



Bayesian evaluation of informative hypotheses for multiple populations

Herbert Hoijtink^{1*} , Xin Gu² and Joris Mulder³

¹Department of Methodology and Statistics, Utrecht University, The Netherlands

²Department of Educational Psychology, East China Normal University, Shanghai, China

³Department of Methodology and Statistics, Tilburg University, The Netherlands

The software package *Bain* can be used for the evaluation of informative hypotheses with respect to the parameters of a wide range of statistical models. For pairs of hypotheses the support in the data is quantified using the approximate adjusted fractional Bayes factor (BF). Currently, the data have to come from one population or have to consist of samples of equal size obtained from multiple populations. If samples of unequal size are obtained from multiple populations, the BF can be shown to be inconsistent. This paper examines how the approach implemented in *Bain* can be generalized such that multiple-population data can properly be processed. The resulting multiple-population approximate adjusted fractional Bayes factor is implemented in the R package *Bain*.

1. Introduction

This paper is the most recent addition to a sequence of papers in which an alternative to null hypothesis significance testing has been developed. Important landmarks in this development are Klugkist, Laudy, and Hoijtink (2005a) and Kuiper, Klugkist, and Hoijtink (2010) who added order constrained hypotheses to the classical null hypothesis and showed in the context of analysis-of-variance models how these can be evaluated using the Bayes factor (Kass & Raftery, 1995); Mulder, Hoijtink, and de Leeuw (2012) who generalized the approach to Bayesian evaluation of informative hypotheses (Hoijtink, 2012), that is, hypotheses specified using equality and inequality (or order) constraints among the parameters of multivariate normal linear models; Gu, Mulder, Dekovic, and Hoijtink (2014) who developed a Bayes factor for the evaluation of inequality-constrained hypotheses in a rather wide range of statistical models; and Mulder (2014) and Gu, Mulder, and Hoijtink (2018) who generalized the latter Bayes factor into the approximate adjusted fractional Bayes factor (AAFBF, henceforth abbreviated to BF) which can be used to evaluate informative hypotheses for one-population data for a wide range of statistical models such as normal linear models, logistic regression models, confirmatory factor analysis, and structural equation models.¹

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Correspondence should be addressed to Herbert Hoijtink, Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC, Utrecht, The Netherlands (email: H.Hoijtink@uu.nl).

¹ The interested reader is referred to <http://informative-hypotheses.sites.uu.nl/> where all the books, dissertations, papers, and software produced during the course of this development are presented.

The copyright line for this article was changed on 28 March 2019 after original online publication.

The BF is simple to compute and the only input needed are estimates of the model parameters, the corresponding covariance matrix, and the sample size. However, as will be discussed in this paper, the BF is inconsistent if samples of unequal size are obtained from multiple populations (similar to O'Hagan, 1995; fractional Bayes factor, as is shown by De Santis & Spezzaferrri, 2001). This paper examines how the BF can be generalized into the multiple-population approximate adjusted fractional Bayes factor (MBF). This Bayes factor is simple to compute too; the only input needed are estimates of the model parameters, separate estimates of the corresponding covariance matrix for each population, and the sample size obtained from each population. As will be shown, the MBF is consistent and can therefore be used for testing informative hypotheses with respect to multiple populations.

With the availability of the MBF (and corresponding software) researchers have a viable alternative for null hypothesis significance testing. In a wide range of statistical models, the null hypothesis can be replaced by informative hypotheses, and the p -values can be replaced by the MBF. The recent and current critical appraisal of null hypothesis significance testing in the literature will not be reiterated here. However, the interested reader is referred to Cohen (1994) who called the null hypothesis the nill hypothesis because he could not come up with research examples in which the null hypothesis might be a realistic representation of the population of interest. This point of view was further elaborated by Royal (1997, pp. 79–81) who claims that the null hypothesis cannot be true, and consequently, that data are not needed in order to be able to reject it. However, the interested reader is also referred to Wainer (1999) who highlights that there are situations where, without dispute, the null hypothesis is relevant. Landmark papers criticizing the use of p -values and significance levels are Ioannides (2005) and Wagenmakers (2007), among others. The latter paper also motivates and illustrates the replacement of p -values by Bayesian hypothesis testing using the Bayesian information criterion (Schwartz, 1978; Raftery, 1995). However, the interested reader is also referred to the American Statistical Association's statement on p -values (Wasserstein & Lazar, 2016) which gives a to-the-point and balanced overview of what can and cannot be done with p -values, and to Benjamin, Berger, Johannesson, Nosek, Wagenmakers, and Johnson (2018) who propose to redefine statistical significance.

The focus of this paper is on the evaluation of informative hypotheses using the Bayes factor. Note that model selection criteria like the Akaike and Bayesian information criteria (Raftery, 1995; Schwartz, 1978) cannot be used (Mulder, Klugkist, Meeus, van de Schoot, Selfhout, & Hoijtink, 2009; Section 3). The penalty for model complexity in both criteria is a function of the number of parameters in the model at hand. Since the number of parameters in an unconstrained hypothesis (e.g. $H_u : \theta_1, \theta_2, \theta_3$) is the same as in a constrained hypothesis (e.g. $H_1 : \theta_1 > \theta_2 > \theta_3$), it does not reflect the fact that H_1 is more parsimonious than H_u . This problem is solved by Kuiper and Hoijtink (2013) who present the generalized order-restricted information criterion (GORIC), which is a generalization of the Akaike information criterion with a penalty term that does properly reflect the fact that H_1 is more parsimonious than H_u . However, the GORIC can only be applied in the context of the multivariate normal linear model, while, as discussed above, the range of application of the (M)BF is not limited to the multivariate normal linear model.

Also, as is elaborated in Van de Schoot, Hoijtink, Romeijn, and Brugman (2012), the penalty for model complexity used by the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van der Linde, 2002) is also not suited to quantifying how parsimonious an informative hypothesis is. Using a modification of the loss function used by the DIC, they obtain the prior information criterion (PIC) which in the examples provided can be used to evaluate informative hypotheses. However, as was shown by

Mulder (2014), using the Bayes factor results in more desirable selection behaviour when testing constrained hypotheses than using the PIC.

Silvapulle and Sen (2004) show how so-called Type A testing problems (evaluating a null hypothesis against an informative hypothesis) and Type B testing problems (evaluating an informative hypothesis against an unconstrained hypothesis) can be evaluated using p -values in a wide range of statistical models. Those in favour of null hypothesis significance testing are well advised to consult this book and the R packages *restriktor* and *ic.infer*. The main limitation of this approach is that it cannot be used to directly compare two competing informative hypotheses.

Stern (2005) proposes using the posterior density of H_k for $k = 1, \dots, K$ to select the best hypothesis. However, as is elaborated in Klugkist, Laudy, and Hoijtink (2005b), this amounts to using f_k to select the best hypothesis, that is, the complexity c_k is ignored. This will work if each hypothesis has the same complexity. However, if, for example, H_u is compared to H_1 , irrespective of the data, H_u will always be preferred because it has by definition a larger fit than H_1 (cf. equation 7).

This paper starts by introducing the BF. Using a simple two-group setup, it will be shown and illustrated that it may show inconsistent behaviour if samples of unequal size are obtained from multiple populations. Subsequently, the BF will be generalized into the MBF and, using the same two-group setup, it will be shown and illustrated that the MBF does not exhibit inconsistent behaviour if samples of unequal size are obtained from multiple populations. Further illustrations of the approach proposed in the context of an analysis-of-covariance (ANCOVA) model and a logistic regression analysis will be provided. Illustrations are executed using the R package² *Bain*.³ The R codes and data used in this paper can be found at the bottom of the *Bain* website (click on the title of this paper). The paper is concluded with a short discussion and contains an Appendix with a further discussion of the consistency of the MBF.

2. The approximate adjusted fractional Bayes factor

Consider a model where θ is a vector of length J containing the structural parameters, and ω a scalar, vector, or matrix containing the nuisance parameters. Hypotheses can be formalized as

$$H_k : S_k \theta = s_k, R_k \theta > r_k, \quad \text{for } k = 1, \dots, K, \quad (1)$$

where S_k is a $p_k \times J$ matrix imposing p_k equality constraints on θ , R_k is a $q_k \times J$ matrix imposing q_k inequality constraints, and s_k and r_k are vectors containing constants of size p_k and q_k , respectively. Additionally of interest is the unconstrained hypothesis $H_u: \theta$, that is, a hypothesis without constraints on the parameters θ . As will be explained below, this hypothesis has a central role in the computation of the Bayes factor.

Mulder (2014), Gu *et al.* (2014), Gu (2016), and Gu *et al.* (2018) show that the relative support in the data for H_k and H_u can be quantified using the approximate adjusted fractional Bayes factor,

$$\text{BF}_{ku} = \frac{f_k}{c_k}, \quad (2)$$

² <https://www.r-project.org>

³ <https://informative-hypotheses.sites.uu.nl/software/bain/>

that is, the ratio of the fit and the complexity of H_k relative to H_u . The interested reader should consult the references given for the derivation of equation (2) from the ratio of the marginal likelihoods of H_k and H_u . The Bayes factor from equation (2) is a quantification of the relative support in the data for H_k against H_u . If, for example, $\text{BF}_{ku} = 5$, the support in the data is five times larger for H_k than for H_u . It will now be shown how the BF can be computed and why it is called the approximate adjusted fractional Bayes factor. It will also be highlighted that the BF is a member of the family of Bayes factors based on encompassing priors (Klugkist, Kato, & Hoijtink, 2005; Wetzels, Grasman, & Wagenmakers, 2010), that is, Bayes factors for which the prior distribution of the model parameters under H_k is derived from the prior distribution under H_u .

Before providing the formulas for f_k and c_k , it must be emphasized that the density of the data can be factored according to O'Hagan (1995) as

$$p(\mathbf{Y}|\theta, \omega, \mathbf{X}) = p(\mathbf{Y}|\theta, \omega, \mathbf{X})^{1-b} p(\mathbf{Y}|\theta, \omega, \mathbf{X})^b, \quad (3)$$

where \mathbf{Y} denotes the data that are modelled (e.g., the dependent variable in a multiple regression) and \mathbf{X} the data that are not modelled and considered to be fixed (e.g., the predictor variables in a multiple regression). The idea of fractional Bayes factors is to use a fraction b of the information in the likelihood function to specify the prior distribution. Usually the fraction b is chosen such that it corresponds to the size of a minimal training sample (Berger & Pericchi, 1996, 2004). For the evaluation of informative hypotheses we implemented in the R package *Bain* $b = J^*/N$, where J^* denotes the number of independent constraints in $[S_1, R_1, \dots, S_K, R_K]$ and N the sample size. This choice can be illustrated using a simple example. If $H_1: \theta_1 > \theta_2 > \theta_3$ and $H_2: \theta_1 = \theta_2 = \theta_3$, the number of independent constraints $J^* = 2$, that is, there are two underlying parameters that are combinations of the target parameters with respect to which hypotheses are formulated: $\theta_1 - \theta_2$ and $\theta_2 - \theta_3$. Our choice is motivated by the fact that in the normal linear model, the minimal training sample needed to obtain a proper posterior distribution is equal to the number of parameters. If, for example, a variable is modelled using a normal distribution with unknown mean μ and variance σ^2 , the minimum training sample needed to obtain a proper posterior based on the prior $b(\mu, \sigma^2) = 1/\sigma^2$ is 2 (cf. Berger & Pericchi, 2004, Example 1). If, a variable is a linear combination of two predictors with normal error, there are four parameters (intercept, two regression coefficients, residual variance) and, consequently, the minimum training sample equals 4.

Gu *et al.* (2014), Gu (2016), and Gu *et al.* (2018) show that, based on equation (3) and an improper uniform prior for θ , a large-sample approximation (see Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2013, Chapter 4) of the posterior distribution of θ under H_u can be obtained:

$$g_u(\theta|\mathbf{Y}, \mathbf{X}) \approx \mathcal{N}(\hat{\theta}, \hat{\Sigma}_\theta), \quad (4)$$

where $\hat{\theta}$ denotes the maximum likelihood estimate of θ and $\hat{\Sigma}_\theta$ the corresponding covariance matrix. Note that the 'approximate' in the name 'approximate adjusted fractional Bayes factor' reflects the fact that for its computation a normal approximation of the posterior distribution is used. An implication of the approximation is that the BF can only be used if a normal approximation to the posterior distribution of θ is reasonable. If the sample size is not too small (see below), this is the case with unbounded parameters such as means and regression coefficients as they appear in generalized linear models and

structural equation models. It is also the case for the fixed regression coefficients (the random effect would be treated as nuisance parameters) in, for example, two-level models. In the latter case, the sample size used is the number of level-two units (and not the number of observations of the dependent variable). This is not necessarily the case with naturally bounded parameters such as variances (naturally bounded to be positive) and probabilities (naturally bounded between 0 and 1), although even there, if the sample size is large, a normal approximation of the posterior distribution may be accurate. The interested reader is referred to Gu *et al.* (2014) who show that, for the evaluation of inequality-constrained hypotheses in the context of a multiple regression with two predictors, the difference between the approximate BF implemented in *Bain* and the corresponding non-approximate BF implemented in *Biems* (Mulder *et al.*, 2012) is negligible if the sample size is at least 20. They also show that inequality constrained hypotheses with respect to the probabilities in a two by two contingency table render an approximate BF that is very similar the the non-approximate BF presented by Klugkist, Laudy, and Hoijtink (2010) if the sample size is at least 40. Although these results give confidence in the performance of the approximate adjusted fractional Bayes factor, further research in the context of different models is needed in order to strengthen these results.

The prior distribution of θ has a covariance matrix which is based on a fraction b of the information in equation (3) and a mean

$$\theta_B \in \{S_k\theta_B = s_k, R_k\theta_B = r_k\}, \quad \text{for } k = 1, \dots, K, \tag{5}$$

that is, θ_B denotes a value of θ on the boundary of all the hypotheses under investigation (Mulder, 2014):

$$b_u(\theta | [\mathbf{Y}, \mathbf{X}]^b) = \mathcal{N}(\theta_B, \hat{\Sigma}_\theta/b), \tag{6}$$

where $[\mathbf{Y}, \mathbf{X}]^b$ stresses that the prior distribution is based on a fraction b of the information in the data. Note that θ_B is called the adjusted mean (Mulder, 2014) of the prior distribution, which explains the ‘adjusted’ in the name ‘approximate adjusted fractional Bayes factor’. As was shown by Mulder (2014), if, for example, $H_1 : \theta > 0$ is compared with $H_u : \theta$, it holds that the more the data support H_1 the smaller the support in the fractional Bayes factor for H_1 ! This phenomenon is addressed if the adjusted fractional Bayes factor is used, that is, if the prior mean is in agreement with equation (5), the more the data are in agreement with H_1 the larger the support in the adjusted fractional Bayes factor for H_1 (see Mulder, 2014, for further details). Note, furthermore, that $b_u(\theta | \cdot)$ is a so-called encompassing prior, that is, the prior distribution of θ under H_k is proportional to $b_u(\theta | \cdot) I_{\theta \in H_k}$, where the indicator function is 1 if the argument is true and 0 otherwise (Klugkist *et al.*, 2005a; Wetzels *et al.*, 2010).

There are situations in which there is no solution to equation (5). For example, if hypotheses are specified using range constraints, for example, $H_1 : |\theta| < 0.2$ (i.e. $H_1 : \theta > -0.2, \theta < 0.2$), there is no solution. *Bain* addresses this problem in the following manner: in equation (5) (and only in this equation) this (part of a) hypothesis is represented as $H_1 : \theta = 0$, that is, θ_B will be equal to the midpoint of the range specified. The rationale is that H_1 essentially implies that $\theta \approx 0$. Another example is given by the hypotheses $H_1 : \theta = 0$ and $H_2 : \theta > 2$. Although each of these hypotheses can be evaluated by itself, they cannot be compared using the approximate adjusted fractional Bayes factor because there is no solution to equation equation (5), that is, both hypotheses are not

compatible because $b_u(\cdot)$ is different for each hypothesis (Hoijtink, 2012; section 9.9.2.1.). Testing non-compatible hypotheses can be done using BIEMS (Mulder *et al.*, 2012) by instructing the program to use the same unconstrained prior for each of the hypotheses under consideration.

Based on equations (4) and (6), the relative fit and complexity from equation (2) are defined as

$$f_k = \int_{\theta \in H_k} g_u(\theta | \mathbf{Y}, \mathbf{X}) d\theta \approx \int_{\theta \in H_k} \mathcal{N}(\theta | \hat{\theta}, \hat{\Sigma}_\theta) d\theta, \quad (7)$$

and

$$c_k = \int_{\theta \in H_k} b_u(\theta | [\mathbf{Y}, \mathbf{X}]^b) d\theta = \int_{\theta \in H_k} \mathcal{N}(\theta | \theta_B, \hat{\Sigma}_\theta / b) d\theta, \quad (8)$$

respectively. The interested reader is referred to Gu (2016, Chapter 3) for the algorithms with which the fit and complexity are computed. The strength of the BF lies in its simplicity. Its computation is based only on maximum likelihood estimates and the corresponding asymptotic covariance matrix, and the choice of the fraction b , which is completely determined by the sample size N and the number of independent constraints J^* .

The approximate adjusted fractional Bayes factor, also in the paragraphs that follow abbreviated as BF, falls in the category of default, automatic, or pseudo Bayes factors because no priors have to be manually specified. Instead, the prior is automatically constructed using a small fraction of the data, while the remaining fraction is used for hypothesis testing, similar to the fractional Bayes factor (O'Hagan, 1995). The BF is coherent in the sense that $\text{BF}_{uk} = 1/\text{BF}_{ku}$ and $\text{BF}_{kk'} = \text{BF}_{ku}/\text{BF}_{k'u}$ (O'Hagan, 1997; sections 3.1 and 3.2). Note that these coherence properties do not necessarily hold for other default Bayes factors (O'Hagan, 1997; Robert, 2007; p. 240).

As further noted by Robert (2007, p. 242), a potential issue of the fractional Bayes factor, and therefore also of the BF in equation (2), is that there is no clear-cut procedure to choose the fraction b . We believe, however, that the use of a minimal fraction is reasonable as it results in a minimally informative default prior while maximal information in the data is used for hypothesis testing (Berger & Mortera, 1995). Furthermore, it has been shown that this choice results in consistent testing behaviour (Mulder, 2014; O'Hagan, 1995). Nevertheless, further research on the choice of b would strengthen the approach we present in this paper. The interested reader is referred to Gu, Hoijtink, and Mulder (2016), for one evaluation of the choice of b . Another potential issue highlighted by Robert (2007, p. 242) is that default Bayes factors can be computationally intensive. The BF procedure that is proposed here, however, is very easy to compute: only the maximum likelihood estimates, error covariance matrix and sample size are needed (Gu *et al.*, 2018).

Finally, it is important to note that default Bayes factors may behave as ordinary Bayes factors based on so-called intrinsic priors (Berger & Pericchi, 1996). Currently, however, intrinsic priors have not yet been explored for the BF. Although this too is a topic worthy of further research, from a pragmatic point of view it is more important to know whether the BF is consistent, that is, whether the support for the true hypothesis goes to infinity when the sample size grows to infinity. According to O'Hagan (1997; section 2.1),

if hypotheses are nested (in the cases we consider all hypotheses are nested within H_u) and if $b \rightarrow 0$ if $N \rightarrow \infty$ (which holds for our b), the fractional Bayes factor is consistent. However, De Santis and Spezzaferri (2001) show that the fractional Bayes factor may show inconsistent behaviour if data from multiple populations are sampled. Similarly, as shown in the next section, the BF is also inconsistent if the data are sampled from multiple populations. In line with the solution proposed by De Santis and Spezzaferri (2001) for the fractional Bayes factor, the MBF is an extension of the BF that is consistent when testing hypotheses in the case of multiple populations.

3. Consistency of the approximate adjusted fractional Bayes factor

The consistency of the (M)BF this will be discussed in terms of (M)BF $_{ku}$ if H_k is specified using only equality constraints. In this case a Bayes factor is called consistent if (M)BF $_{ku} \rightarrow \infty$ (or 0) when H_k (or H_u) is true and $N_g \rightarrow \infty$ at the same rate for all $g = 1, \dots, G$. If H_k is specified using only inequality constraints, it will be discussed in terms of (M)BF $_{kc}$, where H_c is the complement of H_k . In this situation, a Bayes factor is called consistent if BF $_{kc} \rightarrow \infty$ (or 0) when H_k (or H_c) is true, as $N_g \rightarrow \infty$ at the same rate for all $g = 1, \dots, G$. Both scenarios imply that the G populations are treated as one population from which a sample of increasing size (proportionally increasing the sample sizes from each of the G populations) is taken. Note that BF $_{kc} = \text{BF}_{ku}/\text{BF}_{cu}$, where the numerator and denominator can be computed using equation (2). Note, furthermore, that for hypotheses specified using only equality constraints $H_u = H_c$.

When BF $_{ku} \rightarrow \infty$ or BF $_{kc} \rightarrow \infty$ for the same limit, the Bayes factor is called inconsistent. Another form of inconsistency that will be considered in this paper is whether (M)BF $_{ku} \rightarrow \infty$ or 0, and (M)BF $_{kc} \rightarrow \infty$ or 0, as $N_g \rightarrow \infty$ for some populations but not all G populations. This situation applies if a sample of increasing size is obtained from some of the G populations while the sample size from the other populations remains fixed. De Santis and Spezzaferri (2001) showed for this limit that the fractional Bayes factor (O'Hagan, 1995) is inconsistent. In this section it will be illustrated, in line with De Santis and Spezzaferri (2001), that the same holds for the BF. In the next section the MBF will be introduced, which can be seen as an extension of the BF to multiple populations which avoids this form of inconsistency.

Example 1: Comparison of two independent means. Consider the following simple model:

$$y_i = \theta_1 D_{1i} + \theta_2 D_{2i} + \epsilon_i, \quad \text{with } \epsilon_i \sim \mathcal{N}(0, \omega), \tag{9}$$

where D_{1i} equals 1 for $i = 1, \dots, N_1$ and 0 otherwise, D_{2i} equals 1 for $i = N_1 + 1, \dots, N_1 + N_2$ and 0 otherwise (and, consequently, θ_1 and θ_2 denote the means in group 1 and group 2, respectively, and ω the residual variance), and N_1 and N_2 denote the sample sizes of groups 1 and 2, respectively, with $N = N_1 + N_2$. Connecting this notation to that of the previous section renders $\mathbf{Y} = \mathbf{y}$ and $\mathbf{X} = [\mathbf{D}_1, \mathbf{D}_2]$.

Consider testing $H_1 : \theta_1 = \theta_2$ against $H_c : \theta_1 \neq \theta_2$. Note that the marginal likelihood of H_c is equal to the marginal likelihood of the unconstrained hypothesis $H_u : \theta_1, \theta_2$ because $\theta_1 = \theta_2$ has zero probability assuming a bivariate normal prior for θ_1, θ_2 under H_u . For the exposition that follows we arbitrarily assume that $\hat{\omega} = 1$. The approximated unconstrained posterior and prior distribution of θ_1 and θ_2 from equations (4) and (6) are then given by

$$g_u(\theta_1, \theta_2 | \mathbf{y}, \mathbf{D}_1, \mathbf{D}_2) \approx \mathcal{N}\left(\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}, \begin{bmatrix} 1/N_1 & 0 \\ 0 & 1/N_2 \end{bmatrix}\right) \quad (10)$$

and

$$b_u(\theta_1, \theta_2 | [\mathbf{y}, \mathbf{D}_1, \mathbf{D}_2]^b) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/b \times 1/N_1 & 0 \\ 0 & 1/b \times 1/N_2 \end{bmatrix}\right), \quad (11)$$

respectively, where $b = J^*/N = 1/N$. Note that, with respect to H_1 , the prior means for θ are in agreement with equation (5).

If we write $\delta = \theta_1 - \theta_2$, then the BF is given by the Savage–Dickey density ratio (Dickey, 1971; Mulder, Hoijtink, & Klugkist, 2010; Wetzels *et al.*, 2010).

$$\text{BF}_{1u} = \frac{f_1}{c_1} = \frac{g_u(\delta = 0 | \mathbf{y}, \mathbf{D}_1, \mathbf{D}_2)}{b_u(\delta = 0 | [\mathbf{y}, \mathbf{D}_1, \mathbf{D}_2]^b)} = \frac{\mathcal{N}(0 | \hat{\delta}, N_1^{-1} + N_2^{-1})}{\mathcal{N}(0 | 0, (N_1 + N_2)(N_1^{-1} + N_2^{-1}))}.$$

Let us first of all consider the situation in which N_1 and N_2 go to ∞ at the same rate, that is, let $N_g = a_g n$, for some positive constant a_g , $g = 1$ or 2 , and let $n \rightarrow \infty$. If $\hat{\delta} = 0$, then

$$f_1 = \mathcal{N}(0 | 0, N_1^{-1} + N_2^{-1}) \rightarrow \infty, \quad (12)$$

if $n \rightarrow \infty$ and

$$c_1 = \mathcal{N}(0 | 0, (N_1 + N_2)(N_1^{-1} + N_2^{-1})) = \mathcal{N}(0 | 0, (a_1 + a_2)(a_1^{-1} + a_2^{-1})), \quad (13)$$

which is a constant independent of n . Equations (12) and (13) imply that $\text{BF}_{1u} \rightarrow \infty$ if $n \rightarrow \infty$, which is consistent. If $\hat{\delta} \neq 0$,

$$f_1 = \mathcal{N}(0 | \hat{\delta}, N_1^{-1} + N_2^{-1}) \rightarrow 0, \quad (14)$$

if $n \rightarrow \infty$ and c_1 remains as in equation (13). This implies that $\text{BF}_{1u} \rightarrow 0$ if $n \rightarrow \infty$, which is consistent.

Now if we fix N_1 and let $N_2 \rightarrow \infty$, then in the limit equation (14) reduces to

$$f_1 = \mathcal{N}(0 | \hat{\delta}, N_1^{-1}), \quad (15)$$

and the middle part of equation (13) reduces to

$$c_1 = \mathcal{N}(0 | 0, \infty) \rightarrow 0, \quad (16)$$

if $n \rightarrow \infty$. This implies that in the limit $\text{BF}_{1u} \rightarrow \infty$ also if H_u is true, which is inconsistent behaviour.

To get more insight into the (in)consistency, the BF was computed for various numerical examples in Tables 1, 2, and 3. In the case of support for H_0 we set $\hat{\theta}_1 = 0$ and $\hat{\theta}_2 = 0$, and in the case of support for H_u we set $\hat{\theta}_1 = -0.35$ and $\hat{\theta}_2 = 0.35$. In both situations we again let $\hat{\sigma}^2 = 1$. As can be seen in Table 1, when $N_1 = N_2$ and both

increase at the same rate, $BF_{1u} \rightarrow \infty$ if H_1 is true and $BF_{1u} \rightarrow 0$ if H_u is true, that is, the Bayes factor shows consistent behaviour. Table 2 shows that BF_{1u} also shows consistent behaviour if both sample sizes increase at the same rate if $N_1 \neq N_2$. However, as can be seen in Table 3, if there is support for H_1 , BF_{1u} increases if N_2 increases while N_1 remains fixed, but if there is support in the data for H_u , BF_{1u} at first decreases but then starts to increase, which implies that the evidence accumulates in the wrong direction. As N_2 keeps increasing BF_{1u} goes to infinity as shown above. This is a simple illustration of inconsistent behaviour of the BF where multiple populations are considered while the sample size does not increase for all populations. De Santis and Spezzaferrri (2001) show that this behaviour can also be observed for the fractional Bayes factor. The problem is caused by the fact that the prior variances of θ_1 and θ_2 are dependent on the sample sizes in both groups because $b = 1/N$ (Table 3). As N_2 increases, the fraction that is used to construct the default

Table 1. Investigation of consistent behaviour of the one-population Bayes factor (BF) and the multiple-population Bayes factor (MBF) in the case of support for H_1 ($\bar{x}_1 = (0, 0)$) or for H_u ($\bar{x}_2 = (-0.35, 0.35)$) in the case of equal sample sizes for both groups increasing at the same rate

		BF						MBF					
N_1	N_2	b	$\hat{\Sigma}_{11}/b$	$\hat{\Sigma}_{22}/b$	\bar{x}_1 BF_{1u}	\bar{x}_2 BF_{1u}	b_1	b_2	$\hat{\Sigma}_{11}/b_1$	$\hat{\Sigma}_{22}/b_2$	\bar{x}_1 MBF_{1u}	\bar{x}_2 MBF_{1u}	
10	10	.05	2	2	4.47	<i>1.31</i>	.05	.05	2	2	4.47	<i>1.31</i>	
25	25	.02	2	2	7.07	<i>0.33</i>	.02	.02	2	2	7.07	<i>0.33</i>	
50	50	.01	2	2	10.00	<i>0.02</i>	.01	.01	2	2	10.00	<i>0.02</i>	
100	100	.005	2	2	14.14	<i>0.00</i>	.005	.005	2	2	14.14	<i>0.00</i>	

Note. N_1 and N_2 denote the sample sizes in groups 1 and 2, respectively; b denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for groups 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variances of θ_1 and θ_2 from equation (11) and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variances of θ_1 and θ_2 from equation (23). The numbers in italics are referred to in the text.

Table 2. Investigation of consistent behaviour of the one-population Bayes factor (BF) and the multiple-population Bayes factor (MBF) in the case of support for H_1 ($\bar{x}_1 = (0, 0)$) or for H_u ($\bar{x}_2 = (-0.35, 0.35)$) in the case of unequal sample sizes for the two groups increasing at the same rate

		BF						MBF					
N_1	N_2	b	$\hat{\Sigma}_{11}/b$	$\hat{\Sigma}_{22}/b$	\bar{x}_1 BF_{1u}	\bar{x}_2 BF_{1u}	b_1	b_2	$\hat{\Sigma}_{11}/b_1$	$\hat{\Sigma}_{22}/b_2$	\bar{x}_1 MBF_{1u}	\bar{x}_2 MBF_{1u}	
10	50	.017	6	1.2	7.74	1.01	.05	.01	2	2	5.77	0.75	
25	125	.007	6	1.2	12.25	0.07	.02	.004	2	2	9.13	0.06	
50	250	.003	6	1.2	17.32	0.00	.01	.002	2	2	12.91	0.00	

Notes. N_1 and N_2 denote the sample sizes in groups 1 and 2, respectively; b denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for groups 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variances of θ_1 and θ_2 from equation (11) and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variances of θ_1 and θ_2 from equation (23).

Table 3. Investigation of consistent behaviour of the one-population Bayes factor (BF) and the multiple-population Bayes factor (MBF) in the case of support for H_1 ($\bar{\mathbf{x}}_1 = (0, 0)$) or for H_u ($\bar{\mathbf{x}}_2 = (-0.35, 0.35)$) in the case of unequal sample sizes where only the size of group 2 increases

		BF						MBF					
N_1	N_2	b	$\hat{\Sigma}_{11}/b$	$\hat{\Sigma}_{22}/b$	$\bar{\mathbf{x}}_1$ BF $_{1u}$	$\bar{\mathbf{x}}_2$ BF $_{1u}$	b_1	b_2	$\hat{\Sigma}_{11}/b_1$	$\hat{\Sigma}_{22}/b_2$	$\bar{\mathbf{x}}_1$ MBF $_{1u}$	$\bar{\mathbf{x}}_2$ MBF $_{1u}$	
10	10	.05	2	2	4.47	<i>1.31</i>	.05	.05	2	2	4.47	1.31	
10	25	.029	3.5	1.4	5.92	<i>1.03</i>	.05	.02	2	2	5.34	0.93	
10	50	.017	6.0	1.2	7.74	<i>1.01</i>	.05	.01	2	2	5.77	0.75	
10	100	.009	11	1.13	10.48	<i>1.13</i>	.05	.005	2	2	6.03	0.65	
10	200	.005	21	1.05	14.94	<i>1.41</i>	.05	.0025	2	2	6.17	<i>0.60</i>	
10	1000	.001	101	1.01	31.78	<i>2.81</i>	.05	.0005	2	2	6.29	0.56	

Notes. N_1 and N_2 denote the sample sizes in groups 1 and 2, respectively; b denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for groups 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variances of θ_1 and θ_2 from equation (11) and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variances of θ_1 and θ_2 from equation (11). The numbers in italics are referred to in the text.

prior for θ_1 also goes to zero even though the sample size of group 2 does not increase. This undesirable property can be avoided using population-specific fractions in line with Iwaki (1997), Berger and Pericchi (1998), De Santis and Spezzaferri (1999, 2001), and Mulder (2014). In the remainder of this paper it will be detailed how this can be done for the BF to obtain the MBF for multiple populations.

4. The approximate adjusted fractional Bayes factor for multiple populations

In this section the MBF will be introduced. The developments will be illustrated using the comparison of two independent means. Let $g = 1, \dots, G$, where G denotes the number of groups, and N_g the corresponding sample sizes. Let $\theta = [\theta_1, \dots, \theta_g, \dots, \theta_G, \eta]$, where θ_g denotes the structural parameters that are unique to group g and η the structural parameters that are shared by all the groups. Then, in line with De Santis and Spezzaferri (2001), the density of the data of the multiple population model can be factored as

$$p(\mathbf{Y}_1, \dots, \mathbf{Y}_G | \theta_1, \dots, \theta_G, \eta, \omega, \mathbf{X}_1, \dots, \mathbf{X}_G) = \prod_{g=1}^G p_g(\mathbf{Y}_g | \theta_g, \eta, \omega, \mathbf{X}_g)^{1-b_g} p_g(\mathbf{Y}_g | \theta_g, \eta, \omega, \mathbf{X}_g)^{b_g}, \tag{17}$$

where b_g denotes the fraction of the information in the likelihood for population g that will be used for the specification of the prior distribution.

Example 1, continued. The following notation will be used to denote which parts of the data belong to groups 1 and 2. The subscripts 1 and 2 in $\mathbf{y}_1, \mathbf{y}_2$ denote data sampled from populations 1 and 2, respectively. Analogously, the second subscript in $\mathbf{D}_{11}, \mathbf{D}_{12}$, and $\mathbf{D}_{21}, \mathbf{D}_{22}$, denotes data from populations 1 and 2, respectively. Using this notation, the density of the data for the comparison of two independent means can be factored as:

$$\begin{aligned}
 & p(\mathbf{y}_1, \mathbf{y}_2 | \theta_1, \theta_2, \sigma^2, \mathbf{D}_{11}, \mathbf{D}_{12}, \mathbf{D}_{21}, \mathbf{D}_{22}) \\
 &= p(\mathbf{y}_1 | \theta_1, \sigma^2, \mathbf{D}_{11}, \mathbf{D}_{21})^{1-b_1} f(\mathbf{y}_1 | \theta_1, \sigma^2, \mathbf{D}_{11}, \mathbf{D}_{21})^{b_1} \\
 &\quad \times p(\mathbf{y}_2 | \theta_2, \sigma^2, \mathbf{D}_{12}, \mathbf{D}_{22})^{1-b_2} f(\mathbf{y}_2 | \theta_2, \sigma^2, \mathbf{D}_{12}, \mathbf{D}_{22})^{b_2} \\
 &\propto \prod_{i=1}^{N_1} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_1)^2}{\sigma^2/(1-b_1)}\right) \prod_{i=1}^{N_1} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_1)^2}{\sigma^2/b_1}\right) \\
 &\quad \times \prod_{i=N_1+1}^{N_1+N_2} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_2)^2}{\sigma^2/(1-b_2)}\right) \prod_{i=N_1+1}^{N_1+N_2} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_2)^2}{\sigma^2/b_2}\right).
 \end{aligned} \tag{18}$$

The covariance matrix of the parameters in equation (17) can be obtained as a function of the observed or expected Fisher information matrix (the interested reader is referred to Efron and Hinkley (1978), for details of the relative (dis)advantages of both types of information). Using the observed Fisher information, this leads to

$$\begin{aligned}
 \hat{\Sigma}_{\theta, \eta, \omega} &= \begin{bmatrix} \hat{\Sigma}_{\theta_1} & \cdots & \hat{\Sigma}_{\theta_1, \theta_G} & \hat{\Sigma}_{\theta_1, \eta} & \hat{\Sigma}_{\theta_1, \omega} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\Sigma}_{\theta_1, \theta_G} & \cdots & \hat{\Sigma}_{\theta_G} & \hat{\Sigma}_{\theta_G, \eta} & \hat{\Sigma}_{\theta_G, \omega} \\ \hat{\Sigma}_{\theta_1, \eta} & \cdots & \hat{\Sigma}_{\theta_G, \eta} & \hat{\Sigma}_{\eta} & \hat{\Sigma}_{\eta, \omega} \\ \hat{\Sigma}_{\theta_1, \omega} & \cdots & \hat{\Sigma}_{\theta_G, \omega} & \hat{\Sigma}_{\eta, \omega} & \hat{\Sigma}_{\omega} \end{bmatrix} = -1 \times \begin{bmatrix} \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \theta_1} & 0 & 0 & \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \omega} \\ 0 & \cdots & 0 & \cdots & \cdots \\ 0 & 0 & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \theta_G} & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \omega} \\ \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \eta} & \cdots & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \eta \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \omega} & \cdots & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \omega} & \frac{\partial^2 \log p(\cdot)}{\partial \eta \partial \omega} & \frac{\partial^2 \log p(\cdot)}{\partial \omega \partial \omega} \end{bmatrix}^{-1} \\
 &= -1 \times \begin{bmatrix} \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \theta_1} & 0 & 0 & \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \omega} \\ 0 & \cdots & 0 & \cdots & \cdots \\ 0 & 0 & \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \theta_G} & \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \omega} \\ \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \eta} & \cdots & \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \eta} + \cdots + \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \eta} & \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \omega} + \cdots + \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p(\cdot)}{\partial \theta_1 \partial \omega} & \cdots & \frac{\partial^2 \log p(\cdot)}{\partial \theta_G \partial \omega} & \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \omega} + \cdots + \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \omega} & \frac{\partial^2 \log p_1(\cdot)}{\partial \omega \partial \omega} + \cdots + \frac{\partial^2 \log p_G(\cdot)}{\partial \omega \partial \omega} \end{bmatrix}^{-1},
 \end{aligned} \tag{19}$$

where each second-order derivative is to be evaluated using $[\hat{\theta}, \hat{\eta}, \hat{\omega}]$, that is, the unconstrained maximum likelihood estimates of the model parameters obtained using the full density of the data from equation (17). If the expected Fisher information is used, the expected value of each entry in the last part of equation (19) has to be taken. The corresponding normal approximation of the posterior distribution of the structural parameters is

$$g_u(\theta_1, \dots, \theta_G, \eta | \mathbf{Y}_1, \dots, \mathbf{Y}_G, \mathbf{X}_1, \dots, \mathbf{X}_G) \approx \mathcal{N}([\hat{\theta}_1, \dots, \hat{\theta}_G, \hat{\eta}], \hat{\Sigma}_{\theta, \eta}), \tag{20}$$

that is, the multiple-population counterpart of equation (4).

Note that $\hat{\Sigma}_{\theta, \eta}$ can be constructed using the observed Fisher information matrix for the parameters of each group:

$$\hat{\Sigma}_{\theta_g, \eta, \omega} = -1 \times \begin{bmatrix} \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \theta_g} & \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \eta} & \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \omega} \\ \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \eta} & \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \eta} & \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \omega} & \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \omega} & \frac{\partial^2 \log p_g(\cdot)}{\partial \omega \partial \omega} \end{bmatrix}^{-1}, \quad \text{for } g = 1, \dots, G, \quad (21)$$

where each second-order derivative is be evaluated using $\hat{\theta}, \hat{\eta}, \hat{\omega}$, that is, the maximum likelihood estimates of the model parameters obtained using the full density of the data displayed in equation (17) –not only using the data for group g . Analogously, equation (21) can be replaced by the corresponding expected Fisher information matrix. Comparing equation (21) for $g = 1, \dots, G$ to equation (19) shows that the former contains all the elements needed to construct the latter. This is important since the input for *Bain* consists of the covariance matrices for each group, from which *Bain* constructs the overall covariance matrix. As will be detailed in the next paragraph, these group-specific covariance matrices are needed in order to be able to construct the prior distribution based on a fraction b_g of the information of the data in each group.

Once $\hat{\Sigma}_{\theta_g, \eta, \omega}$ for $g = 1, \dots, G$ has been obtained it is straightforward to obtain the multiple-population counterpart of the prior distribution displayed in equation (6) which is based on a covariance matrix using a fraction b_g of the information in $\mathbf{Y}_g, \mathbf{X}_g$ for $g = 1, \dots, G$ (see, equation (17)). Using the mathematical rule that $\partial^2 \log p(v, w)^u \partial v \partial w = u \partial^2 \log p(v, w) \partial v \partial w$, it can be seen that

$$\begin{aligned} \hat{\Sigma}_{\theta_g, \eta, \omega}^{b_g} &= -1 \times \begin{bmatrix} \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \theta_g \partial \theta_g} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \theta_g \partial \eta} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \theta_g \partial \omega} \\ \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \theta_g \partial \eta} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \eta \partial \eta} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \theta_g \partial \omega} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \eta \partial \omega} & \frac{\partial^2 \log p_g(\cdot)^{b_g}}{\partial \omega \partial \omega} \end{bmatrix}^{-1} \\ &= -1 \times \begin{bmatrix} b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \theta_g} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \eta} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \omega} \\ b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \eta} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \eta} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \omega} \\ b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \theta_g \partial \omega} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \eta \partial \omega} & b_g \frac{\partial^2 \log p_g(\cdot)}{\partial \omega \partial \omega} \end{bmatrix}^{-1}. \end{aligned} \quad (22)$$

Reassembling these matrices (cf. equation (19)) renders

$$\hat{\Sigma}_{\theta, \eta, \omega}^b = -1 \times \begin{bmatrix} b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \theta_1} & 0 & 0 & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \eta} & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \omega} \\ 0 & \dots & 0 & \dots & \dots \\ 0 & 0 & b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \theta_G} & b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \eta} & b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \omega} \\ b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \eta} & \dots & b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \eta} & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \eta} + \dots + b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \eta} & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \omega} \\ b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \theta_1 \partial \omega} & \dots & b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \theta_G \partial \omega} & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \omega} & b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \omega \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \omega \partial \omega} \end{bmatrix}^{-1}, \quad (23)$$

where $\hat{\Sigma}_{\theta, \eta, \omega}^{\mathbf{b}}$ denotes a covariance matrix based on fractions $\mathbf{b} = [b_1, \dots, b_G]$ of the information in the data, rendering the multiple-population adjusted fractional prior distribution of the structural parameters

$$b_u(\theta_1, \dots, \theta_G, \eta | [\mathbf{Y}_1, \dots, \mathbf{Y}_G, \mathbf{X}_1, \dots, \mathbf{X}_G]^{\mathbf{b}}) = \mathcal{N}([\theta_{B,1}, \dots, \theta_{B,G}, \eta_B], \hat{\Sigma}_{\theta, \eta}^{\mathbf{b}}). \tag{24}$$

Note that $[\dots]^{\mathbf{b}}$ in equation (24) denotes a prior distribution based on fractions \mathbf{b} of the information in the data. It can be interpreted as a default prior that contains the information of group-specific data fractions, b_g , for the parameters of interest. Note, furthermore, that the subscript B in $\theta_{B,1}, \dots, \theta_{B,G}, \eta_B$ highlights that the prior means of the structural parameters are in agreement with equation (5), that is, centred on the boundary of the hypotheses specified.

The MBF is the counterpart of BF in equation (2) based on the multiple-population posterior and prior distributions displayed in equations (20) and (24):

$$\text{MBF}_{ku} = \frac{\int_{\theta \in H_k} g_u(\theta | \mathbf{Y}, \dots, \mathbf{Y}_G, \mathbf{X}_1, \dots, \mathbf{X}_G) d\theta}{\int_{\theta \in H_k} b_u(\theta | [\mathbf{Y}_1, \dots, \mathbf{Y}_G, \mathbf{X}_1, \dots, \mathbf{X}_G]^{\mathbf{b}}) d\theta} \approx \frac{\int_{\theta \in H_k} \mathcal{N}(\theta | \hat{\theta}, \hat{\Sigma}_{\theta}) d\theta}{\int_{\theta \in H_k} \mathcal{N}(\theta | \theta_B, \hat{\Sigma}_{\theta}^{\mathbf{b}}) d\theta} = \frac{f_k}{c_k}. \tag{25}$$

Example 1, continued. Estimates of $\theta_1, \theta_1, \sigma_2$ are easy to obtain. It is well known that, using the expected Fisher information, the counterpart of equation (20) for the example at hand is

$$\hat{\Sigma}_{\theta_g, \sigma^2} = \begin{bmatrix} \hat{\sigma}^2/N_g & 0 \\ 0 & -1 \times E\left(\frac{d^2 \log f(\cdot)}{d\sigma^2 d\sigma^2}\right)^{-1} \end{bmatrix}, \tag{26}$$

from which, using equation (19), it is straightforward to obtain that

$$\hat{\Sigma}_{\theta_1, \theta_2} = \begin{bmatrix} \hat{\sigma}^2/N_1 & 0 \\ 0 & \hat{\sigma}^2/N_2 \end{bmatrix}. \tag{27}$$

The counterpart of equation (24) for the example at hand has $\theta_B = [0, 0]$ and, applying equation (23),

$$\hat{\Sigma}_{\theta_1, \theta_2}^{\mathbf{b}} = \begin{bmatrix} \hat{\sigma}^2/b_1 \times 1/N_1 & 0 \\ 0 & \hat{\sigma}^2/b_2 \times 1/N_2 \end{bmatrix}. \tag{28}$$

With respect to the computation of equation (21) three situations can be distinguished.

Situation 1. The multivariate normal linear model with group-specific and joint parameters. In the multivariate normal linear model there are $z = 1, \dots, Z$ dependent variables and $p = 1, \dots, P$ predictors with regression coefficients β_{pz} (where the predictor attached to a possible intercept is a column of 1s):

$$\begin{aligned}
 y_{1i} &= \beta_{11}x_{1i} + \dots + \beta_{p1}x_{pi} + \epsilon_{1i}, \\
 &\quad \dots, \\
 y_{zi} &= \beta_{1z}x_{1i} + \dots + \beta_{pz}x_{pi} + \epsilon_{zi} \\
 &\quad \dots, \\
 y_{zi} &= \beta_{1z}x_{1i} + \dots + \beta_{pz}x_{pi} + \epsilon_{zi},
 \end{aligned}
 \tag{29}$$

where $\epsilon_{1i}, \dots, \epsilon_{zi} \sim \mathcal{N}([0, \dots, 0], \mathbf{\Omega})$. Multiple populations arise if two or more of the predictors are used to create groups. Two groups with group-specific intercepts are created if, for example, $x_{1i} = 1$ if person i is a member of group 1 and 0 otherwise and $x_{2i} = 1$ if person i is a member of group 2 and 0 otherwise. Group-specific regression coefficients can additionally be obtained if, for example, $x_{3i} = x_{*i}x_{1i}$ and $x_{4i} = x_{*i}x_{2i}$ (where x_{*i} denotes a continuous predictor for which group-specific regression coefficients are required), that is, the predictor x_{3i} gets a regression coefficient β_{3z} in group 1 and β_{4z} in group 2. With $Z = 1$ the model could be

$$y_i = \beta_1 + \beta_3x_{3i} + \epsilon_i, \tag{30}$$

for group 1, and

$$y_i = \beta_2 + \beta_4x_{3i} + \epsilon_i, \tag{31}$$

for group 2, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

For the multivariate normal linear model,

$$\hat{\Sigma}_{\theta_g, \eta} = \hat{\omega} \otimes [\mathbf{X}_g^t \mathbf{X}_g]^{-1}, \tag{32}$$

where θ_g contains all the group-specific coefficients, η contains the joint coefficients, ω is a matrix containing the covariance matrix of the residuals, and all the data for the predictors for group g are collected in \mathbf{X}_g . Using equation (1), hypotheses with respect to the structural parameters $\theta = \text{vec}(\mathbf{B})$, where \mathbf{B} is a $Z \times P$ matrix containing the regression coefficients β_{pz} , can be formulated. Later in this paper an ANCOVA model will be used in Example 2 to illustrate situation 1. Note that the R function `lm` can be used to estimate the parameters of the multivariate normal linear model.

Situation 2. Models with only group-specific parameters. When all of the parameters (including ω) in the density of the data are group specific, the covariance matrix in equation (19) will be block diagonal with one block for each group. Consequently, it is straightforward to use R packages tailored to the statistical model of interest to obtain estimates $\hat{\theta}_g$ for $g = 1, \dots, G$, and, for each group, the corresponding covariance matrix $\hat{\Sigma}_{\theta_g}$. Note that this does not apply to the example given for situation 1 (equations (30) and (31)) because σ^2 was *not* group specific. This would have applied if, in addition to the intercept and regression coefficient, σ^2 had been group specific too.

Situation 3. All other situations. In all other situations R packages can be used to obtain the estimates $\hat{\theta}, \hat{\eta}, \hat{\omega}$, but the equations rendering $\hat{\Sigma}_{\theta_g, \eta, \omega}$ based on $\hat{\theta}, \hat{\eta}, \hat{\omega}$ and $\mathbf{Y}_g, \mathbf{X}_g$ for $g = 1, \dots, G$ will either have to be programmed in R or obtained through the use of R packages like `numDeriv` which provides numerical approximations of second-order derivatives based on the log density of the data of the statistical model of interest. Later in this paper a logistic regression will be used in Example 3 to illustrate this situation. For users with

limited experience of statistical modelling and R, the third situation will be difficult to handle: the likelihood function of the statistical model at hand has to be formulated and *numDeriv* has to be used to estimate the covariance matrix for each group (and not the overall covariance matrix) using the overall groups estimates of the model parameters. Currently, one annotated example (a logistic regression model) is provided on the *Bain* website. Users requiring support in the context of other models can send an email to the first author of this paper with the request to add additional examples to the website.

5. Choosing b_g

In the case of one population based on Gu *et al.* (2018), the R package *Bain* uses $b = J^*/N$. The remaining question is how to choose b_g for $g = 1, \dots, G$ in the case of multiple populations. If the size of the sample obtained from each population is the same, it should not matter whether BF_{ku} or MBF_{ku} is used, that is, the equality $BF_{ku} = MBF_{ku}$ should hold. Computation of the covariance matrix displayed in equation (23) in the situation where $N_1 = \dots = N_G$ can be done using $b_1 = \dots = b_G = b$. Applying this to the penultimate diagonal entry of the covariance matrix renders

$$\begin{aligned} b_1 \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \eta} + \dots + b_G \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \eta} &= b \frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \eta} + \dots + b \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \eta} \\ &= b \left(\frac{\partial^2 \log p_1(\cdot)}{\partial \eta \partial \eta} + \dots + \frac{\partial^2 \log p_G(\cdot)}{\partial \eta \partial \eta} \right) \\ &= b \frac{\partial^2 \log p(\cdot)}{\partial \eta \partial \eta}, \end{aligned} \tag{33}$$

with

$$b = \frac{J^*}{N_1 + \dots + N_G} = \frac{1}{G} \times J^* \times \frac{G}{N_1 + \dots + N_G}.$$

Therefore a reasonable choice is

$$b_g = \frac{1}{G} \times J^* \times \frac{1}{N_g}. \tag{34}$$

This choice is in keeping with the concept of a minimal fraction from each population to construct an implicit default prior.

6. Consistency of the multiple population approximate adjusted fractional Bayes factor

De Santis and Spezzaferri (2001) show that their generalized fractional Bayes factor is consistent if $N \rightarrow \infty$, that is, if the N_g increase at the same rate (cf. De Santis & Spezzaferri, 2001; Theorem 4.1). It will be illustrated below, via a continuation of Example 1, that the same holds for the MBF, that is, if $N_g \rightarrow \infty$, for all g at the same rate, then $MBF_{ku} \rightarrow \infty$ (or 0) when H_k (or H_u) is true. A more general discussion of the consistency of the MBF is given in the Appendix. The example below also shows that the MBF avoids the inconsistent behaviour shown by the BF when fixing the sample size of one population while letting the sample size of the other population go to infinity.

Example 1, continued. Earlier in this example it was shown that the BF exhibits inconsistent behaviour if the sample size of one group is fixed while the sample size of the other group goes to infinity. When the MBF is used, the posterior distribution is unchanged and identical to equation (10). However, the prior distribution changes from equation (11) to

$$\begin{aligned}
 b_u(\theta_1, \theta_2 | [\mathbf{y}, \mathbf{D}_1, \mathbf{D}_2]^b) &\approx \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/b_1 \times 1/N_1 & 0 \\ 0 & 1/b_2 \times 1/N_2 \end{bmatrix}\right) \\
 &= \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right),
 \end{aligned}
 \tag{35}$$

because $b_1 = \frac{1}{2N_1}$ and $b_2 = \frac{1}{2N_2}$. As can be seen, the prior distribution in equation (35) is independent of N_1 and N_2 . This can be interpreted as the amount of prior information being independent of the sample size, which is a desirable property. Also note that the prior mean does not depend on the information in the data but is chosen to be in agreement with equation (5).

The MBF of H_1 against H_u is given by

$$\begin{aligned}
 \text{MBF}_{1u} &= \frac{f_1}{c_1} = \frac{g_u(\hat{\delta} = 0 | \mathbf{y}, \mathbf{D}_1, \mathbf{D}_2)}{b_u(\hat{\delta} = 0 | [\mathbf{y}, \mathbf{D}_1, \mathbf{D}_2]^b)} = \frac{\mathcal{N}(0 | \hat{\delta}, N_1^{-1} + N_2^{-1})}{\mathcal{N}(0 | 0, 4)} \\
 &= 2(N_1^{-1} + N_2^{-1})^{-\frac{1}{2}} \exp\left\{-\frac{\hat{\delta}^2}{2(N_1^{-1} + N_2^{-1})}\right\} \rightarrow 2N_1^{1/2} \exp\left\{-\frac{N_1 \hat{\delta}^2}{2}\right\},
 \end{aligned}
 \tag{36}$$

as $N_2 \rightarrow \infty$, where $g_u(\cdot)$ and $b_u(\cdot)$ are obtained based on equations (10) and (35), respectively. As can be seen, if $\hat{\delta} = 0$ and $n \rightarrow \infty$, then $f_1 \rightarrow \infty$ and c_1 is constant. This implies that $\text{MBF}_{1u} \rightarrow \infty$. If $\hat{\delta} \neq 0$ and $n \rightarrow \infty$, then $f_1 \rightarrow 0$ and c_1 is constant. This implies that $\text{MBF}_{1u} \rightarrow 0$. Stated otherwise, for $n \rightarrow \infty$, MBF_{1u} is consistent. Furthermore, if $N_2 \rightarrow \infty$ while N_1 is fixed, then if $\hat{\delta} = 0$, in the limit (see the last term of equation (36)) $\text{MBF}_{1u} = 2N_1^{1/2}$, which is larger than 1, that is, correctly expresses support for H_1 . Although for $N_2 \rightarrow \infty$ MBF_{1u} does not approach ∞ , this is reasonable behaviour and the inconsistent behaviour of the BF is avoided. If $\hat{\delta} \neq 0$ the limiting behaviour of MBF_{1u} is shown by the last term of equation (36). If, for example, $N_1 = 25$ and $\hat{\delta} = 0.1$, $\text{MBF}_{1u} = 8.8$, that is, H_1 is supported. This too is reasonable, because both the sample size of group 1 and the effect size are small and therefore the effect is not convincingly different from zero. If both are larger, for example, $N_1 = 49$ and $\hat{\delta} = 0.5$, $\text{MBF}_{1u} = 0.03$, that is, H_u is supported. As is illustrated, the degree support for or against H_1 is based on the sample size and the effect size. This too is reasonable behaviour and again the inconsistent behaviour of the BF is avoided.

As can be seen in the last two columns in the middle and right-hand panels of Tables 1 and 2, if both sample sizes are proportionally increasing, both BF_{1u} and MBF_{1u} show consistent behaviour in the sense that $(\text{M})\text{BF}_{1u} \rightarrow \infty$ if $\theta_1 = \theta_2$ and $(\text{M})\text{BF}_{1u} \rightarrow 0$ if $\theta_1 \neq \theta_2$. Note that, as required by our choice of b_g in equation (34), for equal sample sizes in both groups both Bayes factors are equal (see Table 1).

Furthermore, as can be seen in the last two columns in the middle and right-hand panels of Table 3, if one sample size is fixed and the other is increasing, in contrast to BF_{1u} , MBF_{1u} does not show inconsistent behaviour in the sense that MBF_{1u} is monotonically increasing if $\hat{\theta}_1 = \hat{\theta}_2$ and MBF_{1u} is monotonically decreasing if $\hat{\theta}_1 \neq \hat{\theta}_2$. As can be seen, in

this situation, when only N_2 is increased, MBF_{1u} converges to the upper bound 6.325 (0.546) when $\theta_1 = \theta_2$ ($\theta_1 \neq \theta_2$) based on the limit in equation (36).

As can be seen by comparing the last number on the last row in Table 1 ($N = 200$, $N_1 = N_2 = 100$) with the last number on the penultimate row in Table 3 ($N = 210$, $N_1 = 10$, $N_2 = 200$), it makes a huge difference in outcome whether or not the sample sizes are balanced. Evidence in favor of the true hypothesis is larger with balanced than with unbalanced sample sizes.

7. Example 2: Analysis of covariance

Consider the ANCOVA model

$$y_i = \theta_1 D_{1i} + \dots + \theta_5 D_{5i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad \text{with } \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (37)$$

where D_{1i} is equal to 1 if person i is a member of group 1 and 0 otherwise, the other dummy variables are defined analogously, and both covariates are centred such that $\theta_1, \dots, \theta_5$ denote the covariate adjusted means. Equation (37) can be split into five parts, one for each group:

$$y_i = \theta_g + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad (38)$$

for $g = 1, \dots, G$ and N_g persons in each of the groups. Note that $\theta_1 = \theta_1, \dots, \theta_5 = \theta_5$, $\eta = [\beta_1, \beta_2]$, $\omega = \sigma^2$, and $\mathbf{Y}_g = \mathbf{y}_g$, $\mathbf{X}_g = [\mathbf{D}_{gg}, \mathbf{x}_{1g}, \mathbf{x}_{2g}]$, where the second subscript g denotes that the data correspond to the members of group g .

Applying equation (17), the density of the data of this model can be factored as

$$\begin{aligned} p(\mathbf{Y}_1, \dots, \mathbf{Y}_5 | \theta_1, \dots, \theta_5, \eta, \omega, \mathbf{X}_1, \dots, \mathbf{X}_5) &= p_1(\mathbf{Y}_1 | \theta_1, \eta, \omega, \mathbf{X}_1)^{1-b_1} p_1(\mathbf{Y}_1 | \theta_1, \eta, \omega, \mathbf{X}_1)^{b_1} \\ &\times \dots \times p_5(\mathbf{Y}_5 | \theta_5, \eta, \omega, \mathbf{X}_5)^{1-b_5} p_5(\mathbf{Y}_5 | \theta_5, \eta, \omega, \mathbf{X}_5)^{b_5} \\ &\propto \prod_{i=1}^{N_1} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_1 - \beta_1 x_{1i} - \beta_2 x_{2i})^2}{\sigma^2 / (1 - b_1)}\right) \prod_{i=1}^{N_1} \exp\left(-\frac{1}{2} \frac{(y_i - \theta_1 - \beta_1 x_{1i} - \beta_2 x_{2i})^2}{\sigma^2 / b_1}\right) \times \dots \\ &\times \prod_{i=N_4+1}^N \exp\left(-\frac{1}{2} \frac{(y_i - \theta_5 - \beta_1 x_{1i} - \beta_2 x_{2i})^2}{\sigma^2 / (1 - b_5)}\right) \prod_{i=N_4+1}^N \exp\left(-\frac{1}{2} \frac{(y_i - \theta_5 - \beta_1 x_{1i} - \beta_2 x_{2i})^2}{\sigma^2 / b_5}\right). \end{aligned} \quad (39)$$

Maximum likelihood estimates of the parameters $\theta_1, \dots, \theta_5, \beta_1, \beta_2, \sigma^2$ of the ANCOVA model from equation (37) can, for example, be obtained using the `lm` function from the R package. Subsequently, using a well-known result from the regression literature, the realization of equation (32) for $g = 1, \dots, 5$ is obtained as

$$\hat{\Sigma}_{\theta_g, \beta_1, \beta_2} = \hat{\sigma}^2 \times [\mathbf{X}_g^t \mathbf{X}_g]^{-1}, \quad (40)$$

where $\mathbf{X}_g = [\mathbf{D}_{gg}, \mathbf{x}_{1g}, \mathbf{x}_{2g}]$, in which the second subscript indicates to which group the data elements belong. Equation (40) is obtained using the expected Fisher information. Since the expected value of the second-order derivatives with respect to either $\theta_1, \dots, \theta_5, \beta_1, \beta_2$ on the one hand, or σ^2 on the other hand, are zero, $\hat{\Sigma}_{\theta_1, \dots, \theta_5, \beta_1, \beta_2}$ constructed using equation (19) based on the expected Fisher information for only these parameters is identical to the corresponding part in $\hat{\Sigma}_{\theta_1, \dots, \theta_5, \beta_1, \beta_2, \sigma^2}$ (cf. equations (26) and (27)).

Inverting and multiplication by -1 of $\hat{\Sigma}_{\theta_g, \beta_1, \beta_2}$, for $g = 1, 2$, renders the Fisher information matrices for the groups. Using equation (19), these can be assembled into the overall Fisher information matrix which, after inverting and multiplication by -1 , renders $\hat{\Sigma}_{\theta_1, \theta_2, \beta_1, \beta_2}$. Modifying equation (40) according to equation (22) renders

$$\hat{\Sigma}_{\theta_g, \beta_1, \beta_2}^{b_g} = \hat{\sigma}^2 / b_g \times [\mathbf{X}_g' \mathbf{X}_g]^{-1}, \tag{41}$$

that is, the elements of the expected Fisher information matrix for each group g . Reassembling these elements using equation (23) renders $\hat{\Sigma}_{\theta_1, \theta_2, \beta_1, \beta_2}^b$, that is, the covariance matrix of the prior distribution.

This example concludes using data from Stevens (1996, appendix A) concerning the effect of the first year of the *Sesame Street* series on the knowledge of 240 children in the age range 34–69 months. We will use the following variables: \mathbf{y} , knowledge of numbers after watching *Sesame Street*; \mathbf{x}_1 , the knowledge of numbers before watching *Sesame Street*; \mathbf{x}_2 , a test measuring the mental age of children; and $\mathbf{D}_1, \dots, \mathbf{D}_5$ dummy variables representing the children’s background (1 = disadvantaged inner city, 2 = advantaged suburban, 3 = advantaged rural, 4 = disadvantaged rural, 5 = disadvantaged Spanish-speaking).

The informative hypotheses of interest are

$$H_1 : \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 \tag{42}$$

and

$$H_2 : \{\theta_2, \theta_3\} > \{\theta_1, \theta_4, \theta_5\}. \tag{43}$$

Hypothesis 1 states that the knowledge of numbers after watching *Sesame Street* does not depend on background correcting for initial knowledge and mental age. Hypothesis 2 states that the advantaged children have a greater knowledge after watching *Sesame Street* than the disadvantaged children.

Table 4 presents the input the R package *Bain* needs in order to evaluate H_1 and H_2 , that is, estimates of the adjusted means, regression coefficients, and residual variance, and, for each group, the covariance matrix for the group-specific adjusted mean and both regression coefficients, computed using $\hat{\sigma}^2$ (cf. equation (40)), and the sample size. Table 5 first of all presents the posterior covariance matrix of the structural parameters computed from the group-specific covariance matrices using equation (19). Then the vector \mathbf{b} computed using $b_g = 1/5 \times 4/N_g$ for $g = 1, \dots, G$ is presented. Note that J^* equals 4 because the number of independent constraints in

$$S_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \tag{44}$$

which specifies the equality constraints in H_1 and

Table 4. A five-group ANCOVA model: Input for the R package *Bain*

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
29.16	34.38	28.90	27.12	30.89	0.70	0.05	84.06
N_1	N_2	N_3	N_4	N_5			
60	55	64	43	18			
$\hat{\Sigma}_{\hat{\theta}_1, \beta_1, \beta_2}$				$\hat{\Sigma}_{\hat{\theta}_2, \beta_1, \beta_2}$			
1.62	-0.05	0.05		3.07	-0.01	-0.10	
-0.05	0.02	-0.01		-0.01	0.02	-0.01	
0.05	-0.01	0.01		-0.10	-0.01	0.01	
$\hat{\Sigma}_{\hat{\theta}_3, \beta_1, \beta_2}$				$\hat{\Sigma}_{\hat{\theta}_4, \beta_1, \beta_2}$			
2.32	0.07	0.09		2.21	0.04	0.04	
0.07	0.03	-0.01		0.04	0.03	-0.09	
0.09	-0.01	0.01		0.04	-0.01	0.01	
$\hat{\Sigma}_{\hat{\theta}_5, \beta_1, \beta_2}$							
5.47	0.20	-0.20					
0.20	0.09	-0.05					
-0.20	-0.05	0.05					

Table 5. A five-group ANCOVA model: Output from the R package *Bain*

$\hat{\Sigma}_{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \beta_1, \beta_2}$							
1.45	-0.09	0.03	0.02	-0.04	-0.01	0.01	
-0.09	1.96	-0.24	-0.12	0.14	0.01	-0.03	
0.03	-0.24	1.46	0.07	-0.07	0.00	0.02	
0.02	-0.12	0.07	1.99	-0.03	0.00	0.01	
-0.04	0.14	-0.07	-0.03	4.72	0.01	-0.01	
-0.01	0.01	0.00	0.00	0.01	0.01	-0.00	
0.01	-0.03	0.02	0.01	-0.01	-0.00	0.00	
b							
0.013	0.015	0.012	0.019	0.044			
$\hat{\Sigma}^b_{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \beta_1, \beta_2}$							
108.14	-5.52	2.02	0.89	-2.66	-0.85	0.67	
-5.52	129.41	-13.40	-6.77	7.90	0.46	-1.82	
2.02	-13.40	113.04	4.09	-3.86	0.17	0.85	
0.89	-6.77	4.09	107.20	-1.89	0.14	0.41	
-2.66	7.90	-3.86	-1.89	108.07	0.51	-0.72	
-0.85	0.46	0.17	0.14	0.51	0.31	-0.14	
0.67	-1.82	0.85	0.41	-0.72	-0.14	0.17	
MBF _{1u}	MBF _{2u}	MBF ₁₂					
2.94	1.34	2.21					

Note. The number in italics is referred to in the text.

$$R_2 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \tag{45}$$

which specifies the inequality constraints in H_2 , is equal to 4, that is, the number of independent rows in the combination of S_1 and R_2 is equal to 4. Next, the prior covariance matrix of the structural parameters computed using the group-specific covariance matrices and \mathbf{b} and equations (22) and (23) is displayed. Finally, MBF_{1u} , MBF_{2u} , and MBF_{12} are presented. As can be seen, the support in the data is 2.21 times greater for H_1 than for H_2 , that is, it is slightly more likely that the gain in knowledge of numbers is equal for advantaged and disadvantaged children than that the gain is greater for the advantaged children. More data would be needed to obtain a more decisive conclusion.

8. Example 3: Logistic regression

Example 2 illustrated how $\hat{\Sigma}_{\theta_g, \eta}$ for $g = 1, \dots, G$ can be computed if the statistical model at hand is a member of the (multivariate) normal linear model (previously labelled *situation 1*). In this section it will be illustrated how $\hat{\Sigma}_{\theta_g, \eta}$ for $g = 1, \dots, G$ can be obtained for models outside the (multivariate) normal linear modelling framework (previously labelled *situation 3*) based on the observed Fisher information using the R package *numDeriv*.⁴

Again using the data from Stevens (1996; appendix A), a logistic regression model is specified in which y_i , whether a child is encouraged to watch *Sesame Street* (0 = no, 1 = yes), is predicted from gender (D_{1i} equals 1 for a girl and zero otherwise, D_{2i} equals 1 for a boy and zero otherwise), and centred age \mathbf{x} :

$$p(y_i | D_{1i}, D_{2i}, \mathbf{x}_i, \theta_1, \theta_2, \beta) = \frac{\exp(\theta_1 D_{1i} + \theta_2 D_{2i} + \beta \mathbf{x}_i)}{1 + \exp(\theta_1 D_{1i} + \theta_2 D_{2i} + \beta \mathbf{x}_i)}, \quad \text{for } i = 1, \dots, N. \quad (46)$$

The hypothesis of interest is

$$H_1 : \theta_1 > \theta_2, \beta > 0, \quad (47)$$

that is, girls are more encouraged than boys and older children are more encouraged than younger children.

The top part of Table 6 presents the input the R package *Bain* needs in order to evaluate H_1 . Note that $\hat{\Sigma}_{\theta_1, \beta}$ and $\hat{\Sigma}_{\theta_2, \beta}$ are computed using the observed Fisher information matrix rendered by the R package *numDeriv* using the data for group 1 ($\mathbf{D}_{1g}, \mathbf{x}_g$) and group 2 ($\mathbf{D}_{2g}, \mathbf{x}_g$), respectively. In the bottom part of Table 6 the output resulting from *Bain* is presented. It can be observed that H_1 is not supported by the data, with $\text{MBF}_{1u} = 0.53$. Note that, for the example at hand, $\hat{\Sigma}_{\theta_1, \theta_2, \beta}$ computed using the observed Fisher information matrix is virtually identical to $\hat{\Sigma}_{\theta_1, \theta_2, \beta}$ computed using the expected Fisher information matrix with the R package *glm*. Note, however, that this does not always have to be the case. Researchers preferring the expected Fisher information matrix (but see Efron & Hinkley, 1978) will have to replace the computations with *numDeriv* by formulae for the expected Fisher information for logistic regression models (see, for example, McCullagh & Nelder, 1989, pp. 115–117).

⁴<https://cran.r-project.org/web/packages/numDeriv/>

Table 6. A two-group logistic regression

Input for the R package <i>Bain</i>				
$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\beta}$
0.50		0.60		-0.01
N_1		N_2		
125		115		
$\hat{\Sigma}_{\theta_1, \beta}$			$\hat{\Sigma}_{\theta_2, \beta}$	
0.03		0.00	0.04	-0.00
0.00		0.00	-0.00	0.00
Output from the R package <i>Bain</i>				
$\hat{\Sigma}_{\theta_1, \theta_2, \beta}$				
0.03		-0.00	0.00	
-0.00		0.04	-0.00	
0.00		-0.00	0.00	
b				
0.008		0.009		
$\hat{\Sigma}^b_{\theta_1, \theta_2, \beta}$				
4.28		-0.02	0.03	
-0.02		4.39	-0.04	
0.03		-0.04	0.06	
MBF_{1u}				
0.53				

Note. The number in italics is referred to in the text. Also $\hat{\Sigma}_{\theta_1, \theta_2, \beta}$ does not change when computed using the expected Fisher information.

9. Discussion

In this paper the approximate adjusted fractional Bayes factor BF, which is suited for the evaluation of informative hypotheses if data are sampled from one population, has been generalized to the multiple population approximate adjusted fractional Bayes factor MBF, which is suited for the evaluation of informative hypotheses if data are sampled from one or multiple populations. Both BF and MBF are implemented in the R package *Bain*.

The result is a versatile and generally applicable approach for the evaluation of informative hypotheses by means of the Bayes factor in a wide range of statistical models. However, as mentioned earlier in the paper, there are number of topics that deserve further research. The first topic is which sample sizes are required to obtain an accurate normal approximation of the posterior distribution for a wide range of statistical models. The second topic concerns the choice of **b**, that is, what are the properties of our proposal and what are potential alternatives (the interested reader is referred to Gu *et al.* (2016) for one study on this topic). The third topic is further development of *Bain* such that it is easier for users to deal with, what was previously called situation 3, that is, models for which *numDeriv* or other approaches have to be used to obtain the covariance matrix of the parameters of interest for each of the groups in the data set. The fourth topic is more philosophical in nature. It concerns the question whether there is an intrinsic Bayes factor corresponding to our MBF. The fifth topic concerns a modification of the approach presented in this paper such that it can be applied in variable selection problems (see, for example, O’Hara & Sillanpaa, 2009). The spike-and-slap prior is known to perform well in

variable selection problems with sparse data, for example, regression models with a relatively large number of persons to number of predictors ratio, and in which only a few predictors are expected to have a substantial regression coefficient. Spike-and-slab prior-based variable selection is currently an exploratory approach. In the future we will consider a more confirmatory approach based on an efficient evaluation of sets of informative hypotheses in which it is considered not only if the regression coefficient is substantial, but also its direction, and (partial) orderings of regression coefficients.

Acknowledgments

The first author 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 no. 024.001.003). The third author is supported by a NWO Vidi Grant (number 452-17-006).

References

- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, A., Wagenmakers, E.-J., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Berger, J. O., & Mortera, J. (1995). Discussion to A. Oagan (1995). *Journal of the Royal Statistical Society Series B*, *57*, 99–138.
- Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122. <https://doi.org/10.1080/01621459.1996.10476668>
- Berger, J. O., & Pericchi, L. R. (1998). *On criticisms and comparisons of default Bayes factors for model selection and hypothesis testing*. Proceedings of the Workshop on Model Selection. Bologna, Italy: Pitagora.
- Berger, J. O., & Pericchi, L. R. (2004). Training samples in objective Bayesian model selection. *Annals of Statistics*, *32*, 841–869. <https://doi.org/10.1214/009053604000000229>
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *9*, 997–1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- De Santis, F., & Spezzaferrri, F. (1999). Methods for robust and default Bayesian model comparison: The fractional Bayes factor approach. *International Statistical Review*, *67*(3), 267–286.
- De Santis, F., & Spezzaferrri, 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](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>
- Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*, 457–483. <https://doi.org/10.1093/biomet/65.3.457>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC.
- Gu, X. (2016). *Bayesian evaluation of informative hypotheses*. Doctoral dissertation, University of Utrecht. Retrieved from <http://informative-hypotheses.sites.uu.nl/books-and-theses/>
- Gu, X., Hoijtink, H., & Mulder, J. (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>
- Gu, X., Mulder, J., Dekovic, 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. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, *71*, 229–261. <https://doi.org/10.1111/bmsp.12110>
- Hoijtink, H. (2012). *Informative hypotheses. Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman and Hall/CRC.
- Ioannides, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Iwaki, K. (1997). Posterior expected marginal likelihood for comparison of hypotheses. *Journal of Economics Asia University*, *21*, 105–134.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*, 57–69. <https://doi.org/10.1111/j.1467-9574.2005.00279.x>
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005a). Bayesian eggs and Bayesian omelettes: Reply to Stern (2005). *Psychological Methods*, *10*, 500–503. <https://doi.org/10.1037/1082-989X.10.4.500>
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005b). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*, 477–493. <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>
- Kuiper, R., & Hoijtink, H. (2013). A Fortran 90 program for the generalization of the order-restricted information criterion. *Journal of Statistical Software*, *54*, 1–19. <https://doi.org/10.18637/jss.v054.i08>
- Kuiper, R., Klugkist, I., & Hoijtink, H. (2010). A Fortran 90 program for confirmatory analysis of variance. *Journal of Statistical Software*, *34*, 1–31. <https://doi.org/10.18637/jss.v034.i08>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, UK: Chapman & Hall.
- Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics and Data Analysis*, *71*, 448–463. <https://doi.org/10.1016/j.csda.2013.07.017>
- 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*, 2. <https://doi.org/10.18637/jss.v046.i02>
- Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: 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., Meeus, W., van de Schoot, A., Selfhout, M., & Hoijtink, 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>
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society, Series B*, *57*, 99–138. <http://www.jstor.org/stable/2346088>
- O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, *6*, 101–118. <https://doi.org/10.1007/bf02564428>
- O'Hara, R. B., & Sillanpaa, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, *4*, 85–118. <https://doi.org/10.1214/09-BA403>
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–163). Washington, DC: American Sociological Association.
- Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. New York, NY: Springer.
- Royal, R. (1997). *Statistical evidence. A likelihood paradigm*. London, UK: Chapman & Hall.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Silvapulle, M. J., & Sen, P. K. (2004). *Constrained statistical inference: Order, inequality and shape constraints*. London, UK: Wiley.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, *64*, 583–639. <https://doi.org/10.1111/1467-9868.00353>

Stern, H. S. (2005). Model inference or model selection: Discussion of Klugkist, Laudy, and Hoijtink (2005). *Psychological Methods*, *10*, 494–499. <https://doi.org/10.1037/1082-989X.10.4.494>

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum.

Van de Schoot, R., Hoijtink, H., Romeijn, J.-W., & Brugman, D. (2012). A prior predictive loss function for the evaluation of inequality constrained hypotheses. *Journal of Mathematical Psychology*, *56*, 13–23. <https://doi.org/10.1016/j.jmp.2011.10.001>

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>

Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods*, *4*, 212–213. <https://doi.org/10.1037/1082-989X.4.2.212>

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: Context, process, and purpose. *American Statistician*, *70*, 129–133. <https://doi.org/10.1080/00031305.2016.1154108>

Wetzels, R., Grasman, R. P. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage-Dickey density ratio. *Computational Statistics and Data Analysis*, *54*, 2094–2102. <https://doi.org/10.1016/j.csda.2010.03.016>

Received 5 October 2017; revised version received 10 July 2018

Appendix I

Further discussion of the consistency of the MBF

We consider two different cases. First, we consider the case where hypothesis H_k only contains inequality constraints, and no equality constraints. Second, we consider the case where H_k contains (only) equality constraints. We will discuss both $N \rightarrow \infty$ and one or more but not all of the $N_g \rightarrow \infty$.

As was shown in this paper,

$$MBF_{ku} = \frac{f_k}{c_k} = \frac{\int_{\theta \in H_k} \mathcal{N}(\theta | \hat{\theta}, \hat{\Sigma}_\theta) d\theta}{\int_{\theta \in H_k} \mathcal{N}(\theta | \theta_B, \hat{\Sigma}_\theta^b) d\theta}. \tag{48}$$

If H_k only contains inequality constraints, that is, $H_k : R_k \theta > r_k$, MBF reduces to

$$MBF_{ku} = \frac{\Pr(R_k \theta > r_k | \hat{\theta}, \hat{\Sigma}_\theta)}{\Pr(R_k \theta > r_k | \theta_B, \hat{\Sigma}_\theta^b)}, \tag{49}$$

where the $\Pr(\cdot)$ denote the probabilities that the posterior and prior distribution, respectively, support H_k . To give one simple example, if the prior distribution of $\theta_1, \theta_2, \theta_3$ has $\theta_B = [0, 0, 0]$ and an identity covariance matrix $\hat{\Sigma}_\theta^b$, then $\Pr(R_k \theta > r_k | \theta_B, \hat{\Sigma}_\theta^b) = 1/6$ because there are six possible orderings of three parameters that each have an equal probability.

Again let $N_g = a_g n$, where the a_g represent the relative size of the samples from the G populations and let θ^* denote the true value of θ . If the data support H_k , that is, $\theta^* \in H_k$, then if $n \rightarrow \infty$ then $\hat{\theta} \rightarrow \theta^*$, and the posterior distribution in the numerator of equation (48) is increasingly concentrated around $\hat{\theta}$ and consequently $f_k \rightarrow 1$. Analogously, if the data do not support H_k , that is, $\theta^* \notin H_k$, $f_k \rightarrow 0$. This follows from asymptotic theory; see, for example, Gelman *et al.* (2013, Chapter 4). The prior distribution in the denominator of equation (48) is independent of the N_g and thus independent of n . As can be seen from the combination of equations (22) and (34), for each group the second-order derivatives (which can for the vast majority of statistical models be written as the sum of N_g contributions) are weighted with $b_g = J^*/G \times 1/N_g$, that is, asymptotically each element of equation (22) is independent of N_g . Consequently, asymptotically c_k is a constant that is independent of n . This is exemplified by equation (35).

We now have all the ingredients in place to show that MBF_{kc} , where H_c is the complement of H_k , is consistent. Note that, due to the complementary nature of H_c , $f_c = 1 - f_k$ and $c_c = 1 - c_k$ and thus

$$MBF_{kc} = \frac{f_k}{c_k} \times \frac{c_c}{f_c}. \tag{50}$$

Then if $\theta^* \in H_k$ and $n \rightarrow \infty$, then $MBF_{kc} \rightarrow 1/c_k \times c_c/0 \rightarrow \infty$ and if $\theta^* \notin H_k$ and $n \rightarrow \infty$, then $MBF_{kc} \rightarrow 0/c_k \times c_c/1 \rightarrow 0$, which implies consistency.

Theorem 4.1 from De Santis and Spezzaferri (2001) for the generalized fractional Bayes factor and our exposition in the context of Example 1 for the MBF provide evidence for consistency if $H_k : S_k\theta = s_k$. Further evidence is obtained by realizing that each equality constraint (e.g. $\theta = 0$) can be written as an about-equality constraint $\theta > -z$, $\theta < z$ for $z \rightarrow 0$. If each equality constraint is rewritten in this manner, the exposition given at the beginning of this section applies to $H_k : S_k\theta = s_k$ and also to $H_k : S_k\theta = s_k, R_k\theta > r_k$.

If $N_g \rightarrow \infty$ for some but not all of the G groups, an analogous line of reasoning can be used to show that MBF shows reasonable behaviour. If the data support H_k , that is, $\theta^* \in H_k$ and some of the group sizes increase, then the posterior distribution in the numerator of equation (48) is increasingly concentrated around the parameters corresponding to the groups with increasing group sizes (some of the θ_g^*) and η^* . Consequently, f_k will become larger but will not attain its maximum value 1.0. Analogously, if $\theta^* \notin H_k$, f_k will become smaller, but will not attain its minimum value 0.0. Note that c_k is a constant irrespective of whether $n \rightarrow \infty$ or that some of the group sizes go to infinity. These ingredients can be used to show that the behaviour of the MBF is reasonable. Looking at equation (50), it can be seen that if $\theta^* \in H_k$, MBF will increase (to a boundary value, not to infinity) if some of the group sizes go to infinity; and if $\theta^* \notin H_k$, MBF will decrease (to a boundary value, not to zero). A proof and illustration in the context of a simple model can be found in Example 1.