1   multibridge: An R Package To Evaluate Multinomial Order Constraints

2   Alexandra Sarafoglou[1], Julia M. Haaf[1], Frederik Aust[1], Eric-Jan Wagenmakers[1], & Maarten

3   Marsman[1]

4   [1] University of Amsterdam

5   Abstract

6   The **multibridge** package has been developed to efficiently compute Bayes factors for

7   binomial and multinomial models, that feature inequality constraints, equality constrains

8   and free parameters. By using the bridge sampling algorithm to compute the Bayes factor,

9   **multibridge** facilitates the evaluation of large models with many constraints and models

10 with very small parameter spaces. The package was developed in the R programming

11 language and is freely available from the Comprehensive R Archive Network. We illustrate

12 the functions based on two empirical examples.

13   multibridge: An R Package To Evaluate Multinomial Order Constraints

14   **Introduction**

15 I would first mention for which models the package is, multinomial and binomial and then then constraints.

16   We present **multibridge**, an R package to evaluate informed hypotheses on category
17 variables using Bayesian inference. This package allows users to specify constraints on the
18 underlying category proportions including inequality constraints, equality constraints, free
19 parameters and mixtures between them. This package is available from the Comprehensive R
20 Archive Network (CRAN) at https://CRAN.R-project.org/package=multibridge. Here we
21 introduce the methodology used to evaluate informed hypotheses in multinomial models and
22 models featuring independent binomials and show how to use the implementations provided
23 in **multibridge** through fully reproducible examples.

24   The most common way to analyze categorial variables is to test whether the underlying
25 category proportions are exactly equal or whether they are fixed and follow a predicted
26 pattern. Often however, scientific hypotheses go beyond this standard case and predict for
27 instance ordinal relations among the underlying category proportions, such as increasing or
28 decreasing trends. For instance, to check for irregularities in audit data, one could test
29 whether the leading digits in the data are distributed according to an expected Benford
30 distribution or whether they deviate from it by, for example, showing a general decreasing
31 trend. Informed hypotheses could also feature combinations of equality and inequality
32 constrained parameters, as well as parameters that are free to vary. For instance, when
33 studying the prevalence of statistical reporting errors in articles published in different areas
34 of psychological science, one could hypothesize that articles published in social psychology
35 journals have higher error rates than articles published in other psychological journals while
36 not expressing expectations about the error rate distribution among these other journals
37 (Nuijten, Hartgerink, Assen, Epskamp, & Wicherts, 2016). Generally, testing informed

38 hypotheses allows researchers to specify hypotheses that follow more closely relate to their
39 theories.

> I really like the paragraph below. But I think we have to motivate it a bit better. How
40 do you get from the paragraph above to this one?

41 For the **multibridge** package the emphasis is on the approximation of Bayes factors
42 (Jeffreys, 1935; Kass & Raftery, 1995) to quantify evidence for or against informative models.
43 In the `R` programming language, the package **multinomineq** (Heck & Davis-Stober, 2019) is
44 available to evaluate order constrained hypotheses for multinomial models as well as models
45 that feature independent binomials. **multinomineq** allows users to specify inequality
46 constrained hypotheses but also more general linear inequality constraints. The `BAIN` (Gu,
47 Hoijtink, Mulder, & Rosseel, 2019) package allows for the evaluation of inequality constraints
48 in structural equation models. Outside of `R`, the software package `BIEMS` (Mulder, Hoijtink,
49 Leeuw, & others, 2012) allows for the evaluation of order constraints for multivariate linear
50 models such as MANOVA, repeated measures, and multivariate regression.

> Which language is it? Or is it a GUI?
51

52 All these packages rely on two popular methods to approximate order constrained
53 Bayes factors, the encompassing prior approach (Gu, Mulder, Deković, & Hoijtink, 2014;
54 Hoijtink, 2011; Hoijtink, Klugkist, & Boelen, 2008; Klugkist, Kato, & Hoijtink, 2005) and
55 the conditioning method (Mulder, 2014, 2016; Mulder et al., 2009). But even though these
56 methods are currently very popular and widely used, they are shown to become increasingly
57 unreliable and inefficient as the number of constraints increases or when the parameter space
58 of the constrained model is small (Sarafoglou et al., 2020).

59 The **multibridge** uses a bridge sampling routine that enables users to compute Bayes
60 factors for informed hypotheses more reliably and efficiently (Bennett, 1976; Meng & Wong,
61 1996; Sarafoglou et al., 2020). The workhorse for this analysis, the bridge sampling

62 algorithm, constitutes a special case of the algorithm implemented in the `R` package

63 **bridgesampling** (Gronau, Singmann, & Wagenmakers, 2020). With **bridgesampling**,

64 users are able to estimate the marginal likelihood for a wide variety of models, including

65 models implemented in Stan (Stan Development Team, 2020). However, **bridgesampling** is

66 not suitable for models that include constraints on probability vectors. In **multibridge**, we

67 tailored the bridge sampling algorithm such that it accommodates the specification of

68 informed hypotheses on probability vectors. The general workflow of **multibridge** is
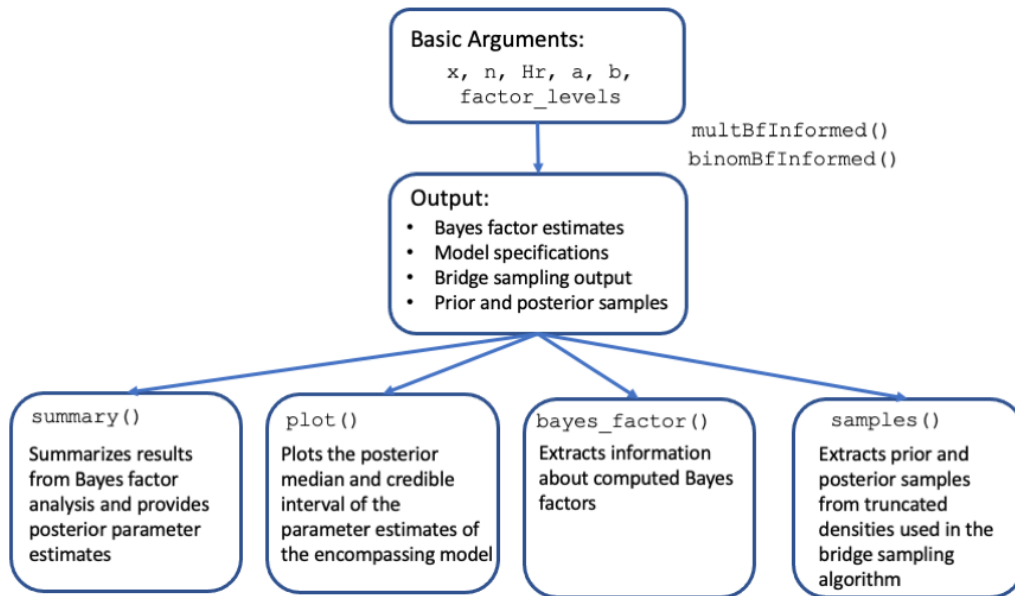
69 illustrated in Figure 1.

Figure 1. The **multibridge** workflow. The user needs to specify the data values (`x` and `n` for binomial models and `x` for multinomial models, respectively), the informed hypothesis (`Hr`), the α and β parameters of the Binomial prior distributions (`a` and `b`) or the concentration parameters for the Dirichlet prior distribution (`a`), respectively, and the factor levels (`factor_levels`). The functions `multBfInformed` and `binomBfInformed` then produce an estimate for the Bayes factor of the informed hypothesis versus the encompassing hypothesis in which all parameters are free to vary. Based on these results different S3 methods can be used to get more detailed information on the individual components of the analysis (`summary`, (\texttt{bayes_factor}), and parameter estimates of the encompassing distribution (`plot`).

The two core functions of **multibridge** –the `multBfInformed`-function and the `binomBfInformed`-function – can be illustrated schematically as follows:

```
multBfInformed(x, Hr, a factor_levels)
binomBfInformed(x, n, Hr, a, b, factor_levels)
```

The basic required arguments for these functions are listed in Table 1.

The package produces an estimate for the Bayes factor in favor of or against the informed hypothesis. The resulting Bayes factor compares the evidence for the informed hypotheses to the encompassing hypothesis that imposes no constraints on the underlying category proportions. Given this result, the user can then either receive a visualization of the prior and posterior parameter estimates using the `plot`-method, or get more detailed information on how the Bayes factors is composed using the `summary`-method. For hypotheses that include mixtures between equality and inequality informed hypotheses the `bayes_factor` method shows the conditional Bayes factor for the inequality constraints given the equality constraints and a Bayes factor for the equality constraints. Table 2 summarizes all S3 methods currently implemented in **multibridge**.

This remainder of this article is organized as follows: In the methods section, we describe the Bayes factor identity for informed hypotheses in binomial and multinomial models, and present the bridge sampling routine implemented in the **multibridge** package including details of the necessary transformations required for this routine. In Section 3, we will schematically introduce the most relevant functions in **multibridge** and their arguments. Section 4 illustrates how to use the **multibridge** package to estimate parameters, and compute Bayes factors using two examples.

stopped here.

Methods

multibridge allows users to specify informed hypotheses in multinomial models and

models that feature independent binomial probabilities. In the multinomial model, twe

assumes that the vector of observations $x_1, \cdots, x_K$ in the K categories follow a multinomial

distribution. The parameter vector of the multinomial model, $\theta_1, \cdots, \theta_K$, contains the

probabilities of observing a value in a particular category. The parameter vector $\theta_1, \cdots, \theta_K$

is drawn from a Dirichlet distribution with concentration parameters $\alpha_1, \cdots, \alpha_K$. Formally,

the model can be described as follows:

$$x_1, \cdots, x_K \sim \text{Multinomial}(\sum_{k=1}^{K} x_k, \theta_1, \cdots, \theta_K) \tag{1}$$

$$\theta_1, \cdots, \theta_K \sim \text{Dirichlet}(\alpha_1, \cdots, \alpha_K). \tag{2}$$

$$\tag{3}$$

In the binomial model we assume that the elements in the vector of successes

$x_1, \cdots, x_K$ and the elements in the vector of total number of observations $n_1, \cdots, n_K$ in the

K categories follow independent binomial distributions. As in the multinomial model, the

parameter vector of the binomial success probabilities, $\theta_1, \cdots, \theta_K$, contains the probabilities

of observing a value in a particular category. The parameter vector $\theta_1, \cdots, \theta_K$ are drawn

from independent beta distributions with parameters $\alpha_1, \cdots, \alpha_K$ and $\beta_1, \cdots, \beta_K$. The

model can be described as follows:

$$x_1 \cdots x_K \sim \prod_{k=1}^{K} \text{Binomial}(\theta_k, n_k) \tag{4}$$

$$\theta_1 \cdots \theta_K \sim \prod_{k=1}^{K} \text{Beta}(\alpha_k, beta_k) \tag{5}$$

$$\tag{6}$$

¹⁰⁶ **Bayes factor**

¹⁰⁷   When evaluating informed hypotheses that feature mixtures between inequality and

¹⁰⁸ equality constraints it is important to realize that the Bayes factor, further denoted as BF_{me},

¹⁰⁹ factors follows:

$$\mathrm{BF}_{me} = \mathrm{BF}_{0e} \times \mathrm{BF}_{re} \mid \mathrm{BF}_{0e},$$

¹¹⁰ where the subscript m denotes a hypothesis that features mixtures of inequality and equality

¹¹¹ constraints. A Bayes factor for mixtures thus factors into a Bayes factor for the equality

¹¹² constraints, BF_{0e}, and a conditional Bayes factor for the inequality constraints given the

¹¹³ equality constraints $\mathrm{BF}_{re} \mid \mathrm{BF}_{0e}$.

¹¹⁴ **The Bayes Factor For Equality Constraints**

    For binomial models, the (marginal) Bayes factor for the equality constraints can be

computed analytically and implemented in the function `binomBfEquality`. Assuming that

the first i binomial probabilities in a model are equality constraint, the Bayes factor BF_{0e} is

defined as:

$$\mathrm{BF}_{0e} = \quad = \frac{\prod_{i<k} \mathrm{B}(\alpha_i,\, \beta_i)}{\prod_{i<k} \mathrm{B}(\alpha_i + x_i,\, \beta_i + n_i - x_i)} \times \frac{\mathrm{B}(\alpha_+ + x_+ - i + 1,\, \beta_+ + n_+ - x_+ - i + 1)}{\mathrm{B}(\alpha_+ - i + 1,\, -i + 1)}$$

¹¹⁵ where $\sum_{i<k} \alpha_i = \alpha_+$, $\sum_{i<k} \beta_i = \beta_+$, $\sum_{i<k} x_i = x_+$ and $\sum_{i<k} n_i = n_+$. The latter factor

¹¹⁶ introduces a correction for marginalizing which stems from the change in degrees of freedom,

¹¹⁷ when we collapse i equality constraint parameters, that is, for i collapsed categories, $i - 1$

¹¹⁸ degrees of freedom are lost which are subtracted from the prior parameters in the

¹¹⁹ corresponding Binomial distribution.

    For multinomial models, the (marginal) Bayes factor for the equality constraints is also

analytically available and implemented in the function `multBayesBfEquality`. Assuming

again that the first i category probabilities in a model are equality constraint, the Bayes

factorBF$_{0e}$ is defined as:

$$\mathrm{BF}_{e0} = \frac{\mathrm{B}(\boldsymbol{\alpha})}{\mathrm{B}(\boldsymbol{\alpha} + \mathbf{x})} \left(\frac{1}{i}\right)^{\sum_{i<k} x_i} \frac{\mathrm{B}\left(\sum_{i<k} \alpha_i + x_i - i + 1,\, \alpha_k + x_k,\, \ldots,\, \alpha_K + x_K\right)}{\mathrm{B}\left(\sum_{i<k} \alpha_i - i + 1,\, \alpha_k,\, \ldots,\, \alpha_K\right)},$$

The Bayes Factor For Inequality Constraints

For inequality constrained hypotheses, Klugkist et al. (2005) has derived the following identity of the Bayes factor BF_{re}:

$$\mathrm{BF}_{re} = \frac{\overbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathbf{x}, \mathcal{H}_e)}^{\substack{\text{Proportion of posterior parameter} \\ \text{space consistent with the restriction}}}}{\underbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)}_{\substack{\text{Proportion of prior parameter} \\ \text{space consistent with the restriction}}}}, \tag{7}$$

where in BF_{re}, the subscript r denotes the inequality constrained hypothesis and the subscript e denotes the encompassing hypothesis that lets all parameters free to vary. Recently, however, Sarafoglou et al. (2020) showed that the Bayes factor BF_{re} can also be interpreted as ratio of two marginal likelihoods:

$$\mathrm{BF}_{re} = \frac{\overbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathbf{x}, \mathcal{H}_e)}^{\substack{\text{Marginal likelihood of} \\ \text{posterior distribution}}}}{\underbrace{p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)}_{\substack{\text{Marginal likelihood of} \\ \text{prior distribution}}}}. \tag{8}$$

In this identity, $p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathbf{x}, \mathcal{H}_e)$ denotes the marginal likelihood of the constrained posterior distribution and $p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)$ denotes the marginal likelihood of the constrained prior distribution. Even though both identities are mathematically equivalent, the methods to estimate these identities are very different. In the first case, for instance, the number of samples from the encompassing distribution in accordance with the inequality constrained

132 hypothesis, serve as an estimate for the proportion of prior parameter space consistent with

133 the restriction. On the flip side, however, this means that the accuracy of this estimate is

134 strongly dependent on the number of the constrained parameters in the model and the size

135 of the constrained parameter space. That is, the smaller the constrained parameter space is,

136 the less likely it is that draws from the encompassing distribution will fall in this region, so

137 that in some cases the estimation of the Bayes factor becomes practically impossible

138 (Sarafoglou et al., 2020).

139   However, when we interpret the Bayes factor BF_{re} as ratio of marginal likelihoods and

140 we are able to sample from the constrained prior and posterior distributions, numerical

141 sampling methods such as bridge sampling to obtain the estimates. Crucially, in this

142 approach, it does not matter how small the constrained parameter space is in proportion to

143 the encompassing density. This gives the method a decisive advantage in terms of accuracy

144 and efficiency.

The Bridge Sampling Method

146   Bridge sampling is a method to estimate the ratio of two marginal likelihoods which

147 yield the Bayes factor (Bennett, 1976; Meng & Wong, 1996). In the **multibridge** package

148 we implemented a version of bridge sampling that estimates one marginal likelihood at the

149 time since it increases the accuracy of the method without considerably increasing its

150 computational efficiency (Overstall & Forster, 2010). Specifically, we subsequently estimate

151 the marginal likelihood for the constrained prior distribution and the marginal likelihood of

152 the constrained posterior distribution.

153   When applying this modified version of the bridge sampling method, we estimate a

154 marginal likelihood by means of a so-called proposal distribution. In **multibridge** this

155 proposal distribution is the multivariate normal distribution. To estimate the marginal

156 likelihood, bridge sampling only requires samples from the distribution of interest –the

157 so-called target distribution– and samples from the proposal distribution. In **multibridge**,

158 the samples from the target distribution –that is the constrained prior and posterior

159 Dirichlet distribution for multinomial models and constrained prior and posterior beta

160 distributions for binomial models– are drawn through the Gibbs sampling algorithms

161 proposed by Damien and Walker (2001). For binomial models, we apply the suggested Gibbs

162 sampling algorithm for constrained beta distributions. In the case of the multinomial models,

163 however, we apply an algorithm that simulates values from constrained Gamma distributions

164 which are then transformed into Dirichlet random variables. To sample efficiently from these

165 distributions, **multibridge** uses a `C++` routine for this algorithm.

166   The efficiency of the bridge sampling method is guaranteed only if the target and

167 proposal distribution (1) operate on the same parameter space and (2) have sufficient

168 overlap. To meet these requirements, **multibridge** applies the appropriate probit

169 transformations on the samples of the constrained distributions to move the samples from

170 the probability space to the entire real line. Details on these transformations are provided in

171 the appendix. To ensure sufficient overlap, half of the draws are then used to construct the

172 proposal distribution using the method of moments. Samples from the proposal distribution

173 can be generated using the standard `rmvnorm()`-function from the `R` package `stats`. For the

174 marginal likelihood of the constrained prior distribution, the modified bridge sampling

175 identity is then defined as:

$$p(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e) = \frac{\mathbb{E}_{g(\boldsymbol{\theta})}\left(p(\boldsymbol{\theta} \mid \mathcal{H}_e)\mathbb{I}(\boldsymbol{\theta} \in \mathcal{R}_r)h(\boldsymbol{\theta})\right)}{\mathbb{E}_{\text{prior}}\left(g(\boldsymbol{\theta})h(\boldsymbol{\theta})\right)}, \tag{9}$$

176 where the term $h(\boldsymbol{\theta})$ refers to the bridge function proposed by Meng and Wong (1996) which

177 minimized the relative mean square error of the estimate and $g(\boldsymbol{\theta})$ refers to the proposal

178 distribution. The numerator evaluates the unnormalized density for the constrained prior

179 distribution with samples from the proposal distribution. The denominator evaluates the

180 normalized proposal distribution with samples from the constrained prior distribution. The

181 expression for the marginal likelihood for the constrained posterior distribution can be

182 described in a similar way. As final step, we apply the iterative scheme proposed by Meng

183 and Wong (1996) to receive the bridge sampling estimator:

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

184 where N_1 denotes the number of samples drawn from the constrained distribution, that is,

185 $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta} \mid \mathcal{H}_r)$, N_2 denotes the number of samples drawn from the proposal distribution, that

186 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:

$$\ell_{1,n} = \frac{q_{1,1}}{q_{1,2}} = \frac{p(\boldsymbol{\theta}_{\boldsymbol{n}}^* \mid \mathcal{H}_e)\mathbb{I}(\boldsymbol{\theta}_{\boldsymbol{n}}^* \in \mathcal{R}_r)}{g(\boldsymbol{\xi_n}^*)}, \tag{10}$$

$$\ell_{2,m} = \frac{q_{2,1}}{q_{2,2}} = \frac{p(\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}} \mid \mathcal{H}_e)\mathbb{I}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}} \in \mathcal{R}_r)}{g(\tilde{\boldsymbol{\xi}}_{\boldsymbol{m}})}, \tag{11}$$

187 where $\boldsymbol{\xi_n}^* = \Phi^{-1}\left(\dfrac{\boldsymbol{\theta}_{\boldsymbol{n}}^* - \mathbf{l}}{\mathbf{u} - \mathbf{l}}\right)$, and $\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}} = ((\mathbf{u} - \mathbf{l})\Phi(\tilde{\boldsymbol{\xi}}_{\boldsymbol{m}}) + \mathbf{l})\,|J|)$. The quantity $q_{1,1}$ refers to the

188 evaluations of the constrained distribution for constrained samples and $q_{1,2}$ refers to the

189 proposal evaluations for constrained samples, respectively. The quantities $q_{2,1}$ refers to

190 evaluations of the constrained distribution for samples from the proposal and $q_{2,2}$ refers to

191 the proposal evaluations for samples from the proposal, respectively. Note that the quantities

192 $\ell_{1,n}$ and $\ell_{2,m}$ have been adjusted to account for the necessary parameter transformations to

193 create overlap between the constrained distributions and the proposal distribution.

194 **multibridge** runs the iterative scheme until the tolerance criterion suggested by Gronau et

195 al. (2017) is reached, that is, $\dfrac{\mid \hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)^{(t+1)} - \hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)^{(t)} \mid}{\hat{p}(\boldsymbol{\theta} \in \mathcal{R}_r \mid \mathcal{H}_e)^{(t+1)}} \leq 10^{-10}.$

196   The bridge sampling estimate for the log marginal likelihood of the constrained

197   distribution and its associate relative mean square error, the number of iterations, and the

198   quantities $q_{1,2}$, $q_{1,2}$, $q_{1,2}$, and $q_{1,2}$ are included in the standard output in **multibridge**. The

199   function to compute the relative mean square error was taken from the R package

200   **bridgesampling**.

Usage and Examples

202   The **multibridge** package can be installed from the Comprehensive R Archive

203   Network (CRAN) at https://CRAN.R-project.org/package=multibridge:

```
# install.packages('multibridge')

library('multibridge')
```

204   Note that the following examples make use of multiple functions that we implemented

205   in **multibridge**. A list of all functions and datasets currently available are given in Table 3.

206   Additional examples are available as vignettes at:

207   https://cran.r-project.org/package=multibridge.

Example 1: Appling A Benford Test to Greek Fiscal Data

209   The first digit phenomenon, otherwise known as Benford's law (Benford, 1938;

210   Newcomb, 1881) states that the expected proportion of leading digits in empirical data can

211   be formalized as follows: for any given leading digit $d, d = (1, \cdots, 9)$ the expected proportion

212   is approximately equal to

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

213   This means that a number in a empirical dataset has leading digit 1 in 30.1% of the cases,

214   and leading digit 2 in 17.61% of the cases; leading digit 9 is the least frequent digit with an

215   expected proportion of only 4.58% (see Table 4 for an overview of the expected proportions).

216   Benford (1938) showed that his law applies to a broad range of real-world data; among

²¹⁷ others, it applies to data on population sizes, death rates, baseball statistics, atomic weights

²¹⁸ of elements, and physical constants. In contrast, generated data, such as telephone numbers,

²¹⁹ do in general not obey Benford's law (Hill, 1995). Since Benford's law proved to be highly

²²⁰ suitable to discriminate between empirical data and generated data, a so-called Benford test

²²¹ can be used in fields like accounting and auditing as an indication for poor data quality (for

²²² an overview, see e.g., Durtschi, Hillison, and Pacini (2004), Nigrini and Mittermaier (1997),

²²³ Nigrini (2012)). A Benford test typically checks whether observed frequencies of first digits,

²²⁴ for instance, from fiscal statements, obey Benford's law. Data that do not pass the Benford

²²⁵ test, should raise audit risk concerns, meaning that, it is recommended that the data

²²⁶ undergo additional follow-up checks (Nigrini, 2019).

²²⁷   In the following, we discuss three possible Bayesian adaptations of Benford's test. In a

²²⁸ first scenario we simply conduct Bayesian multinomial test in which we test the point-null

²²⁹ hypothesis $\mathcal{H}_0$ which predicts a Benford distribution against the encompassing hypothesis $\mathcal{H}_e$

²³⁰ which lets all model parameters free to vary. Testing against the encompassing hypothesis is

²³¹ considered standard practice, yet, it leads to an unfair comparison to the detriment of the

²³² null hypothesis. In general, if we are dealing with a high-dimensional parameter space and

²³³ the competing hypotheses differ largely in their complexity, the Bayes factor generally favors

²³⁴ the less complex hypothesis even if the data follow the predicted trend of the more complex

²³⁵ hypothesis considerably well. In a second scenario we therefore test the null hypothesis

²³⁶ against an alternative hypothesis, denoted as $\mathcal{H}_{r1}$, which predicts a decreasing trend in the

²³⁷ proportions of leading digits. The hypothesis $\mathcal{H}_{r1}$ is considerably more complex than $\mathcal{H}_e$ and

²³⁸ is a suitable choice if our primary goal is to distinguish whether data comply with Benford's

²³⁹ law or whether the data only follow a similar trend. In a third scenario we could be

²⁴⁰ interested in testing the null hypothesis against an alternative hypothesis, which predicts a

²⁴¹ trend that is characteristic for manipulated data. This alternative hypothesis, which we

²⁴² denote as $\mathcal{H}_{r2}$, could be derived from empirical research on fraud or be based on observed

²⁴³ patterns from former fraud cases. For instance, Hill (1988) instructed students to produce a

²⁴⁴ series random numbers; in the resulting data the proportion of the leading digit 1 occurred

²⁴⁵ most often and the digits 8 and 9 occurred least often which is consistent with the general

²⁴⁶ pattern of Benford's law. However, the proportion for the remaining leading digits were

²⁴⁷ approximately equal. We do want to note, that the predicted distribution derived from Hill

²⁴⁸ (1988) is not currently used as a test to detect manipulated data patterns. However, for the

²⁴⁹ sake of simplicity, if we assume that this pattern could be an indication for completely

²⁵⁰ invented auditing data, the Bayes factor could quantify the evidence of whether the

²⁵¹ proportion of first digits resemble authentic or invented data.

²⁵²   **Data and Hypothesis.**   The data we use to illustrate the computation of Bayes

²⁵³ factors were originally published by the European statistics agency "Eurostat" and served as

²⁵⁴ basis for reviewing the adherence to the Stability and Growth Pact of EU member states.

²⁵⁵ Rauch, Göttsche, Brähler, and Engel (2011) conducted a Benford test on data related to

²⁵⁶ budget deficit criteria, i.e., public deficit, public dept and gross national products. This data

²⁵⁷ used for this example contains fiscal data from Greece related in the years between 1999 and

²⁵⁸ 2010; a total of $N = 1497$ numerical data were included in the analysis. We choose this data,

²⁵⁹ since the Greek government deficit and debt statistics states has been repeatedly criticized

²⁶⁰ by the European Commission in this timespan (European Commision, 2004,

²⁶¹ @europeanCommision2010). In particular, the commission has accused the Greek statistical

²⁶² authorities, to have misreported deficit and debt statistics. For further details on the dataset

²⁶³ see Rauch et al. (2011). The observed proportions are displayed in Table 4, the figure

²⁶⁴ displaying the observed versus the expected proportions are displayed in Figure **??**.

²⁶⁵   In this example, the parameter vector of the multinomial model, $\theta_1, \cdots, \theta_K$, reflects

²⁶⁶ the probabilities of a leading digit in the Greek fiscal data being a number from 1 to 9. Thus,

²⁶⁷ we can formalize the discussed hypotheses as follows. The null hypothesis specifies that the

²⁶⁸ proportions of first digits obeys Benford's law:

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

We are testing the null hypothesis against the following alternative hypotheses:

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

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

$$\mathcal{H}_{r2} : \theta_1 > (\theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6 = \theta_7) > (\theta_8, \ \theta_9).$$

Note that the `multibridge` package exclusively computes Bayes factors of a informed hypothesis against the encompassing hypothesis. In cases, in which we are interested in computing two informed hypotheses with each other, we need to make use of the transitivity property of the Bayes factor. For instance, if we would like to compare the inequality-constrained hypothesis $\mathcal{H}_r$ against the null hypothesis $\mathcal{H}_0$, we would first compute BF_{er} and BF_{e0} and then yield BF_{r0} as follows:

$$\text{BF}_{re} \times \text{BF}_{e0} = \text{BF}_{r0}.$$

Method.   We can compare $\mathcal{H}_0$ and $\mathcal{H}_e$ by means of a Bayesian multinomial test, that is, we stipulate equality constraints on the entire parameter vector $\boldsymbol{\theta}$. The corresponding Bayes factor is thus computationally straightforward; we can calculate BF_{0e} by applying the function `multBfEqualtiy()`. To evaluate $\mathcal{H}_0$, we only need to specify (1) a vector with observed counts, (2) a vector with concentration parameters, and (3) the vector of predicted proportions. Since we have no specific expectations about the distribution of leading digits in the Greek fiscal data, we choose in all subsequent analyses the uniform Dirichlet distribution as prior for the vector of model parameters.

```
# Observed counts
x <- c(509, 353, 177, 114,  77,  77,  53,  73,  64)
# Prior specification
a <-  rep(1, 9)
# Expected proportions
p <- log10((1:9 + 1)/1:9)
```

```
# Execute the analysis

results_H0_He  <- multBfEquality(x = x, a = a, p = p)
```

Since the hypotheses $\mathcal{H}_{r1}$ and $\mathcal{H}_{r2}$ contain inequality constraints, we use the function multBfInformed to compute the Bayes factor of the informed hypotheses to the encompassing hypothesis. To evaluate $\mathcal{H}_{r1}$ and $\mathcal{H}_{r2}$, we need to specify (1) and a vector containing the number of observations, (2) the inequality-constrained hypotheses, (3) a vector with concentration parameters, and (4) the categories of interest (i.e., leading digits):

```
# Categories of interest

factor_levels <- 1:9


# Specifying the informed Hypothesis (step 3)

Hr1 <- c('1 > 2 > 3 > 4 > 5 > 6 > 7 > 8 > 9')

Hr2 <- c('1 > 2 = 3 = 4 = 5 = 6 = 7 > 8 > 9')


# Execute the analysis

results_He_Hr1 <- multBfInformed(x = x,

                                 Hr = Hr1,

                                 a = a,

                                 factor_levels = factor_levels,

                                 seed = 2020)

results_He_Hr2 <- multBfInformed(x = x,

                                 Hr = Hr2,

                                 a = a,

                                 factor_levels = factor_levels,

                                 seed = 2020)
```

```r
logbf <- c(log(results_H0_He$bf$BF0e),
           log(results_H0_He$bf$BF0e * results_He_Hr1$bf_list$bf$BFer),
           log(results_H0_He$bf$BF0e * results_He_Hr2$bf_list$bf$BFer))
bayes_factor_table <- data.frame(
    BFType = c('LogBF0e', 'LogBF0r1', 'LogBF0r2'),
    LogBF  = logbf)
bayes_factor_table
```

As the evidence is extreme in all three cases, we reported all Bayes factors on the log scale which allows us to compare the numbers more easily. The log Bayes factor $log(\mathrm{BF}_{e0})$ suggests extreme evidence against the hypothesis that the first digits in the Greek fiscal data follow a Benford's distribution; $\log(\mathrm{BF}_{0e})$ = -17.67. The log Bayes factor $\log(\mathrm{BF}_{er1})$ indicates extreme evidence in favor for a decreasing trend, $\log(\mathrm{BF}_{0r1})$ = -25.09. Only for the hypothesis that the data follow a pattern of fraudulent data, we yield extreme evidence in favor for the null hypothesis, that is, $\log(\mathrm{BF}_{er2})$ = 154.57. Overall, these results suggest that the data deviate from the Benford distribution, in the sense, that the proportion of leading digits are decreasing, instead of all parameters varying freely ($\log(\mathrm{BF}_{r1e})$ = 7.42), or being distributed as one could expect from completely invented data ($\log(\mathrm{BF}_{r1r2})$ = 180).

```
## Bayes factor analysis
##
##   Hypothesis H_e:
##   All parameters are free to vary.
##
##   Hypothesis H_r:
##   1 > 2 > 3 > 4 > 5 > 6 > 7 > 8 > 9
##
## Bayes factor estimate LogBFer:
```

```
307  ##
308  ## -7.4168
309  ##
310  ## Based on 1 independent inequality-constrained hypothesis.
311  ##
312  ## Posterior Median and Credible Intervals Of Marginal Beta Distributions:
313  ##         alpha      beta    2.5%     50%  97.5%
314  ## 1 1 1 + 509 8 + 988  0.3150 0.3390 0.3630
315  ## 2 2 1 + 353 8 + 1144 0.2140 0.2350 0.2570
316  ## 3 3 1 + 177 8 + 1320 0.1020 0.1180 0.1350
317  ## 4 4 1 + 114 8 + 1383 0.0635 0.0762 0.0903
318  ## 5 5 1 + 77  8 + 1420 0.0412 0.0516 0.0635
319  ## 6 6 1 + 77  8 + 1420 0.0412 0.0516 0.0635
320  ## 7 7 1 + 53  8 + 1444 0.0271 0.0357 0.0458
321  ## 8 8 1 + 73  8 + 1424 0.0388 0.0489 0.0606
322  ## 9 9 1 + 64  8 + 1433 0.0335 0.0430 0.0540
```

Discussion. In this example we tested the data quality of Greek fiscal data in the years 1999 to 2009 by conducting three variations of a Bayesian Benford test. More precise, we evaluated the null hypothesis that the data conform to Benfords law. We tested this hypothesis against three alternatives. The first alternative hypothesis, $\mathcal{H}_e$ relaxed the constraints imposed by the null hypothesis and let all model parameters free to vary. The second alternative hypothesis, $\mathcal{H}_{r1}$ predicted a decreasing trend in the proportion of leading digits. The third alternative hypothesis $\mathcal{H}_{r2}$ predicted a trend that Hill (1988) observed in manipulated data. Our result suggest that the leading digits in the fiscal statistics do not follow a Benford distribution; in fact, we collected extreme evidence against Benford's law compared to all two out of three of the alternative hypotheses. When comparing the alternative hypotheses directly to each other, the data show most evidence in favor for a
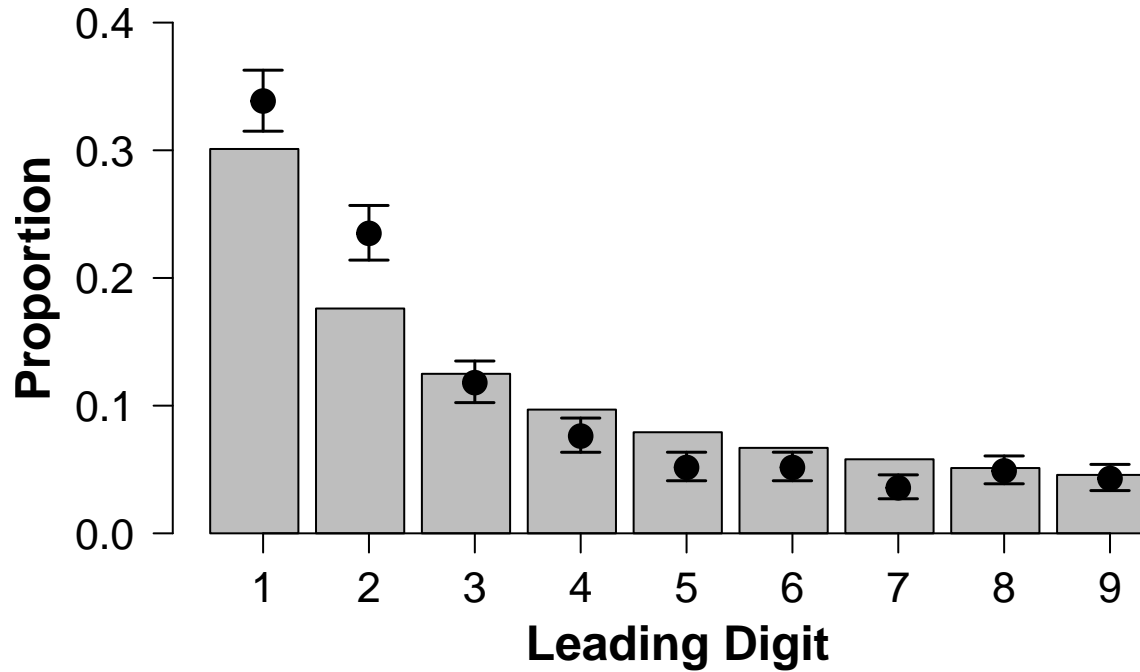
Figure 2. The bargraph displays the expected proportions of leading digits according to Benford's law. The black dots indicate for the actual fiscal statistics from Greece the posterior estimates for the proportion of leading digits and the corresponding 95% credible intervals. Only three out of nine estimates cover the expected proportions.

decreasing trend. A Benford test when used in to check fiscal statements can be a helpful tool to detect poor data quality and suspicious numbers. In follow-up checks of these numbers, it could then be examined for instance, whether financial statements were actually materially misstated (by, for instance, rounding up or down numbers, avoiding certain thresholds etc., Nigrini, 2019).

Example 2: Prevalence of Statistical Reporting Errors

In any scientific article that uses null hypothesis significance testing, there is a chance that the reported test statistic and degrees of freedom, do not match the reported p-value. In most cases this is because researchers copy the relevant test statistics by hand into their articles and there are no automatic checks to detect these mistakes. Therefore, Epskamp and

³⁴⁴ Nuijten (2014) developed the R package `statcheck`, which only requires the PDF of a given

³⁴⁵ scientific article to detect these reporting errors automatically and efficiently. This package

³⁴⁶ allowed Nuijten et al. (2016) to get an overview about the prevalence of statistical reporting

³⁴⁷ errors in the field of psychology. In total, the authors investigated a sample of $30,717$ articles

³⁴⁸ (which translates to over a quarter of a million p-values) published in eight major

³⁴⁹ psychological journals between 1985 to 2013: *Developmental Psychology* (DP), the *Frontiers*

³⁵⁰ *in Psychology* (FP), the *Journal of Applied Psychology* (JAP), the *Journal of Consulting and*

³⁵¹ *Clinical Psychology* (JCCP), *Journal of Experimental Psychology: General* (JEPG), the

³⁵² *Journal of Personality and Social Psychology* (JPSP), the *Public Library of Science* (PLoS),

³⁵³ *Psychological Science* (PS).

³⁵⁴   Besides the overall prevalence of statistical reporting errors across these journals, the

³⁵⁵ authors were interested whether there is a higher prevalence for reporting inconsistencies in

³⁵⁶ certain subfields in psychology compared to others. In this context the assumption was raised

³⁵⁷ that there exists a relationship between the prevalence for reporting inconsistencies and

³⁵⁸ questionable research practices. Specifically, the authors argued that besides honest mistakes

³⁵⁹ when transferring the test statistics into the manuscript, statistical reporting error occur

³⁶⁰ when authors misreport p-values, for instance, by incorrectly rounding them down below 0.05.

³⁶¹ Based on this assumption Nuijten et al. (2016) predicted that the proportion of statistical

³⁶² reporting errors should be highest in articles published in the *Journal of Personality and*

³⁶³ *Social Psychology* (JPSP), compared to other journals, since researchers in social psychology

³⁶⁴ were shown to have the highest prevalence for questionable research practices (John,

³⁶⁵ Loewenstein, & Prelec, 2012). Specifically, John et al. (2012) found that researchers from

³⁶⁶ the area of social psychology assessed questionable research practices both as more defensible

³⁶⁷ and more applicable for their research compared to other research areas in psychology.

³⁶⁸   **Data and Hypothesis.**   We will use the original data published in the article by

³⁶⁹ Nuijten et al. (2016) and which we included under the name of `journals` in the package

³⁷⁰ **multibridge**.

```
# load the package and data
data(journals)
```

371   The hypothesis of interest, $\mathcal{H}_r$, formulated by Nuijten et al. (2016) states that the

372 prevalence for statistical reporting errors for articles published in social psychology journals

373 (i.e., JPSP) is higher than for articles published in other journals. We will test this

374 hypothesis against the the null hypothesis $\mathcal{H}_0$ that all journals have the same prevalence for

375 statistical reporting errors. In this example, the parameter vector of the binomial success

376 probabilities, $\boldsymbol{\theta}$, it reflects the probabilities of a statistical reporting error in one of the 8

377 journals. Thus, we can formalize the discussed hypotheses as follows:

$$\mathcal{H}_r : \theta_{\text{DP}}, \theta_{\text{FP}}, \theta_{\text{JAP}}, \theta_{\text{JCCP}}, \theta_{\text{JEPG}}, \theta_{\text{PLoS}}, \theta_{\text{PS}}) < \theta_{\text{JPSP}}$$

$$\mathcal{H}_0 : \theta_{\text{DP}} = \theta_{\text{FP}} = \cdots = \theta_{\text{JPSP}}.$$

378   **Method.**   As before, we can compute BF_{0r} through the transitivity of the Bayes

379 factor, that is, $\text{BF}_{r0} = \text{BF}_{re} \times \text{BF}_{e0}$. The Bayes factor BF_{e0} can be computed by using the

380 `binomBfEquality()` function.

```
# Since percentages are rounded to two decimal values, we round the articles
# with an error to receive 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
# We assign a uniform beta distribution to each binomial propotion
a <- rep(1, 8)
```

```r
b <- rep(1, 8)


results_He_H0   <- binomBfEquality(x=x, n=n, a=a, b=b)

eq_bayesfactors <- results_He_H0$bf
```

381   The data suggest that the null hypothesis is highly unlikely; we collected extreme

382 evidence against the null hypothesis with a log Bayes factor $\log(\text{BF}_{0e})$ of -156.

383   To compute the Bayes factor BF_{re} we need to specify (1) a vector with observed

384 successes, and (2) a vector containing the total number of observations, (3) the informed

385 hypothesis, (4) a vector with prior parameters alpha for each binomial proportion, (5) a

386 vector with prior parameters beta for each binomial proportion, and (6) the categories of

387 interest (i.e., journal names). With this information, we can now conduct the analysis with

388 the function `binomBfInformed`.

```r
# Specifying the informed Hypothesis
Hr <- c('JAP , PS , JCCP , PLOS , DP , FP , JEPG < JPSP')


# Categories of interest
journal_names <- journals$journal


# Execute the analysis
results_Hr_He <- binomBfInformed(x = x, n = n, Hr = Hr, a = a, b = b,
                                 factor_levels = journal_names,
                                 bf_type = 'BFre', seed = 2020)
```

389   We collected moderate evidence for the informed hypothesis. Specifically, the results

390 suggest that the data are 7.43 more likely under the informed hypothesis than under the

391 hypothesis that all parameters are free to vary. As final step, we compare the informed and

392 the null hypothesis directly with each other.

```
BFre <- results_Hr_He$bf_list$bf$BFre

BFe0 <- eq_bayesfactors[['BFe0']]

BFr0 <- BFre * BFe0
```

393 The Bayes factor $log(\text{BF}_{r0})$ suggests extreme evidence for the informed hypothesis;
394 $\log(\text{BF}_{r0}) = 158$.

395 **Discussion.** In this example we tested whether the prevalence for statistical
396 reporting errors for articles published in social psychology journals (i.e., JPSP) is higher
397 than for articles published in other journals. We tested this hypothesis against the null
398 hypothesis that the prevalence for statistical reporting errors is equal across all journals. The
399 resulting Bayes factor of $\log(\text{BF}_{r0}) = 5.48e+68$ provides extreme evidence for the informed
400 hypothesis. However, this result should be interpreted with caution and be considered more
401 differentiated. It seems that the result is above all an indication that the null hypothesis is
402 highly misspecified and that the prevalence for a statistical reporting error varies greatly
403 from journal to journal. Evidence that JPSP stands out and has a higher prevalence than
404 the other journals is relatively small; the data provided only moderate evidence against the
405 encompassing hypotheses.

<h1 align="center">Summary</h1>

406

407 The R package **multibridge** facilitates the computation of Bayes factors for informed
408 hypotheses in multinomial models. The underlying algorithm is based on a bridge sampling
409 routine that was recently developed and was shown to be more efficient and produces more
410 reliable estimates than comparable methods. The current version of **multibridge**, the user
411 can specify hypotheses that feature equality constraints, inequality constraints, and free
412 parameters as well as mixtures between them. The core functions of the software package
413 were illustrated with two empirical examples. The **multibridge** package is under continuous
414 development. In future versions of the package, we aim to implement methods that allow for

the evaluation of hierarchical multinomial models. In addition, we want to allow users to

specify order constraints that are more complex, including hypotheses on the size ratios of

the parameters of interest or the difference between category proportions.

References

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 551–572.

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*, 245–268.

Damien, P., & Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, *10*, 206–215.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, *5*, 17–34.

Epskamp, S., & Nuijten, M. (2014). Statcheck: Extract statistics from articles and recompute p values (R package version 1.0.0.).

European Commision. (2004). Report by Eurostat on the revision of the Greek government deficit and debt figures. Eurostat Report, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/GREECE.

European Commision. (2010). Report on Greek government deficit and debt statistics. Eurostat Report, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/COM_2010_REPORT_GREEK.

Frigyik, B. A., Kapila, A., & Gupta, M. R. (2010). *Introduction to the Dirichlet distribution and related processes.* Department of Electrical Engineering, University of Washington.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ...
       Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97.

Gronau, Q. F., Singmann, H., & Wagenmakers, E. (2020). Bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software, Articles*, *92*(10), 1–29.

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*, 1526–1553.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*, 511–527.

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.

Hill, T. P. (1988). Random-number guessing and the first digit phenomenon. *Psychological Reports*, *62*, 967–971.

Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science*, 354–363.

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists.* Boca Raton, FL: Chapman & Hall/CRC.

Hoijtink, H., Klugkist, I., & Boelen, P. (Eds.). (2008). *Bayesian evaluation of informative hypotheses.* New York: Springer Verlag.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. In

461   *Proceedings of the Cambridge Philosophy Society* (Vol. 31, pp. 203–222).

462   John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable

463   research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

464   Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical

465   Association*, *90*, 773–795.

466   Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing

467   priors. *Statistica Neerlandica*, *59*, 57–69.

468   Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple

469   identity: A theoretical exploration. *Statistica Sinica*, *6*, 831–860.

470   Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in) equality constrained

471   hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463.

472   Mulder, J. (2016). Bayes factors for testing order–constrained hypotheses on correlations.

473   *Journal of Mathematical Psychology*, *72*, 104–115.

474   Mulder, J., Hoijtink, H., Leeuw, C. de, & others. (2012). BIEMS: A Fortran 90 program for

475   calculating Bayes factors for inequality and equality constrained models. *Journal of

476   Statistical Software*, *46*, 1–39.

477   Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H.

478   (2009). Bayesian model selection of informative hypotheses for repeated

479   measurements. *Journal of Mathematical Psychology*, *53*, 530–546.

480   Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers.

481   *American Journal of Mathematics*, *4*, 39–40.

482   Nigrini, M. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud*

detection (Vol. 586). Hoboken, New Jersey: John Wiley & Sons.

Nigrini, M. J. (2019). The patterns of the numbers used in occupational fraud schemes. *Managerial Auditing Journal*, *34*, 602–622.

Nigrini, M. J., & Mittermaier, L. J. (1997). The use of benford's law as an aid in analytical procedures. *Auditing*, *16*, 52.

Nuijten, M. B., Hartgerink, C. H., Assen, M. A. van, Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*, 1205–1226.

Overstall, A. M., & Forster, J. J. (2010). Default Bayesian model determination methods for generalised linear mixed models. *Computational Statistics & Data Analysis*, *54*, 3269–3288.

Rauch, B., Göttsche, M., Brähler, G., & Engel, S. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, *12*, 243–255.

Sarafoglou, A., Haaf, J. M., Ly, A., Gronau, Q. F., Wagenmakers, E., & Marsman, M. (2020). Evaluating multinomial order restrictions with bridge sampling. *PsyArXiv*. Retrieved from https://psyarxiv.com/bux7p/

Stan Development Team. (2020). *Stan modeling language user's guide and reference manual, version 2.23.0*. R Foundation for Statistical Computing. Retrieved from http://mc-stan.org/

Appendix: Transforming An Ordered Probability Vector To The Real Line

Since we choose the multivariate normal as proposal distribution, the mapping between the proposal and target distribution requires us to move $\boldsymbol{\theta}$ to the real line. Crucially this transformation needs to retain the inequality constraints imposed on the parameters. To

506 achieve this goal, **multibridge** uses a probit transformation which subsequently transforms

507 the elements in $\boldsymbol{\theta}$ moving from its lowest to its highest value. In the binomial model, we

508 move all elements in $\boldsymbol{\theta}$ to the real line and thus construct a new vector $\boldsymbol{y} \in \mathbb{R}^K$. For

509 multinomial models, however, it follows from the unit constraint that the vector $\boldsymbol{\theta}$ is

510 completely determined by its first $K-1$ elements of $\boldsymbol{\theta} : \theta_1 \leq \theta_2 \leq \cdots \leq 1 - \sum_{k=1}^{K} \theta_k$. Hence,

511 for the transformation we will only consider the first $K-1$ elements of $\boldsymbol{\theta}$ and we will

512 transform them to $K-1$ elements of a new vector $\boldsymbol{y} \in \mathbb{R}^{K-1}$.

513   Let ϕ denote the density of a normal variable with a mean of zero and a variance of

514 one, Φ denote its cumulative density function, and Φ^{-1} denote the inverse cumulative density

515 function. Then for each element θ_k, the transformation is

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

516 The inverse transformation is given by

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

517 The Jacobian of this transformation is:

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

518   To perform the transformations, we thus need to determine the lower bound l_k and the

519 upper bound u_k of each θ_k. Assuming $\theta_{k-1} < \theta_k$ for $k \in \{1 \cdots, K\}$ the lower bound for any

520 element in $\boldsymbol{\theta}$ is defined as

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

521   This definition holds for both binomial models and multinomial models. Differences in

522 these two models appear only when determining the upper bound for each parameter, since

523 parameters in a multinomial models are unit constrained. For binomial models, the upper

524 bound for each θ_k is simply 1. For multinomial models, however, the upper bound for each

525 θ_k depends on the size of smaller elements as well as on the number of remaining larger

526 elements in $\boldsymbol{\theta}$. To determine the upper bound for multinomial parameters we are using a

527 stick-breaking method (Frigyik, Kapila, & Gupta, 2010; Stan Development Team, 2020).

528 The stick-breaking approach represents $\boldsymbol{\theta}$ as unit-length stick which we subsequently divide

529 into K elements. By this definition, the upper bound or any θ_k is:

$$u_k = \begin{cases} \dfrac{1}{K} & \text{if } k = 1 \\[2ex] \dfrac{1 - \sum_{i<k} \theta_i}{ERS} & \text{if } 1 < k < K, \end{cases} \tag{12}$$

530 where $1 - \sum_{i<k} \theta_i$ represents the length of the remaining stick, that is, the proportion of the

531 unit-length stick that still needs to be divided among the remaining elements in $\boldsymbol{\theta}$. The

532 elements in the remaining stick are denoted as ERS, and are computed as follows:

$$ERS = K - 1 + k$$

533 .

534   The transformations outlined above are suitable for binomial and multinomial models

535 featuring hypotheses in which all parameters are inequality constrained. However, when

536 hypotheses feature a combination of equality and inequality constrained parameters, as well

537 as parameters that are free to vary we need to modify the formula to compute the upper and

538 lower bounds:

$$l_k = \begin{cases} 0 & \text{if } k = 1 \\[2ex] \frac{\theta_{k-1}}{e_{k-1}} \times e_k & \text{if } 1 < k < K, \end{cases} \tag{13}$$

539 where e_{k-1} refers to the number of equality constrained parameters that are collapsed in θ_{k-1}.

540   The upper bound for parameters in the binomial models still remains 1. For

541 multinomial models, the upper bound is then defined as:

$$
u_k = \begin{cases}
\dfrac{1}{K} - (f_k \times l_k) & \text{if } k = 1 \\[2ex]
\left(\dfrac{1 - \sum_{i<k} \theta_i}{ERS} - (f_k \times l_k) \right) \times e_k & \text{if } 1 < k < K \text{ and } u_k \geq \max(\theta_{i<k}), \\[3ex]
\left(\left(2 \times \left(\dfrac{1 - \sum_{i<k} \theta_i}{ERS} - (f_k \times l_k) \right) - \max(\theta_{i<k}) \right) \times e_k \right. & \text{if } 1 < k < K \text{ and } u_k < \max(\theta_{i<k}),
\end{cases}
$$

(14)

542 where f_k represents the number of free parameters that share common upper and lower

543 bounds. The elements in the remaining stick are then computed as follows

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

544 .The rationale behind these modifications will be described in more detail in the following

545 sections. In **multibridge** information that is relevant for the transformation of the

546 parameter vectors is stored in the generated `restriction_list` which is returned by the

547 main functions `binomBfInformed` and `multBfInformed` but can also be generated

548 separately with the function `generateRestrictionList`. This restriction list features the

549 sublist `inequality_constraints` which encodes the number of equality constraints

550 collapsed in each parameter in `nr_mult_equal`. Similarly the number of free parameters

551 that share a common bounds are encoded under `nr_mult_free`.

    Equality Constrained Parameters.   When informed hypotheses feature a mix of

equality and inequality constrained parameters, we collapse in the constrained prior and

posterior distributions all equality constrained parameters into one category. When

transforming the samples from these distributions, we need to account for the fact that

inequality constraints on the collapsed parameters might not hold even though the constraint is valid under the original parameter values. For instance, for $\theta_1 = \theta_2 = \theta_3 \leq \theta_4 \leq \theta_5$, where the elements in $\boldsymbol{\theta}$ take the values $(0.15, 0.15, 0.15, 0.25, 0.3)$, the inequality constraint does not hold for the collapsed parameters (i.e., $\theta^* \not\leq \theta_4 \leq \theta_5$ since $0.45 \not\leq 0.25 \leq 0.3$). For these cases, the upper and lower bounds for the parameters need to be adjusted as follows:

$$
l_k = \begin{cases} 0 & \text{if } k = 1 \\[2mm] \dfrac{\theta_{k-1}}{e_{k-1}} \times e_k & \text{if } 1 < k < K, \end{cases} \tag{15}
$$

where e_{k-1} and e_k refer to the number of equality constrained parameters that are collapsed in θ_{k-1} and θ_k, respectively. The upper bound is defined as

$$
u_k = \begin{cases} \dfrac{1}{ERS} \times e_k & \text{if } k = 1 \\[3mm] \dfrac{1 - \sum_{i<k} \theta_i}{ERS} \times e_k & \text{if } 1 < k < K, \end{cases} \tag{16}
$$

where $1 - \sum_{i<k} \theta_i$ represents the length of the remaining stick and the number of elements in the remaining stick are computed as follows: $ERS = \sum_k^K e_k$. The upper bound is then multiplied by the number of equality constrained parameters within the current constraint.

Concretely, for the constraint above, that is $\theta^* \leq \theta_4$, the lower bound for θ^* would be 0. The upper bound is computed by taking account the number of equality constrained parameters, such that $u_k = {}^1\!/_5 \times 3 = 0.6$. For θ_4 the lower bound is ${}^{\theta^*}\!/_3 = 0.15$, since 3 parameters are collapsed in θ^*. The upper bound for θ_4 is then $\dfrac{(1 - \theta^*)}{2} = 0.275$ and θ_5 is $1 - \theta^* - \theta_4 = 1 - 0.45 - 0.25 = 0.3$.

Corrections for Free Parameters. Different adjustments are required for a sequence of inequality constraint parameters that have shared upper and lower bounds, but can vary freely within certain upper and lower bounds. For instance, the hypothesis

$$
\mathcal{H}_r : \theta_1 < \theta_2, \theta_3
$$

specifies that θ_2 and θ_3 have the shared lower bound θ_1 and the shared upper bound 1, however, θ_2 can be larger than θ_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, however the upper bounds need to be adjusted as follows:

$$u_k = \begin{cases} \dfrac{1}{K} - (f_k \times l_k) & \text{if } k = 1 \\ \dfrac{1 - \sum_{i<k} \theta_i}{ERS} - (f_k \times l_k) & \text{if } 1 < k < K, \end{cases} \tag{17}$$

where f_k represents the number of free parameters that share common upper and lower bounds. Here, the number of elements in the remaining stick are computed as follows: $ERS = 1 + \sum_{j>k} f_j$. Subtracting the lower bound for the remaining free parameters from the upper bound of the current parameter secures a minimum stick-length for the remaining free parameters to comply with the constraint. A further correction is required, if a preceding free parameter (i.e., a free parameter that was already accounted for in the stick) is larger than the upper bound of the current parameter. In that case, we need we subtract the difference between the largest preceding free parameter in the sequence with the current upper bound. Thus if $u_k < \max(\theta_{i<k})$, the upper bound becomes:

$$u_k = u_k - (\max(\theta_{i<k}) - u_k) \tag{18}$$

$$= 2 \times u_k - \max(\theta_{i<k}). \tag{19}$$

To outline when such a correction is necessary, consider the constraint $\theta_1 \leq \theta_2, \theta_3 \leq \theta_4$, where the elements in $\boldsymbol{\theta}$ take on the values $(0.1, 0.35, 0.15, 0.40)$. When transforming the parameters, the lower bound for θ_1 is 0, the upper bound $^1/_4$. The parameters θ_2 and θ_3 share the same lower bound, which is, $\theta_1 = 0.1$. The upper bound for θ_2, is the length of the remaining stick divided by the elements of the remaining stick, that is, $^{0.9}/_2 = 0.45$. From the resulting upper bound, we subtract the lower bound for the remaining free parameters of the

sequence, which yields an upper bound for θ_2 of $0.45 - 0.1 = 0.35$. Since θ_2 is the first free parameter in the sequence that is evaluated an additional downward correction is not necessary. The upper bound for θ_3 is $(1 - 0.1 - 0.35)/2 = 0.275$. However, if θ_3 would actually take on the value 0.275, θ_4 would need to be 0.275 too, which would violate the constraint (i.e., $0.1 \leq 0.35, 0.275 \not\leq 0.275$). Therefore, we apply the additional correction, such that

$$u_k = 2 \times u_k - \max(\theta_{i<k}) \tag{20}$$

$$= 2 \times 0.275 - 0.35 \tag{21}$$

$$= 0.2, \tag{22}$$

which secures the proper ordering for the remainder of the parameters, since $\theta_4 = 0.2$ would yield $0.1 \leq 0.35, 0.2 \leq 0.35$.

Table 1

To estimate the Bayes factor in favor for or against the specified informed hypothesis, the user provides the core functions `multBfInformed` *and* `binomBfInformed` *with the following basic required arguments*

Argument	Description
`x`	a 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)
`n`	numeric. Vector of counts of trials. Must be the same length as `x`. Ignored if `x` is a matrix or a table
`Hr`	string or character. Encodes the user specified informed hypothesis. Users can either use the specified `factor_levels` or indexes to refer to parameters.
`a`	numeric. Vector with concentration parameters of Dirichlet distribution (for multinomial models) or alpha parameters for independent beta distributions (for binomial models). Default sets all parameters to 1
`b`	numeric. Vector with beta parameters. Must be the same length as `x`. Default sets all beta parameters to 1
`factor_levels`	character. Vector with category names. Must be the same length as `x`

Table 2

*S3 methods implemented in **multibridge***

Function Name(s)	S3 Method	Description
`multBfInformed`, `binomBfInformed`	`print`	Prints model specifications and descriptives.
	`summary`	Prints and returns the Bayes factor and associated hypotheses for the full model, and all equality and inequality constraints.
	`plot`	Plots the posterior median and credible interval of the parameter estimates of the encompassing model.
	`bayes_factor`	Contains all Bayes factors and log marginal likelihood estimates for inequality constraints.
	`samples`	Extracts prior and posterior samples from constrained distribution (if bridge sampling was applied).
	`bridge_output`	Extracts bridge sampling output and associated error measures.
	`restriction_list`	Extracts restriction list and associated informed hypothesis.
`binomBfInequality`, `binomBfInequality`	`print`	Prints the bridge sampling estimate for the log marginal likelihood and the corresponding percentage error.
	`summary`	Prints and returns the bridge sampling estimate for the log marginal likelihood and associated error terms.

Table 3

Core functions implemented in **multibridge**

Function Name(s)	Description
`multBfInformed`	Evaluates informed hypotheses on multinomial parameters.
`multBfInequality`	Estimates the marginal likelihood of a constrained prior or posterior Dirichlet distribution.
`multBfEquality`	Computes Bayes factor for equality constrained multinomial parameters using the standard Bayesian multinomial test.
`multTruncatedSampling`	Samples from truncated prior or posterior Dirichlet density.
`lifestresses`, `peas`	Datasets associated with informed hypotheses in multinomial models.
`binomBfInformed`	Evaluates informed hypotheses on binomial parameters.
`binomBfInequality`	Estimates the marginal likelihood of constrained prior or posterior beta distributions.
`binomBfEquality`	Computes Bayes factor for equality constrained binomial parameters.
`binomTruncatedSampling`	Samples from truncated prior or posterior beta densities.
`journals`	Dataset associated with informed hypotheses in binomial models.
`generateRestrictionList`	Encodes the informed hypothesis.

Table 4

The Table shows the Observed Counts, Observed Proportions, and Expected Proportions of first digits in Greece governmental data. The total sample size was $N = 1497$ 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