, 6, 8, 7, 7)'$	0.0347872	0.24	4.17
$\mathbf{x}_4 = (3, 6, 9, 6, 3)'$	0.00196566	4.24	0.24
$\mathbf{x}_5 = (3, 6, 9, 12, 10)'$	0.0937483	0.089	11.25
$\mathbf{x}_6 = (3, 6, 9, 8, 2)'$	0.000707877	11.78	0.085
$\mathbf{x}_7 = (3, 6, 9, 12, 15)'$	0.255149	0.033	30.62
$\mathbf{x}_8 = (3, 6, 9, 12, 15, 18)'$	0.149099	0.01	107.35

Note. The exact normalizing constant of the restricted prior distribution is .008333 for \mathbf{x}_1 and \mathbf{x}_3 – \mathbf{x}_7 , and .001389 for \mathbf{x}_2 and \mathbf{x}_8 .

(Appendix continues)

posterior distribution, resulting in a total of 200,000 draws for \mathbf{x}_1 and \mathbf{x}_3 – \mathbf{x}_7 , and 240,000 draws for \mathbf{x}_2 and \mathbf{x}_8 . For the unconditional encompassing method, we drew 5 million samples from the unconstrained posterior distribution.

Results

Figures A1–A3 show violin plots that display the Bayes factors for the three estimation methods for the eight data sets. Two results stand out in this simulation: First, in general, all estimation methods approximate the exact Bayes factor, with the conditional encompassing method showing the highest variability. Second, the advantage of bridge sampling becomes most evident for data sets that show evidence for the encompassing hypothesis. Especially for data set \mathbf{x}_2 , which provides extreme evidence against the inequality-constrained hypothesis, bridge sampling is able to accurately estimate the exact Bayes factor, whereas the conditional encompassing method yields highly variable results and the encompassing prior method fails to estimate any realistic Bayes factor at all: None of the posterior draws were consistent with the restrictive hypothesis, yielding a Bayes factor of 0 for all 100 estimates (Panel b in Figure A1). For \mathbf{x}_6 , the variability of the Bayes factors in the conditional encompassing method might even lead to different statistical decisions: The range of the Bayes factors is between eight (which is considered moderate evidence) to 18 (which is considered strong evidence, see Panel d in Figure A2).

Models With a Higher Number of Categories

Method

To further understand how the behavior of the bridge sampling method and the conditional encompassing method differ with increasing model size, we estimated Bayes factors for a model with $K = 18$ categories as applied to five additional data sets:

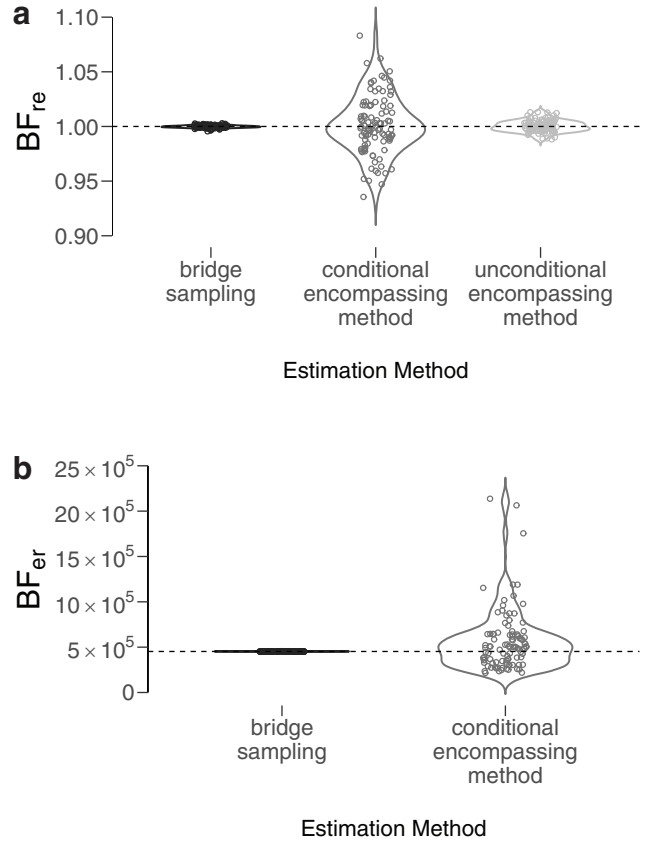
- $\mathbf{x}_9 : (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$
- $\mathbf{x}_{10} : (1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 8, 7, 6, 5, 4, 3, 5, 6)'$
- $\mathbf{x}_{11} : (1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 8, 7, 6, 5, 4, 4, 4, 3)'$
- $\mathbf{x}_{12} : (1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 8, 7, 6, 5, 4, 3, 6, 7)'$
- $\mathbf{x}_{13} : (1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 8, 7, 6, 5, 3, 2, 4, 4)'$

As in the first simulation study we drew 20,000 samples from the constrained posterior distribution for the bridge sampling method. For the conditional encompassing method the marginal probabilities of each constraint holding were estimated using 40,000 draws from the posterior distribution, resulting in a total of 720,000 draws.

Results

For each of the five data sets, Table A2 shows the means and range of the Bayes factors for the two estimation methods. The overall results are consistent with those from the first simulation study. Specifically, the bridge sampling method provides

Figure A1
Violin Plots Display the Estimated Bayes Factors for the Bridge Sampling Method (Black), the Conditional Encompassing Method (Dark Grey), and the Unconditional Encompassing Method (Light Grey) for Data Sets \mathbf{x}_1 and \mathbf{x}_2



Note. The dashed horizontal line indicates the exact Bayes factor that the y-axis always shows the Bayes factor in favor of the preferred hypothesis. (a) Distribution of Bayes factors for each estimation method for \mathbf{x}_1 for which the exact Bayes factor BF_{re} is 1. (b) Distribution of Bayes factors for each estimation method for \mathbf{x}_2 for which the exact Bayes factor BF_{er} is 452, 373. For these data, the unconditional encompassing method does not succeed in estimating any Bayes factor.

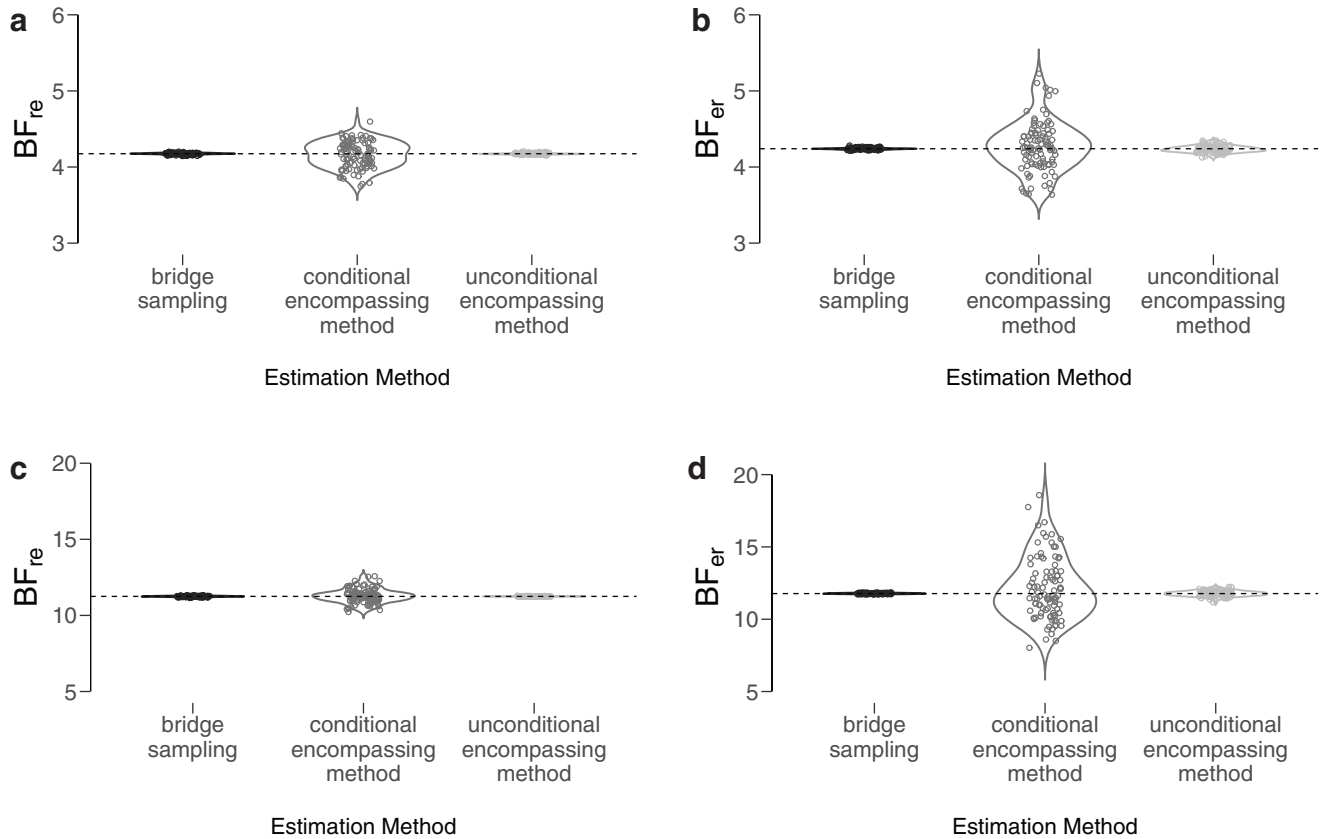
estimates that are less variable than those of the conditional encompassing method. Here again, for the data sets that provide evidence in favor of the encompassing hypothesis (i.e., \mathbf{x}_{11} and \mathbf{x}_{13}), the variability of the Bayes factors in the conditional encompassing method might lead to different statistical decisions: the Bayes factors for \mathbf{x}_{11} range between 1.8 (which is considered anecdotal evidence) to 5.8 (which is considered moderate evidence) and the Bayes factors for \mathbf{x}_{13} range between 7.7 (which is considered moderate evidence) to 38 (which is considered strong evidence). For comparison, with bridge sampling the estimated Bayes factor for \mathbf{x}_{13} ranged

(Appendix continues)

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure A2

Violin Plots Display the Estimated Bayes Factors for the Bridge Sampling Method (Black), the Conditional Encompassing Method (Dark Grey), and the Unconditional Encompassing Method (Light Grey) for Data Sets \mathbf{x}_3 – \mathbf{x}_6



Note. The dashed horizontal line indicates the exact Bayes factor. The Bayes factor estimates of the conditional encompassing method and the encompassing prior method are less variable if the data provides evidence for the restricted hypothesis (Panels a and c) than if the data provides evidence for the encompassing hypothesis (Panels b and d). (a) Distribution of Bayes factors for each estimation method for \mathbf{x}_3 for which the exact Bayes factor BF_{re} is 4.17. (b) Distribution of Bayes factors for each estimation method for \mathbf{x}_4 for which the exact Bayes factor BF_{er} is 4.24. (c) Distribution of Bayes factors for each estimation method for \mathbf{x}_5 for which the exact Bayes factor BF_{re} is 11.25. (d) Distribution of Bayes factors for each estimation method for \mathbf{x}_6 for which the exact Bayes factor BF_{er} is 11.78.

between 14.9 and 15.5. In general, for both methods the variability in the estimates increases with the strength of the evidence, either in favor or against the inequality-constrained hypothesis.

Conclusion

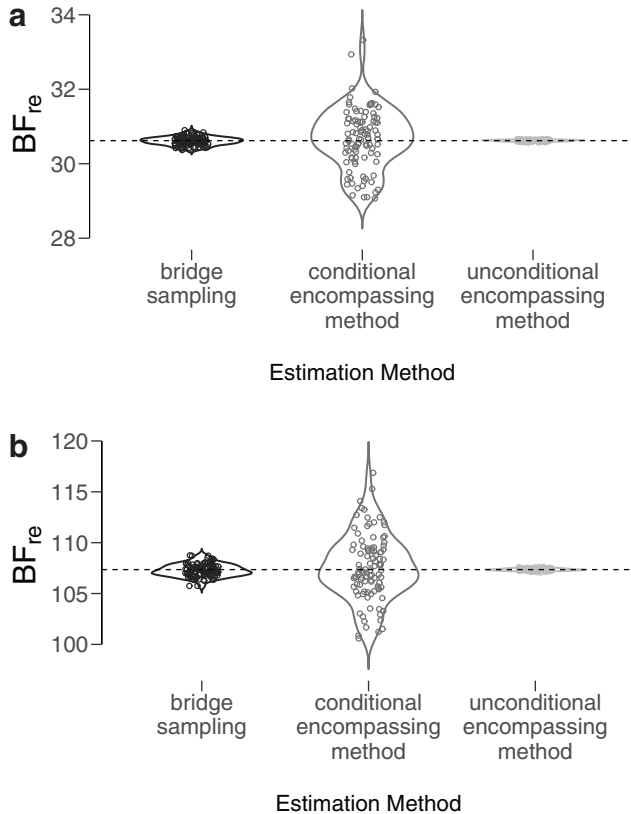
In two simulation studies we assessed the accuracy of the bridge sampling method, the conditional encompassing method, and the unconditional encompassing method. In the first simulation study, we obtained the exact Bayes factor for eight data sets and then estimated the Bayes factor using the

three estimation methods. In the second simulation study, we specified a larger model and estimated Bayes factors for five additional data sets using bridge sampling and the conditional encompassing method. The first simulation shows that the unconditional encompassing method is the most accurate for four of the eight data sets. However, the performance of this method depends heavily on the size of the model: For a model with 18 categories, the unconditional encompassing method could not be applied anymore. The method also quickly deteriorates when the data show evidence against the inequality-constrained hypothesis. In contrast, the conditional encompassing method is more responsive to fluctuations in evidence. To

(Appendix continues)

Figure A3

Violin Plots Display the Estimated Bayes Factors for the Bridge Sampling Method (Black), the Conditional Encompassing Method (Dark Grey), and the Unconditional Encompassing Method (Light Grey) for Data Sets \mathbf{x}_7 and \mathbf{x}_8 .



Note. The dashed horizontal line indicates the exact Bayes factor. (a) Distribution of Bayes factors for each estimation method for \mathbf{x}_7 for which the exact Bayes factor BF_{re} is 30.62. (b) Distribution of Bayes factors for each estimation method for \mathbf{x}_8 for which the exact Bayes factor BF_{re} is 107.35.

improve the accuracy of the method one could increase the number of samples. However, one should take into account that an increase in the number of samples comes at the expense of runtime, which is already many times higher than that of the other two methods.

Based on the recommendation of one of the reviewers we used the bridge sampling method to compute Bayes factors for models with 30 and 50 categories in order to assess runtime and check whether accuracy decreases with increasing dimensionality. Even in these relatively extreme scenarios the variability of the bridge sampling estimates remained relatively low: In the 30 category scenario the mean Bayes factor was $M = 11.01$ and ranged between 10.55 and 11.45 whereas in the

Table A2

Data Sets, Bayes Factor Types, and Mean Bayes Factors for Models With 18 Categories

Data set	Bayes factor	Bridge sampling	Conditional encompassing method
\mathbf{x}_9	BF_{re}	$M = 1.00$ [0.98, 1.01]	$M = 0.99$ [0.77, 1.32]
\mathbf{x}_{10}	BF_{re}	$M = 3.63$ [3.55, 3.71]	$M = 3.75$ [2.00, 7.40]
\mathbf{x}_{11}	BF_{er}	$M = 3.02$ [2.95, 3.08]	$M = 3.19$ [1.84, 5.82]
\mathbf{x}_{12}	BF_{re}	$M = 10.96$ [10.72, 11.20]	$M = 11.02$ [4.94, 17.25]
\mathbf{x}_{13}	BF_{er}	$M = 15.23$ [14.87, 15.55]	$M = 17.03$ [7.68, 38.33]

Note. Square brackets indicate the minimum and maximum Bayes factor estimate.

50 category scenario the mean Bayes factor was $M = 9.51$ and ranged between 8.71 and 10.24. Regarding computational efficiency, the model with 30 categories took an average of about 65 s to compute a Bayes factor, whereas the model with 50 categories took an average of about 1 min and 43 s to compute a Bayes factor. The increase in runtime between the 30 and the 50 category scenarios was largely due to the increase in time it takes to sample from the restricted distribution.

Overall, the bridge sampling routine shows a relatively good trade-off between accuracy and efficiency. The variability of the estimates remain in an acceptable range and bridge sampling outperforms the other methods especially when the data provides evidence against the inequality-constraint and when the models feature many categories. The reliability of bridge sampling, which was already on display in the empirical application, was again confirmed in these simulation studies. At this point, we would like to refer the interested reader to our online appendix for a more extended simulation study. This additional simulation study further describes under which conditions the unconditional encompassing method and sometimes even the conditional encompassing method fail to estimate a realistic Bayes factor.

Received August 19, 2020
Revision received April 2, 2021
Accepted April 21, 2021 ■