

<sup>1</sup> multibridge: An R Package To Evaluate Informed Hypotheses in Binomial and Multinomial  
<sup>2</sup> Models

<sup>3</sup> Alexandra Sarafoglou, Frederik Aust, Maarten Marsman, Eric-Jan Wagenmakers, & Julia M.  
<sup>4</sup> Haaf

<sup>5</sup> University of Amsterdam

<sup>6</sup> Correspondence concerning this article should be addressed to: Alexandra Sarafoglou,  
<sup>7</sup> Department of Psychology, PO Box 15906, 1001 NK Amsterdam, The Netherlands, E-mail:  
<sup>8</sup> alexandra.sarafoglou@gmail.com

9

## Abstract

10 The **multibridge** R package allows a Bayesian evaluation of informed hypotheses  $\mathcal{H}_r$   
11 applied to frequency data from an independent binomial or multinomial distribution.  
12 **multibridge** uses bridge sampling to efficiently compute Bayes factors for the following  
13 hypotheses concerning the latent category proportions  $\boldsymbol{\theta}$ : (a) hypotheses that postulate  
14 equality constraints (e.g.,  $\theta_1 = \theta_2 = \theta_3$ ); (b) hypotheses that postulate inequality constraints  
15 (e.g.,  $\theta_1 < \theta_2 < \theta_3$  or  $\theta_1 > \theta_2 > \theta_3$ ); (c) hypotheses that postulate mixtures of inequality  
16 constraints and equality constraints (e.g.,  $\theta_1 < \theta_2 = \theta_3$ ); and (d) hypotheses that postulate  
17 mixtures of (a)–(c) (e.g.,  $\theta_1 < (\theta_2 = \theta_3), \theta_4$ ). Any informed hypothesis  $\mathcal{H}_r$  may be compared  
18 against the encompassing hypothesis  $\mathcal{H}_e$  that all category proportions vary freely, or against  
19 the null hypothesis  $\mathcal{H}_0$  that all category proportions are equal. **multibridge** facilitates the  
20 fast and accurate comparison of large models with many constraints and models for which  
21 relatively little posterior mass falls in the restricted parameter space. This paper describes  
22 the underlying methodology and illustrates the use of **multibridge** through fully  
23 reproducible examples.

24 multibridge: An R Package To Evaluate Informed Hypotheses in Binomial and Multinomial  
25 Models

26 **Introduction**

27 The most common way to analyze categorical variables is to conduct either binomial tests,  
28 multinomial tests, or chi-square goodness of fit tests. These tests compare the encompassing  
29 hypothesis to a null hypothesis that all underlying category proportions are either exactly  
30 equal, or follow a specific distribution. Accordingly, these tests are suitable when theories  
31 predict either the invariance of all category proportions or specific values. For instance,  
32 chi-square goodness of fit tests are commonly used to test Benford's law, which predicts the  
33 distribution of leading digits in empirical datasets (Benford, 1938; Newcomb, 1881). Often,  
34 however, the predictions that researchers are interested in are of a different kind. Consider  
35 for instance the weak-order mixture model of decision-making (Regenwetter & Davis-Stober,  
36 2012). The theory predicts that individuals' choice preferences are weakly ordered at all  
37 times, that is, if they prefer choice *A* over *B* and *B* over *C* then they will also prefer *A* over  
38 *C* (Regenwetter, Dana, & Davis-Stober, 2011)—a well-constrained prediction of behavior.  
39 The theory is, however, silent about the exact values of each choice preference. Hence, the  
40 standard tests that compare  $\mathcal{H}_e$  to  $\mathcal{H}_0$  are unsuited to test the derived predictions. Instead,  
41 the predictions need to be translated into an informed hypothesis  $\mathcal{H}_r$  that reflects the  
42 predicted ordinal relations among the parameters. Only then is it possible to adequately test  
43 whether the theory of weakly-ordered preference describes participants' choice behavior. Of  
44 course, researchers may be interested in more complex hypotheses, including ones that  
45 feature combinations of equality constraints, inequality constraints, and unconstrained  
46 category proportions. For instance, Nuijten, Hartgerink, Assen, Epskamp, and Wicherts  
47 (2016) hypothesized that articles published in social psychology journals would have higher  
48 error rates than articles published in other psychology journals. As in the previous example,  
49 the authors had no expectations about the exact error rate distribution across journals.

50 Here, again, the standard tests are inadequate. Generally, by specifying informed hypotheses  
 51 researchers and practitioners are able to “add theoretical expectations to the traditional  
 52 alternative hypothesis” (Hoijtink, Klugkist, & Boelen, 2008, p. 2) and thus test hypotheses  
 53 that relate more closely to their theories (Haaf, Klaassen, & Rouder, 2019; Rijkeboer & van  
 54 den Hout, 2008).

In the Bayesian framework, researchers may test hypotheses of interest by means of Bayes factors (Jeffreys, 1935; Kass & Raftery, 1995). Bayes factors quantify the extent to which the data change the prior model odds to the posterior model odds, that is, the extent to which one hypothesis outpredicts the other. Specifically, Bayes factors are the ratio of marginal likelihoods of the respective hypotheses. For instance, the Bayes factor for the informed hypothesis versus the encompassing hypothesis is defined as:

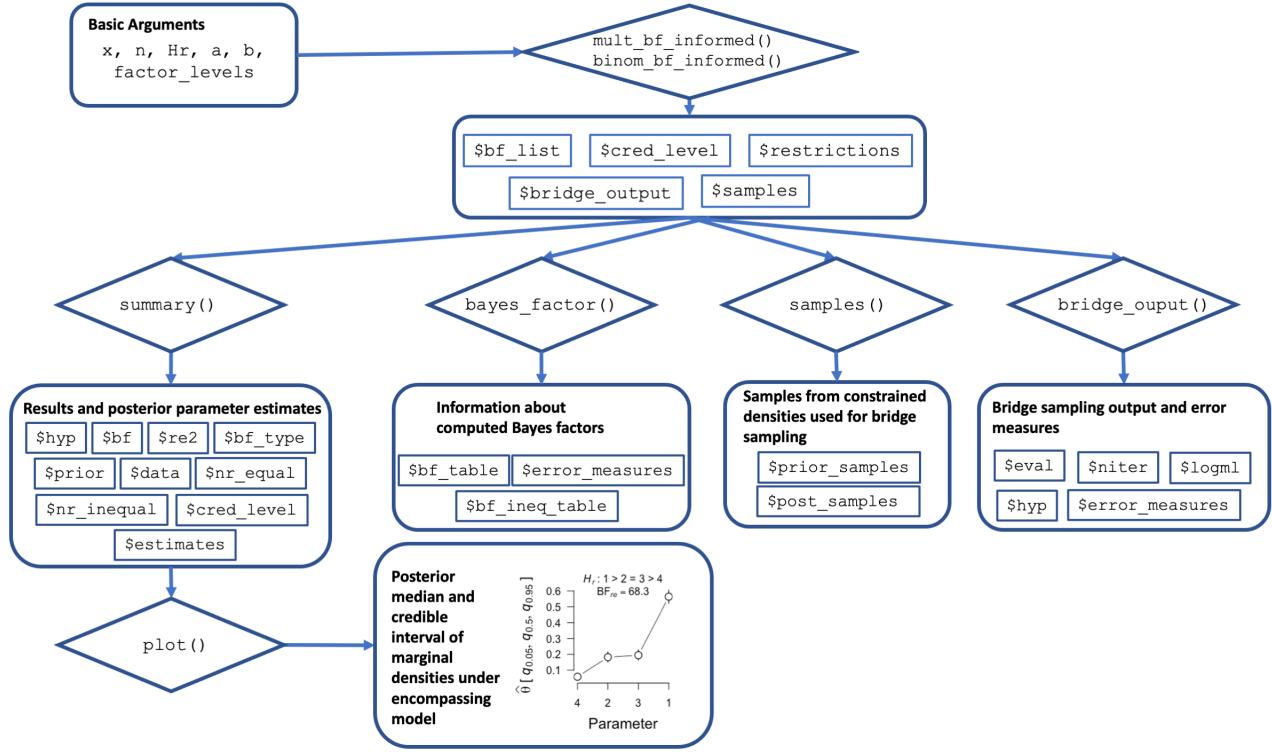
$$\text{BF}_{re} = \frac{\overbrace{p(\mathbf{x} | \mathcal{H}_r)}^{\text{Marginal likelihood under } \mathcal{H}_r}}{\underbrace{p(\mathbf{x} | \mathcal{H}_e)}_{\text{Marginal likelihood under } \mathcal{H}_e}},$$

55 where the subscript  $r$  denotes the informed hypothesis and  $e$  denotes the encompassing  
 56 hypothesis. Several available R packages compute Bayes factors for informed hypotheses. For  
 57 instance, the package **multinomineq** (Heck & Davis-Stober, 2019) evaluates informed  
 58 hypotheses for multinomial models as well as models that feature independent binomials.  
 59 The package **BFpack** (Mulder et al., in press) evaluates informed hypotheses for statistical  
 60 models such as univariate and multivariate normal linear models, generalized linear models,  
 61 special cases of linear mixed models, survival models, and relational event models. The  
 62 package **BAIN** (Gu, Hoijtink, Mulder, & Rosseel, 2019) evaluates informed hypotheses for  
 63 structural equation models. Outside of R, the Fortran 90 program **BIEMS** (Mulder,  
 64 Hoijtink, & de Leeuw, 2012) evaluates informed hypotheses for multivariate linear models  
 65 such as MANOVA, repeated measures, and multivariate regression. All these packages rely  
 66 on one of two implementations of the encompassing prior approach (Klugkist, Kato, &

67 Hoijsink, 2005; Sedransk, Monahan, & Chiu, 1985) to approximate order constrained Bayes  
68 factors: the unconditional encompassing method (Hoijsink, 2011; Hoijsink et al., 2008;  
69 Klugkist et al., 2005 ) and the conditional encompassing method (Gu, Mulder, Deković, &  
70 Hoijsink, 2014; Laudy, 2006; Mulder, 2014, 2016; Mulder et al., 2009). Even though the  
71 encompassing prior approach is currently the most common method to evaluate informed  
72 hypotheses, it becomes increasingly unreliable and inefficient as the number of restrictions  
73 increases or the parameter space of the restricted model decreases (Sarafoglou et al., in  
74 press).

75 As alternative to the encompassing prior approach, Sarafoglou et al. (in press) recently  
76 proposed a bridge sampling routine (Bennett, 1976; Meng & Wong, 1996) that computes  
77 Bayes factors for informed hypotheses more reliably and efficiently. This routine is  
78 implemented in **multibridge** (<https://CRAN.R-project.org/package=multibridge>) and is  
79 suitable to evaluate inequality constraints for multinomial and binomial models. When an  
80 informed hypothesis includes mixtures of equality and inequality constraints, the core  
81 functions in **multibridge** split the hypothesis to compute Bayes factors separately for  
82 equality constraints (for which the Bayes factor has an analytic solution) and inequality  
83 constraints (for which the Bayes factor is estimated using bridge sampling). The core  
84 functions of **multibridge**, that is `mult_bf_informed` and `binom_bf_informed`, return the  
85 Bayes factor estimate in favor of or against the informed hypothesis (see Table 2 for a  
86 summary of the basic required arguments of the two core functions). In addition, users can  
87 visualize the posterior parameter estimates under the encompassing hypothesis using the  
88 `plot`-method, or get more detailed information on how the Bayes factor is composed using  
89 the `summary`-method. For hypotheses that include mixtures between equality and inequality  
90 constrained hypotheses the `bayes_factor` method separately returns the Bayes factor for  
91 the equality constraints and the conditional Bayes factor for the inequality constraints given  
92 the equality constraints. The informed hypothesis can be conveniently specified using a  
93 string or character vector. Furthermore, the transitivity property of Bayes factors can be

- 94 used to test two informed hypotheses against each other (see Example 1 for an illustration).  
 95 The general workflow of **multibridge** is illustrated in Figure 1. A list of all currently  
 96 available functions and data sets is given in Table 1.



*Figure 1.* The **multibridge** workflow. When calling `mult_bf_informed` or `binom_bf_informed`, the user specifies the data values (`x` and `n` for binomial models and `x` for multinomial models, respectively), the informed hypothesis (`Hr`), the  $\alpha$  and  $\beta$  parameters of the binomial prior distributions (`a` and `b`) or the concentration parameters for the Dirichlet prior distribution (`a`), respectively, and the category labels of the factor levels (`factor_levels`). The functions then return the estimated Bayes factor for the informed hypothesis relative to the encompassing or the null hypothesis. Based on these results different S3 methods can be used to get more detailed information on the individual components of the analysis (e.g., `summary`, `bayes_factor`), and parameter estimates of the encompassing distribution (`plot`).

Table 1

*Core functions available in **multibridge**.*

Function Name(s)	Description
<code>mult_bf_informed</code>	Evaluates informed hypotheses on multinomial parameters.
<code>mult_bf_inequality</code>	Estimates the marginal likelihood of a constrained prior or posterior Dirichlet distribution.
<code>mult_bf_equality</code>	Computes Bayes factor for equality constrained multinomial parameters using the standard Bayesian multinomial test.
<code>mult_tsampling</code>	Samples from constrained prior or posterior Dirichlet density.
<code>lifestresses, peas</code>	Data sets associated with informed hypotheses in multinomial models.
<code>binom_bf_informed</code>	Evaluates informed hypotheses on binomial parameters.
<code>binom_bf_inequality</code>	Estimates the marginal likelihood of constrained prior or posterior beta distributions.
<code>binom_bf_equality</code>	Computes Bayes factor for equality constrained binomial parameters.
<code>binom_tsampling</code>	Samples from constrained prior or posterior beta densities.
<code>journals</code>	Data set associated with informed hypotheses in binomial models.
<code>generate_restriction_list</code>	Encodes the informed hypothesis.

<sup>97</sup> This paper showcases how the proposed bridge sampling routine by Sarafoglou et al. (in  
<sup>98</sup> press) can be applied in a user-friendly way with **multibridge**. In the remainder of this  
<sup>99</sup> article, we will describe the Bayes factor identity for informed hypotheses in binomial and  
<sup>100</sup> multinomial models, and briefly describe the bridge sampling method. Then, we illustrate the

101 core functions of **multibridge** package using two examples and end with a brief summary.

102

## Methods

103 In this section we formalize multinomial models and models that feature independent  
104 binomial probabilities as they have been implemented in **multibridge**. In the multinomial  
105 model, we assume that the vector of observations  $\mathbf{x}$  in the  $K$  categories follows a multinomial  
106 distribution in which the parameters of interest,  $\boldsymbol{\theta}$ , represent the underlying category  
107 proportions. Since the  $K$  categories are dependent, the vector of probability parameters is  
108 constrained to sum to one, such that  $\sum_{k=1}^K (\theta_1, \dots, \theta_K) = 1$ . Therefore, a suitable choice for  
109 a prior distribution for  $\boldsymbol{\theta}$  is the Dirichlet distribution with concentration parameter vector  $\boldsymbol{\alpha}$ :

$$x_1, \dots, x_K \sim \text{Multinomial}\left(\sum_{k=1}^K x_k, \theta_1, \dots, \theta_K\right) \quad (1)$$

$$\theta_1, \dots, \theta_K \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K), \quad (2)$$

110 where  $\boldsymbol{\alpha}$  can be interpreted as vector of *a priori* category counts. The formalization of the  
111 model for independent binomial probabilities is similar since the multinomial model above  
112 constitutes a generalization of the binomial model (for  $K \geq 2$ ). In the binomial model, we  
113 assume that the elements in the vector of successes  $\mathbf{x}$  and the elements in the vector of total  
114 number of observations  $\mathbf{n}$  in the  $K$  categories follow independent binomial distributions. As  
115 in the multinomial model, the parameter vector of the binomial success probabilities  $\boldsymbol{\theta}$   
116 contains the underlying category proportions, however, in this model we assume that  
117 categories are independent which removes the sum-to-one constraint. Therefore, a suitable  
118 choice for a prior distribution for  $\boldsymbol{\theta}$  is a vector of independent beta distributions with  
119 parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ :

$$x_1 \cdots x_K \sim \prod_{k=1}^K \text{Binomial}(\theta_k, n_k) \quad (3)$$

$$\theta_1 \cdots \theta_K \sim \prod_{k=1}^K \text{Beta}(\alpha_k, \beta_k), \quad (4)$$

120 where  $\boldsymbol{\alpha}$  can be interpreted as vector of *a priori* successes that observations fall within the  
 121 various categories and  $\boldsymbol{\beta}$  can be interpreted as vector of *a priori* failures.

122 **Bayes factor**

123 **multibridge** features two different methods to compute Bayes factors: one method  
 124 computes Bayes factors for equality constrained parameters and one method computes Bayes  
 125 factors for inequality constrained parameters. Both methods will be outlined below. In cases  
 126 where informed hypotheses feature mixtures between inequality and equality constraints, we  
 127 compute the overall Bayes factor  $\text{BF}_{re}$  by multiplying the individual Bayes factors for both  
 128 constraint types. This is motivated by the fact that the Bayes factor for mixtures will factor  
 129 into a Bayes factor for the equality constraints and a conditional Bayes factor for the  
 130 inequality constraints given the equality constraints (see Sarafoglou et al., in press, for the  
 131 proof).

**The Bayes Factor For Equality Constraints.** In **multibridge** the Bayes factor for the equality constraints can be computed analytically both for binomial and multinomial models using the functions `binom_bf_equality` and `mult_bf_equality`. For binomial models, assuming that the all binomial probabilities in a model are exactly equal, the Bayes factor is defined as:

$$\text{BF}_{0e} = \frac{\prod_{k=1}^K \text{B}(\alpha_k, \beta_k)}{\prod_{k=1}^K \text{B}(\alpha_k + x_k, \beta_k + n_k - x_k)} \times \frac{\text{B}(\alpha_+ + x_+ + 1, \beta_+ + n_+ - x_+ + 1)}{\text{B}(\alpha_+ + 1, \beta_+ + 1)},$$

where  $\text{B}(\cdot)$  denotes the beta function and  $\alpha_+ = \sum_{k=1}^K \alpha_k$ ,  $\beta_+ = \sum_{k=1}^K \beta_k$ ,  $x_+ = \sum_{k=1}^K x_k$  and  $n_+ = \sum_{k=1}^K n_k$ . If all binomial probabilities in a model are assumed to be exactly equal *and*

equal to a predicted value  $\theta_0$ , the Bayes factor is defined as:

$$\text{BF}_{0e} = \frac{\prod_{k=1}^K \text{B}(\alpha_k, \beta_k)}{\prod_{k=1}^K \text{B}(\alpha_k + x_k, \beta_k + n_k - x_k)} \times \theta_0^{x_+} (1 - \theta_0)^{n_+ - x_+}.$$

132 Note that **multibridge** only supports the specification of one predicted value for all  
133 binomial probabilities. The package does not support the specification of different predicted  
134 values for different binomial probabilities. The reason for this is theoretical: we believe that  
135 such hypotheses are better tested using a hierarchical structure (thus modeling the binomial  
136 probabilities as dependent).

For multinomial models, assuming that all category proportions in a model are equality constrained, the Bayes factor  $\text{BF}_{0e}$  is defined as:

$$\text{BF}_{0e} = \frac{\text{B}(\alpha_1, \dots, \alpha_K)}{\text{B}(\alpha_1 + x_1, \dots, \alpha_K + x_K)} \times \frac{\text{B}(\boldsymbol{\alpha} + \mathbf{x})}{\text{B}(\boldsymbol{\alpha})} \times \prod_{k=1}^K \theta_{0k}^{x_k},$$

137 where  $\theta_{0k}$  represent the predicted category proportions. When all category proportions are  
138 assumed to be exactly equal all  $\theta_{0k}$  are set to  $\frac{1}{K}$ . Otherwise,  $\boldsymbol{\theta}_0$  is replaced with the  
139 user-specified predicted values.

140 **The Bayes Factor For Inequality Constraints.** To approximate the Bayes  
141 factor for informed hypotheses, Klugkist et al. (2005) derived an identity that defines the  
142 Bayes factor  $\text{BF}_{re}$  as the ratio of proportions of posterior and prior parameter space  
143 consistent with the restriction. This identity forms the basis of the encompassing prior  
144 approach. Recently, Sarafoglou et al. (in press) highlighted that these proportions can be  
145 reinterpreted as the marginal likelihoods (i.e., the normalizing constants) of the constrained  
146 posterior and constrained prior distribution:

$$\text{BF}_{re} = \frac{\overbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathbf{x}, \mathcal{H}_e)}^{\text{Marginal likelihood of constrained posterior distribution}}}{\overbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r | \mathcal{H}_e)}^{\text{Marginal likelihood of constrained prior distribution}}}. \quad (5)$$

<sup>147</sup> The benefit of reinterpreting the identity by Klugkist et al. (2005) is that we can estimate the  
<sup>148</sup> Bayes factor by utilizing numerical sampling methods such as bridge sampling. For that we  
<sup>149</sup> only need to be able to sample from the constrained densities. Crucially, when using bridge  
<sup>150</sup> sampling, it does not matter how small the constrained parameter space is in proportion to  
<sup>151</sup> the encompassing density. This gives the method a decisive advantage over the encompassing  
<sup>152</sup> prior approach in terms of accuracy and efficiency especially (1) when binomial and  
<sup>153</sup> multinomial models with moderate to high number of categories (i.e.,  $K > 10$ ) are evaluated  
<sup>154</sup> and (2) when relatively little posterior mass falls in the constrained parameter space.

<sup>155</sup> The bridge sampling algorithm implemented in **multibridge** estimates one marginal  
<sup>156</sup> likelihood at the time (cf., Gronau et al., 2017; Overstall & Forster, 2010). Specifically, we  
<sup>157</sup> separately estimate the marginal likelihood for the constrained prior distribution and the  
<sup>158</sup> marginal likelihood of the constrained posterior distribution. Here we describe how to  
<sup>159</sup> estimate the marginal likelihood for the constrained prior distribution; the steps presented  
<sup>160</sup> can then be applied accordingly to the posterior distribution. It should be noted that the  
<sup>161</sup> bridge sampling algorithm implemented in **multibridge** is an adapted version of the  
<sup>162</sup> algorithm implemented in the R package **bridgesampling** (Gronau, Singmann, &  
<sup>163</sup> Wagenmakers, 2020) and allows for the specification of informed hypotheses on probability  
<sup>164</sup> vectors.<sup>1</sup> The bridge sampling identity for the marginal likelihood of the constrained prior  
<sup>165</sup> distribution is defined as:

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

<sup>166</sup> where the term  $h(\boldsymbol{\theta})$  refers to the bridge function proposed by Meng and Wong (1996),  $g(\boldsymbol{\theta})$   
<sup>167</sup> refers to a so-called proposal distribution, and  $p(\boldsymbol{\theta} \mid \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r)$  is the part of the prior  
<sup>168</sup> parameter space under the encompassing hypothesis that is in accordance with the

---

<sup>1</sup>In addition, the function to compute the relative mean square error for bridge sampling estimates in **multibridge** is based on the code of the `error_measures`-function from the **bridgesampling** package.

constraint. To estimate the marginal likelihood, bridge sampling requires samples from the target distribution, that is, the constrained Dirichlet distribution for multinomial models and constrained beta distributions for binomial models, and samples from the proposal distribution which in principle can be any distribution with a known marginal likelihood; in **multibridge** the proposal distribution is the multivariate normal distribution. Samples from the target distribution are generated using the Gibbs sampling algorithms proposed by Damien and Walker (2001). For binomial models, we apply the suggested Gibbs sampling algorithm for constrained beta distributions. In the case of the multinomial models, we apply an algorithm that simulates values from constrained Gamma distributions which are then transformed into Dirichlet random variables. To sample efficiently from these distributions, **multibridge** provides a C++ implementation of this algorithm. Samples from the proposal distribution are generated using the standard `rmvnorm`-function from the R package **mvtnorm** (Genz et al., 2020).

The efficiency of the bridge sampling method is optimal only if the target and proposal distribution operate on the same parameter space and have sufficient overlap. We therefore probit transform the samples of the constrained distributions to move the samples from the probability space to the entire real line. Subsequently, we use half of these draws to construct the proposal distribution using the method of moments. Details on the probit transformations are provided in the appendix.

The numerator in Equation 6 evaluates the unnormalized density for the constrained prior distribution with samples from the proposal distribution. The denominator evaluates the normalized proposal distribution with samples from the constrained prior distribution. Using this identity, we obtain the bridge sampling estimator for the marginal likelihood of the constrained prior distribution by applying the iterative scheme proposed by Meng and Wong (1996):

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

<sup>194</sup> where  $N_1$  denotes the number of samples drawn from the constrained distribution, that is,  
<sup>195</sup>  $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta} \mid \mathcal{H}_r)$ ,  $N_2$  denotes the number of samples drawn from the proposal distribution, that  
<sup>196</sup> is  $\tilde{\boldsymbol{\theta}} \sim g(\boldsymbol{\theta})$ ,  $s_1 = \frac{N_1}{N_2+N_1}$ , and  $s_2 = \frac{N_2}{N_2+N_1}$ . The quantities  $\ell_{1,n}$  and  $\ell_{2,m}$  are defined as follows:

$$\ell_{1,n} = \frac{q_{1,1}}{q_{1,2}} = \frac{p(\boldsymbol{\theta}_n^* \mid \mathcal{H}_e) \mathbb{I}(\boldsymbol{\theta}_n^* \in \mathcal{R}_r)}{g(\boldsymbol{\xi}_n^*)}, \quad (7)$$

$$\ell_{2,m} = \frac{q_{2,1}}{q_{2,2}} = \frac{p(\tilde{\boldsymbol{\theta}}_m \mid \mathcal{H}_e) \mathbb{I}(\tilde{\boldsymbol{\theta}}_m \in \mathcal{R}_r)}{g(\boldsymbol{\xi}_m)}, \quad (8)$$

where  $\boldsymbol{\xi}_n^* = \Phi^{-1} \left( \frac{\boldsymbol{\theta}_n^* - \mathbf{l}}{\mathbf{u} - \mathbf{l}} \right)$ , and  $\tilde{\boldsymbol{\theta}}_m = ((\mathbf{u} - \mathbf{l}) \Phi(\boldsymbol{\xi}_m) + \mathbf{l}) \mid J \mid$ . The quantity  $q_{1,1}$  refers to the evaluations of the constrained distribution for constrained samples and  $q_{1,2}$  refers to the proposal distribution evaluated at the probit-transformed samples from the constrained distribution, respectively. The quantity  $q_{2,1}$  refers to evaluations of the constrained distribution at the inverse probit-transformed samples from the proposal distribution and  $q_{2,2}$  refers to the proposal evaluations for samples from the proposal, respectively. Note that the quantities  $\ell_{1,n}$  and  $\ell_{2,m}$  have been adjusted to account for the necessary parameter transformations to create overlap between the constrained distributions and the proposal distribution. **multibridge** runs the iterative scheme until the tolerance criterion suggested by Gronau et al. (2017) is reached, that is:

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

<sup>197</sup> The sampling from the target and proposal distribution, the transformations and  
<sup>198</sup> computational steps are performed automatically within the core functions of **multibridge**.

199 The user only needs to provide the functions with the data, a prior and a specification of the  
 200 informed hypothesis. As part of the standard output of `binom_bf_informed` and  
 201 `mult_bf_informed`, the functions return the bridge sampling estimate for the log marginal  
 202 likelihood of the target distribution, its associate relative mean square error, the number of  
 203 iterations, and the quantities  $q_{1,1}$ ,  $q_{1,2}$ ,  $q_{2,1}$ , and  $q_{2,2}$ .

204

## Usage and Examples

205 In the following, we will outline two examples on how to use `multibridge` to compare an  
 206 informed hypothesis to a null or encompassing hypothesis. The first example concerns  
 207 multinomial data and the second example concerns independent binomial data. Additional  
 208 examples are available as vignettes (see `vignette(package = "multibridge")`).

209 The two core functions of `multibridge`—`mult_bf_informed` and the  
 210 `binom_bf_informed`—can be illustrated schematically as follows:

```
mult_bf_informed(x, Hr, a, factor_levels)
binom_bf_informed(x, n, Hr, a, b, factor_levels)
```

211 To compute a Bayes factor, the core functions require the observed counts, the  
 212 informed hypothesis, the parameters of the prior distribution under  $\mathcal{H}_e$ , and the category  
 213 labels. An overview of these arguments are provided in Table 2.

Table 2

To estimate the Bayes factor in favor for or against the specified informed hypothesis, the user provides the core functions ***mult\_bf\_informed*** and ***binom\_bf\_informed*** with the basic required arguments listed below.

Argument	Description
<b>x</b>	<b>numeric</b> . Vector with data (for multinomial models) or a vector of counts of successes, or a two-dimensional table (or matrix) with 2 columns, giving the counts of successes and failures, respectively (for binomial models).
<b>n</b>	<b>numeric</b> . Vector with counts of trials. Must be the same length as <b>x</b> . Ignored if <b>x</b> is a matrix or a table. Included only in <b>binom_bf_informed</b> .
<b>Hr</b>	<b>string</b> or <b>character</b> . String or vector with the user specified informed hypothesis. Parameters may be referenced by the specified <b>factor_levels</b> or by numerical indices.
<b>a</b>	<b>numeric</b> . Vector with concentration parameters of Dirichlet distribution (for multinomial models) or $\alpha$ parameters for independent beta distributions (for binomial models). Must be the same length as <b>x</b> . Default sets all parameters to 1.
<b>b</b>	<b>numeric</b> . Vector with $\beta$ parameters. Must be the same length as <b>x</b> . Default sets all $\beta$ parameters to 1. Included only in <b>binom_bf_informed</b> .
<b>factor_levels</b>	<b>character</b> . Vector with category labels. Must be the same length as <b>x</b> .

<sup>214</sup>

The package also includes S3 methods that, among other things, summarize the results,

215 plot the parameter estimates under  $\mathcal{H}_e$ , or extract the Bayes factors. Table 3 summarizes all  
216 S3 methods currently available in **multibridge**.

Table 3

*S3 methods available in **multibridge**.*

Function Name(s)	S3 Method	Description
<code>mult_bf_informed,</code>	<code>print</code>	Prints model specifications and descriptives.
<code>binom_bf_informed</code>	<code>summary</code>	Prints and returns the Bayes factor and associated hypotheses for the full model, and all equality and inequality constraints.
	<code>plot</code>	Plots the posterior median and credible interval of the parameter estimates of the encompassing model. Default sets credible interval to 95%.
	<code>bayes_factor</code>	Contains all Bayes factors and log marginal likelihood estimates for inequality constraints.
	<code>samples</code>	Extracts prior and posterior samples from constrained densities (if bridge sampling was applied).
	<code>bridge_output</code>	Extracts bridge sampling output and associated error measures.
	<code>restriction_list</code>	Extracts restriction list and associated informed hypothesis.
<code>mult_bf_inequality,</code>	<code>print</code>	Prints the bridge sampling estimate for the log marginal likelihood and the corresponding percentage error.
<code>binom_bf_inequality</code>	<code>summary</code>	Prints and returns the bridge sampling estimate for the log marginal likelihood and associated error terms.

<sup>217</sup> **Example 1: Applying A Benford Test to Greek Fiscal Data**

<sup>218</sup> The first-digit phenomenon, otherwise known as Benford's law (Benford, 1938; Newcomb,  
<sup>219</sup> 1881) states that the expected proportion of leading digits in empirical data can be  
<sup>220</sup> formalized as follows: for any given leading digit  $d$ ,  $d = (1, \dots, 9)$  the expected proportion is  
<sup>221</sup> approximately equal to

$$\mathbb{E}_{\theta_d} = \log_{10}((d + 1)/d).$$

<sup>222</sup> This means that in an empirical data set, numbers with smaller leading digits are more  
<sup>223</sup> common than numbers with larger leading digits. Specifically, a number has leading digit 1  
<sup>224</sup> in 30.1% of the cases, and leading digit 2 in 17.61% of the cases; leading digit 9 is the least  
<sup>225</sup> frequent digit with an expected proportion of only 4.58% (see Table 4 for an overview of the  
<sup>226</sup> expected proportions). Empirical data for which this relationship holds include population  
<sup>227</sup> sizes, death rates, baseball statistics, atomic weights of elements, and physical constants  
<sup>228</sup> (Benford, 1938). In contrast, artificially generated data, such as telephone numbers, do in  
<sup>229</sup> general not obey Benford's law (Hill, 1995). Given that Benford's law applies to empirical  
<sup>230</sup> data but not artificially generated data, a so-called Benford test can be used in fields like  
<sup>231</sup> accounting and auditing to check for indications for poor data quality (for an overview, see  
<sup>232</sup> e.g., Durtschi, Hillison, & Pacini, 2004; Nigrini, 2012; Nigrini & Mittermaier, 1997). Data  
<sup>233</sup> that do not pass the Benford test, should raise audit risk concerns, meaning that it is  
<sup>234</sup> recommended that they undergo additional follow-up checks (Nigrini, 2019).

<sup>235</sup> Below we discuss four possible Bayesian adaptations of the Benford test. In a first  
<sup>236</sup> scenario we simply conduct a Bayesian multinomial test in which we test the point-null  
<sup>237</sup> hypothesis  $\mathcal{H}_0$  which predicts a Benford distribution against the encompassing hypothesis  
<sup>238</sup>  $\mathcal{H}_e$ . In a second scenario we test the null hypothesis against an alternative hypothesis,  
<sup>239</sup> denoted as  $\mathcal{H}_{r1}$ , which predicts a decreasing trend in the proportions of leading digits. The  
<sup>240</sup> hypothesis  $\mathcal{H}_{r1}$  exerts considerably more constraint than  $\mathcal{H}_e$  and provides a more sensitive  
<sup>241</sup> test if our primary goal is to test whether data comply with Benford's law or whether the

242 data follow a similar but different trend. In the next two scenarios, our main goal is to  
243 identify fabricated data. The third scenario therefore tests the null hypothesis against the  
244 hypothesis that all proportions occur equally often. This hypothesis  $\mathcal{H}_{r2}$  could be considered  
245 if it is suspected that the data were generated randomly. In a fourth scenario we test the  
246 null hypothesis against a hypothesis which predicts a trend that is characteristic for  
247 manipulated data. This hypothesis, which we denote as  $\mathcal{H}_{r3}$ , could be derived from empirical  
248 research on fraud or be based on observed patterns from former fraud cases. For instance,  
249 Hill (1995) instructed students to produce a series of random numbers; in the resulting data  
250 the proportion of the leading digit 1 occurred most often and the digits 8 and 9 occurred  
251 least often which is consistent with the general pattern of Benford's law. However, the  
252 proportion for the remaining leading digits were approximately equal. Note that the  
253 predicted distribution derived from Hill (1995) is not currently used as a test to detect fraud.  
254 However, for the sake of simplicity, if we assume that this pattern could be an indication of  
255 manipulated auditing data, the Bayes factor  $BF_{0r3}$  would quantify the evidence of whether  
256 the proportion of first digits resemble authentic or fabricated data.

257 **Data and Hypothesis.** The data we use to illustrate the computation of Bayes  
258 factors were originally published by the European statistics agency “Eurostat” and served as  
259 basis for reviewing the adherence to the Stability and Growth Pact of EU member states.  
260 Rauch, Götsche, Brähler, and Engel (2011) conducted a Benford test on data related to  
261 budget deficit criteria, that is, public deficit, public dept and gross national products. The  
262 data used for this example features the proportion of first digits from Greek fiscal data in the  
263 years between 1999 and 2010; a total of  $N = 1,497$  numerical data were included in the  
264 analysis. We choose this data, since the Greek government deficit and debt statistics states  
265 has been repeatedly criticized by the European Commission in this time span (European  
266 Commision, 2004, 2010). In particular, the commission has accused the Greek statistical  
267 authorities to have misreported deficit and debt statistics. For further details on the data set  
268 see Rauch et al. (2011). The observed and expected proportions are displayed in Table 4;

<sup>269</sup> the expected proportions versus the posterior parameter estimates under the encompassing  
<sup>270</sup> hypothesis are displayed in Figure 2.

Table 4

*Observed counts, observed proportions, and expected proportions of first digits in the Greek fiscal data set. The total sample size was  $N = 1,497$  observations. Note that the observed proportions and counts deviate slightly from those reported in Rauch et al. (2011) (probably due to rounding errors).*

Leading digit	Observed Counts	Observed Proportions	Expected Proportions:
			Benford's Law
1	509	0.340	0.301
2	353	0.236	0.176
3	177	0.118	0.125
4	114	0.076	0.097
5	77	0.051	0.079
6	77	0.051	0.067
7	53	0.035	0.058
8	73	0.049	0.051
9	64	0.043	0.046

<sup>271</sup> In this example, the parameter vector of the multinomial model,  $\theta_1, \dots, \theta_K$ , reflects  
<sup>272</sup> the probabilities of a leading digit in the Greek fiscal data being a number from 1 to 9. The  
<sup>273</sup> hypotheses introduced above can then be formalized as follows. The null hypothesis specifies  
<sup>274</sup> that the proportions of first digits obeys Benford's law:

$$\mathcal{H}_0 : \boldsymbol{\theta}_0 = (0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046).$$

This null hypothesis can then be tested against each of the following four alternative

hypotheses:

$$\mathcal{H}_e : \boldsymbol{\theta} \sim \text{Dirichlet}(\mathbf{1}),$$

$$\mathcal{H}_{r1} : \theta_1 > \theta_2 > \theta_3 > \theta_4 > \theta_5 > \theta_6 > \theta_7 > \theta_8 > \theta_9,$$

$$\mathcal{H}_{r2} : \boldsymbol{\theta}_0 = \left( \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9} \right),$$

$$\mathcal{H}_{r3} : \theta_1 > (\theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6 = \theta_7) > (\theta_8, \theta_9).$$

<sup>275</sup> The comparison of any two informed hypotheses with one another follows from the fact that  
<sup>276</sup> Bayes factors are transitive. For instance, the Bayes factor comparison between  $\mathcal{H}_0$  and  $\mathcal{H}_{r1}$   
<sup>277</sup> can be obtained by first computing  $\text{BF}_{e0}$  and  $\text{BF}_{er1}$ , and then dividing out the common  
<sup>278</sup> hypothesis  $\mathcal{H}_e$ :

$$\text{BF}_{0r1} = \frac{\text{BF}_{e0}}{\text{BF}_{er1}}.$$

<sup>279</sup> An overview of the relative plausibility of all  $M = 5$  models simultaneously may be  
<sup>280</sup> obtaining by presenting the posterior model probabilities  $p(\mathcal{H}_i | x)$  (Berger & Molina, 2005).  
<sup>281</sup> Denoting the prior model probability for model  $\mathcal{H}_i$  by  $p(\mathcal{H}_i)$ , the posterior model probability  
<sup>282</sup> for  $\mathcal{H}_0$  is given by:

$$p(\mathcal{H}_0 | \mathbf{x}) = \frac{\frac{p(\mathbf{x} | \mathcal{H}_0)}{p(\mathbf{x} | \mathcal{H}_e)} \times p(\mathcal{H}_0)}{\sum_{i=1}^M \frac{p(\mathbf{x} | \mathcal{H}_i)}{p(\mathbf{x} | \mathcal{H}_e)} \times p(\mathcal{H}_i)}.$$

<sup>283</sup> When all hypotheses are equally likely *a priori*, this simplifies to:

$$p(\mathcal{H}_0 | \mathbf{x}) = \frac{\text{BF}_{0e}}{\text{BF}_{0e} + \text{BF}_{r1e} + \text{BF}_{r2e} + \text{BF}_{r3e} + \text{BF}_{ee}}.$$

<sup>284</sup> **Method.** Both  $\text{BF}_{0e}$  and  $\text{BF}_{r2e}$  may be readily computed by means of a Bayesian  
<sup>285</sup> multinomial test which is implemented in the function `mult_bf_equality`. This function

286 requires (1) a vector with observed counts, (2) a vector with concentration parameters of the  
 287 Dirichlet prior distribution under  $\mathcal{H}_e$ , and (3) the vector of expected proportions under  $\mathcal{H}_0$   
 288 and under  $\mathcal{H}_{r2}$ . We do not incorporate specific expectations about the distribution of leading  
 289 digits in the Greek fiscal data and therefore set all concentration parameters under  $\mathcal{H}_e$  to 1  
 290 (i.e., we assign  $\boldsymbol{\theta}$  a uniform Dirichlet prior distribution).

```
# Observed counts
x <- c(509, 353, 177, 114, 77, 77, 53, 73, 64)

# Prior specification for Dirichlet prior distribution under H_e
a <- rep(1, 9)

# Expected proportions for H_0 and H_r2
p0 <- log10((1:9 + 1)/1:9)
pr2 <- rep(1/9, 9)

# Execute the analysis
results_H0_He <- mult_bf_equality(x = x, a = a, p = p0)
results_Hr2_He <- mult_bf_equality(x = x, a = a, p = pr2)

logBFe0 <- results_H0_He$bf$LogBFe0
logBFer2 <- results_Hr2_He$bf$LogBFe0
```

291 The hypotheses  $\mathcal{H}_{r1}$  and  $\mathcal{H}_{r3}$  contain inequality constraints, and this necessitates the  
 292 use of the function `mult_bf_informed` to compute the Bayes factors  $\text{BF}_{r1e}$  and  $\text{BF}_{r3e}$ . This  
 293 function requires (1) a vector with observed counts, (2) a vector with concentration  
 294 parameters of the Dirichlet prior distribution under  $\mathcal{H}_e$ , (3) labels for the categories of  
 295 interest (i.e., leading digits), and (4) the informed hypothesis  $\mathcal{H}_{r1}$  or  $\mathcal{H}_{r3}$  (e.g., as a string).  
 296 In addition to the basic required arguments, we use two additional arguments here. The first  
 297 argument sets the Bayes factor type, that is, whether the output should print the Bayes  
 298 factor in favor of the informed hypothesis (i.e.,  $\text{BF}_{re}$ ) or in favor of the encompassing

<sup>299</sup> hypothesis (i.e.,  $\text{BF}_{er}$ ). It is also possible to compute the log Bayes factor in favor of the  
<sup>300</sup> hypothesis, which is the setting we choose for this example. The purpose of the second  
<sup>301</sup> argument `seed` is to make the results reproducible:

```
# Observed counts
x <- c(509, 353, 177, 114, 77, 77, 53, 73, 64)

# Prior specification for Dirichlet prior distribution under H_e
a <- rep(1, 9)

# Labels for categories of interest
factor_levels <- 1:9

# Specifying the informed hypotheses as a string
Hr1 <- c('1 > 2 > 3 > 4 > 5 > 6 > 7 > 8 > 9')
Hr3 <- c('1 > 2 = 3 = 4 = 5 = 6 = 7 > 8 > 9')

# Execute the analysis
results_He_Hr1 <- mult_bf_informed(x = x, Hr = Hr1, a = a,
                                      factor_levels = factor_levels,
                                      bf_type = 'LogBFer', seed = 2020)

logBFer1 <- summary(results_He_Hr1)$bf

results_He_Hr3 <- mult_bf_informed(x = x, Hr = Hr3, a = a,
                                      factor_levels = factor_levels,
                                      bf_type = 'LogBFer', seed = 2020)

logBFer3 <- summary(results_He_Hr3)$bf
```

<sup>302</sup> We may now exploit transitivity to compare all alternative hypotheses to the Benford  
<sup>303</sup> null hypothesis  $\mathcal{H}_0$ . We also compute the posterior model probabilities for all hypotheses.  
<sup>304</sup> The results are shown in Table 5.

Table 5

*Prior model probabilities, posterior model probabilities, and Bayes factors for five rival accounts of first digit frequencies in the Greek fiscal data set.*

Hypothesis	$p(\mathcal{H}_.)$	$p(\mathcal{H}_.   \mathbf{x})$	$\log(\text{BF}_{.0})$
$\mathcal{H}_0$	0.2	$1.27 \times 10^{-11}$	0
$\mathcal{H}_{r1}$	0.2	0.9994	25.09
$\mathcal{H}_e$	0.2	0.0006	17.67
$\mathcal{H}_{r3}$	0.2	$9.46 \times 10^{-79}$	-154.57
$\mathcal{H}_{r2}$	0.2	$2.71 \times 10^{-212}$	-462.06

305        The results indicate strong support for  $\mathcal{H}_{r1}$  –the model in which the proportions are  
 306    assumed to decrease monotonically— over all other models. The log Bayes factor of  $\mathcal{H}_{r1}$   
 307    against Benford’s law  $\mathcal{H}_0$  is an overwhelming 25.09; the evidence for  $\mathcal{H}_{r1}$  is even stronger  
 308    when it is compared against models that feature equality constraints (i.e.,  $\mathcal{H}_{r2}$  and  $\mathcal{H}_{r3}$ ).  
 309    Finally,  $\mathcal{H}_{r1}$  also outperforms model  $\mathcal{H}_e$ , the unconstrained model in which all parameters  
 310    are free to vary. The latter result demonstrates how a parsimonious model that makes  
 311    precise predictions can be favored over a model that is more complex (e.g., Jefferys & Berger,  
 312    1992). The strong Bayes factor support for  $\mathcal{H}_{r1}$  translates to a relatively extreme posterior  
 313    model probability of 0.9994.

314        To summarize, the data offer overwhelming support for hypothesis  $\mathcal{H}_{r1}$ , which  
 315    postulates a decreasing trend in the digit proportions. This model outperformed both  
 316    simpler models (e.g., the Benford model) and a more complex model in which the  
 317    proportions were free to vary. Detailed follow-up analyses are needed to discover why the  
 318    data follow a monotonically decreasing pattern but not any of the two specific patterns that  
 319    were put to the test (Nigrini, 2019).

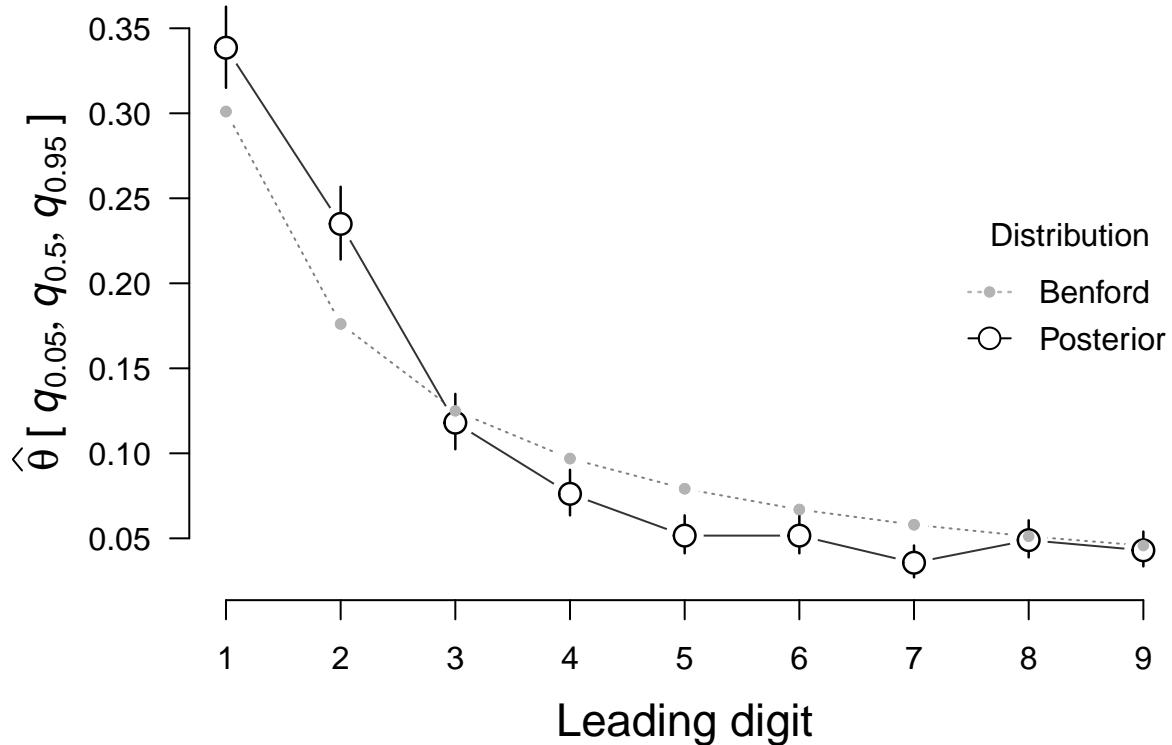


Figure 2. Predictions from Benford’s law (in grey) show together with the posterior medians (black circles) for the category proportions estimated under the encompassing model  $\mathcal{H}_e$ . The circle skewers show the 95% credible intervals. Only three of nine intervals encompass the expected proportions, suggesting that the data do not follow Benford’s law. This plot was created using the `plot-S3-method` for `summary.bmult` objects in **multibridge**.

### 320 Example 2: Prevalence of Statistical Reporting Errors

321 This section illustrates how **multibridge** may be used to evaluate models for independent  
 322 binomial data rather than multinomial data. Our example concerns the prevalence of  
 323 statistical reporting errors across eight different psychology journals. In any article that uses  
 324 null hypothesis significance testing, there is a chance that the reported test statistic and  
 325 degrees of freedom do not match the reported  $p$ -value, possibly because of copy-paste errors.  
 326 To flag these errors, Epskamp and Nuijten (2014) developed the R package **statcheck**,  
 327 which scans the PDF of a given scientific article and automatically detects statistical

328 inconsistencies. This package allowed Nuijten et al. (2016) to estimate the prevalence of  
 329 statistical reporting errors in the field of psychology. In total, the authors investigated a  
 330 sample of 30,717 articles (which translates to over a quarter of a million  $p$ -values) published  
 331 in eight major psychology journals between 1985 to 2013: *Developmental Psychology* (DP),  
 332 the *Frontiers in Psychology* (FP), the *Journal of Applied Psychology* (JAP), the *Journal of*  
 333 *Consulting and Clinical Psychology* (JCCP), *Journal of Experimental Psychology: General*  
 334 (*JEPG*), the *Journal of Personality and Social Psychology* (JPSP), the *Public Library of*  
 335 *Science* (PLoS), *Psychological Science* (PS).

336 Based on several background assumptions, Nuijten et al. (2016) predicted that the  
 337 proportion of statistical reporting errors is higher for articles published in the *Journal of*  
 338 *Personality and Social Psychology* (JPSP) than for articles published in the seven other  
 339 journals.

340 **Data and Hypothesis.** Here we reuse the original data published by Nuijten et al.  
 341 (2016), which we also distribute with the package **multibridge** under the name **journals**.

```
data(journals)
```

342 The Nuijten et al. (2016) hypothesis of interest,  $\mathcal{H}_r$ , states that the prevalence for  
 343 statistical reporting errors is higher for JPSP than for the other journals.<sup>2</sup> We will consider  
 344 two specific versions of the Nuijten et al. (2016)  $\mathcal{H}_r$  hypothesis. The first hypothesis,  $\mathcal{H}_{r1}$ ,  
 345 stipulates that JPSP has the highest prevalence of reporting inconsistencies, whereas the  
 346 other seven journals share a prevalence that is lower. The second hypothesis,  $\mathcal{H}_{r2}$ , also  
 347 stipulates that JPSP has the highest prevalence of reporting inconsistencies, but does not  
 348 commit to any particular structure on the prevalence for the other seven journals.

349 The **multibridge** package can be used to test  $\mathcal{H}_{r1}$  and  $\mathcal{H}_{r2}$  against the null hypothesis

---

<sup>2</sup>Nuijten et al. (2016) did not report inferential tests because they had sampled the entire population. We do report inferential tests here because we wish to learn about the latent data-generating process.

350  $\mathcal{H}_0$  that all eight journals have the same prevalence of statistical reporting errors. In addition,  
351 we will compare  $\mathcal{H}_{r1}$ ,  $\mathcal{H}_{r2}$ , and  $\mathcal{H}_0$  against the encompassing hypothesis  $\mathcal{H}_e$  that makes no  
352 commitment whatsoever about the prevalence of reporting inconsistencies across the eight  
353 journals. In this example, the parameter vector of the binomial success probabilities,  $\theta$ ,  
354 reflects the probabilities that articles contain at least one statistical reporting inconsistency  
355 across journals. Thus, the above hypotheses can be formalized as follows:

$$\mathcal{H}_0 : \theta_{\text{JAP}} = \theta_{\text{PS}} = \theta_{\text{JCCP}} = \theta_{\text{PLOS}} = \theta_{\text{DP}} = \theta_{\text{FP}} = \theta_{\text{JEPG}} = \theta_{\text{JPSP}}$$

$$\mathcal{H}_{r1} : (\theta_{\text{JAP}} = \theta_{\text{PS}} = \theta_{\text{JCCP}} = \theta_{\text{PLOS}} = \theta_{\text{DP}} = \theta_{\text{FP}} = \theta_{\text{JEPG}}) < \theta_{\text{JPSP}}$$

$$\mathcal{H}_{r2} : (\theta_{\text{JAP}}, \theta_{\text{PS}}, \theta_{\text{JCCP}}, \theta_{\text{PLOS}}, \theta_{\text{DP}}, \theta_{\text{FP}}, \theta_{\text{JEPG}}) < \theta_{\text{JPSP}}$$

$$\mathcal{H}_e : \theta_{\text{JAP}} \cdots \theta_{\text{JPSP}} \sim \prod_{k=1}^K \text{Beta}(\alpha_k, \beta_k).$$

356 **Method.** To compute the Bayes factor  $\text{BF}_{0r}$  we need to specify (1) a vector with  
357 observed successes (i.e., the number of articles that contain a statistical inconsistency), (2) a  
358 vector containing the total number of observations (i.e., the number of articles), (3) a vector  
359 with prior parameter  $\alpha_k$  for each binomial proportion of the beta prior distribution under  
360  $\mathcal{H}_e$ , (4) a vector with prior parameter  $\beta_k$  for each binomial proportion of the beta prior  
361 distribution under  $\mathcal{H}_e$ , (5) the category labels (i.e., journal names), and (6) the informed  
362 hypothesis  $\mathcal{H}_{r1}$  or  $\mathcal{H}_{r2}$  (e.g., as a string). We also change the Bayes factor type to `LogBFr0`  
363 so that the function returns the log Bayes factor in favor for the informed hypothesis  
364 compared to the null hypothesis. Since we have no specific expectations about the  
365 distribution of statistical reporting errors in any given journal, we set all parameters  $\alpha_k$  and  
366  $\beta_k$  to one which corresponds to uniform beta distributions. With this information, we can  
367 now conduct the analysis with the function `binom_bf_informed`.

```

# Since percentages are rounded to two decimal values, we round the
# articles with an error to obtain integer values

x <- round(journals$articles_with_NHST *
            (journals$perc_articles_with_errors/100))

# Total number of articles

n <- journals$articles_with_NHST

# Prior specification for beta prior distributions under H_e

a <- rep(1, 8)

b <- rep(1, 8)

# Labels for categories of interest

journal_names <- journals$journal

# Specifying the informed Hypothesis

Hr1 <- c('JAP = PS = JCCP = PLOS = DP = FP = JEPG < JPSP')

Hr2 <- c('JAP , PS , JCCP , PLOS , DP , FP , JEPG < JPSP')

# Execute the analysis for Hr1

results_H0_Hr1 <- binom_bf_informed(x = x, n = n, Hr = Hr1, a = a, b = b,
                                       factor_levels = journal_names,
                                       bf_type = 'LogBFr0', seed = 2020)

# Execute the analysis for Hr2

results_H0_Hr2 <- binom_bf_informed(x = x, n = n, Hr = Hr2, a = a, b = b,
                                       factor_levels = journal_names,
                                       bf_type = 'LogBFr0', seed = 2020)

LogBFe0  <- results_H0_Hr1$bf_list$bf0_table[['LogBFe0']]
LogBFr10 <- summary(results_H0_Hr1)$bf

```

```
LogBFr20 <- summary(results_H0_Hr2)$bf
```

Table 6

*Prior model probabilities, posterior model probabilities, and Bayes factors for four hypotheses concerning the prevalence of statistical reporting errors across psychology journals.*

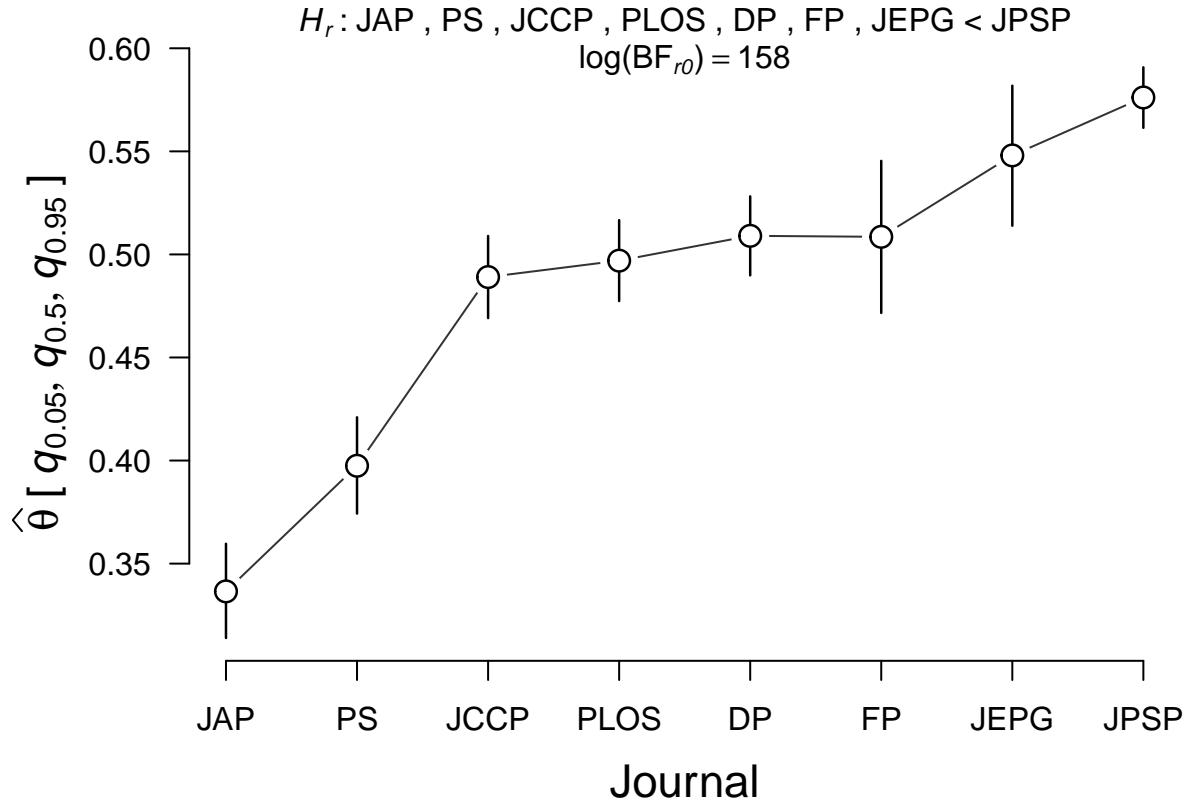
Hypothesis	$p(\mathcal{H}_.)$	$p(\mathcal{H}_.   \mathbf{x})$	$\log(\text{BF}_{.0})$
$\mathcal{H}_0$	0.25	$1.6073 \times 10^{-69}$	0
$\mathcal{H}_{r2}$	0.25	0.8814	158.28
$\mathcal{H}_e$	0.25	0.1186	156.27
$\mathcal{H}_{r1}$	0.25	$1.9517 \times 10^{-37}$	73.88

368 As the evidence is extreme in all four cases, we again report all Bayes factors on the log scale.  
 369 The Bayes factor  $\log(\text{BF}_{r20})$  indicates overwhelming evidence for the informed hypothesis  
 370 that JPSP has the highest prevalence for statistical reporting inconsistencies compared to  
 371 the null hypothesis that the statistical reporting errors are equal across all eight journals;  
 372  $\log(\text{BF}_{r20}) = 158.28$ .

373 For a clearer picture about the ordering of the journals we can investigate the posterior  
 374 distributions for the prevalence rates obtained under the encompassing model.

```
plot(summary(results_H0_Hr2), xlab = "Journal")
```

375 The posterior medians and 95% credible intervals are returned by the **summary**-method  
 376 and are shown in Figure 3. The figure strongly suggests that the prevalence of reporting  
 377 inconsistencies is not equal across all eight journals. This impression may be quantified by  
 378 comparing the null hypothesis  $\mathcal{H}_0$  to the encompassing hypothesis  $\mathcal{H}_e$ . The corresponding  
 379 Bayes factor equals  $\log(\text{BF}_{e0}) = 156.27$ , which confirms that the data dramatically undercut  
 380 the null hypothesis that the prevalence of statistical reporting inconsistencies is equal across  
 381 journals.



*Figure 3.* Posterior medians for the prevalence of statistical reporting inconsistencies across eight psychology journals, as obtained using the encompassing model. The circle skewers show the 95% credible intervals. Analysis based on data from Nuijten et al. (2016). This plot was created using the plot-S3-method for `summary.bmult` objects.

382 The data offer most support for the Nuijten hypothesis  $\mathcal{H}_{r2}$ , which posits that JPSP

383 has the highest prevalence but does not commit to any restriction on the prevalences for the

384 remaining seven journals. This hypothesis may be compared to the encompassing hypothesis

385  $\mathcal{H}_e$ , which yields  $\log(\text{BF}_{r2e}) = 2.01$ . This means that the observed data are  $\exp(2.01) \approx 7.45$

386 times more likely under  $\mathcal{H}_{r2}$  than under  $\mathcal{H}_e$ ; this is moderate evidence for the restriction

387 suggested by Nuijten et al. (2016). Under equal prior probability for the models, this Bayes

388 factor translates to a posterior probability on  $\mathcal{H}_e$  of 0.119, an amount that researchers may

389 deem too large to discard in an all-or-none fashion.

390 To summarize, the data provide moderate evidence for the hypothesis stated by

<sup>391</sup> Nuijten et al. (2016) that the prevalence of statistical reporting inconsistencies in JPSP is  
<sup>392</sup> higher than that in seven other psychology journals.

<sup>393</sup> **Summary**

<sup>394</sup> The R package **multibridge** facilitates the estimation of Bayes factors for informed  
<sup>395</sup> hypotheses in both multinomial and independent binomial models. The efficiency gains of  
<sup>396</sup> **multibridge** are particularly pronounced when the parameter restrictions are highly  
<sup>397</sup> informative or when the number of categories is large.

<sup>398</sup> **multibridge** supports the evaluation of informed hypotheses that feature equality  
<sup>399</sup> constraints, inequality constraints, and free parameters, as well as mixtures between them.  
<sup>400</sup> Moreover, users can choose to test the informative hypothesis against an encompassing  
<sup>401</sup> hypothesis that lets all parameters vary freely or against the null hypothesis that states that  
<sup>402</sup> category proportions are exactly equal. Beyond the core functions currently implemented in  
<sup>403</sup> **multibridge**, there are several natural extensions we aim to include in future versions of  
<sup>404</sup> this package. For instance, to compare several models with each other we plan to implement  
<sup>405</sup> functions that compute the posterior model probabilities. Another extension is to facilitate  
<sup>406</sup> the specification of hierarchical binomial and multinomial models which would allow users to  
<sup>407</sup> analyze data where responses are nested within a higher-order structure such as participants,  
<sup>408</sup> schools, or countries. Hierarchical multinomial models can be found, for instance, in source  
<sup>409</sup> memory research where people need to select a previously studied item from a list (e.g.,  
<sup>410</sup> Arnold, Heck, Bröder, Meiser, & Boywitt, 2019). In addition, we aim to enable the  
<sup>411</sup> specification of informed hypotheses that are more complex, including hypotheses on the size  
<sup>412</sup> ratios of the parameters (e.g.,  $\theta_1 < 2 \times \theta_2$ ) of interest or the difference between category  
<sup>413</sup> proportions such that informed hypotheses can also be specified on odds ratios (e.g.,  
<sup>414</sup>  $\frac{\theta_1}{(\theta_1 + \theta_2)} < \frac{\theta_3}{(\theta_3 + \theta_4)}$ ).

415

## Declarations

416 **Availability of data and code**

417 The source code of the R package is available at:

418 <https://github.com/ASarafoglou/multibridge/>. In addition, readers can access the code for  
419 reproducing all analyses and plots via our project folder on the Open Science Framework:  
420 <https://osf.io/2wf5y/>.

421 **Funding**

422 This research was supported by a Netherlands Organisation for Scientific Research  
423 (NWO) grant to AS (406-17-568), a Veni grant from the NWO to MM (451-17-017), a Vici  
424 grant from the NWO to EJW (016.Vici.170.083), as well as a European Research Council  
425 (ERC) grant to EJW (283876). This paper was written in Rmarkdown, using the R package  
426 **papaja** (Aust & Barth, 2020).

427 **Author contributions**

428 The authors made the following contributions. Alexandra Sarafoglou:

429 Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology,  
430 Project Administration, Software, Validation, Visualization, Writing - Original Draft  
431 Preparation, Writing - Review & Editing; Frederik Aust: Conceptualization, Software,  
432 Supervision, Validation, Visualization, Writing - Original Draft Preparation, Writing -  
433 Review & Editing; Maarten Marsman: Funding Acquisition, Conceptualization,  
434 Methodology, Supervision, Validation, Writing - Review & Editing; Eric-Jan Wagenmakers:  
435 Funding Acquisition, Methodology, Supervision, Validation, Writing - Review & Editing;

<sup>436</sup> Julia M. Haaf: Conceptualization, Formal Analysis, Methodology, Software,  
<sup>437</sup> Supervision, Validation, Writing - Original Draft Preparation, Writing - Review & Editing.

<sup>438</sup> **Conflicts of interest**

<sup>439</sup> The authors declare that there were no conflicts of interest with respect to the  
<sup>440</sup> authorship or the publication of this article.

<sup>441</sup> **Ethical Approval**

<sup>442</sup> This is a methodological contribution which requires no ethical approval.

**References**

- 443      Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing  
444      hypotheses about binding in context memory with a hierarchical multinomial  
445      modeling approach. *Experimental Psychology*, 66, 239–251.
- 446      Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R*  
447      *Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 448      Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American  
449      Philosophical Society*, 78, 551–572.
- 450      Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data.  
451      *Journal of Computational Physics*, 22, 245–268.
- 452      Damien, P., & Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities.  
453      *Journal of Computational and Graphical Statistics*, 10, 206–215.
- 454      Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in  
455      detecting fraud in accounting data. *Journal of Forensic Accounting*, 5, 17–34.
- 456      Epskamp, S., & Nuijten, M. (2014). *Statcheck: Extract statistics from articles and recompute*  
457      *p values (R package version 1.0.0.)*. Comprehensive R Archive Network. Retrieved  
458      from <https://cran.r-project.org/web/packages/statcheck>
- 459      European Commision. (2004). *Report by Eurostat on the revision of the Greek government  
460      deficit and debt figures* [Eurostat Report].  
461      <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/GREECE>.
- 462      European Commision. (2010). *Report on Greek government deficit and debt statistics*  
463      [Eurostat Report]. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/>
- 464

- 465 /COM\_2010\_REPORT\_GREEK.
- 466 Frigyik, B. A., Kapila, A., & Gupta, M. R. (2010). *Introduction to the Dirichlet distribution*  
467 *and related processes*. Department of Electrical Engineering, University of  
468 Washington.
- 469 Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., & Hothorn, F. S. T. (2020). *Mvtnorm:*  
470 *Multivariate normal and t distributions*. Retrieved from  
471 <http://CRAN.R-project.org/package=mvtnorm>
- 472 Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ...  
473 Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical*  
474 *Psychology*, 81, 80–97.
- 475 Gronau, Q. F., Singmann, H., & Wagenmakers, E. (2020). Bridgesampling: An R package for  
476 estimating normalizing constants. *Journal of Statistical Software, Articles*, 92, 1–29.
- 477 Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing  
478 of order constrained hypotheses in structural equation models. *Journal of Statistical*  
479 *Computation and Simulation*, 89, 1526–1553.
- 480 Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality  
481 constrained hypotheses. *Psychological Methods*, 19, 511–527.
- 482 Haaf, J. M., Klaassen, F., & Rouder, J. (2019). Capturing ordinal theoretical constraint in  
483 psychological science. *PsyArXiv*. Retrieved from  
484 <https://doi.org/10.31234/osf.io/a4xu9>
- 485 Heck, D. W., & Davis-Stober, C. P. (2019). Multinomial models with linear inequality  
486 constraints: Overview and improvements of computational methods for Bayesian  
487 inference. *Journal of Mathematical Psychology*, 91, 70–87.

- 488 Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science*, 10,  
489 354–363.
- 490 Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social*  
491 scientists. Boca Raton, FL: Chapman & Hall/CRC.
- 492 Hoijtink, H., Klugkist, I., & Boelen, P. (Eds.). (2008). *Bayesian evaluation of informative*  
493 *hypotheses*. New York: Springer Verlag.
- 494 Jeffreys, W. H., & Berger, J. O. (1992). Ockham’s razor and Bayesian analysis. *American*  
495 *Scientist*, 80, 64–72.
- 496 Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability.  
497 *Proceedings of the Cambridge Philosophy Society*, 31, 203–222.
- 498 Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical*  
499 *Association*, 90, 773–795.
- 500 Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing  
501 priors. *Statistica Neerlandica*, 59, 57–69.
- 502 Laudy, O. (2006). *Bayesian inequality constrained models for categorical data* (PhD thesis).  
503 Utrecht University.
- 504 Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple  
505 identity: A theoretical exploration. *Statistica Sinica*, 6, 831–860.
- 506 Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in) equality constrained  
507 hypotheses. *Computational Statistics & Data Analysis*, 71, 448–463.
- 508 Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations.  
509 *Journal of Mathematical Psychology*, 72, 104–115.

- 510 Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., ...  
511 van Lissa, C. (in press). BFpack: Flexible Bayes factor testing of scientific theories in  
512 R. *Journal of Statistical Software*.
- 513 Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for  
514 calculating Bayes factors for inequality and equality constrained models. *Journal of*  
515 *Statistical Software*, 46, 1–39.
- 516 Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H.  
517 (2009). Bayesian model selection of informative hypotheses for repeated  
518 measurements. *Journal of Mathematical Psychology*, 53, 530–546.
- 519 Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers.  
520 *American Journal of Mathematics*, 4, 39–40.
- 521 Nigrini, M. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud*  
522 *detection* (Vol. 586). Hoboken, New Jersey: John Wiley & Sons.
- 523 Nigrini, M. J. (2019). The patterns of the numbers used in occupational fraud schemes.  
524 *Managerial Auditing Journal*, 34, 602–622.
- 525 Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's law as an aid in analytical  
526 procedures. *Auditing*, 16, 52–67.
- 527 Nuijten, M. B., Hartgerink, C. H., Assen, M. A. van, Epskamp, S., & Wicherts, J. M. (2016).  
528 The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior*  
529 *Research Methods*, 48, 1205–1226.
- 530 Overstall, A. M., & Forster, J. J. (2010). Default Bayesian model determination methods for  
531 generalised linear mixed models. *Computational Statistics & Data Analysis*, 54,  
532 3269–3288.

- 533 Rauch, B., Götsche, M., Brähler, G., & Engel, S. (2011). Fact and fiction in  
534 EU-governmental economic data. *German Economic Review*, 12, 243–255.
- 535 Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences.  
536 *Psychological Review*, 118, 42–56.
- 537 Regenwetter, M., & Davis-Stober, C. P. (2012). Behavioral variability of choices versus  
538 structural inconsistency of preferences. *Psychological Review*, 119, 408–416.
- 539 Rijkeboer, M., & van den Hout, M. (2008). A psychologists's view on Bayesian evaluation of  
540 informative hypotheses. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian  
541 evaluation of informative hypotheses* (pp. 299–309). Berlin: Springer Verlag.
- 542 Sarafoglou, A., Haaf, J. M., Ly, A., Gronau, Q. F., Wagenmakers, E., & Marsman, M. (in  
543 press). Evaluating multinomial order restrictions with bridge sampling. *Psychological  
544 Methods*.
- 545 Sedransk, J., Monahan, J., & Chiu, H. (1985). Bayesian estimation of finite population  
546 parameters in categorical data models incorporating order restrictions. *Journal of the  
547 Royal Statistical Society. Series B (Methodological)*, 47, 519–527.

548

## Transforming an Ordered Probability Vector to the Real Line

549

The bridge sampling routine in **multibridge** uses the multivariate normal distribution as proposal distribution, which requires moving the target distribution  $\boldsymbol{\theta}$  to the real line.

551

Crucially, the transformation needs to retain the ordering of the parameters, that is, it needs to take into account the lower bound  $l_k$  and the upper bound  $u_k$  of each  $\theta_k$ . To meet these requirements, **multibridge** uses a probit transformation, as proposed in Sarafoglou et al. (in press), and subsequently transforms the elements in  $\boldsymbol{\theta}$ , moving from its lowest to its highest value. In the binomial model, we move all elements in  $\boldsymbol{\theta}$  to the real line and thus construct a new vector  $\mathbf{y} \in \mathbb{R}^K$ . For multinomial models it follows from the sum-to-one constraint that the vector  $\boldsymbol{\theta}$  is completely determined by its first  $K - 1$  elements, where  $\theta_K$  is defined as  $1 - \sum_{k=1}^{K-1} \theta_k$ . Hence, for multinomial models we will only consider the first  $K - 1$  elements of  $\boldsymbol{\theta}$  and we will transform them to  $K - 1$  elements of a new vector  $\mathbf{y} \in \mathbb{R}^{K-1}$ .

560

Let  $\phi$  denote the density of a normal variable with a mean of zero and a variance of one,  $\Phi$  denote its cumulative density function, and  $\Phi^{-1}$  denote the inverse cumulative density function. Then for each element  $\theta_k$ , the transformation is

$$\xi_k = \Phi^{-1} \left( \frac{\theta_k - l_k}{u_k - l_k} \right),$$

563 The inverse transformation is given by

$$\theta_k = (u_k - l_k)\Phi(\xi_k) + l_k.$$

564

To perform the transformations, we need to determine the lower bound  $l_k$  and the upper bound  $u_k$  of each  $\theta_k$ . Assuming  $\theta_{k-1} < \theta_k$  for  $k \in \{2, \dots, K\}$  the lower bound for any element in  $\boldsymbol{\theta}$  is defined as

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

567 This definition holds for both binomial models and multinomial models. Differences in  
 568 these two models appear only when determining the upper bound for each parameter. For  
 569 binomial models, the upper bound for each  $\theta_k$  is simply 1. For multinomial models, however,  
 570 due to the sum-to-one constraint the upper bounds depend on the values of smaller elements  
 571 as well as on the number of remaining larger elements in  $\boldsymbol{\theta}$ . To be able to determine the  
 572 upper bounds, we represent  $\boldsymbol{\theta}$  as unit-length stick which we subsequently divide into  $K$   
 573 elements (Frigyik, Kapila, & Gupta, 2010, p. stan2020). By using this so-called  
 574 stick-breaking method we can define the upper bound for any  $\theta_k$  as follows:

$$u_k = \begin{cases} \frac{1}{K} & \text{if } k = 1 \\ \frac{1 - \sum_{i < k} \theta_i}{ERS} & \text{if } 1 < k < K, \end{cases} \quad (\text{C1})$$

575 where  $1 - \sum_{i < k} \theta_i$  represents the length of the remaining stick, that is, the proportion of the  
 576 unit-length stick that has not yet been accounted for in the transformation. The elements in  
 577 the remaining stick are denoted as  $ERS$ , and are computed as follows:

$$ERS = K - 1 + k.$$

578 The transformations outlined above are suitable only for ordered probability vectors,  
 579 that is, for informed hypotheses in binomial and multinomial models that only feature  
 580 inequality constraints. However, when informed hypotheses also feature equality constrained  
 581 parameters, as well as parameters that are free to vary we need to modify the formula.  
 582 Specifically, to determine the lower bounds for any  $\theta_k$ , we need to take into account how

many parameters were set equal to it (denoted as  $e_k$ ) and how many parameters were set equal to its preceding value  $\theta_{k-1}$  (denoted as  $e_{k-1}$ ):

$$l_k = \begin{cases} 0 & \text{if } k = 1 \\ \frac{\theta_{k-1}}{e_{k-1}} \times e_k & \text{if } 1 < k < K. \end{cases} \quad (\text{C2})$$

The upper bound for parameters in the binomial models still remains 1. To determine the upper bound for multinomial models we must, additionally for each element  $\theta_k$ , take into account the number of free parameters that share common upper and lower bounds (denoted with  $f_k$ ). The upper bound is then defined as:

$$u_k = \begin{cases} \frac{1 - (f_k \times l_k)}{K} = \frac{1}{K} & \text{if } k = 1 \\ \left( \frac{1 - \sum_{i < k} \theta_i - (f_k \times l_k)}{ERS} \right) \times e_k & \text{if } 1 < k < K \text{ and } u_k \geq \max(\theta_{i < k}), \\ \left( 2 \times \left( \frac{1 - \sum_{i < k} \theta_i - (f_k \times l_k)}{ERS} \right) - \max(\theta_{i < k}) \right) \times e_k & \text{if } 1 < k < K \text{ and } u_k < \max(\theta_{i < k}). \end{cases} \quad (\text{C3})$$

The elements in the remaining stick are then computed as follows

$$ERS = e_k + \sum_{j > k} e_j \times f_j.$$

The rationale behind these modifications will be described in more detail in the following sections. In **multibridge**, information that is relevant for the transformation of the parameter vectors is stored in the generated `restriction_list` which is returned by the main functions `binom_bf_informed` and `mult_bf_informed` but can also be generated separately with the function `generate_restriction_list`. This restriction list features the

595 sublist `inequality_constraints` which encodes the number of equality constraints  
 596 collapsed in each parameter in `nr_mult_equal`. Similarly the number of free parameters  
 597 that share common bounds are encoded under `nr_mult_free`.

## 598 Equality Constrained Parameters

599 In cases where informed hypotheses feature a mix of equality and inequality  
 600 constrained parameters, we compute the Bayes factor  $\text{BF}_{re}$ , by multiplying the individual  
 601 Bayes factors for both constraint types with each other:

$$602 \quad \text{BF}_{re} = \text{BF}_{1e} \times \text{BF}_{2e} \mid \text{BF}_{1e},$$

602 where the subscript 1 denotes the hypothesis that only features equality constraints and the  
 603 subscript 2 denotes the hypothesis that only features inequality constraints. To receive  
 604  $\text{BF}_{2e} \mid \text{BF}_{1e}$ , we collapse all equality constrained parameters in the constrained prior and  
 605 posterior distributions into one category. This collapse has implications on the performed  
 606 transformations.

607 When transforming the samples from the collapsed distributions, we need to account  
 608 for the fact that the inequality constraints imposed under the original parameter values  
 609 might not hold for the collapsed parameters. Consider, for instance, a multinomial model in  
 610 which we specify the following informed hypothesis

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

where samples from the encompassing distribution take the values

(0.05, 0.15, 0.15, 0.15, 0.23, 0.27). For these parameter values the inequality constraints hold since 0.05 is smaller than 0.15, 0.23, and 0.27. However, the same constraint does not hold when we collapse the categories  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$  into  $\theta_*$ . That is, the collapsed parameter

$\theta_* = 0.15 + 0.15 + 0.15 = 0.45$  is now larger than 0.23 and 0.27. In general, to determine the lower bound for a given parameter  $\theta_k$  we thus need to take into account both the number of collapsed categories in the preceding parameter  $e_{k-1}$  as well as the number of collapsed categories in the current parameter  $e_k$ . Thus, lower bounds for the parameters need to be adjusted as follows:

$$l_k = \begin{cases} 0 & \text{if } k = 1 \\ \frac{\theta_{k-1}}{e_{k-1}} \times e_k & \text{if } 1 < k < K, \end{cases}$$

611 which leads to Equation C2. In this equation,  $e_{k-1}$  and  $e_k$  refer to the number of equality  
 612 constrained parameters that are collapsed in  $\theta_{k-1}$  and  $\theta_k$ , respectively. In the example above,  
 613 this means that to determine the lower bound for  $\theta_*$  we multiply the preceding value  $\theta_1$  by  
 614 three, such that the lower bound is  $(\frac{0.05}{1}) \times 3 = 0.15$ . In addition, to determine the lower  
 615 bound of  $\theta_5$  we divide the preceding value  $\theta_*$  by three, that is,  $(\frac{0.45}{3}) \times 1 = 0.15$ . Similarly,  
 616 to determine the upper bound for a given parameter value  $\theta_k$ , we need to multiple the upper  
 617 bound by the number of parameters that are collapsed within it:

$$u_k = \begin{cases} \frac{1}{ERS} \times e_k & \text{if } k = 1 \\ \frac{1 - \sum_{i < k} \theta_i}{ERS} \times e_k & \text{if } 1 < k < K, \end{cases} \quad (\text{C4})$$

618 where  $1 - \sum_{i < k} \theta_i$  represents the length of the remaining stick and the number of elements in  
 619 the remaining stick are computed as follows:  $ERS = \sum_k^K e_k$ . For the example above, the

620 upper bound for  $\theta_*$  is  $\frac{1 - 0.05}{5} \times 3 = 0.57$ . The upper bound for  $\theta_5$  is then

621  $\frac{(1 - 0.05 - 0.45)}{2} \times 1 = 0.25$ .

622 Corrections for Free Parameters

623 Different adjustments are required for a sequence of inequality constrained parameters  
 624 that share upper and lower bounds. Consider, for instance, a multinomial model in which we  
 625 specify the informed hypothesis

$$\mathcal{H}_r : \theta_1 < (\theta_2, \theta_3) < \theta_4.$$

This hypothesis specifies that  $\theta_2$  and  $\theta_3$  have the shared lower bound  $\theta_1$  and the shared upper bound  $\theta_4$ , however,  $\theta_2$  can be larger than  $\theta_3$  or vice versa. To integrate these cases within the stick-breaking approach one must account for these potential changes of order. For these cases, the lower bounds for the parameters remain unchanged. To determine the upper bound for  $\theta_k$ , we need to subtract from the length of the remaining stick the lower bound from the parameters that are free to vary. However, only those parameters are included in this calculation that have not yet been transformed:

$$u_k = \begin{cases} \frac{1 - (f_k \times l_k)}{K} & \text{if } k = 1 \\ \frac{1 - \sum_{i < k} \theta_i - (f_k \times l_k)}{ERS} & \text{if } 1 < k < K, \end{cases} \quad (C5)$$

626 where  $f_k$  represents the number of free parameters that share common bounds with  $\theta_k$  and  
 627 that have been not yet been transformed. Here, the number of elements in the remaining  
 628 stick is defined as the number of all parameters that are larger than  $\theta_k$ :  $ERS = 1 + \sum_{j > k} f_j$ .  
 629 To illustrate this correction, assume that samples from the encompassing distribution take  
 630 the values  $(0.15, 0.29, 0.2, 0.36)$ . The upper bound for  $\theta_1$  is simply  $\frac{1}{4}$ . For  $\theta_2$ , we need to take  
 631 into account that  $\theta_2$  and  $\theta_3$  share common bounds. To compute the upper bound for  $\theta_2$ , we  
 632 subtract from the length of the remaining stick the lower bound of  $\theta_3$ :

$$633 \frac{1 - 0.15 - (1 \times 0.15)}{1 + 1} = 0.35.$$

A further correction is required if a preceding free parameter (i.e., a parameter with common bounds that was transformed already) is larger than the upper bound of the current

parameter. For instance, in our example the upper bound for  $\theta_3$  would be

$\frac{1 - 0.44 - 0}{1 + 1} = 0.28$ , which is smaller than the value of the preceding free parameter, which was 0.29. If in this case  $\theta_3$  would actually take on the value close to its upper bound, for instance  $\theta_3 = 0.275$ , then—due to the sum-to-one constraint— $\theta_4$  would violate the constraint (i.e.,  $0.15 < (0.29, 0.275) \not< 0.285$ ). In these cases, the upper bound for the current  $\theta_k$  needs to be corrected downwards. To do this, we subtract from the current upper bound the difference to the largest preceding free parameter. Thus, if  $u_k < \max(\theta_{i < k})$ , the upper bound becomes:

$$u_k = u_k - (\max(\theta_{i < k}) - u_k) \quad (\text{C6})$$

$$= 2 \times u_k - \max(\theta_{i < k}). \quad (\text{C7})$$

- 634 For our example the corrected upper bound for  $\theta_3$  would become  $2 \times 0.28 - 0.29 = 0.27$
- 635 which secures the proper ordering for the remainder of the parameters. If in this case  $\theta_3$
- 636 would take on the value close to its upper bound, for instance  $\theta_3 = 0.265$ ,  $\theta_4$ —due to the
- 637 sum-to-one constraint—would take on the value 0.295 which would be in accordance with
- 638 the constraint (i.e.,  $0.15 < (0.29, 0.265) < 0.295$ ).