Journal of the Korean Statistical Society 45 (2016) 518–525

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Asymptotic equivalence between the default Bayes factors and the ordinary Bayes factors with intrinsic priors[☆]

Seong W. Kim[a], Jinheum Kim[b,*]

[a] Department of Applied Mathematics, Hanyang University, Gyeonggi-Do 15588, South Korea

[b] Department of Applied Statistics, University of Suwon, Gyeonggi-Do 18323, South Korea

ARTICLE INFO

Article history:
Received 28 July 2015
Accepted 1 March 2016
Available online 8 April 2016

AMS 2000 subject classifications:
primary 62C10
secondary 62F15

Keywords:
Asymptotic equivalence
Fractional Bayes factor
Intrinsic Bayes factor
Intrinsic prior
Model selection

ABSTRACT

In Bayesian model selection or testing problems, default priors are typically improper; that is, the resulting Bayes factor is not well defined. To circumvent this problem, two methodologies, namely, intrinsic and fractional Bayes factors are proposed and developed. Further, these two Bayes factors are asymptotically equivalent to the ordinary Bayes factors computed with proper priors called intrinsic priors. However, it seems that there are some necessary conditions to satisfy asymptotic equivalence. Such conditions are derived and justified in this article and illustrative examples are provided. Simulations are performed to demonstrate the results.

1. Introduction

Suppose that there are two models M_1 and M_2 contending with each other. Under model M_i, data $\boldsymbol{z} = (x_1, \ldots, x_n)'$ follow a parametric distribution with probability density function (*pdf*) $f_i(\boldsymbol{z}|\boldsymbol{\theta}_i)$ for $i = 1, 2$, where $\boldsymbol{\theta}_i$ is a vector of unknown parameters. Let $\pi_i^N(\boldsymbol{\theta}_i)$ be an improper prior density, and Θ_i be the parameter space for $\boldsymbol{\theta}_i$. The Bayes factor B_{21}^N of model M_2 to model M_1 is

$$B_{21}^N = \frac{m_2^N(\boldsymbol{z})}{m_1^N(\boldsymbol{z})} = \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^N(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^N(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}, \tag{1}$$

where $m_i^N(\boldsymbol{z})$ is the marginal or predictive density under M_i. We often call $\pi_i^N(\boldsymbol{\theta}_i)$ a 'starting prior'. Since $\pi_i^N(\boldsymbol{\theta}_i)$ is improper, the integral $\int_{\Theta_i} \pi_i^N(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$ diverges. This implies that there is no normalization of $\pi_i^N(\boldsymbol{\theta}_i)$ available. Thus, it is defined up to an arbitrary multiplicative constant such as k_i. Subsequently, the Bayes factor in (1) contains a ratio of unspecified constants, say k_2/k_1.

☆ This research was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0028933) (Seong W. Kim) and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A2056869) (Jinheum Kim).

* Corresponding author.
E-mail address: jkimdt65@gmail.com (J. Kim).

http://dx.doi.org/10.1016/j.jkss.2016.03.002
1226-3192/

This issue has been addressed by several authors including Geisser and Eddy (1979), Spiegelhalter and Smith (1982), and San Martini and Spezzaferri (1984). Berger and Pericchi (1996) introduced a new model selection criterion called the intrinsic Bayes factor (IBF), using a data-splitting idea called the training sample method, which would remove the arbitrariness of improper priors. One of the shortcomings of IBF approach is due to its considerable amount of computational expenses. To mitigate this drawback, O'Hagan (1995) proposed another criterion, called the fractional Bayes factor (FBF). It is computed by exponentiating the likelihood to a power δ for $0 \le \delta \le 1$. One of its advantages is that it does not require heavy computation.

Meanwhile, Berger and Pericchi (1996) mentioned that if we have reasonable proper priors for each model, the training sample computation is not needed. Berger and Pericchi (1996) suggested how to derive proper priors, called intrinsic priors. Although intrinsic priors are neither necessarily unique nor proper, the motivation is based on the conjecture that the ordinary Bayes factor with a set of intrinsic priors would provide results which are asymptotically equivalent to the IBF or the FBF. However, there should be an additional condition needed in order for the two Bayes factors to be close to each other asymptotically.

The rest of the paper is organized as follows. In Section 2, the intrinsic and fractional Bayes factors are reviewed. Also, we present how to derive intrinsic priors in conjunction with these two Bayes factors. In Section 3, we provide a justification of necessary conditions under which the asymptotic equivalence is satisfied. In Section 4, several intrinsic priors are derived for testing two exponential means. In Section 5, a class of intrinsic priors is derived for testing two normal variances based on the fractional approach. Numerical results based on various simulations are given in Section 6. We finish this article with concluding remarks in Section 7.

2. Default Bayes factors and intrinsic priors

It has been seen that the Bayes factor B_{21}^N in (1) involves arbitrary constants. We discuss the method for removing this arbitrariness in (1) by a training sample. Let $\boldsymbol{x}(l)$ be a training sample and let $\boldsymbol{x}(-l)$ be the remainder of the data. First, compute the posterior $\pi_i^N(\boldsymbol{\theta}_i|\boldsymbol{x}(l))$ and then compute the Bayes factor with $\boldsymbol{x}(-l)$ as data, using $\pi_i^N(\boldsymbol{\theta}_i|\boldsymbol{x}(l))$ as the prior. Then, by the Bayes theorem the Bayes factor is given by

$$B_{21}(l) = B_{21}^N \cdot B_{12}^N(\boldsymbol{x}(l)), \tag{2}$$

where $B_{12}^N(\boldsymbol{x}(l))$ is the Bayes factor computed with the training sample $\boldsymbol{x}(l)$. In practice, $\boldsymbol{x}(l)$ is chosen to be minimal in the sense that the marginal $m_i^N(\boldsymbol{x}(l))$ is finite and no subset of $\boldsymbol{x}(l)$ gives finite marginals. Note that $B_{21}(l)$ in (2) does not depend on arbitrary constants, and thus is well defined. Furthermore, the Bayes factor defined by (2) depends on the choice of the minimal training sample. To avoid this dependence, Berger and Pericchi (1996) suggested computing the average of $B_{21}(l)$ over all $\boldsymbol{x}(l)$.

Definition 2.1. The IBF of M_2 to M_1 is defined by

$$B_{21}^I = \frac{1}{L}\sum_{l=1}^{L} B_{21}(l) = B_{21}^N \cdot CFA_{12}, \tag{3}$$

where L is the number of all possible minimal training samples, and the correction factor CFA_{12} is given by

$$CFA_{12} = \frac{1}{L}\sum_{l=1}^{L} B_{12}^N(\boldsymbol{x}(l)).$$

On the other hand, we introduce the method for removing the arbitrariness in (1) by a portion of the likelihood with fraction δ. O'Hagan (1995) proposed the following FBF.

Definition 2.2. The FBF of model M_2 to model M_1 is defined by

$$B_{21}^F = B_{21}^N \cdot CFR_{12}(\delta), \tag{4}$$

where the correction factor $CFR_{12}(\delta)$ is given by

$$CFR_{12}(\delta) = \frac{\int_{\Theta_1} [f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)]^\delta \pi_1^N(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int_{\Theta_2} [f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)]^\delta \pi_2^N(\boldsymbol{\theta}_2) d\boldsymbol{\theta}_2}.$$

A commonly suggested choice of δ is $\delta = m/n$, where m is the size of the minimal training sample proposed by Berger and Pericchi (1996), and n is the size of the whole sample.

Under the regularity conditions in Berger and Pericchi (1996), a set of intrinsic priors defined by (π_1^I, π_2^I) is a solution of the following system of equations:

$$\begin{cases} \dfrac{\pi_2^I(\boldsymbol{\theta}_2)\pi_1^N(\psi_1(\boldsymbol{\theta}_2))}{\pi_2^N(\boldsymbol{\theta}_2)\pi_1^I(\psi_1(\boldsymbol{\theta}_2))} = B_2^*(\boldsymbol{\theta}_2), \\ \dfrac{\pi_2^I(\psi_2(\boldsymbol{\theta}_1))\pi_1^N(\boldsymbol{\theta}_1)}{\pi_2^N(\psi_2(\boldsymbol{\theta}_1))\pi_1^I(\boldsymbol{\theta}_1)} = B_1^*(\boldsymbol{\theta}_1). \end{cases} \tag{5}$$

Here,

$$\psi_i(\boldsymbol{\theta}_j) = \lim_{n\to\infty} E_{\boldsymbol{\theta}_j}^{M_j}(\hat{\boldsymbol{\theta}}_i), \quad \text{for } i \neq j,$$

where $\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}_i$ under model M_i, and

$$B_i^*(\boldsymbol{\theta}_i) = \lim_{L\to\infty} E_{\boldsymbol{\theta}_i}^{M_i}[CFA_{12}] \quad \text{or} \quad B_i^*(\boldsymbol{\theta}_i) = \lim_{n\to\infty} E_{\boldsymbol{\theta}_i}^{M_i}[CFR_{12}(\delta)].$$

Berger and Pericchi (1996) noted that the solutions are not necessarily unique nor necessarily proper.

3. Generalization of asymptotic equivalence to B_{21}^I and $B_{21}^F(\delta)$

Let (π_1^I, π_2^I) denote a set of intrinsic priors satisfying (5). Let c_i be a normalizing constant corresponding to π_i^I for $i = 1, 2$, and let r_{21} be a ratio c_2/c_1 of two normalizing constants. Berger and Pericchi (1996) suggested that the ordinary Bayes factors with the normalized intrinsic priors, namely, π_1^{I*} and π_2^{I*}, are asymptotically equivalent to those obtained from (3) or (4) when the sample size is large enough. Here, $\pi_i^{I*} = c_i^{-1}\pi_i^I$ denotes the normalized intrinsic prior of π_i^I for $i = 1, 2$. That is, $\int_{\Theta_i} \pi_i^{I*}(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i = 1$. Note that their assertion is true as long as r_{12} is equal to one. So, we investigate whether their conjecture can be generalized to the case that r_{12} is other than one or not.

Proposition 1. *Under the regularity conditions in Berger and Pericchi (1996), as the sample size n goes to infinity, for the intrinsic priors π_1^I and π_2^I satisfying (5), B_{21}^I in (3) or B_{21}^F in (4) can be approximated by $r_{21}B_{21}^{I*}$, where*

$$B_{21}^{I*} = \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^{I*}(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^{I*}(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}. \tag{6}$$

Proof. Following the arguments of Berger and Pericchi (1996) and Moreno (1997), respectively, we have

$$CF = \frac{\pi_2^I(\hat{\boldsymbol{\theta}}_2)\pi_1^N(\hat{\boldsymbol{\theta}}_1)}{\pi_2^N(\hat{\boldsymbol{\theta}}_2)\pi_1^I(\hat{\boldsymbol{\theta}}_1)}(1 + o(1)),$$

with CF being CFA_{12} or $CFR_{12}(\delta)$. Thus,

$$\begin{aligned} B_{21}^I(\text{or } B_{21}^F) &= B_{21}^N \frac{\pi_2^I(\hat{\boldsymbol{\theta}}_2)\pi_1^N(\hat{\boldsymbol{\theta}}_1)}{\pi_2^N(\hat{\boldsymbol{\theta}}_2)\pi_1^I(\hat{\boldsymbol{\theta}}_1)}(1 + o(1)) \\ &= \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^N(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^N(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1} \cdot \frac{\pi_2^I(\hat{\boldsymbol{\theta}}_2)\pi_1^N(\hat{\boldsymbol{\theta}}_1)}{\pi_2^N(\hat{\boldsymbol{\theta}}_2)\pi_1^I(\hat{\boldsymbol{\theta}}_1)}(1 + o(1)). \end{aligned} \tag{7}$$

Taking the limit of (7) becomes

$$\begin{aligned} \lim_{L\to\infty} B_{21}^I(\text{or } \lim_{n\to\infty} B_{21}^F) &= \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^N(\boldsymbol{\theta}_2)\frac{\pi_2^I(\boldsymbol{\theta}_2)}{\pi_2^N(\boldsymbol{\theta}_2)}d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^N(\boldsymbol{\theta}_1)\frac{\pi_1^I(\boldsymbol{\theta}_1)}{\pi_1^N(\boldsymbol{\theta}_1)}d\boldsymbol{\theta}_1} \\ &= \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^I(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^I(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1} \\ &= \frac{c_2}{c_1}\frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\boldsymbol{\theta}_2)\pi_2^{I*}(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\boldsymbol{\theta}_1)\pi_1^{I*}(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1} \\ &= r_{21}B_{21}^{I*}. \end{aligned}$$

This completes the proof. □

Remark 1. The implications of Proposition 1 can be justified in the following way. Even though proper intrinsic priors are derived, asymptotic equivalence should not be guaranteed unless the ratio r_{21} is one.

4. Testing exponential means

Sun and Kim (1999) derived a general set of intrinsic priors for testing k ordered exponential means with Jefferys priors being used as starting priors for both models. It turned out that the normalizing constants were 1 for both π_1^I and π_2^I. Subsequently, the Bayes factors computed with these intrinsic priors get closer to the intrinsic Bayes factors as the sample size grows. Kim (2000) derived several intrinsic priors in testing two exponential means, where Jefferys priors were employed for both null and two-sided alternative hypotheses. The corresponding results are fairly similar to those of Sun and Kim (1999). Moreover, Kim and Kim (2000) derived intrinsic priors for testing two exponential means with the fractional Bayes approach. However, a lack of asymptotic equivalence was noted because the ratio of two normalizing constants was not one, but this issue was not extensively justified in that paper. We now use different starting priors to derive several intrinsic priors under different alternative hypotheses, where the ratio r_{21} in Proposition 1 is not one.

4.1. One-sided test

Let $\mathrm{Exp}(\mu)$ denote the exponential distribution with mean μ. Suppose that we have independent observations $X_{ij} \sim \mathrm{Exp}(\mu_i)$ for $i = 1, 2; j = 1, \dots, n_i$. We use the following notation throughout this paper. Let $n = n_1 + n_2$. Assume that $n_1/n \to a \in (0, 1)$ as $n \to \infty$. Consider two models (hypotheses),

$$M_1 : \mu_1 = \mu_2 \quad \text{and} \quad M_2 : \mu_1 < \mu_2.$$

Let μ denote the common value of μ_1 and μ_2 under M_1. Uniform and Jeffreys priors are used as starting priors for each hypothesis. That is, $\pi_1^N(\mu) = 1$ and $\pi_2^N(\mu_1, \mu_2) = 1/(\mu_1\mu_2)$. After some algebra described in Kim (2000) along with the strong consistency of the MLE, the system of equations in (5) becomes

$$\begin{cases} \dfrac{\pi_2^I(\mu_1, \mu_2)}{\pi_1^I(a\mu_1 + (1-a)\mu_2)/(\mu_1\mu_2)} = \mu_1, & 0 < \mu_1 < \mu_2 < \infty, \\ \dfrac{\pi_2^I(\mu, \mu)}{\pi_1^I(\mu)/\mu^2} = \mu, & 0 < \mu < \infty. \end{cases} \tag{8}$$

Theorem 1. *For any legitimate pdf $g(t)$ for $t > 0$, i.e. $\int_0^\infty g(t)dt = 1$, the set of intrinsic priors is composed of*

$$\pi_1^I(\mu) = g(\mu), \quad 0 < \mu < \infty,$$

and

$$\pi_2^I(\mu_1, \mu_2) = \frac{1}{\mu_2} g(a\mu_1 + (1-a)\mu_2), \quad 0 < \mu_1 < \mu_2 < \infty.$$

Further, π_2^I is a proper density.

Proof. By Lemma 1 of Kim (2000), (π_1^I, π_2^I) is a solution of (8). To prove the propriety of π_2^I, let $s = \mu_1/\mu_2$ and $t = \mu_2$. Then

$$\begin{aligned} \int_0^\infty \int_0^{\mu_2} \pi_2^I(\mu_1, \mu_2) d\mu_1 d\mu_2 &= \int_0^1 \int_0^\infty g(t(as + 1 - a))dtds \\ &= -\frac{1}{a}\log(1-a) \equiv c_2. \end{aligned}$$

This implies $r_{21} = c_2$. □

4.2. Two-sided test

Consider the following hypotheses,

$$M_1 : \mu_1 = \mu_2 \quad \text{and} \quad M_2 : \mu_1 \neq \mu_2.$$

Again, let μ denote the common value of μ_1 and μ_2 under M_1. We use uniform and Jeffreys priors as starting priors for each hypothesis. That is, $\pi_1^N(\mu) = 1$ and $\pi_2^N(\mu_1, \mu_2) = 1/(\mu_1\mu_2)$. After some algebra described in Kim (2000) along with the strong consistency of the MLE, the system of equations in (5) becomes

$$\begin{cases} \dfrac{\pi_2^I(\mu_1, \mu_2)}{\pi_1^I(a\mu_1 + (1-a)\mu_2)/(\mu_1\mu_2)} = B_2^*(\mu_1, \mu_2), & 0 < \mu_1, \mu_2 < \infty, \\ \dfrac{\pi_2^I(\mu, \mu)}{\pi_1^I(\mu)/\mu^2} = \dfrac{\mu}{3}, & 0 < \mu < \infty, \end{cases} \tag{9}$$

where

$$B_2^*(\mu_1,\mu_2)=\frac{(\mu_1\mu_2)^2}{(\mu_2-\mu_1)^3}\left\{\frac{(\mu_1+\mu_2)(\mu_2-\mu_1)}{\mu_1\mu_2}+2\log\left(\frac{\mu_1}{\mu_2}\right)\right\}.$$

Theorem 2. *For any legitimate pdf $g(t)$ for $t>0$, the set of intrinsic priors is composed of*

$$\pi_1^I(\mu)=g(\mu),\quad 0<\mu<\infty,$$

and

$$\pi_2^I(\mu_1,\mu_2)=\frac{1}{\mu_1\mu_2}B_2^*(\mu_1,\mu_2)g(a\mu_1+(1-a)\mu_2),\quad 0<\mu_1,\mu_2<\infty.$$

Further, π_2^I is a proper density.

Proof. By Lemma 1 of Kim (2000), (π_1^I,π_2^I) is a solution of (8). To prove the propriety of π_2^I, let $s=\mu_1/\mu_2$ and $t=\mu_2$. Then

$$\begin{aligned}\int_0^\infty\int_0^\infty \pi_2^I(\mu_1,\mu_2)d\mu_2 d\mu_1 &= \int_0^\infty\int_0^\infty \frac{s}{(1-s)^3}\left[\frac{(1+s)(1-s)}{s}+2\log s\right]g(t(as+1-a))dtds\\ &=\int_0^\infty \frac{s}{(1-s)^3(as+1-a)}\left[\frac{(1+s)(1-s)}{s}+2\log s\right]ds.\end{aligned}$$

Let

$$q(s)=\frac{s}{(1-s)^3(as+1-a)}\left[\frac{(1+s)(1-s)}{s}+2\log s\right].$$

Since $\int_0^1 q(s)ds=\int_1^\infty q(s)ds$, it follows from the Mclaurin series expansion that

$$\begin{aligned}\int_0^1 q(s)ds &= 2\sum_{j=0}^\infty \frac{1}{(j+2)(j+3)}\int_0^1 \frac{(1-s)^j}{as+1-a}ds\\ &=0.7269.\end{aligned}$$

This implies $r_{21}=1.4538$. □

Remark 2. Similar to the results proposed by Kim (2000), we have seen that there is a class of proper intrinsic priors for testing two exponential means. Specifically, the inverse gamma distribution is used for $\pi_1^I(\mu)$ in our computation.

5. Testing two normal variances

Suppose we have independent observations $X_{ij}\sim N(0,\tau_i)$, $i=1,2; j=1,\ldots,n_i$. Consider the following hypotheses,

$$M_1:\tau_1=\tau_2 \quad\text{and}\quad M_2:\tau_1\neq\tau_2.$$

Let τ denote the common value of τ_1 and τ_2 under M_1. We use Jeffreys prior for M_1 as a starting prior. So, $\pi_1^N(\tau)=1/\tau$ for $\tau>0$. We use a flexible starting prior for M_2. That is, the starting prior for M_2 has the following form: for $0\le\alpha,\beta<\infty$,

$$\pi_2^N(\tau_1,\tau_2)=\frac{1}{\tau_1^{1+\alpha}\tau_2^{1+\beta}},\quad 0<\tau_1,\tau_2<\infty. \tag{10}$$

When $\alpha=\beta=0$, $\pi_2^N(\tau_1,\tau_2)$ becomes the Jeffreys prior. Kim and Kim (2002) derived several intrinsic priors under these Jeffreys priors with both intrinsic and fractional approaches. It turned out that r_{21} was one for both approaches. When $\alpha=1/2$ and $\beta=1$, it becomes the reference prior of Berger and Bernardo (1992) in case of τ_1 being the parameter of interest. Finally, when $\alpha=\beta=1/2$, it is also the reference prior in case of the covariance being the parameter of interest. Although the covariance is known with zero due to independence assumption, we use this starting prior for illustration.

We only consider the fractional approach in testing two normal variances. Simply following the procedures in Kim and Kim (2002) with $\delta=2/n$, we can derive a general form of intrinsic priors

$$\begin{cases}\pi_1^I(\tau)=g(\tau), & 0<\tau<\infty,\\ \pi_2^I(\tau_1,\tau_2)=\dfrac{a\tau_1+(1-a)\tau_2}{\tau_1^{a+\alpha}\tau_2^{1-a+\beta}}B_2^*(\tau_1,\tau_2)\pi_1^I(a\tau_1+(1-a)\tau_2), & 0<\tau_1,\tau_2<\infty,\end{cases} \tag{11}$$

where $g(\tau)$ is any legitimate *pdf* with support of $\tau>0$, and

$$B_2^*(\tau_1,\tau_2)=\frac{(a\tau_1)^{a+\alpha}((1-a)\tau_2)^{1-a+\beta}}{\Gamma(a+\alpha)\Gamma(1-a+\beta)(a\tau_1+(1-a)\tau_2)}.$$

Theorem 3. *The intrinsic prior* $\pi_2^I(\tau_1, \tau_2)$ *in* (11) *is proper.*

Proof. Let $s = \tau_1/\tau_2$, $t = \tau_2$. Then

$$\begin{aligned}\int_0^\infty \int_0^\infty \pi_2^F(\tau_1, \tau_2) d\tau_1 d\tau_2 &= \frac{a^{a+\alpha}(1-a)^{1-a+\beta}}{\Gamma(a+\alpha)\Gamma(1-a+\beta)} \int_0^\infty \int_0^\infty s^{a-1} g((as+1-a)t) dt ds \\ &= \frac{a^{a+\alpha}(1-a)^{1-a+\beta}}{\Gamma(a+\alpha)\Gamma(1-a+\beta)} \int_0^\infty \frac{s^{a-1}}{as+1-a} ds \\ &= a^\alpha (1-a)^\beta \frac{\Gamma(a)\Gamma(1-a)}{\Gamma(a+\alpha)\Gamma(1-a+\beta)} \equiv c_2.\end{aligned}$$

This implies $r_{21} = c_2$. □

Proposition 2. *When* $g(\tau)$ *in* (11) *is the pdf of inverse gamma distribution with hyperparameters* λ *and* η, *i.e.* $g(\tau) \propto \tau^{-(\lambda+1)} e^{-\eta/\tau}$, $\tau > 0$, *the ordinary Bayes factor with this set of intrinsic priors in* (6) *is given by*

$$B_{21}^{I*} = C \cdot \frac{a^a (1-a)^{1-a}}{\Gamma(a)\Gamma(1-a)} \left(\frac{\sum_{j=1}^{n_1} x_{1j}^2 + \sum_{j=1}^{n_2} x_{2j}^2 + 2\eta}{2} \right)^{n/2+\lambda}, \tag{12}$$

where

$$C = \int_0^\infty \frac{s^{a-1-n_1/2}}{(as+1-a)^{\lambda+1}} \cdot \left(\frac{\sum_{j=1}^{n_1} x_{1j}^2}{2s} + \frac{\sum_{j=1}^{n_2} x_{2j}^2}{2} + \frac{\eta}{as+1-a} \right)^{-n/2-\lambda} ds.$$

Remark 3. Proposition 2 can justify the result in Proposition 1 with the following sense. Note that the Bayes factor in (12) is computed with the normalized version of (11). However, this Bayes factor does not depend on α and β, but does depend only on the data and hyperparameters λ and η. Thus, it is virtually impossible to make an asymptotic equivalence between B_{21}^F and B_{21}^{I*}, at least with the inverse gamma prior because B_{21}^F depends on the starting prior π_2^N which involves α and β.

6. Numerical results

In this section, simulation study is conducted to verify the findings in Proposition 1 based on finite samples. Given that legitimate intrinsic priors are used, i.e. $c_1 = 1$ for model M_1 on both exponential and normal populations, we only need to check the behaviors of asymptotic equivalence associated with the normalizing constant c_2 of model M_2. Thus, we compute two quantities of the relative differences (RD) between the IBF (or FBF) and the ordinary Bayes factor with intrinsic priors to investigate how the value of c_2 contributes in preserving asymptotic equivalence. More precisely, we calculate two RDs:

$$\mathrm{RD}_1 = \frac{|B_{21}^{I*} - B_{21}^I|}{B_{21}^I} \quad \left(\text{or } \frac{|B_{21}^{I*} - B_{21}^F|}{B_{21}^F} \right) \quad \text{and} \quad \mathrm{RD}_2 = \frac{|c_2 B_{21}^{I*} - B_{21}^I|}{B_{21}^I} \quad \left(\text{or } \frac{|c_2 B_{21}^{I*} - B_{21}^F|}{B_{21}^F} \right). \tag{13}$$

In (13), the former is suggested by Berger and Pericchi (1996), and the latter is based on our suggestion.

Example 1. We conducted several simulations for testing two exponential means. Three cases, namely, $(\mu_1, \mu_2) = (1, 1)$, $(\mu_1, \mu_2) = (1, 2)$, and $(\mu_1, \mu_2) = (1, 3)$, were examined with some choices of n_1 and n_2. The RD_1 and RD_2 described in (13) and their standard deviations (SD) were computed based on 200 replications. We denote the Berger and Pericchi's suggestion by 'B&P' and our proposed assertion in Section 3 by 'Proposed' in Tables 1 and 2. Note that B_{21}^{I*} in (13) is the ordinary Bayes factor computed with the inverse gamma prior with the hyperparameters λ and η. More precisely, for testing one-sided alternative B_{21}^{I*} is

$$B_{21}^{I*} = c_2^{-1} \left(\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_2} x_{2j} + \eta \right)^{n+\lambda} \int_0^1 \frac{(as+1-a)^{n-1} s^{n_2+\lambda}}{\left[\left(as+1-a\right) \left(\sum_{j=1}^{n_1} x_{1j} + s \sum_{j=1}^{n_2} x_{2j} \right) + \eta s \right]^{n+\lambda}} ds,$$

where $c_2 = -\log(1-a)/a$. Similarly, for two-sided alternative B_{21}^{I*} is

$$B_{21}^{I*} = c_2^{-1} \left(\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_2} x_{2j} + \eta \right)^{n+\lambda} \int_0^\infty \frac{(as+1-a)^{n-1} s^{n_2+\lambda+1} r(s)}{(1-s)^3 \left[\left(as+1-a\right) \left(\sum_{j=1}^{n_1} x_{1j} + s \sum_{j=1}^{n_2} x_{2j} \right) + \eta s \right]^{n+\lambda}} ds,$$

Table 1
The relative differences (RD) and their standard deviations (SD) based on the Berger and Pericchi's suggestion (B&P) and our proposed assertion (Proposed) with 200 replications. The normalizing constant for $\pi_2^I(\mu_1, \mu_2)$ is $c_2 = -\log(1-a)/a$.

(μ_1, μ_2)	(n_1, n_2)	B&P		Proposed	
		RD_1	SD	RD_2	SD
(1, 1)	(5, 5)	0.7042	0.0017	0.2471	0.0001
	(10, 10)	0.6660	0.0066	0.2056	0.0050
	(10, 20)	0.4672	0.0043	0.2150	0.0029
	(20, 10)	0.8285	0.0094	0.1103	0.0058
	(30, 30)	0.5439	0.0035	0.1142	0.0026
(1, 2)	(5, 5)	0.4205	0.0437	0.1437	0.0231
	(10, 10)	0.3986	0.0034	0.0715	0.0020
	(10, 20)	0.2021	0.0068	0.0666	0.0017
	(20, 10)	0.6665	0.0037	0.0275	0.0005
	(30, 30)	0.3692	0.0005	0.0176	0.0001
(1, 3)	(5, 5)	0.2816	0.0064	0.1369	0.0090
	(10, 10)	0.3108	0.0025	0.0696	0.0007
	(10, 20)	0.1235	0.0052	0.0822	0.0047
	(20, 10)	0.6257	0.0010	0.0197	0.0011
	(30, 30)	0.3501	0.0001	0.0263	0.0001

Table 2
The relative differences (RD) and their standard deviations (SD) based on the Berger and Pericchi's suggestion (B&P) and our proposed assertion (Proposed) with 200 replications. The normalizing constant for $\pi_2^I(\mu_1, \mu_2)$ is $c_2 = 1.4538$.

(μ_1, μ_2)	(n_1, n_2)	B&P		Proposed	
		RD_1	SD	RD_2	SD
(1, 1)	(5, 5)	0.5353	0.0099	0.1208	0.0121
	(10, 10)	0.4930	0.0045	0.0809	0.0073
	(10, 20)	0.2833	0.0006	0.0787	0.0030
	(20, 10)	0.8130	0.0010	0.0746	0.0069
	(30, 30)	0.4735	0.0152	0.0449	0.0020
(1, 2)	(5, 5)	0.5117	0.0276	0.1233	0.0072
	(10, 10)	0.4973	0.0122	0.0810	0.0006
	(10, 20)	0.2404	0.0110	0.0693	0.0065
	(20, 10)	0.8623	0.0312	0.0779	0.0105
	(30, 30)	0.4629	0.0188	0.0452	0.0076
(1, 3)	(5, 5)	0.4639	0.0140	0.1144	0.0084
	(10, 10)	0.4706	0.0186	0.0775	0.0009
	(10, 20)	0.2211	0.0064	0.0609	0.0041
	(20, 10)	0.8508	0.0206	0.0741	0.0036
	(30, 30)	0.4665	0.0056	0.0426	0.0007

where $c_2 = 1.4538$ and $r(s) = (1+s)(1-s)/s + 2\log s$. Note that one-dimensional numerical integration is needed in calculating these two ordinary Bayes factors. We only report the results with the choice of $(\lambda, \eta) = (0.1, 0.1)$. Similar results were achieved when $(\lambda, \eta) = (0.01, 0.01)$ and $(1.0, 1.0)$.

Table 1 shows the results for testing the one-sided alternative, and those for the two-sided alternative are provided in Table 2. We can see that the values of 'B&P' are huge with random in all cases, whereas those of 'Proposed' are relatively small. In addition, as the sample size increases, the RD_2 decreases, which is what we would expect from an asymptotic sense.

Example 2. We performed simulations for testing two normal variances. The cases of $\tau_1 = \tau_2 = 1$ and $\tau_1 = 1$ and $\tau_2 = 1.5$ were examined with some choices of n_1 and n_2. Like did in Example 1, the RDs and their corresponding SDs were computed using the normalizing constant c_2 for $\pi_2^I(\tau_1, \tau_2)$ given by Theorem 3. We also use the hyperparameters of $(\lambda, \eta) = (0.1, 0.1)$ in this computation.

Numerical results are reported in Table 3. When $(\alpha, \beta) = (0, 0)$, the corresponding normalizing constant becomes one. So, two RDs should be the same accordingly. Except for the third column, we have similar results appeared in exponential distributions. That is, all the results associated with 'B&P' are bad, whereas those associated with 'Proposed' are fairly nice and consistent as the sample size increases. In particular, when $(\alpha, \beta) = (1/2, 1/2)$, the corresponding results seem to be the best compared to others.

Table 3
The relative differences (RD) and their standard deviations (SD) based on the Berger and Pericchi's suggestion (B&P) and our proposed assertion (Proposed) with 200 replications. The normalizing constant for $\pi_2^I(\mu_1, \mu_2)$ is c_2 given by Theorem 3.

(τ_1, τ_2)	(n_1, n_2)	(α, β)									
		(0, 0)		(1/2, 1/2)				(1/2, 1)			
		B&P		B&P		Proposed		B&P		Proposed	
		RD_1	SD	RD_1	SD	RD_2	SD	RD_1	SD	RD_2	SD
(1, 1)	(5, 5)	0.0907	0.0032	0.5334	0.0058	0.0524	0.0017	0.2877	0.0047	0.0579	0.0016
	(10, 10)	0.0513	0.0025	0.5680	0.0035	0.0198	0.0009	0.2827	0.0028	0.0298	0.0015
	(10, 20)	0.0370	0.0005	0.6294	0.0008	0.0155	0.0007	0.3844	0.0007	0.0184	0.0001
	(20, 10)	0.0361	0.0005	0.6275	0.0010	0.0156	0.0007	0.2178	0.0007	0.0253	0.0004
	(30, 30)	0.0163	0.0004	0.5701	0.0007	0.0050	0.0001	0.2632	0.0005	0.0088	0.0005
(1, 1.5)	(5, 5)	0.1094	0.0023	0.5648	0.0024	0.0472	0.0015	0.3129	0.0021	0.0685	0.0031
	(10, 10)	0.0607	0.0006	0.5845	0.0009	0.0205	0.0003	0.2962	0.0007	0.0365	0.0004
	(10, 20)	0.0383	0.0013	0.6329	0.0020	0.0153	0.0002	0.3873	0.0017	0.0185	0.0008
	(20, 10)	0.0491	0.0014	0.6498	0.0022	0.0170	0.0006	0.2345	0.0016	0.0366	0.0014
	(30, 30)	0.0219	0.0005	0.5786	0.0008	0.0056	0.0002	0.2701	0.0006	0.0134	0.0005

7. Concluding remarks

Both IBF and FBF are well defined, reasonable, and applicable in general settings. However, when we deal with non-exchangeable cases, such as time series analysis, the training sample computation might be difficult. Moreover, it is often difficult to obtain a minimal training sample in non-linear models. If we have reasonable proper priors for each model (hypothesis), the training sample computation is not necessary. Berger and Pericchi (1996) suggested how to derive the proper priors, called intrinsic priors. The ordinary Bayes factor computed with the set of intrinsic priors is asymptotically equivalent to the corresponding IBF or FBF only when the ratio of two normalizing constants of intrinsic priors is one.

In our study, a justification of necessary conditions was provided to attain the asymptotic equivalence between B_{21}^{I*} and B_{21}^{I} (or B_{21}^{F}) regardless of whether the ratio is one or not. Such was illustrated with three examples. In addition, simulations showed that all the results based on the assertion of Berger and Pericchi (1996) are not ideal, whereas those associated with our generalization are fairly nice and consistent as the sample size increases.

References

Berger, J. O., & Bernardo, J. (1992). On the development of the reference priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 35–60). London: Oxford University Press.
Berger, J. O., & Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122.
Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*, 153–160.
Kim, S. W. (2000). Intrinsic priors for testing exponential means. *Statistics & Probability Letters*, *46*, 195–201.
Kim, S. W., & Kim, H. (2000). Intrinsic priors for testing two exponentil means with the fractional Bayes factor. *Journal of the Korean Statistical Society*, *29*, 395–405.
Kim, S. W., & Kim, D. H. (2002). Intrinsic priors for two-sample tests in normal populations. *Communications in Statistics. Theory and Methods*, *31*, 1091–1105.
Moreno, E. (1997). Bayes factors for intrinsic and fractional priors in nested models. Bayesian robustness. *IMS Lecture Notes-Monograph Series*, *31*, 257–270.
O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of Royal Statistical Society. Series B*, *57*, 99–138.
San Martini, A., & Spezzaferri, S. F. (1984). A predictive model selection criterion. *Journal of Royal Statistical Society. Series B*, *46*, 296–303.
Spiegelhalter, D. J., & Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of Royal Statistical Society. Series B*, *44*, 377–387.
Sun, D., & Kim, S. W. (1999). *Intrinsic priors for testing ordered exponential means. Technical report # 99*. National Institute of Statistical Sciences.