British Journal of Mathematical and Statistical Psychology (2018), 71, 229–261

The British Psychological Society
www.wileyonlinelibrary.com

Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses

Xin Gu[1,2]*, Joris Mulder[3] and Herbert Hoijtink[1,4]

[1]Department of Methodology and Statistics, Utrecht University, The Netherlands
[2]Department of Geography and Planning, University of Liverpool, UK
[3]Department of Methodology and Statistics, Tilburg University, The Netherlands
[4]CITO Institute for Educational Measurement, Arnhem, The Netherlands

Informative hypotheses are increasingly being used in psychological sciences because they adequately capture researchers' theories and expectations. In the Bayesian framework, the evaluation of informative hypotheses often makes use of default Bayes factors such as the fractional Bayes factor. This paper approximates and adjusts the fractional Bayes factor such that it can be used to evaluate informative hypotheses in general statistical models. In the fractional Bayes factor a fraction parameter must be specified which controls the amount of information in the data used for specifying an implicit prior. The remaining fraction is used for testing the informative hypotheses. We discuss different choices of this parameter and present a scheme for setting it. Furthermore, a software package is described which computes the approximated adjusted fractional Bayes factor. Using this software package, psychological researchers can evaluate informative hypotheses by means of Bayes factors in an easy manner. Two empirical examples are used to illustrate the procedure.

1. Introduction

One of the objectives of psychological studies is to test hypotheses that represent scientific expectations. The main tool available for this purpose is null hypothesis significance testing where the goal is to falsify a null hypothesis of 'no effect'. On the other hand, psychologists may expect, for example, that the learning ability of children is stronger than the learning ability of adolescents, which in turn is stronger than the learning ability of adults, or it is expected that a patient's psychological disease would decrease after the first therapy, and decrease further after subsequent therapies. These expectations cannot be formulated by the traditional null hypothesis. Instead, such expectations can be translated to so-called informative hypotheses which assume a specific structure of the model parameters (Hoijtink, 2012). An informative hypothesis consists of equality and/or inequality constraints among the parameters of interest in a statistical model. For example, three equal parameters can be represented by an equality constrained hypothesis $H_1 : \theta_1 = \theta_2 = \theta_3$, and three ordered parameters can be represented by an inequality constrained hypothesis $H_2 : \theta_1 < \theta_2 < \theta_3$. This class of informative

*Corresponding should be addressed to Xin Gu, Department of Geography and Planning, University of Liverpool, Roxby Building, Liverpool L69 7ZT, UK (email: xin.gu@liverpool.ac.uk).

DOI:10.1111/bmsp.12110

hypotheses covers a much broader range of scientific expectations than the class of standard null hypotheses. In addition, by testing competing informative hypotheses directly against each other a researcher obtains a direct answer as to which scientific theory is most supported by the data. The interested reader is referred to http://informative-hypotheses.sites.uu.nl/ for an overview of psychological research in which informative hypotheses were used.

Informative hypothesis testing has drawn a lot of attention both in frequentist statistics (Barlow, Bartholomew, Bremner, & Brunk, 1972; Silvapulle & Sen, 2004) and in Bayesian statistics (Hoijtink, 2012). In the frequentist framework, hypothesis testing with inequality constraints has been studied for over 50 years, starting with (Bartholomew, 1959). Some recent contributions can be found in van de Schoot, Hoijtink, and Deković (2010), and Klugkist, Bullens, and Postma (2012). Bayesian evaluation of informative hypotheses by means of the Bayes factor is relatively new. A decade ago, Klugkist, Laudy, and Hoijtink (2005) started using Bayes factors to evaluate inequality constrained hypotheses in ANOVA models. Follow-up research appeared in Klugkist and Hoijtink (2007) for Bayesian testing of inequality and about equality constrained hypotheses, in Mulder, Klugkist, van de Schoot, Meeus, Selfhout, and Hoijtink (2009) for Bayesian informative hypothesis testing in repeated measures models, in Klugkist, Laudy, and Hoijtink (2010) for Bayesian evaluation of equality and inequality constrained hypotheses in contingency tables, and in Mulder, Hoijtink, and Klugkist (2010) for Bayesian model selection of equality and inequality constrained hypotheses in the context of multivariate normal linear models. Developments in the use of Bayes factors for informative hypothesis testing are summarized in Hoijtink (2012). However, these studies are limited to assessing informative hypotheses in specific models and cannot yet be applied in other models, such as confirmatory factor analysis or logistic regression. More recently, van de Schoot, Hoijtink, Hallquist, and Boelen (2012) have enabled researchers to test inequality constrained hypotheses in structural equation models, Gu, Mulder, Deković, and Hoijtink (2014) have shown how to evaluate inequality constrained hypothesis in general statistical models, and Böing-Messing, van Assen, Hofman, Hoijtink, and Mulder (2017) have enabled researchers to test informative hypotheses on group variances. Furthermore, the usefulness of the Bayes factor for testing hypotheses in psychological research has been highlighted in various studies in a special issue on the topic (Mulder & Wagenmakers, 2016). Although these studies enable hypothesis testing in a large number of statistical models using the Bayes factor, the available methods for testing hypotheses with both equality constraints and inequalities are still limited.

The incessant debate between frequentist and Bayesian hypothesis testing (Wagenmakers, 2007) has highlighted an advantage of the Bayes factor: it quantifies the relative support in the data for one hypothesis against another (Kass & Raftery, 1995). This cannot be done using classical p-values. Psychological researchers can quantify how much the data favour a hypothesis relative to another hypothesis by means of the Bayes factor. However, the popularity of the Bayes factor is limited for two reasons: the specification of the prior can be a difficult task, especially when prior information is weak or completely unavailable; and the computation can be very intensive when the statistical model is complex. To overcome these barriers, Bayesian statisticians have presented several default Bayes factors based on default priors. Default priors usually do not reflect subjective prior beliefs and have distributional forms chosen such that the Bayes factor can easily be computed. Examples of default Bayes factors are the JZS Bayes factor (Jeffreys, 1961; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Zellner & Siow, 1980), partial Bayes factors (de Santis & Spezzaferri, 1999), the Bayes factor based on expected posterior priors (Pérez & Berger, 2002), the intrinsic Bayes factor (Berger & Pericchi, 1996) and the fractional Bayes factor (O'Hagan, 1995). The last two Bayes factors are closely related to the partial Bayes factor.

In the partial Bayes factor the data are split into two parts: one part is used as a training sample to update an improper non-informative prior distribution, and the remaining part is used to compute the Bayes factor. The training sample is proper if it renders a proper updated prior. Furthermore, the training sample is called minimal if any of its subsets is not proper (Berger & Pericchi, 2004). Both the intrinsic Bayes factor and the fractional Bayes factor use the partial Bayes factor method (de Santis & Spezzaferri, 1997, 1999). The intrinsic Bayes factor is an average of the partial Bayes factors based on all possible minimal training samples. Because of the use of all possible minimal training samples, the computation of the intrinsic Bayes factor can be intensive especially when the sample size and the size of the minimal training sample are large. Alternatively, the fractional Bayes factor takes a small fraction b of the likelihood of the complete data (O'Hagan, 1995). The updated proper prior in the fractional Bayes approach is then implicitly specified from a non-informative prior and a fraction of full likelihood (de Santis & Spezzaferri, 1999; Gilks, 1995; Moreno, 1997; Mulder, 2014b). In this paper, we shall refer to updated priors following from the fractional Bayes methodology as fractional priors. The remaining fraction of the likelihood is then used for testing the hypotheses of interest. As will be shown in this paper, the fractional Bayes factor is computationally easy. Recently, Fouskakis, Ntzoufras, and Draper (2015) presented power expected posterior priors, which are similar to fractional priors in the sense that both of them are specified using a fraction of a likelihood function. The main difference is that the fractional prior comes from a fraction of the likelihood of the observed data, whereas the power expected posterior prior follows from a fraction of the likelihood of imaginary training data coming from a prior predictive distribution.

In this paper we focus on the fractional Bayes factor as it stands out for its convenience of evaluating informative hypotheses (Mulder, 2014b). Recently, Mulder (2014b) proposed an adjustment of the fractional Bayes factor where the fractional prior was shifted around the null value. This approach resulted in an adjusted fractional Bayes factor that converges faster to a true inequality constrained hypothesis. However, the current applications of (adjusted) fractional Bayes factors in informative hypothesis testing are still within the class of multivariate normal linear models.

This paper proposes an approximation of a fractional Bayes factor to extend its applicability to testing informative hypotheses for more general models. These models can be generalized linear (mixed) models (McCullogh & Searle, 2001) such as logistic regression models and multilevel models, and structural equation models (Kline, 2011) such as path models, confirmatory factor analysis models and latent class models. Due to large-sample theory (Gelman, Carlin, Stern, & Rubin, 2004, pp. 101–107), the posterior distribution of the parameters in each model can be approximated by a (multivariate) normal distribution. This paper also approximates the implicit fractional prior with a (multivariate) normal distribution as a general methodology to ensure a fast computation of the (adjusted) fractional Bayesian factor. Based on these approximations, we can approximate a fractional Bayes factor to evaluate informative hypotheses in general statistical models. In addition, we discuss different choices of the fraction (Gu, Mulder, & Hoijtink, 2016; O'Hagan, 1995), which is a tuning parameter in the fractional prior, and provide a guideline for choosing this fraction. Furthermore, an important issue in Bayesian hypothesis testing is the consistency of the Bayesian procedure. Previous studies have discussed the consistency of the intrinsic Bayes factor (Casella, Giron, & Moreno, 2009), the fractional Bayes factor (de Santis & Spezzaferri, 2001; O'Hagan, 1997), and posterior model probabilities (Moreno, Giron, & Casella, 2015). In this paper, the consistency of the approximate adjusted fractional Bayes factor (AAFBF) will be elaborated and illustrated.

This paper is organized as follows. Section 2 introduces the informative hypothesis in general statistical models, and illustrates how the informative hypothesis is constructed based on researchers' expectations by means of two empirical examples. Section 3 elaborates the specification of the adjusted fractional prior and the posterior distribution using normal approximations. Based on the specified prior and posterior distributions, the AAFBF is derived and a software package is presented for the evaluation of informative hypotheses in general statistical models. In Section 4 we discuss different choices of the fraction, and conduct a sensitivity study for the fractional Bayes factors with those choices. Section 5 revisits the two empirical examples to show how to evaluate informative hypotheses using the proposed fractional Bayes factors. Section 6 concludes.

2. Informative hypotheses in general statistical models

A statistical model is described by the likelihood function $f(\mathbf{X}|\boldsymbol{\theta}, \boldsymbol{\zeta})$, where $\mathbf{X}$ denotes the data, $\boldsymbol{\theta}$ contains the parameters that are used to specify informative hypotheses, and $\boldsymbol{\zeta}$ contains the nuisance parameters. Informative hypotheses are constructed using equality and/or inequality constraints based on the theories or expectations of researchers. The general form of the informative hypothesis is given by

$$H_i : \mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}, \mathbf{R}_{i_1}\boldsymbol{\theta} > \mathbf{r}_{i_1}, \tag{1}$$

where $\mathbf{R}_{i_0}$ and $\mathbf{R}_{i_1}$ are the restriction matrices for equality and inequality constraints in H_i, respectively, and $\mathbf{r}_{i_0}$ and $\mathbf{r}_{i_1}$ contain constants. Note that the number of rows in $\mathbf{R}_{i_0}$ equals the number of equality constraints, the number of rows in $\mathbf{R}_{i_1}$ equals the number of inequality constraints, and the numbers of columns in $\mathbf{R}_{i_0}$ and $\mathbf{R}_{i_1}$ equal the length of $\boldsymbol{\theta}$.

For example, hypothesis $H_1 : \theta_1 = 2\theta_2 = 3\theta_3 > 4\theta_4 < 5$ corresponds to

$$\mathbf{R}_{1_0}\boldsymbol{\theta} = \begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 2 & -3 & 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{r}_{1_0},$$

$$\mathbf{R}_{1_1}\boldsymbol{\theta} = \begin{bmatrix} 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} > \begin{bmatrix} 0 \\ -5 \end{bmatrix} = \mathbf{r}_{1_1}.$$

Note that a range constraint, in which the parameters of interest are constrained between two values, can be written as two inequality constraints. For example, hypothesis $H_2 : 0 < \theta < 1$ can be expressed by $H_2 : \mathbf{R}_{2_1}\theta > \mathbf{r}_{2_1}$, where $\mathbf{R}_{2_1} = (1, -1)^T$ and $\mathbf{r}_{2_1} = (0, -1)^T$. This hypothesis can be seen as one where it is expected that θ is approximately equal to 0.5 with maximal deviation of 0.5, that is, $\theta \approx 0.5 \Leftrightarrow |\theta - 0.5| < 0.5$, where the maximal deviation of 0.5 should be specified subjectively by the user.

An informative hypothesis H_i can be tested against the unconstrained hypothesis

$$H_u : \boldsymbol{\theta} \text{ is unconstrained,} \tag{2}$$

against its complement

$$H_{i_c} : \text{not } H_i, \tag{3}$$

which expresses what a researcher does not expect, or against another informative hypothesis

$$H_{i'} : \mathbf{R}_{i'_0}\boldsymbol{\theta} = \mathbf{r}_{i'_0}, \mathbf{R}_{i'_1}\boldsymbol{\theta} > \mathbf{r}_{i'_1}. \tag{4}$$

It should be noted that when an informative hypothesis H_i contains at least one equality constraint, the complement of H_i is the same as the unconstrained hypothesis H_u.

Before evaluating the informative hypotheses, the parameters of interest may need to be standardized in some situations. The need for standardization depends on the statistical model and informative hypothesis under evaluation. On the one hand, the parameters have to be standardized when comparing, for example, coefficients in regression models and factor loadings in confirmatory factor analysis. For example, testing whether the regression coefficient θ_1 is larger than θ_2 requires the standardization of θ_1 and θ_2, because a large coefficient can also result from a large scale of the corresponding predictor. On the other hand, it may not be necessary to standardize the parameters $\boldsymbol{\theta}$ if they are compared to constants, and it is undesirable to standardize the parameters $\boldsymbol{\theta}$ if they represent means. For instance, testing whether a regression coefficient is larger than 0 or testing whether the mean of group 1 is smaller than the mean of group 2 does not require standardization. If standardization is required, Gu *et al.* (2014) discussed two ways to do this: (1) standardize all observed and latent variables, or (2) use standardized parameters. In the situation considered by Gu *et al.* (2014), there was little difference between the performances of the two methods. Therefore, researchers can use either of them if necessary.

In what follows, we will use two empirical examples to illustrate how researchers' expectations can be expressed by informative hypotheses.

2.1. Example 1: Multiple regression

The first example concerns a multiple regression model used in Guber (1999) to investigate the relation between the educational costs of a school and the academic performance of the students. The data were collected in 50 US states (available at www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm). The performance of the students is measured by the average total SAT score y_i, ranging from 400 to 1,600. Its predictors are the average public school expenditure x_{1i}, the percentage of students taking the SAT exams x_{2i}, and the average pupil–teacher ratio x_{3i}. The descriptives for the dependent variable y_i and independent variables x_{1i}, x_{2i} and x_{3i} are shown in Table 1. The relationship between student performance and its predictors is given in a regression model,

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i} + \epsilon_i, \tag{5}$$

Table 1. Descriptives for variables in regression model

	y_i	x_{1i}	x_{2i}	x_{3i}
Mean	965.92	5.91	35.24	16.86
Standard deviation	74.82	1.36	26.76	2.27

where θ_0 is the intercept, θ_1, θ_2 and θ_3 are the regression coefficients, and $\varepsilon_i \sim N(0, \sigma^2)$, denotes the residuals, with σ^2 being their residual variance. For this regression model, the likelihood is

$$f(\mathbf{X}|\boldsymbol{\theta}, \zeta) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \theta_3 x_3)^2\right\}, \tag{6}$$

where $n = 50$ denotes the sample size, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ and $\boldsymbol{\zeta} = (\theta_0, \sigma^2)$.

Guber (1999) theorized that higher education expenditures result in better student performance in SAT exams, which implies that the coefficient θ_1 of the predictor x_{1i} is positive. In addition, in those states with a small percentage of students taking SATs, the students are expected to do well because they have self-selected into the SAT exam, which is only required by universities with high prestige. This implies that the coefficient θ_2 of the predictor x_{2i} is negative. Furthermore, although a lower pupil–teacher ratio would be associated with better performance, a school needs to spend more money on education and therefore this predictor overlaps with the expenditures. This suggests that the coefficient θ_3 of predictor x_{3i} is zero. Consequently, we specify the informative hypothesis

$$H_1 : \theta_1 > 0,\ \theta_2 < 0,\ \theta_3 = 0, \tag{7}$$

with $\mathbf{R}_{1_0} = (0, 0, 1)$, $\mathbf{R}_{1_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$, $r_{1_0} = 0$, and $\mathbf{r}_{1_1} = (0, 0)^T$ in $H_1 : \mathbf{R}_{1_0}\boldsymbol{\theta} = r_{1_0}, \mathbf{R}_{1_1}\boldsymbol{\theta} > \mathbf{r}_{1_1}$. Hypothesis H_1 can be tested against its complement

$$H_{1_c} : \text{not}\, H_1. \tag{8}$$

2.2. Example 2: Repeated measures ANOVA

We reanalyse the example of the repeated measures ANOVA used in Howell (2012, p. 462) based on an experiment with relaxation therapy. The experiment investigated the duration of nine patients' migraine headaches before and after relaxation training. The duration of headaches is measured by the number of hours per week. Our example uses the data for the last 2 weeks of the baseline where patients received no training and the last 2 weeks of training. Therefore, the data shown in Table 2 consists of four dependent variables: the number of hours with a headache per week for nine patients in 4 weeks. The random effects model for these dependent variables is (Hox, 2010, p. 83).

$$y_{ij} = \mu + \eta_i + \tau_j + \epsilon_{ij}, \tag{9}$$

Table 2. Data in repeated measures ANOVA

	Baseline		Training	
Subject	Week 1	Week 2	Week 3	Week 4
1	21	22	6	6
2	20	19	4	4
3	17	15	4	5
4	25	30	12	17
5	30	27	8	6
6	19	27	7	4
7	26	16	2	5
8	17	18	1	5
9	26	24	8	9

where y_{ij}, for i = 1, ... , 9 and j = 1, ... , 4, denotes the four dependent variables, μ denotes the grand mean, $\eta_i \sim N(0, \sigma_\eta^2)$ denotes the random difference for person i which is constant for different j, τ_j denotes the fixed measurement difference for week j which is constant for different i, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is the measurement error with respect to person i and week j. To investigate the effect of relaxation training, we specify the individual differences with a random effect and the treatment differences with a fixed effect. Thus, the mean for each measurement is

$$\theta_j = \mu + \tau_j \tag{10}$$

and $\Sigma_{j=1}^4 \tau_j = 0$.

The researchers expected a reduction of the duration of headaches after relaxation training. Furthermore, it is reasonable to expect that the mean durations are equal in the first 2 weeks of baseline and in the last 2 weeks of training to ensure that other factors do not influence the duration of headaches. These expectations can be expressed by the informative hypothesis

$$H_2 : \theta_1 = \theta_2 > \theta_3 = \theta_4, \tag{11}$$

with $\mathbf{R}_{2_0} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$, $\mathbf{R}_{2_1} = [0, 1, -1, 0]$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$, $\mathbf{r}_{2_0} = (0, 0)^T$, and $r_{2_1} = 0$ in $H_2 : \mathbf{R}_{2_0}\boldsymbol{\theta} = r_{2_0}, \mathbf{R}_{2_1}\boldsymbol{\theta} > \mathbf{r}_{2_1}$. We compare this hypothesis to another informative hypothesis that the mean number of headache hours continually declines in the 4 weeks:

$$H_{2'} : \theta_1 > \theta_2 > \theta_3 > \theta_4, \tag{12}$$

which only contains inequality constraints $\mathbf{R}_{2'_1}\boldsymbol{\theta} > \mathbf{r}_{2'_1}$ with $\mathbf{r}_{2'_1} = (0, 0, 0)^T$ and

$$\mathbf{R}_{2'_1} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

The informative hypotheses constructed in these examples can be evaluated using Bayes factors, which will be elaborated in the next section. We will revisit these examples in Section 5 to display the results of the evaluation of these informative hypotheses.

3. Approximated adjusted fractional Bayes factors

The Bayes factor is the cornerstone of Bayesian hypothesis testing. It quantifies the relative evidence in the data for one hypothesis against another. The Bayes factor of an informative hypothesis H_i against another informative hypothesis $H_{i'}$ is defined by their marginal likelihood ratio (Jeffreys, 1961; Kass & Raftery, 1995):

$$\mathrm{BF}_{ii'} = \frac{m(\mathbf{X}|H_i)}{m(\mathbf{X}|H_{i'})}. \tag{13}$$

In Bayesian hypothesis testing, the Bayes factor has a direct interpretation as the relative evidence from the data for one hypothesis against another. If $\mathrm{BF}_{ii'} > 1$ ($\mathrm{BF}_{ii'} < 1$), this implies that hypothesis H_i ($H_{i'}$) receives more support from the data. Specifically, if $\mathrm{BF}_{ii'} = 5$, then the support for H_i is five times larger than for $H_{i'}$. For researchers who are new to Bayes factors we recommend using the guidelines for their interpretation as provided by Kass and Raftery (1995). The degree of evidence in favour of H_i can be classified as unconvincing for $1 < \mathrm{BF}_{ii'} < 3$, positive for $\mathrm{BF}_{ii'} > 3$, strong for $\mathrm{BF}_{ii'} > 20$, and very strong for $\mathrm{BF}_{ii'} > 150$. However, these rules for interpreting Bayes factors are not strict and can differ in different contexts.

The informative hypothesis H_i is nested in the unconstrained hypothesis H_u which does not contain any constraints on $\boldsymbol{\theta}$. When comparing H_i to H_u we can use the encompassing prior approach of Klugkist *et al.* (2005) where a prior is constructed under H_i via a truncation of the unconstrained (or encompassing) prior $\pi_u(\boldsymbol{\theta}, \zeta)$ under H_u. The prior under H_i is then given by $\pi_i(\boldsymbol{\theta}, \zeta) = c_i^{-1}\pi_u(\theta, \zeta)\mathbf{1}_{\boldsymbol{\Theta}_i}(\boldsymbol{\theta})$, where $c_i = \iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}, \zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta$ is a normalizing constraint, and $\boldsymbol{\Theta}_i = \{\boldsymbol{\theta}|\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}, \mathbf{R}_{i_1}\boldsymbol{\theta} > \mathbf{r}_{i_1}\}$ is the parameter space of $\boldsymbol{\theta}$ in agreement with the informative hypothesis H_i. Consequently, the Bayes factor for the informative hypothesis against the unconstrained hypothesis can be expressed as

$$\begin{aligned}
\mathrm{BF}_{iu} &= \frac{\iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_i(\boldsymbol{\theta}, \zeta) f(\mathbf{X}|\theta, \zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}{\iint \pi_u(\boldsymbol{\theta}, \zeta) f(\mathbf{X}|\boldsymbol{\theta}, \zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta} \\
&= \iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \frac{\pi_u(\boldsymbol{\theta}, \zeta) f(\mathbf{X}|\boldsymbol{\theta}, \zeta) \cdot c_i^{-1}}{\iint \pi_u(\boldsymbol{\theta}, \zeta) f(\mathbf{X}|\boldsymbol{\theta}, \zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta} \mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta \\
&= c_i^{-1} \iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}, \zeta|\mathbf{X})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta \\
&= \frac{\iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}, \zeta|\mathbf{X})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}{\iint_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}, \zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta},
\end{aligned} \tag{14}$$

where $\pi_u(\boldsymbol{\theta}, \zeta|\mathbf{X})$ is the posterior distribution of $\boldsymbol{\theta}$ and ζ under H_u. For example, for hypothesis H_1: $\theta_1 > 0$, $\theta_2 < 0$, $\theta_3 = 0$ in (7) with equality and inequality constraints, where we denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ and $\zeta = (\theta_0, \sigma^2)^T$, the Bayes factor of H_1 against the unconstrained alternative in (14) comes down to

$$\begin{aligned}\mathrm{BF}_{1u} &= \frac{\iint_{\theta_1>0,\theta_2>0}\pi_u((\theta_1,\theta_2,0)^T,\zeta|\mathbf{X})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}{\iint_{\theta_1>0,\theta_2>0}\pi_u((\theta_1,\theta_2,0)^T,\zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}\\ &= \frac{\int_{\theta_1>0,\theta_2>0}\pi_u(\theta_1,\theta_2|\theta_3=0,\mathbf{X})\pi_u(\theta_3=0|\mathbf{X})\mathrm{d}\boldsymbol{\theta}}{\int_{\theta_1>0,\theta_2>0}\pi_u(\theta_1,\theta_2|\theta_3=0)\pi_u(\theta_3=0)\mathrm{d}\boldsymbol{\theta}}\\ &= \frac{\Pr(\theta_1>0,\theta_2>0|\theta_3=0,\mathbf{X})}{\Pr(\theta_1>0,\theta_2>0|\theta_3=0)}\frac{\pi_u(\theta_3=0|\mathbf{X})}{\pi_u(\theta_3=0)}.\end{aligned} \tag{15}$$

Note further that for a hypothesis with only equality constraints, such as H_0: $\theta_1 = \theta_2 = \theta_3 = 0$, expression (14) is equal to the well-known Savage–Dickey density ratio (Dickey, 1971; Wetzels, Grasman, & Wagenmakers, 2010),

$$\mathrm{BF}_{0u} = \frac{\pi_u(\boldsymbol{\theta}=\mathbf{0}|\mathbf{X})}{\pi_u(\boldsymbol{\theta}=\mathbf{0})}. \tag{16}$$

Finally, for a hypothesis with only inequality constraints, say, H_2: $\theta_1 > \theta_2 > \theta_3 > 0$, expression (14) is equal to the ratio of posterior and prior probabilities that the inequality constraints hold under H_u,

$$\mathrm{BF}_{2u} = \frac{P(\theta_1>\theta_2>\theta_3>0|\mathbf{X})}{P(\theta_1>\theta_2>\theta_3>0)}. \tag{17}$$

Thus, in order to compute the Bayes factor the unconstrained prior and corresponding unconstrained posterior need to be determined, and subsequently the unconstrained prior and posterior need to be integrated over the constrained region under the informative hypothesis. In this section we propose a novel and general approach using normal distributions to approximate the unconstrained posterior and the unconstrained fractional prior to compute default Bayes factors.

3.1. Fractional prior and posterior

To avoid *ad hoc* or subjective specification of the unconstrained prior, we consider the approach of O'Hagan (1995), referred to as the fractional Bayes factor. A proper default prior is automatically generated by updating a non-informative improper prior $\pi_u^N(\boldsymbol{\theta},\zeta)$ using a fraction b of the likelihood (Gilks, 1995). In the fractional Bayes factor the marginal likelihood of the hypothesis H_u is defined by

$$\begin{aligned}m_b^N(\mathbf{X}|H_u) &= \frac{m^N(\mathbf{X}|H_u)}{m^N(\mathbf{X}^b|H_u)} = \frac{\iint\pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}{\iint\pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)^b\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}\\ &= \iint f(\mathbf{X}|\boldsymbol{\theta},\zeta)^{1-b}\frac{\pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)^b}{\iint\pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)^b\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta\\ &= \iint f(\mathbf{X}|\theta,\zeta)^{1-b}\pi_u(\boldsymbol{\theta},\zeta|\mathbf{X}^b)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta,\end{aligned} \tag{18}$$

where the proper default prior is defined by

$$\pi_u(\boldsymbol{\theta},\zeta|\mathbf{X}^b) = \frac{\pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)^b}{\iint \pi_u(\boldsymbol{\theta},\zeta)^N f(\mathbf{X}|\boldsymbol{\theta},\zeta)^b \mathrm{d}\boldsymbol{\theta}\mathrm{d}\zeta}. \tag{19}$$

We shall refer to (19) as the fractional prior. Note that the marginal likelihood in the fractional Bayes factor in (18) is closely related to the marginal likelihood in the partial Bayes factor, where a proper default prior is obtained by training a non-informative prior with a small subset of the data, called a training sample, $\mathbf{X}(l)$, while the remaining part of the data, say, $\mathbf{X}(-l)$, is used for computing the marginal likelihood. The marginal likelihood in the fractional Bayes factor also follows this idea, but takes a fraction b of the data, denoted by $\mathbf{X}^b$, to train a non-informative prior and then uses the remaining fraction of the data, $\mathbf{X}^{1-b}$, for computing the marginal likelihood in (18). The advantage of the fractional Bayes factor is that it does not depend on the exact choice of the subset of the data because a fraction of the complete data is used (de Santis & Spezzaferri, 1999; O'Hagan, 1995).

Following similar steps to (14) and integrating out the nuisance parameters, the fractional Bayes factor of an informative hypothesis against the unconstrained hypothesis is given by (Mulder, 2014b)

$$\mathrm{FBF}_{iu} = \frac{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}|\mathbf{X})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}|\mathbf{X}^b)\mathrm{d}\boldsymbol{\theta}}. \tag{20}$$

3.2. Normal approximations of the fractional prior and posterior distributions

Due to large-sample theory (e.g., Gelman *et al.*, 2004, p. 101), the marginal posterior in the numerator of (20) can be approximated using a normal distribution where the mean is equal to the maximum likelihood estimate and the covariance matrix is equal to the inverse of the Fisher information matrix,

$$\pi_u(\boldsymbol{\theta}|\mathbf{X}) \approx N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_\theta), \tag{21}$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_\theta$ denote the maximum likelihood estimate and covariance matrix of $\boldsymbol{\theta}$, respectively. Note that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_\theta$ can be obtained using statistical software such as Mplus (Muthén & Muthén, 2010) or the R package lavaan (Rosseel, 2012). This will be further elaborated when we return to the empirical examples in Section 5.

The fractional prior in the denominator of (20) is also centred around the maximum likelihood estimate. However, it is based on a fraction b of the data, which implies an approximated covariance matrix of $\hat{\boldsymbol{\Sigma}}_\theta/b$. Consider, for example, a normally distributed data set $x_i \sim N(\theta, \sigma^2)$ with known σ^2. The posterior of θ is given by $\pi_u(\theta|X) = N(\hat{\theta}, \hat{\sigma}_\theta^2)$, where $\hat{\theta}$ equals the sample mean $\bar{x}$ and $\hat{\sigma}_\theta^2 = \sigma^2/n$. In this setting the fractional prior of θ would be $\pi_u(\theta|X^b) = N(\hat{\theta}, \hat{\sigma}_\theta^2/b) = N(\bar{x}, \sigma^2/nb)$. For this reason we propose to approximate the fractional prior according to

$$\pi_u(\boldsymbol{\theta}|\mathbf{X}^b) \approx N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_\theta/b). \tag{22}$$

3.3. Adjusting the prior mean

Various authors have suggested centring the prior distribution of $\boldsymbol{\theta}$ around the focal point of interest; see, for example, Zellner and Siow (1980) and Jeffreys (1961, pp. 268–274) for null hypothesis testing, and Mulder (2014b) for testing informative hypotheses. Suppose, for example, we evaluate H_1: $\theta \leq 0$ against its complement H_2: $\theta > 0$. By constructing the priors for θ under H_1 and H_2 as a truncation of an unconstrained prior that is centred around the focal point 0, the prior distributions for θ under both hypotheses are essentially equivalent; the only difference is the sign. Furthermore, by centring the prior at 0 it is assumed that small effects are more likely *a priori* than large effects, which is often the case in practice. A more detailed discussion on centring prior means can be found in Mulder (2014b). In this paper, we adjust the prior in (22) as follows:

$$\pi_u^*(\boldsymbol{\theta}|\mathbf{X}^b) = N(\boldsymbol{\theta}^*, \hat{\boldsymbol{\Sigma}}_\theta/b), \tag{23}$$

where the adjusted prior mean is given by $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_i^* = \{\boldsymbol{\theta}|\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}, \mathbf{R}_{i_1}\boldsymbol{\theta} = \mathbf{r}_{i_1}\}$. For each informative hypothesis, one can define a parameter space $\boldsymbol{\Theta}_i^*$ which contains one or more $\boldsymbol{\theta}$*. For example, $H_1 : \theta_1 > 2\theta_2 > 4$ results in $\boldsymbol{\theta}^* = (4, 2)^T$, and $H_2 : \theta_1 = \theta_2$ results in $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_i^* = \{\theta_1, \theta_2|\theta_1 = \theta_2\}$ in which $\theta_1^* = \theta_2^*$ can be any value. Note the suggestion that the prior mean for parameters in a range constrained hypothesis is in the middle of the range space (Mulder, Hoijtink, & de Leeuw, 2012), because a range constraint basically implies an approximate equality, which in terms of a restriction for the prior mean becomes an equality. For example, the range constraint $-0.2 < \theta < 0.2$ corresponds to the approximate equality $\theta \approx 0$ with maximal deviation of 0.2. Thus, the focal point is 0, and therefore we set the prior mean to $\theta^* = 0$. Below we will deal with the choice of $\boldsymbol{\theta}$*.

The prior distribution proposed in (23) depends on the informative hypothesis under evaluation, because the prior mean $\boldsymbol{\theta}$* is located on the boundary of the constrained region of the informative hypothesis. When two or more informative hypotheses are under comparison, the intersection of their constrained regions must be non-empty so that a common unconstrained prior mean $\boldsymbol{\theta}$* exists to evaluate all informative hypotheses against the unconstrained hypothesis. A set of informative hypotheses H_i, $i = 1, \ldots, I$, are comparable if there exists at least one solution of $\boldsymbol{\theta}$ to the set of equations

$$\begin{bmatrix}\mathbf{R}_{1_0}\\ \mathbf{R}_{1_1}\end{bmatrix}\boldsymbol{\theta} = \begin{bmatrix}\mathbf{r}_{1_0}\\ \mathbf{r}_{1_1}\end{bmatrix}, \ldots, \begin{bmatrix}\mathbf{R}_{I_0}\\ \mathbf{R}_{I_1}\end{bmatrix}\boldsymbol{\theta} = \begin{bmatrix}\mathbf{r}_{I_0}\\ \mathbf{r}_{I_1}\end{bmatrix} \tag{24}$$

(Mulder *et al.*, 2010). The solution of $\boldsymbol{\theta}$ for these equations defines the parameter space $\boldsymbol{\Theta}^*$. Examples of comparable hypotheses are H_1: $\theta = 0$ versus H_2: $\theta > 0$ and H_3: $\theta_1 > \theta_2 > \theta_3 > 0$ versus H_4: $\theta_3 > \theta_2 > \theta_1$. Hypotheses H_5: $\theta_1 = \theta_2$ versus H_6: $\theta_1 > \theta_2 + 1$ are not comparable because there is no solution of θ_1 and θ_2 for equations $\theta_1 = \theta_2$ and $\theta_1 = \theta_2 + 1$. It should be noted that the hypothesis H_7: $\theta_1 > 0, \theta_2 > 0, \theta_2 > \theta_1-1$ cannot be properly evaluated yet because a solution does not exist for equations $\theta_1 = 0, \theta_2 = 0$, and $\theta_2 = \theta_1-1$.

Adjusting the prior mean from $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}$* results in a slight change of the posterior for $\boldsymbol{\theta}$. In particular, the posterior mean of $\hat{\boldsymbol{\theta}}$ would be slightly shifted towards the prior mean $\boldsymbol{\theta}$*. Large-sample theory, however, dictates that the prior has a negligible effect on the posterior for large samples. Therefore, we leave the approximated posterior for $\boldsymbol{\theta}$, given by $N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_\theta)$, unaltered. Note that a similar argument is used in the Bayesian information criterion approximation of the Bayes factor (Kass & Raftery, 1995; Schwarz, 1978).

Based on the adjusted fractional prior distribution (23) and the posterior distribution (21), the AAFBF for an informative hypothesis versus the unconstrained hypothesis can be defined as

$$\mathrm{AAFBF}_{iu} = \frac{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u(\boldsymbol{\theta}|\mathbf{X})\mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} \pi_u^*(\boldsymbol{\theta}|\mathbf{X}^b)\mathrm{d}\boldsymbol{\theta}}, \tag{25}$$

where the parameter space $\boldsymbol{\Theta}_i = \{\boldsymbol{\theta}|\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}, \mathbf{R}_{i_1}\boldsymbol{\theta} > \mathbf{r}_{i_1}\}$ is in agreement with the informative hypothesis H_i. The computation of the AAFBF will be elaborated next.

3.4. Bayes factor computation

To compute the AAFBF, we first need to determine the adjusted prior mean $\boldsymbol{\theta}^*$ in (23). Finding the parameter space $\boldsymbol{\Theta}_i^*$ can be difficult for complicated informative hypotheses (Mulder *et al.*, 2012). However, if we transform the parameters of interest using $\boldsymbol{\beta}_0 = \mathbf{R}_{i_0}\boldsymbol{\theta} - \mathbf{r}_{i_0}$ and $\boldsymbol{\beta}_1 = \mathbf{R}_{i_1}\boldsymbol{\theta} - \mathbf{r}_{i_1}$, then the informative hypothesis under consideration becomes $H_i : \boldsymbol{\beta}_0 = 0,\ \boldsymbol{\beta}_1 > 0$ such that we can simply specify the prior mean vector equal to zero for the new parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_1^T)^T$. Note that the range constrained hypothesis (e.g., H_1: $0 < \theta < 1$) is an exception because, as elaborated earlier, the prior mean for θ is centred at $\theta^* = 0.5$, which requires $\beta_{11}^* = \theta^* = 0.5$ and $\beta_{12}^* = 1 - \theta^* = 0.5$. The specification of the prior mean for range constraints is given in Appendix A. This parameter transformation was also used in Mulder (2016) for hypotheses with only inequality constraints on correlations. Here we generalize it to equality and inequality constraints on parameters in general statistical models. The parameter transformation of $\boldsymbol{\theta}$ to $\boldsymbol{\beta}$ simplifies the form of the hypothesis without changing the expectation of researchers. For instance, testing whether two parameters are equal ($\theta_1 = \theta_2$) is identical to testing whether their difference is 0 (i.e., $\beta_0 = \theta_1 - \theta_2 = 0$). Consequently, the adjusted fractional prior distribution and posterior distribution for the new parameter $\boldsymbol{\beta}$ are given by

$$\pi_u^*(\boldsymbol{\beta}|\mathbf{X}^b) = N(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_\beta/b) \tag{26}$$

and

$$\pi_u(\boldsymbol{\beta}|\mathbf{X}) = N(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_\beta), \tag{27}$$

respectively, where $\hat{\boldsymbol{\beta}} = \mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}$ and $\hat{\boldsymbol{\Sigma}}_\beta = \mathbf{R}\hat{\boldsymbol{\Sigma}}_\theta\mathbf{R}^T$ with $\mathbf{R} = (\mathbf{R}_{i_0}^T, \mathbf{R}_{i_1}^T)^T$ and $\mathbf{r} = (\mathbf{r}_{i_0}^T, \mathbf{r}_{i_1}^T)^T$. Specifically, $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0^T, \hat{\boldsymbol{\beta}}_1^T)^T$ where $\hat{\boldsymbol{\beta}}_0 = \mathbf{R}_{i_0}\hat{\boldsymbol{\theta}} - \mathbf{r}_{i_0}$ and $\hat{\boldsymbol{\beta}}_1 = \mathbf{R}_{i_1}\hat{\boldsymbol{\theta}} - \mathbf{r}_{i_1}$, and $\hat{\boldsymbol{\Sigma}}_\beta = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\beta_0} & \hat{\boldsymbol{\Sigma}}_{01} \\ \hat{\boldsymbol{\Sigma}}_{10} & \hat{\boldsymbol{\Sigma}}_{\beta_1} \end{bmatrix}$ where $\hat{\boldsymbol{\Sigma}}_{\beta_0} = \mathbf{R}_{i_0}\hat{\boldsymbol{\Sigma}}_\theta\mathbf{R}_{i_0}^T$ and $\hat{\boldsymbol{\Sigma}}_{\beta_1} = \mathbf{R}_{i_1}\hat{\boldsymbol{\Sigma}}_\theta\mathbf{R}_{i_1}^T$.

This parameter transformation from $\boldsymbol{\theta}$ to $\boldsymbol{\beta}$ simplifies the computation of the AAFBF. First, the AAFBF for an informative hypothesis with only equality constraints (i.e., H_i: $\boldsymbol{\beta}_0 = 0$), compared to the unconstrained hypothesis, can be obtained using the Savage–Dickey density ratio (Dickey, 1971; Mulder, 2014b; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010):

$$\mathrm{AAFBF}_{iu}^{0} = \frac{\pi_u(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X})}{\pi_u^*(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}^b)}, \tag{28}$$

where $\pi_u^*(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}^b)$ and $\pi_u(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X})$ are the densities of the prior (26) and posterior (27), respectively, for $\boldsymbol{\beta}_0$ at the point $\boldsymbol{\beta}_0 = 0$ under H_u. Second, the AAFBF for an informative hypothesis with only inequality constraints (i.e., H_i: $\boldsymbol{\beta}_1 > \mathbf{0}$), compared to the unconstrained hypothesis, is given by (Hoijtink, 2012; Mulder, 2014b)

$$\mathrm{AAFBF}_{iu}^{1} = \frac{\int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u(\boldsymbol{\beta}_1|\mathbf{X})\mathrm{d}\boldsymbol{\beta}_1}{\int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u^*(\boldsymbol{\beta}_1|\mathbf{X}^b)\mathrm{d}\boldsymbol{\beta}_1}, \tag{29}$$

where $\pi_u^*(\boldsymbol{\beta}_1|\mathbf{X}^b)$ and $\pi_u(\boldsymbol{\beta}_1|\mathbf{X})$ are the prior (26) and posterior (27), respectively, for $\boldsymbol{\beta}_1$. Finally, the AAFBF for an informative hypothesis with both equality and inequality constraints (i.e., H_i: $\boldsymbol{\beta}_0 = \mathbf{0}$, $\boldsymbol{\beta}_1 > 0$), compared to the unconstrained hypothesis, can be obtained via

$$\mathrm{AAFBF}_{iu} = \frac{\pi_u(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X})}{\pi_u^*(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}^b)} \cdot \frac{\int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X})\mathrm{d}\boldsymbol{\beta}_1}{\int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u^*(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}^b)\mathrm{d}\boldsymbol{\beta}_1}, \tag{30}$$

where $\pi_u^*(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}^b)$ and $\pi_u(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X})$ are the prior and posterior distributions of $\boldsymbol{\beta}_1$ given $\boldsymbol{\beta}_0 = \mathbf{0}$, respectively. Note that $\pi_u^*(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}^b) = N(\mathbf{0}, (\hat{\boldsymbol{\Sigma}}_{\beta_1} - \hat{\boldsymbol{\Sigma}}_{10}\hat{\boldsymbol{\Sigma}}_{\beta_0}^{-1}\hat{\boldsymbol{\Sigma}}_{01})/b)$ and $\pi_u(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}) = N(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\Sigma}}_{10}\hat{\boldsymbol{\Sigma}}_{\beta_0}^{-1}\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\Sigma}}_{\beta_1} - \hat{\boldsymbol{\Sigma}}_{10}\hat{\boldsymbol{\Sigma}}_{\beta_0}^{-1}\hat{\boldsymbol{\Sigma}}_{01})$.

We let $c_i^0 = \pi_u^*(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}^b)$ and $c_i^1 = \int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u^*(\boldsymbol{\beta}_1|\mathbf{X}^b)\mathrm{d}\boldsymbol{\beta}_1$, which can be interpreted as the relative complexities of the equality constrained hypothesis and inequality constrained hypothesis, respectively, compared to H_u under prior (26). Then, in general,

$$c_i = \pi_u^*(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}^b) \cdot \int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u^*(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}^b)\mathrm{d}\boldsymbol{\beta}_1 \tag{31}$$

represents the relative complexity of informative hypothesis H_i (Hoijtink, 2012; Mulder, 2014a), which is a relative measure of the size of the parameter space under an informative hypothesis in comparison to the unconstrained parameter space. For example, the relative complexity of '$\theta_1 > \theta_2$, and θ_3 unconstrained' is larger than the relative complexity of '$\theta_1 > \theta_2 > \theta_3$'. This can be understood from the fact that the parameter space of the latter is a subset of the parameter space of the former. Similarly, the relative complexity of '$\theta_1 = 0$, θ_2 unconstrained' is larger than the relative complexity of '$\theta_1 = 0$, $\theta_2 = 0$'. It is interesting to note that the relative complexity c_i^0 of an equality constrained hypothesis H_i: $\boldsymbol{\beta} = \mathbf{0}$ becomes smaller when the prior variance of $\boldsymbol{\beta}$ under H_u becomes larger. The reason is that a larger variance of the unconstrained prior implies that a larger region of the unconstrained parameter space is likely *a priori*, which means that H_i is simpler relative to the unconstrained hypothesis. Furthermore, we let $f_i^0 = \pi_u(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X})$ and $f_i^1 = \int_{\boldsymbol{\beta}_1 > \mathbf{0}} \pi_u(\boldsymbol{\beta}_1|\mathbf{X})\mathrm{d}\boldsymbol{\beta}_1$, which can be interpreted as the measures of relative fit of the equality constrained hypothesis and inequality constrained hypothesis, respectively, compared to H_u. Then

$$f_i = \pi_u(\boldsymbol{\beta}_0 = \mathbf{0}|\mathbf{X}) \cdot \int_{\boldsymbol{\beta}_1 > 0} \pi_u(\boldsymbol{\beta}_1|\boldsymbol{\beta}_0 = \mathbf{0}, \mathbf{X}) d\boldsymbol{\beta}_1 \tag{32}$$

expresses the relative fit of H_i (Hoijtink, 2012; Mulder, 2014a), which implies how well a hypothesis is supported by the data compared to the unconstrained hypothesis. The relative complexity and fit in the AAFBF can be estimated based on a similar procedure presented in Gu *et al.* (2014) which only considers inequality constraints. We generalize the method to hypotheses with inequality as well as equality constraints to cover a very large spectrum of informative hypotheses that can be tested.

The computation of the AAFBF is implemented in the software package BaIn (Bayesian evaluation of informative hypotheses) available at http://informative-hypotheses.sites.uu.nl/software/. A user manual for BaIn is given in Appendix B. The input of BaIn needs the maximum likelihood estimate and covariance matrix of the parameters of interest, which can be obtained using other software packages such as Mplus (Muthén & Muthén, 2010) or the free R package lavaan (Rosseel, 2012). Executing BaIn renders the AAFBF for each informative hypothesis H_i under evaluation.

The Bayes factor of an informative hypothesis H_i against its complement H_{i_c} is

$$\text{AAFBF}_{ii_c} = \frac{f_i/c_i}{(1-f_i)/(1-c_i)} \tag{33}$$

if H_i does not contain equality constraints. Otherwise $\text{AAFBF}_{ii_c} = \text{AAFBF}_{iu}$ because the marginal likelihood of the complement of a hypothesis which contains equality constraints is equal to the marginal likelihood of the unconstrained hypothesis. For the comparison of two informative hypotheses H_i and $H_{i'}$, the AAFBF for H_i against $H_{i'}$ can be obtained as

$$\text{AAFBF}_{ii'} = \frac{\text{AAFBF}_{iu}}{\text{AAFBF}_{i'u}}. \tag{34}$$

Running BaIn for H_i and $H_{i'}$ renders AAFBF_{iu} and $\text{AAFBF}_{i'u}$ such that $\text{AAFBF}_{ii'}$ can be computed using (34).

4. Choices for *b*

This section discusses the choices of the fraction b for the specification of fractional priors. We first show the influence of the choices of b on the AAFBF when evaluating informative hypotheses because, as with the original fractional Bayes factor (Conigliani & O'Hagan, 2000), the choice of the fraction b also plays a crucial in the AAFBF. Then we present two traditional choices and one novel choice of b. At the end of this section, we conduct a sensitivity study to investigate the approximation error of the AAFBF relative to the actual adjusted fractional Bayes factor. It should be noted that this paper uses one common fraction b of the likelihood for prior specification. For this reason the AAFBF should only be used for testing hypotheses based on data that come from one population or balanced data with equal group sizes in the case of multiple populations, similar to the fractional Bayes factor (de Santis & Spezzaferri, 2001).

4.1. The role of b in AAFBF

The influence of the fraction b on the AAFBF is different for the evaluation of equality constraints $\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}$ and of inequality constraints $\mathbf{R}_{i_1}\boldsymbol{\theta} > \mathbf{r}_{i_1}$. First of all, b is a very influential parameter when evaluating equality constraints $\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}$. The underlying reason is that a small (large) b implies a prior with large (small) variance such that the prior density evaluated at $\mathbf{R}_{i_0}\boldsymbol{\theta} = \mathbf{r}_{i_0}$ or $\boldsymbol{\beta}_0 = \mathbf{0}$ in (28) is small (large). This can be illustrated in Figure 1 in which the solid line represents the density of prior distribution $\pi_u^*(\theta|x^b) = N(0, \sigma_\theta^2/b)$ with $\sigma_\theta^2 = 0.02$ under (a) $b = 0.05$ and (b) $b = 0.2$. As can be seen, when testing the hypothesis H_1: $\theta = 0$ against H_u, the prior density at $\theta = 0$ is 0.63 under $b = 0.05$ in Figure 1a, half the value 1.26 under $b = 0.2$ in Figure 1b. Given an estimate of $\hat{\theta} = 0.2$, the resulting AAFBF for H_1 against H_u under $b = 0.05$ is $\text{AAFBF}_{1u} = 1.64$, whereas under $b = 0.2$ it is $\text{AAFBF}_{1u} = 0.82$ according to equation (28).

Secondly, for range constrained hypotheses the effect of b is similar to that for an equality constrained hypothesis: a small (large) b implies a large (small) AAFBF for the range constrained hypothesis against the unconstrained hypothesis. For example, the shaded area in Figure 1 represents the prior probability in line with the range constrained hypothesis H_2: $-0.5 < \theta < 0.5$, which implies that the absolute effect is expected to be

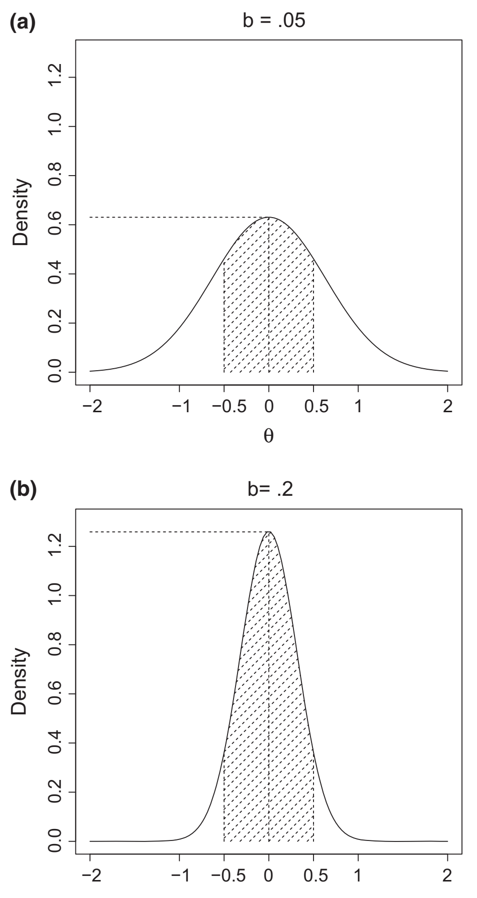

Figure 1. Relative complexities under different values of *b*.

smaller than 0.5. For a small $b = 0.05$ the prior probability of $-0.5 < \theta < 0.5$ shown in Figure 1a is 0.57, whereas for a large $b = 0.2$ the prior probability in Figure 1b is 0.89. Based on $\hat{\theta} = 0.2$ and equation (29) the AAFBF for H_2 against H_u under $b = 0.05$ is $\text{AAFBF}_{2u} = 1.72$, which is different from $\text{AAFBF}_{2u} = 1.11$ under $b = 0.2$.

Thirdly, the AAFBF is independent of the choice of b for inequality constrained hypotheses which do not contain range constraints. This property was proven in Mulder (2014b) and can also be seen in Figure 1 where the prior probability that the constraint of H_3: $\theta > 0$ holds under H_u is equal to 0.5 for both choices of b.

The influence of b on the AAFBF is illustrated in Figure 2 when comparing the equality constrained hypothesis H_1: $\theta = 0$, the range constrained hypothesis H_2: $-0.5 < \theta < 0.5$, and the inequality constrained hypothesis H_3: $\theta > 0$ to the unconstrained hypothesis H_u. Given the estimate $\hat{\theta} = 0.2$ and variance $\hat{\sigma}^2_\theta = 0.02$ for θ, Figure 2 shows the AAFBF for each informative hypothesis under various $b \in (0, .5]$. As can be seen, the AAFBF for H_1 decreases as b increases, the AAFBF for H_2 behaves similarly to that for H_1, and the AAFBF for H_3 is stable as b changes. This illustrates that the fraction b has to be carefully specified when equality constrained hypotheses and range constrained hypotheses are of interest to the researcher, while any fraction b can be used when only inequality constrained hypotheses without range constraints are formulated by the user. In what follows we will specify b in three different ways.

4.2. Traditional choices for b

Previous studies have recommended two choices for b for the fractional Bayes factor. The first one comes from Berger and Pericchi (1996) and O'Hagan (1995) who suggested using the minimal training sample for prior specification to leave maximal information in the data for hypothesis testing. This corresponds to $b = m/n$ in the fractional prior, where m is the size of the minimal training sample that makes all parameters identifiable. For example, for the one-sample t test of H_0: $\theta = 0$ where the data are $x_i \sim N(\theta, \sigma^2)$, the actual adjusted fractional prior distribution for θ is $\pi^*_u(\theta|x^b) = t(0, s^2/(nb-1), nb-1)$, that is, a Student t density with mean 0, scale parameter $s^2/(nb-1)$, and degrees of freedom $nb-1$. In this case, the minimal m is 2 because $m = 1$ results in $b = 1/n$ and degrees of freedom 0, which is not allowed.

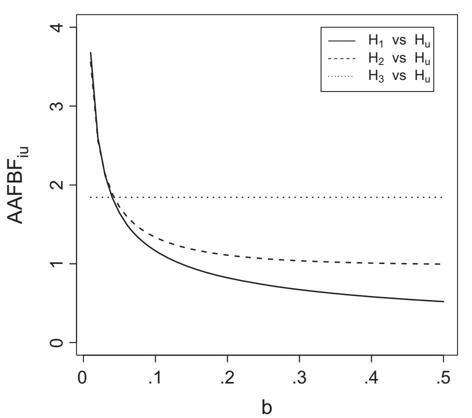

Figure 2. Influence of b on AAFBF.

For the AAFBF we propose a similar approach to determine our first choice of b. To estimate $\boldsymbol{\beta}$ (with length J) we need at least $J + 1$ observations. Therefore, our first choice of the fraction equals

$$b_{\min} = (J+1)/n, \tag{35}$$

where J is the number of independent constraints in all the informative hypotheses under investigation, that is, J equals the rank of $\mathbf{R} = (\mathbf{R}_{1_0}^T, \mathbf{R}_{1_1}^T, \ldots, \mathbf{R}_{I_0}^T, \mathbf{R}_{I_1}^T)^T$ for a set of informative hypotheses H_i for $i = 1, \ldots, I$. Thus, if H_3: $\theta_1 = 0$ and H_4: $\theta_1 > 0$, $\theta_2 > 0$ are under evaluation, for example, $J = 2$ when computing the AAFBF for each informative hypothesis against the unconstrained hypothesis because there are two independent constraints.

For multiple regression model (5) in Section 2, $J = 3$ because H_1: $\theta_1 > 0, \theta_2 < 0, \theta_3 = 0$ can be formulated using a vector $\boldsymbol{\beta}$ of length 3. With sample size $n = 50$, the first choice of the fraction b can be set to $b_{\min} = 2/25$. For repeated measures model (10), $J = 3$ based on a vector $\boldsymbol{\beta}$ of length 3 in $H_2 : \theta_1 = \theta_2 > \theta_3 = \theta_4$ and $H_{2'} : \theta_1 > \theta_2 > \theta_3 > \theta_4$, and therefore $b_{\min} = 1/9$ based on sample size $n = 36$.

The second way of choosing b is (O'Hagan, 1995)

$$b_{\text{robust}} = \max\{(J+1)/n, 1/\sqrt{n}\}, \tag{36}$$

which is in general larger than the first choice. O'Hagan (1995) stated that a larger b can reduce the sensitivity of the fractional Bayes factor to the distributional form of the prior. Conigliani and O'Hagan (2000) further derived a measure of the sensitivity of the fractional Bayes factor and proved that this measure is a decreasing function of b. The second choice of b can also be applied to the AAFBF defined in (25). When setting a larger b, the AAFBF becomes more similar to the non-AAFBF. Thus, the AAFBF is less sensitive to the prior distribution given larger b. We will more to say on this topic in Section 4.4. Given the sample size $n = 50$ in the regression model in Section 2, $b_{\text{robust}} = 1/\sqrt{50}$ is specified to evaluate hypothesis H_1. In the case of the repeated measures model with sample size $n = 36$, one can set $b_{\text{robust}} = 1/6$ for the comparison of H_2 and $H_{2'}$.

4.3. A frequentist choice for b

Gu *et al.* (2016) recently proposed another method of specifying b by taking into account the frequentist error probabilities. In Bayesian hypothesis testing, the probability of a Bayes factor favouring H_u when H_i is true is

$$p_1 = P(\mathrm{BF}_{iu} < 1 | H_i) \tag{37}$$

which corresponds to the Type I error probability if H_i is a traditional null hypothesis, and the probability of a Bayes factor favouring H_i when H_u is true is

$$p_2 = P(\mathrm{BF}_{iu} > 1 | H_u). \tag{38}$$

which then corresponds to the Type II error probability. Gu *et al.* (2016) found that these probabilities are often quite different when using traditional choices of b in the one-sample t test. This may not be preferable from a frequentist point of view where the goal typically is to control the error probabilities. Here we show how to specify b to control the error probabilities under certain conditions. First, we shall use a one-sample t test to illustrate the procedure for specifying b based on this method, and then apply it to the AAFBF (28) for general statistical models. Finally, a rule for choosing b is proposed.

4.3.1. One-sample t test

Consider a one-sample t test for which data come from $x_i \sim N(\theta, \sigma^2)$, where θ denotes the population mean and σ^2 denotes the population variance, and the hypotheses under consideration are H_1: $\theta = 0$ against H_u: θ. The AAFBF for H_1 against H_u can be derived using equation (28):

$$\text{AAFBF}_{1u} = b^{-1/2}\exp\left(-\frac{1}{2}n(\bar{x}/s)^2\right), \tag{39}$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$. For this AAFBF the error probabilities eqns (37) and (38) become

$$\begin{aligned} p_1 = P(\text{AAFBF}_{1u} < 1|H_1) &= P\left(|\frac{\bar{x}}{s}| > \sqrt{-\log b/n}|H_1\right) \\ &\approx \frac{1}{L}\sum_{l=1}^{L} I\left(|\frac{\bar{x}^{(1l)}}{s^{(1l)}}| > \sqrt{-\log b/n}\right) \end{aligned} \tag{40}$$

and

$$\begin{aligned} p_2 = P(\text{AAFBF}_{1u} > 1|H_u) &= P\left(|\frac{\bar{x}}{s}| < \sqrt{-\log b/n}|H_u\right) \\ &\approx \frac{1}{L}\sum_{l=1}^{L} I\left(|\frac{\bar{x}^{(2l)}}{s^{(2l)}}| < \sqrt{-\log b/n}\right), \end{aligned} \tag{41}$$

where $\bar{x}^{(1l)}$ and $s^{(1l)}$, for $l = 1, \ldots, L$, are the mean and standard deviation of data $x_i^{(1l)}$ sampled from H_1, $\bar{x}^{(2l)}$ and $s^{(2l)}$ are the mean and standard deviation of data $x_i^{(2l)}$ sampled from H_u, and $I(\cdot)$ is the indicator function which is 1 if the argument is true and 0 otherwise. When sampling data from H_u, an expected standardized effect size, denoted by β_e, needs to be specified under H_u, namely, H_u: $\theta = \beta_e\sigma$, so that the scaled data are sampled from $y_i \sim N(\beta_e, 1)$ under H_u, where $y_i = x_i/\sigma$. Note that sampling $\bar{x}^{(2l)}/s^{(2l)}$ based on $x_i \sim N(\theta, \sigma^2)$, where $\theta/\sigma = \beta_e$, is identical to sampling the mean $\bar{y}^{(2l)}$ based on $y_i \sim N(\beta_e, 1)$. The specification of the standardized effect size β_e will be discussed in Section 4.3.3.

In the one-sample t test, $\bar{x}/s$ is the observed standardized effect size known as Cohen's d (Cohen, 1992). It has sampling distributions under H_1 and H_u which can be obtained using $\bar{x}^{(1l)}/s^{(1l)}$ and $\bar{x}^{(2l)}/s^{(2l)}$, respectively. Figure 3 shows the distributions of $\bar{x}/s$ under H_1: $\theta = 0$ (solid line) and H_u: $\theta = \beta_e$ (dashed line) given $\sigma^2 = 1$ and $n = 20$, where $\beta_e = .5$ is the pre-specified standardized effect size under H_u. Note that according to Cohen (1992), $\beta_e = .2$, .5, and .8 correspond to small, medium, and large effects, respectively. If

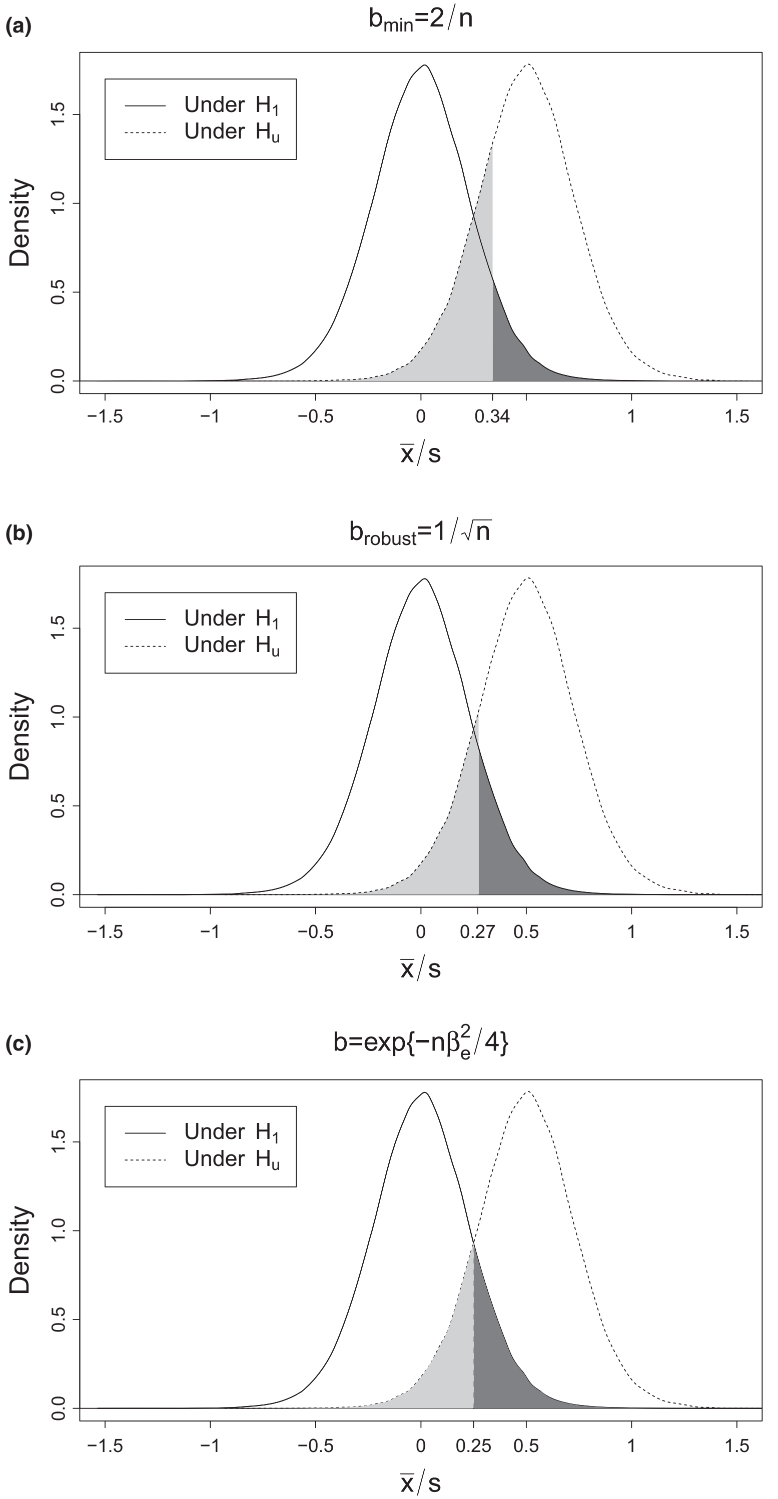

Figure 3. Sampling distributions of observed effect size $\bar{x}/s$ in one-sample t test for n = 20 and β_e = 0.5 under H_u.

we use $b_{\min} = 2/n$ for the one-sample t test, the error probabilities in (40) and (41) become $p_1 = P(|\bar{x}/s| > .34|H_1) = .073$ and $p_2 = P(|\bar{x}/s| < .34|H_u) = .241$, whereas if we specify $b_{\text{robust}} = 1/\sqrt{n}$, the error probabilities are $p_1 = P(|\bar{x}/s| > .27|H_1) = .122$ and $p_2 = P(|\bar{x}/s| < .27|H_u) = .159$. These error probabilities are marked in Figure 3a for $b_{\min}$ and Figure 3b for b_{robust}, where the dark grey area represents p_1 and the light grey area represents p_2. As can be seen, $p_1 < p_2$ under both $b_{\min}$ and b_{robust}, which means that we are more likely to incorrectly prefer H_1 when H_u is true than incorrectly prefer H_u when H_1 is true.

In order to correct for this, Gu *et al.* (2016) showed how to choose b such that $p_1 = p_2$ given sample size n and effect size β_e under H_u. A direct way of obtaining such a b is proposed by Morey, Wagenmakers, and Rouder (2016) and illustrated in Figure 3c. As can be seen, the distributions of $\bar{x}/s$ under H_1: $\theta = 0$ and H_u: $\theta = \beta_e$ are symmetric on $\beta_e/2$. This implies that we can simply specify $\sqrt{-\log b/n} = \beta_e/2$ or equivalently $b = \exp(-n\beta_e^2/4)$ to attain equal error probabilities, because $p_1 = P(|\bar{x}/s| > \beta_e/2|H_1)$ is equal to $p_2 = P(|\bar{x}/s| < \beta_e/2|H_u)$. For example, given $n = 20$ and $\beta_e = .5$ under H_u in Figure 3c, the dark grey area for p_1 is the same size as the light grey area for p_2 when setting $b = \exp(-n\beta_e^2/4) = .287$. The error probabilities under this setting are $p_1 = p_2 = .139$.

4.3.2. General case

The method of choosing b based on equal error probabilities can be generalized to the AAFBF of any H_i: $\boldsymbol{\beta}_0 = \mathbf{0}$ against H_u: $\boldsymbol{\beta}_0 \neq \mathbf{0}$. Based on the adjusted fractional prior (26) and approximated posterior (27), the AAFBF in (28) is

$$\text{AAFBF}_{iu}^0 = b^{-1/2} \exp\left(-\frac{1}{2}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_\beta^{-1}\hat{\boldsymbol{\beta}}^T\right). \tag{42}$$

It is interesting to note that $\sqrt{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_\beta^{-1}\hat{\boldsymbol{\beta}}^T}$ in (42) is the test statistic in the Wald test (Engle, 1984) which assumes that $\boldsymbol{\beta}$ is approximately normally distributed. The test statistic is not only the cornerstone in frequentist hypothesis testing, but also important in default Bayes factors. For example, the Bayes factor proposed by Rouder *et al.* (2009) for the t test is a function of the t statistic, and the Bayes factor based on Zellner's g prior (Zellner & Siow, 1980) in regression models is a function of the F statistic. The standardized effect size is often defined as a test statistic divided by $\sqrt{n}$ to offset the influence of the sample size (Cohen, 1992), because the effect size should not be affected by the sample size as it expresses the degree to which H_u differs from H_i. Thus, the observed standardized effect size in this case can be defined as

$$\hat{\beta}_e = \sqrt{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_\beta^{-1}\hat{\boldsymbol{\beta}}^T/n}. \tag{43}$$

Then using the steps as in (40) and (41) for the one-sample t test, the error probabilities of AAFBFs are defined as

$$p_1 = P(\text{AAFBF}_{iu}^0 < 1|H_i) = P(\hat{\beta}_e > \sqrt{-\log b/n}|H_i) \tag{44}$$

and

$$p_2 = P(\mathrm{AAFBF}_{iu}^{0} > 1|H_u) = P(\hat{\beta}_e < \sqrt{-\log b/n}|H_u). \tag{45}$$

The observed standardized effect size $\hat{\beta}_e$ is usually within the interval [0,1] for equality constrained hypothesis testing, because $\hat{\beta}_e$ can be interpreted analogously as Cohen's d or Cohen's f^2 (Cohen, 1992), which rarely exceeds 1. First, for a one-sample t test $x_i \sim N(\theta, \sigma^2)$, and $H_1 : \theta = 0$ versus H_u: θ, the maximum likelihood estimate of $\beta = \theta$ is $\hat{\beta} = \bar{x}$ and the standard deviation is $\hat{\sigma}_\beta = s/\sqrt{n}$. Then the observed standardized effect size (43) becomes $\hat{\beta}_e = (\hat{\beta}/\hat{\sigma}_\beta)/\sqrt{n} = \bar{x}/s$ which is the same as Cohen's d. Second, we consider the F test of $H_2 : \theta_1 = 0$ against $H_u : \theta_1$ in a simple linear regression model $y_i = \theta_0 + \theta_1 x_i + \epsilon_i$, where θ_0 is the intercept, θ_1 is the regression coefficient, and $\epsilon_i \sim N(0, \sigma^2)$ is the residual. The maximum likelihood estimate of $\beta = \theta_1$ is $\hat{\beta} = r_{xy}s_y/s_x$ and the standard deviation is $\hat{\sigma}_\beta = (\sigma/s_x)/\sqrt{n}$, where s_x and s_y are the standard derivations of x_i and y_i, and r_{xy} is the correlation coefficient between x_i and y_i. Note that r_{xy}^2 is equal to the coefficient of determination R^2 in the case of the simple linear regression model. Thus, because the coefficient of determination is equal to $R^2 = 1 - \sigma^2/s_y^2$, the observed standardized effect size in (43) becomes $\hat{\beta}_e = (\hat{\beta}/\hat{\sigma}_\beta)/\sqrt{n} = r_{xy}s_y/\sigma = \sqrt{R^2/(1-R^2)}$, which is the square root of Cohen's $f^2 = R^2/\,(1-R^2)$.

Analogously to the effect size $\bar{x}/s$ in the one-sample t test, the observed standardized effect size $\hat{\beta}_e$ also has sampling distributions under H_i and H_u, which are symmetric around half of the pre-specified standardized effect size β_e under H_u. Therefore, by setting $\sqrt{-\log b/n} = \beta_e/2$, or equivalently

$$b = \exp\,(-n\beta_e^2/4), \tag{46}$$

the test for H_i against H_u using the AAFBF has equal error probability:

$$p_1 = P(\hat{\beta}_e > \beta_e/2|H_i) = P(\hat{\beta}_e < \beta_e/2|H_u) = p_2. \tag{47}$$

We now turn to how to specify β_e in (46).

4.3.3. A new rule for choosing b

Before presenting the new choice of b based on equal error probabilities, we need to deal with two issues: the range of b for consistent Bayes factors and the specification of standardized effect size β_e under H_u. The consistency of the Bayes factor is an important property in Bayesian hypothesis testing. The Bayes factor for $H_i : \boldsymbol{\beta} = 0$ against $H_u : \boldsymbol{\beta} \neq \mathbf{0}$ is consistent if it goes to infinity as sample size goes to infinity when H_i is true, and goes to 0 when H_u is true. Morey *et al.* (2016) found that the prior specification based on frequentist error probabilities may result in inconsistent Bayes factors. Gu *et al.* (2016) showed how to resolve this by restricting b to $b \geq 2/n$ in the one-sample t test. As stated in Section 4.2, $b = (J+1)/n$ is based on the minimal number of observations to specify proper priors, and therefore we will always constrain $b \geq (J+1)/n$ in the AAFBF. Furthermore, we also suggest constraining $b \leq 1/2$ because $b > 1/2$ implies that more than half of the likelihood is used for prior specification, which is undesirable in Bayesian tests (Berger & Pericchi, 1996). Consequently, the range of b is set to $b \in [(J+1)/n, 1/2]$.

To obtain b in (46) for equal error probabilities, the standardized effect size β_e under H_u has to be specified. Given any specific β_e, a fraction b in (46) can be obtained such that $p_1 = p_2$. However, in practice β_e is unknown. Therefore, a distribution for β_e is specified that covers a range of realistic effect sizes (i.e., $\beta_e \in [0, 1]$ as already discussed). Here we consider a uniform distribution $\pi^*(\beta_e) = U(0, 1)$ in which every effect size from small to large is equally likely within the interval [0,1] (Gu *et al.*, 2016). Note that this choice for b would be the same as when using $\pi^*(\beta_e) = U(-1, 1)$ because the choice of b is independent of the sign of the effect.

Based on the distribution of effects $\pi^*(\beta_e) = U(0, 1)$, the third choice for b for equal error probabilities is given by

$$b_{\text{freq}} = E_{\pi^*(\beta_e)}[\exp(-n\beta_e^2/4)] = \int_0^1 \exp(-n\beta_e^2/4)\mathrm{d}\beta_e. \tag{48}$$

The integration in (48) can be calculated numerically (see Gu *et al.*, 2016). Although b_{freq} cannot always achieve equal error probabilities as we constrain $b \in [(J + 1)/n, 1/2]$ and specify $\pi^*(\beta_e) = U(0, 1)$, Gu *et al.* (2016) show that this choice results in error probabilities that are often about equal for the one-sample t test. It was shown that the difference between the Type I and Type II error probabilities was typically smaller for this choice than when using the more traditional choices for b. We recommend the choice b_{freq} when the sample size is small, because in this case the error probabilities p_1 and p_2 are relatively large and the difference between p_1 and p_2 can be quite severe. In the following subsection, we will discuss the sensitivity of AAFBF based on different choices of b.

4.4. Sensitivity to prior distributions

In Section 3 we specified the normal prior (26) for $\boldsymbol{\beta}$ in general statistical models. However, the adjusted fractional prior for the parameters in a specific model is often not normally distributed. Thus, when using a normal approximation of the fractional prior, as in the case of the AAFBF, we may misspecify the prior distribution for the parameters of interest. For example, if the parameter is a probability which is bounded in [0,1] in a binomial model, the (implicit) fractional prior has a beta distribution. Therefore the use of the AAFBF, where the fractional prior is approximated using a normal distribution, may be different from the non-AAFBF. Thus, it is useful to investigate the sensitivity of the AAFBF when the fractional prior is far from normally distributed.

O'Hagan (1995) argued that the sensitivity of the fractional Bayes factor depends on the magnitude of b. This dependence was proved by Conigliani and O'Hagan (2000). Increasing b reduces the sensitivity to the distributional form of the fractional prior. This is also the case for the adjusted fractional Bayes factor (AFBF) of Mulder (2014b), because a larger b implies that more information in the data is used for prior specification, which makes the distribution of the adjusted fractional prior in the AFBF more similar to a normal distribution. This section will use two simple examples to illustrate how much difference there is between the AAFBF using the normal prior and the AFBF using the actual fractional prior. Furthermore, it is shown that the AAFBFs based on the different fractions show consistent behavior. In these examples, we will only focus on equality constrained hypotheses because, as explained earlier, the AFBF for inequality constrained hypotheses is independent of b.

The first example again concerns the one-sample t test, where data come from $x_i \sim N(\theta, \sigma^2)$ with unknown mean and variance, and the hypotheses under consideration are $H_1 : \theta = 0$ against $H_u : \theta$. In the AAFBF, the default prior (26) for $\beta = \theta$ is $\pi_u^*(\beta|X^b) = N(0, s^2/nb)$, while the actual adjusted fractional prior for a normal mean has a t distribution $\pi_u^*(\beta|X^b) = t(0, s^2/(nb-1), nb-1)$ with mean 0, variance $s^2/(nb-1)$, and degrees of freedom $nb - 1$. It is well known that the t distribution has heavier tails than the normal distribution, such that the density at the mode $\beta = 0$ from the normal distribution is larger than the density from the t distribution. Furthermore, as b increases, the degrees of freedom $nb - 1$ increase such that the t distribution $t(0, s^2/(nb-1), nb-1)$ becomes more similar to the normal distribution $N(0, s^2/nb)$. This implies that for a larger b the AAFBF where the default prior has a normal distribution performs more similarly to the AFBF under the actual fractional prior. This is illustrated in Figure 4.

Figure 4 shows the logarithms of AFBFs and AAFBFs for H_1 against H_u for different observed effect sizes $\bar{x}/s = 0, 0.1, 0.2$, and different fractions b_{min}, b_{robust}, and b_{freq}.

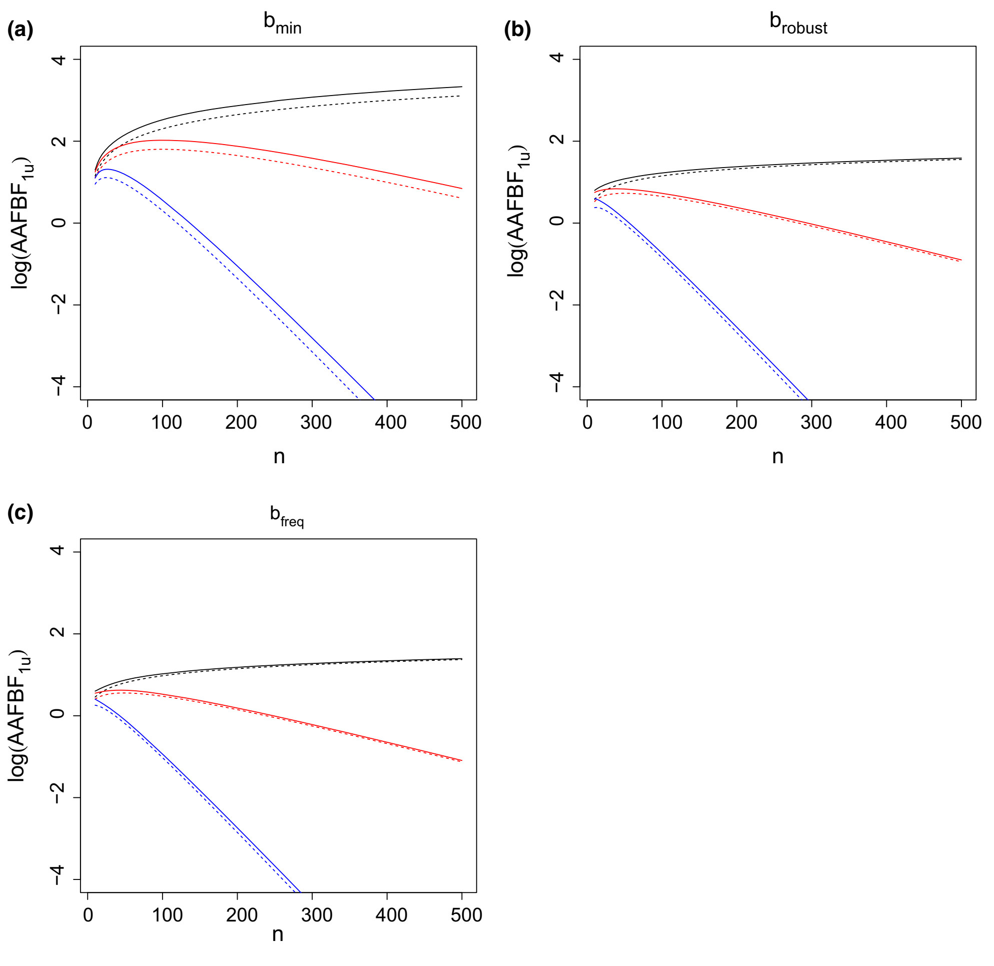

Figure 4. The logarithms of the adjusted fractional Bayes factor with a Student t prior (solid line) and the AAFBF with a normal prior (dashed line). The black, red, and blue lines correspond to the logarithms of Bayes factors for observed effect sizes $\bar{x}/s = 0$, 0.1, and 0.2, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

The sample size n varies from 10 to 500. First, as can be seen in Figure 4a, based on b_{min} the logarithms of AAFBFs under the normal prior distribution (dashed line) differ substantially from the logarithms of AFBFs under the t prior distribution (solid line). This difference does not decrease as n increases because when setting $b_{\text{min}} = 2/n$ the degrees of freedom in the t distribution are 1, which is independent of n. This suggests high sensitivity to the functional form of the prior distribution. Second, Figure 4b shows that based on b_{robust} there is not much difference between the logarithms of AAFBFs and AFBFs. This implies that the choice of b_{robust} results in less sensitivity to the functional form of the prior distribution than b_{min}. Third, Figure 4c shows the logarithms of AAFBFs and AFBFs under b_{freq}. As can be seen, with b_{freq} there is no sensitivity either.

It is interesting to note that Figure 4 also illustrates the consistency of AAFBFs. The consistency in this example requires that as sample size goes to infinity, the AAFBF for H_1 against H_u approaches infinity when the observed effect size is equal to 0 and goes to zero when the observed effect size is not equal to 0. As can be seen in Figure 4, for an observed effect size $\bar{x}/s = 0$ the logarithm of the AAFBF (black lines) in each figure goes to infinity as sample size n increases. Conversely, the logarithms of the AAFBF based on an observed effect size of $\bar{x}/s = 0.1$ (red lines) and $\bar{x}/s = 0.2$ (blue lines) diverge to minus infinity, which implies decisive evidence for the true unconstrained hypothesis as the sample size goes to infinity.

Next, we consider a binomial model, where data come from $x \sim Bin(n,p)$. The hypotheses under evaluation are $H_2 : p = 0.4$ against $H_u : 0 \leq p \leq 1$. Since H_2 is nested in H_u, we can use the AAFBF (28) to evaluate H_2 against H_u. Given data $x \sim Bin(n,p)$, the estimate of $\beta = p - 0.4$ is $\hat{\beta} = x/n - 0.4$ and the variance is $\hat{\sigma}^2_{\beta} = x(n-x)/(n^2(n+1))$, and therefore the normal adjusted fractional prior (26) is $\pi^*_u(\beta|X^b) = N(0, bx(n-x)/(n^2(n+1)))$. On the other hand, following the idea of adjusted fractional Bayes factors, the fractional prior has a beta distribution, $p = \beta + 0.4 \sim Beta(0.4nb, 0.6nb)$, which has a mean of 0.4 and thus β has a prior mean of 0. Note that this prior is centred on the focal point of 0.4 in H_2.

Figure 5 plots the logarithms of the AFBFs and AAFBFs for H_2 against H_u as the sample size n increases from 10 to 500. The observed data are $x = 0.4n, 0.5n, 0.6n$. As can be seen in Figure 5 there is a considerably smaller approximation error of the AAFBF with respect to the AFBF in comparison to the first example in Figure 4. Again, the difference is largest for b_{min} because this fraction is always smaller than b_{robust} and b_{freq}. Finally, note that the AAFBFs show consistent behaviour for this testing problem.

These two examples include the evaluation of equality constrained hypotheses in both continuous data and discrete data. Although the models used are simple, the results of the sensitivity study of adjusted fractional Bayes factors can be applied in the multivariate normal model where the parameters (e.g., the group means in the ANOVA model, the coefficients in the regression model) have a multivariate t distribution, and in the multinomial model where the parameters (e.g., the probabilities in contingency tables) have a Dirichlet distribution, which is the multivariate generalization of the beta distribution. Furthermore, in more complicated settings such as structural equation models and generalized linear models, it can be anticipated that the larger b will result in less sensitive AFBFs because this implies that more data are used to specify the fractional prior such that the normal approximation to the prior has better performance based on large-sample theory.

Based on the discussion in this section, we propose the following scheme for specifying b in the AAFBF:

- Choose $b_{\text{min}} = (J+1)/n$ to have a default prior that is based on the idea of a minimal training sample.

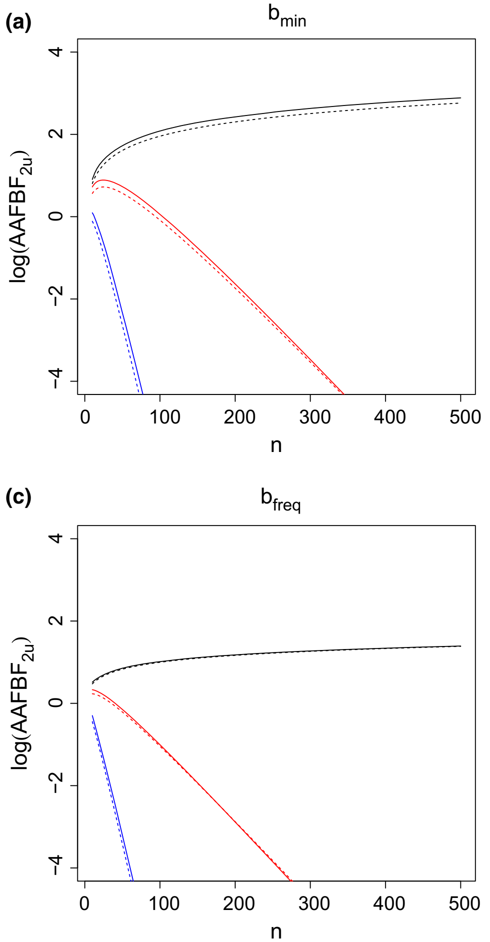

Figure 5. The logarithms of the adjusted fractional Bayes factor with a beta prior (solid line) and the AAFBF with a normal prior (dashed line). The black, red, and blue lines correspond to the logarithms of Bayes factors under observed effect sizes $\bar{x}/s = 0.4n$, $0.5n$, and $0.6n$, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

- Choose $b_{\text{robust}} = \max\{(J+1)/n, \sqrt{n}/n\}$ to ensure that the default prior is close to normal.
- Choose $b_{\text{freq}} = \int_0^1 \exp(-n\beta_e^2/4)\mathrm{d}\beta_e$ to control the frequentist error probabilities when testing an equality constrained hypothesis against the unconstrained alternative.

Note that n and J denote the sample size and the number of independent constraints for all the informative hypotheses, respectively.

5. Results for empirical examples

Let us revisit the examples introduced in Section 2 to illustrate how the AAFBF can be used to evaluate informative hypotheses. In the regression model, three parameters with respect to the regression coefficients are considered in the informative hypothesis $H_1: \theta_1 > 0,\ \theta_2 < 0,\ \theta_3 = 0$. The first step is to specify the prior and posterior

Table 3. Result for regression model example

	$b_{\min} = .080$	$b_{\text{robust}} = .141$	$b_{\text{freq}} = .216$
AAFBF_{11_c}	6.04	4.46	3.55

distributions in (26) and (27), which needs the estimates $\hat{\boldsymbol{\theta}}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_\theta$ of the parameters. These can be obtained by analysing the regression model with the data in Table 1 using a number of statistical software packages, such as Mplus (Muthén & Muthén, 2010) and R package lavaan (Rosseel, 2012). Note that we do not need to standardize the three coefficients as they are compared with zero. Analysis of the data in lavaan gives the maximum likelihood estimates of the parameters, $\hat{\theta}_1 = 11.01$, $\hat{\theta}_2 = -2.85$, $\hat{\theta}_3 = -2.03$, and the covariance matrix

$$\hat{\boldsymbol{\Sigma}}_\theta = \begin{bmatrix} 18.236 & -0.500 & 2.812 \\ -0.500 & 0.043 & -0.004 \\ 2.812 & -0.004 & 4.481 \end{bmatrix}.$$

To obtain the AAFBF for H_1 against H_{1_c}, the fraction b has to be specified. Based on the sample size of $n = 50$ and the length of vector $\boldsymbol{\beta}$ of $J = 3$ in this example, the three choices of fraction are $b_{\text{min}} = .080$, $b_{\text{robust}} = .141$, and $b_{\text{freq}} = .216$. Running BaIn with the estimates and covariance matrix of parameters of interest yields the AAFBF displayed in Table 3. As can be seen, AAFBF_{11_c} is greater than 3 under each choice of b, which implies positive evidence in the data for H_1 against H_{1_c} according to Kass and Raftery (1995) rule.

The hypothesis in the repeated measures ANOVA model consists of four parameters of which the estimates are $\hat{\theta}_1 = 22.33$, $\hat{\theta}_2 = 22$, $\hat{\theta}_3 = 5.78$ and $\hat{\theta}_4 = 6.78$, and the covariance matrix is

$$\hat{\boldsymbol{\Sigma}}_\theta = \begin{bmatrix} 5.18 & 4.86 & 2.61 & 2.86 \\ 4.86 & 5.13 & 2.90 & 3.03 \\ 2.61 & 2.90 & 1.93 & 1.97 \\ 2.86 & 3.03 & 1.97 & 2.39 \end{bmatrix}.$$

Given sample size $n = 36$ and length of vector $\boldsymbol{\beta}$ of $J = 3$, three choices of b are automatically specified in BaIn as $b_{\text{min}} = .111$, $b_{\text{robust}} = .167$, and $b_{\text{freq}} = .255$. Based on these specifications, BaIn renders the AAFBFs AAFBF_{2u} for H_2 versus H_u and $\text{AAFBF}_{2'u}$ for $H_{2'}$ versus H_u. The results are shown in Table 4. As can be seen, $\text{AAFBF}_{2'u}$ is independent of b because the AAFBF for inequality constrained hypotheses is invariant to the choice of b. Then the AAFBF $\text{AAFBF}_{22'}$ for H_2 versus $H_{2'}$ can be computed by $\text{AAFBF}_{2u}/\text{AAFBF}_{2'u}$ which is shown in the last row in Table 4. The result of $\text{AAFBF}_{22'}$ in the last row suggests positive evidence in the data for H_2 against $H_{2'}$.

Table 4. Result for repeated measures ANOVA example

	$b_{\min} = .111$	$b_{\text{robust}} = .167$	$b_{\text{freq}} = .255$
AAFBF_{2u}	4.60	3.07	2.01
$\text{AAFBF}_{2'u}$	0.24	0.24	0.24
$\text{AAFBF}_{22'}$	19.2	12.8	8.38

6. Conclusion

This paper has presented a new approximate Bayesian procedure for the evaluation of informative hypotheses that can be used for virtually any model. The methodology is based on the prior adjusted default Bayes factor of Mulder (2014b). Furthermore, normal approximations were used to ensure fast computations. Numerical results showed that the approximation is close to the prior adjusted fractional Bayes factor. This implies that the proposed AAFBF provides an accurate quantification of the relative evidence between informative hypotheses. Furthermore, different choices were given for the fraction b, similar as in the fractional Bayes factor of O'Hagan (1995). The first choice relies on the concept of priors containing minimal information. The second choice uses a robustness argument resulting in a default prior distribution that is close to normal. The third choice is based on a frequency argument to control the classical error probabilities. The choice can be made by the user depending on the property which he/she finds most important. By computing the AAFBF for each choice of b we get a complete picture how much support there is in the data between two hypotheses when taking into account different philosophies.

We provide a software package BaIn, with a user manual in Appendix B, to evaluate the informative hypotheses which only needs the maximum likelihood estimates and covariance matrix of the parameters of interest, denoted by $\boldsymbol{\theta}$ in this paper. BaIn computes the AAFBF for an informative hypothesis against an unconstrained hypothesis. By computing these quantities for each informative hypothesis against the unconstrained hypothesis, psychology researchers can straightforwardly compute the relative support in the data for pairs of informative hypotheses.

The study in this paper contributes to the quantitative techniques in psychology research in three respects. First, the proposed Bayesian test stimulates psychologists to translate scientific expectations into informative hypotheses that can be tested with the data in a direct manner. Second, the approximate Bayesian procedure allows psychologists to test their informative hypotheses in virtually any statistical model. Third, the software package allows psychologists to apply the new methodology to their own data in an easy manner.

Acknowledgements

This research is supported by the Consortium on Individual Development (CID) which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003).

References

Barlow, R., Bartholomew, D., Bremner, J., & Brunk, H. (1972). *Statistical inference under order restrictions: Theory and application of isotonic regression*. New York, NY: John Wiley & Sons.

Bartholomew, D. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, *46*, 36–48. https://doi.org/10.2307/2332806

Berger, J., & Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122. https://doi.org/10.2307/2291387

Berger, J., & Pericchi, L. (2004). Training samples in objective Bayesian model selection. *Annals of Statistics*, *32*, 841–869. https://doi.org/10.1214/009053604000000229

Böing-Messing, F., van Assen, M., Hofman, A., Hoijtink, H., & Mulder, J. (2017). Bayesian evaluation of constrained hypotheses on variances of multiple independent groups. *Psychological Methods*, *22*, 262–287. https://doi.org/10.1037/met0000116

Casella, G., Giron, F., & Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *Annals of Statistics*, *37*, 1207–1228. https://doi.org/10.1214/08-AOS606

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Conigliani, C., & O'Hagan, A. (2000). Sensitivity of the fractional Bayes factor to prior distributions. *Canadian Journal of Statistics*, *28*, 343–352. https://doi.org/10.2307/3315983

de Santis, F., & Spezzaferri, F. (1997). Alternative Bayes factors for model selection. *Canadian Journal of Statistics*, *25*, 503–515. https://doi.org/10.2307/3315344

de Santis, F., & Spezzaferri, F. (1999). Methods for default and robust Bayesian model comparison: The fractional Bayes factor approach. *International Statistical Review*, *67*, 1–20. https://doi.org/10.2307/1403706

de Santis, F., & Spezzaferri, F. (2001). Consistent fractional Bayes factor for nested normal linear models. *Journal of Statistical Planning and Inference*, *97*, 305–321. https://doi.org/10.1016/S0378-3758(00)00240-8

Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *Annals of Mathematical Statistics*, *42*, 204–223. https://doi.org/10.1214/aoms/1177693507

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of Econometrics*, *2*, 775–826. https://doi.org/10.1016/S1573-4412(84)02005-5

Fouskakis, D., Ntzoufras, I., & Draper, D. (2015). Power-expected-posterior priors for variable selection in Gaussian linear models. *Bayesian Analysis*, *10*, 75–107. https://doi.org/10.1214/14-BA887

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gilks, W. R. (1995). Discussion to fractional Bayes factors for model comparisons' (by O'Hagan). *Journal of the Royal Statistical Society, Series B*, *57*, 118–120.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*, 511–527. https://doi.org/10.1037/met0000017

Gu, X., Mulder, J., & Hoijtink, H. (2016). Error probabilities in default Bayesian hypothesis testing. *Journal of Mathematical Psychology*, *72*, 130–143. https://doi.org/10.1016/j.jmp.2015.09.001

Guber, D. (1999). Getting what you pay for: The debate over equity in public school expenditures. *Journal of Statistics Education*, *7*, 1–8.

Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman & Hall/CRC.

Howell, D. (2012). *Statistical methods for psychology* (8th ed.). Belmont, CA: Cengage Learning.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795. https://doi.org/10.2307/2291091

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Klugkist, I., Bullens, J., & Postma, A. (2012). Evaluating order constrained hypotheses for circular data using permutation tests. *British Journal of Mathematical and Statistical Psychology*, *65*, 222–236. https://doi.org/10.1111/j.2044-8317.2011.02018.x

Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*, 6367–6379. https://doi.org/10.1016/j.csda.2007.01.024

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Baysian approach. *Psychological Methods*, *10*, 447–493. https://doi.org/10.1037/1082-989X.10.4.477

Klugkist, I., Laudy, O., & Hoijtink, H. (2010). Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables. *Psychological Methods*, *15*, 281–299. https://doi.org/10.1037/a0020137

McCullogh, C. E., & Searle, S. R. (2001). *Generalized linear and mixed models*. New York, NY: John Wiley & Sons.

Moreno, E. (1997). Bayes factor for intrinsic and fractional priors in nested models: Bayesian robustness. In D. Yadolah (Ed.), *L1-statistical procedures and related topics* (pp. 257–270). Hayward, CA: Institute of Mathematical Statistics.

Moreno, E., Giron, F., & Casella, G. (2015). Posterior model consistency in variable selection as the model dimension grows. *Statistical Science*, *30*, 228–241. https://doi.org/10.1214/14-STS508

Morey, R., Wagenmakers, E.-J., & Rouder, J. (2016). "Calibrated" Bayes factors should not be used: A reply to Hoijtink, van Kooten, and Hulsker. *Multivariate Behavioral Research*, *51*, 11–19. https://doi.org/10.1080/00273171.2015.1052710

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, *67*, 153–171. https://doi.org/10.1111/bmsp.12013

Mulder, J. (2014b). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463. https://doi.org/10.1016/j.csda.2013.07.017

Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, *72*, 104–115. https://doi.org/10.1016/j.jmp.2014.09.004

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, *46*, 1–39. https://doi.org/10.18637/jss.v046.i02

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*, 887–906. https://doi.org/10.1016/j.jspi.2009.09.022

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W., Selfhout, M., & Hoi-jtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, *53*, 530–546. https://doi.org/10.1016/j.jmp.2009.09.003

Mulder, J., & Wagenmakers, E.-J. (2016). Editor's introduction to the special issue on 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments'. *Journal of Mathematical Psychology*, *72*, 1–5. https://doi.org/10.1016/j.jmp.2016.01.002

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons (with discussion). *Journal of the Royal Statistical Society, Series B*, *57*, 99–138.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, *6*, 101–118. https://doi.org/10.1007/BF02564428

Pérez, J., & Berger, J. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, *89*, 491–512. https://doi.org/10.1093/biomet/89.3.491

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t*-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Silvapulle, M., & Sen, P. (2004). *Constrained statistical inference: Order, inequality, and shape constraints*. New York, NY: John Wiley & Sons.

van de Schoot, R., Hoijtink, H., & Deković, M. (2010). Testing inequality constrained hypotheses in SEM models. *Structural Equation Modeling*, *19*, 1–17. https://doi.org/10.1080/10705511.2010.489010 van de Schoot, R., Hoijtink, H., Hallquist, M., & Boelen, P. (2012). Bayesian evaluation of inequality constrained hypotheses in SEM models using Mplus. *Structural Equation Modeling*, *17*, 443–463. https://doi.org/10.1080/10705511.2012.713267

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin and Review*, *14*, 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage Dickey method. *Cognitive Psychology*, *60*, 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E. J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio test. *Computational Statistics & Data Analysis*, *38*, 666–690. https://doi.org/10.1016/j.csda.2010.03.016

Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. DeGroot, D. V. Lindley & A. F. M. Smith (Eds.), *Bayesian statistics: Proceedings of the first in-ternational meeting held in Valencia* (pp. 585–603). Valencia, Spain: University of Valencia.

Received 8 November 2016; revised version received 5 January 2017

Appendix A: Adjusting the prior mean for range constraints

The specification of the prior mean for $\boldsymbol{\beta}_1 = \mathbf{R}_{i_1}\boldsymbol{\theta} - \mathbf{r}_{i_1}$ in range constrained hypotheses consists of two steps:

1. Find the range constraints in the hypotheses under investigation. A hypothesis contains range constraint(s) if there exist lines in $\mathbf{R}_{i_1}$ of which the sum is the zero vector. If there is more than one range constraint in the same hypothesis, then there are multiple sets of two or more lines that are added to zero. For example, the hypothesis $H_1 : 0 < \theta_1 < \theta_2 < 1$ with $\mathbf{R}_{i_1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$ and $\mathbf{r}_{i_1} = (0, 0, -1)^T$ contains a range constraint, because $(\sum_{k=1}^{3} \mathbf{R}_{i_1}(k, 1), \sum_{k=1}^{3} \mathbf{R}_{i_1}(k, 2)) = (0, 0)$.
2. Specify the prior mean of $\boldsymbol{\beta}_1 = \mathbf{R}_{i_1}\boldsymbol{\theta} - \mathbf{r}_{i_1}$ for the range constraints. $\boldsymbol{\beta}_1$ contains the elements related to the range constraints and other inequality constraints. The prior means for those elements of $\boldsymbol{\beta}_1$ that represent the edges of a range constraint are specified as $\beta_1^* = -\sum_{k=1}^{K} r_{i_1}(k)/2$ where K is the number of lines in $\mathbf{R}_{i_1}$ for each range constraint and $r_{i_1}(k)$ is the constant for this range constraint, whereas the prior means for other elements of $\boldsymbol{\beta}_1$ are $\mathbf{0}$, which is not different from that for equality and inequality constrained hypotheses. For example, for the hypothesis $H_1 : 0 < \theta_1 < \theta_2 < 1$ the edges of the range constraint are $\beta_{11} = \theta_1 > 0$ and $\beta_{13} = 1 - \theta_2 > 0$. Thus, β_{11} and β_{13} have prior means of .5, whereas $\beta_{12} = \theta_2 - \theta_1$ has a prior mean of 0.

Appendix B: BaIn user manual

The software package BaIn is written in Fortran 90 with the IMSL 5.0 numerical library. It computes Bayes factors to evaluate any informative hypotheses (Section 2) and compare pairs from a set of informative hypotheses if they are comparable (Section 3.3). BaIn can be freely downloaded from the website http://informative-hypotheses.sites.uu.nl/software/bain/. The downloaded folder consists of an executable file (BaIn.exe), an input file

(Input.txt), and an output file (Output.txt). Running BaIn.exe with Input.txt located in the same folder produces Output.txt. This appendix shows how to fill in Input.txt so that BaIn.exe can properly read the information. Input.txt mainly contains the estimates and covariance matrix of parameters $\boldsymbol{\theta}$ for prior and posterior specification, and the restriction matrix and constant vector for each informative hypothesis.

The repeated measures ANOVA example in Section 2.2 is used to illustrate the valid specification of input file. We will first display and then explain the context below from Input.txt when evaluating the informative hypotheses H_2 (11) and $H_{2'}$ (12).

```
1    #Number of parameters of interest; Number of informative hypotheses;
     Sample size
2    4 2 36
3    #Estimates of parameters
4    22.33 22 5.78 6.78
5    #Covariance matrix of parameters
6    5.18 4.86 2.61 2.86
7    4.86 5.13 2.90 3.03
8    2.61 2.90 1.93 1.97
9    2.86 3.03 1.97 2.39
10   #Numbers of equality and inequality constraints in H1
11   2 1
12   #Restriction matrix (R0|r0) for equality constraints
13   1 -1  0  0   0
14   0  0  1 -1   0
15   #Restriction matrix (R1|r1) for inequality constraints
16   0  1 -1  0   0
17   #Numbers of equality and inequality constraints in H2
18   0 3
19   #Restriction matrix (R0|r0) for equality constraints
20   #Restriction matrix (R1|r1) for inequality constraints
21   1 -1  0  0   0
22   0  1 -1  0   0
23   0  0  1 -1   0
```

The input text has strictly fixed structure. There are annotation lines starting with # below which the corresponding information (numbers) has to be given. The first line is the annotation for the number of structural parameters, number of informative hypotheses, and sample size, which means we need to write three numbers in the second line (i.e., 4, 2 and 9). Because the number of structural parameters is 4, four numbers for the estimates of parameters are presented in line 4, and a 4×4 covariance matrix is written in lines 6–9. Furthermore, because the number of informative hypotheses is 2, two hypotheses are specified. For the first hypothesis, line 11 specifies 2 and 1 for the numbers of equality and inequality constraints, respectively. Therefore, the augmented restriction matrix with constant vector for equality constraints has two rows shown in lines 13 and 14, and one row for inequality constraints in line 16. For the second hypothesis, the numbers of equality and inequality constraints are 0 and 3 given in line 18, respectively. As can be seen, there is no line with numbers immediately after line 19 because this hypothesis does not contain any equality constraints. In lines 21–23 the augmented restriction matrix for three inequality constraints is displayed.

The estimates and covariance matrix of structural parameters can be obtained from other statistical software, such as Mplus (Muthén & Muthén, 2010) and R package lavaan (Rosseel, 2012), and the augmented restriction matrix (R0|r0) and (R1|r1) can be specified based on the informative hypotheses under evaluation. Executing BaIn.exe with this information renders the relative complexities, fits and Bayes factors for informative hypotheses under different choices of *b* in Output.txt. The results for the repeated measures ANOVA example are as follows:

```
Result for H1

Equality constraints
   Fit    Complexity (b1)   Complexity (b2)   Complexity (b3)
  0.091        0.049             0.059             0.096

Inequality constraints (conditional on equality constraints)
   Fit    Complexity (b1)   Complexity (b2)   Complexity (b3)
  1.000        0.500             0.500             0.500

 Number of iterations
  3000          3000              3000              3000

 BF1u (b1=0.111)    BF1u (b2=0.167)    BF1u (b3=0.255)
      4.603              3.069              2.006

 BF1c (b1=0.111)    BF1c (b2=0.167)    BF1c (b3=0.255)
      4.603              3.069              2.006


 Result for H2

 Equality constraints
   Fit    Complexity (b1)   Complexity (b2)   Complexity (b3)
  1.000        1.000             1.000             1.000

 Inequality constraints (conditional on equality constraints)
   Fit    Complexity (b1)   Complexity (b2)   Complexity (b3)
  0.023        0.096             0.098             0.097
 Number of iterations
  46000         9000              9000              9000

 BF2u (b1=0.111)    BF2u (b2=0.167)    BF2u (b3=0.255)
      0.240              0.237              0.238

 BF2c (b1=0.111)    BF2c (b2=0.167)    BF2c (b3=0.255)
      0.223              0.219              0.220
```

The results contain the relative fits and complexities for both equality and inequality constraints, as well as the Bayes factors under different *b*s in each hypothesis. For equality constraints, the relative fit and complexity are the normal posterior and prior densities in (28), and thus can be directly computed. However, the computation of relative fit and complexity for inequality constraints is often difficult and needs to sample from the posterior and prior distributions using Markov chain Monte Carlo methods (Gu *et al.*, 2014). BaIn uses an efficient algorithm, which requires fewer iterations (displayed below fit and complexities) in the Markov chains to accurately estimate the relative fit and complexity. Note that the Bayes factor for informative hypothesis H_1 against H_2 can be computed using (34) with BF_{1u} and BF_{2u}.