NISS

Intrinsic Priors for Testing Ordered Exponential Means

Dongchu Sun and Seong W. Kim

Technical Report Number 99
April, 1999

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Intrinsic Priors for Testing Ordered Exponential Means

DONGCHU SUN AND SEONG W. KIM

Department of Statistics, The University of Missouri, Columbia, MO 65211, USA

SUMMARY

In Bayesian model selection or testing problems, Bayes factors under proper priors have been very successful. In practice, however, limited information and time constraints often require us to use noninformative priors which are typically improper and are defined only up to arbitrary constants. The resulting Bayes factors are then not well defined. A recently proposed model selection criterion, the intrinsic Bayes factor, overcomes such problems by using a part of the sample as a training sample to get a proper posterior and then use the posterior as the prior for the remaining observations to compute the Bayes factor. Surprisingly, such a Bayes factor can also be computed directly from the full sample by using some proper priors, namely intrinsic priors. The present paper explains how to derive intrinsic priors for ordered exponential means. Some simulation results are also given to illustrate the method and compare it with classical methods.

Some Keywords: Intrinsic Bayes factor, Intrinsic priors, Jeffreys prior, Noninformative priors, Restricted maximum likelihood estimator.

1 Introduction

In reliability theory or survival analysis for comparing treatments, we often need to test the following hypotheses

$$\begin{aligned} M_1 &: \mu_1 = \mu_2 = \cdots = \mu_k, \text{ vs} \\ M_2 &: \mu_1 \le \mu_2 \le \cdots \le \mu_k, \end{aligned} \tag{1}$$

where the μ_i's are the means of certain distributions such as exponential distributions. Robertson, Wright and Dykstra (1988) found the asymptotic distribution of the generalized likelihood ratio test statistic for M_1 versus $M_2 - M_1$ for the exponential family using level probabilities. However, for small sample sizes, the results from asymptotic approximations are often undesirable.

It has been noticed that the generalized likelihood ratio test or the most powerful test could be misleading, even when the sample sizes are large. For example, Berger, Brown and Wolpert (1994) showed that the test for a simple hypothesis against a simple alternative on the normal means with known variance rejects the null hypothesis systematically while the Bayes factor is just 1. This motivated the following example.

Example 1 . Let $Exp(\mu)$ denote the exponential distribution with mean μ. Suppose that we have independent observations X_{ij}, where X_{ij} is $Exp(\mu_i)$, $i = 1, 2;\ j = 1, \ldots, n$. Consider the following testing problem,

$$H_0 : \frac{\mu_1}{\mu_2} = 2, \text{ vs } H_1 : \frac{\mu_1}{\mu_2} = \frac{1}{2}.$$

Define $\boldsymbol{X} = (X_{11}, \cdots, X_{1n}, X_{21}, \cdots, X_{2n})^t$, $\overline{X}_i = \sum_{j=1}^{n} X_{ij}/n$ and $G(\boldsymbol{X}) = \overline{X}_1/\overline{X}_2$. The generalized likelihood ratio test will reject H_0 if $G(\boldsymbol{X}) < C$, and accept H_0 if $G(\boldsymbol{X}) \geq C$. Let α and β denote the probabilities of Type I and Type II errors respectively. If we assume that $\alpha = \beta$, then the critical value C should be 1, and $\alpha = \beta = F(0.5; 2n, 2n)$, where $F(\cdot; 2n, 2n)$ is the cdf (cumulative distribution function) of an F distribution with $2n$ and $2n$ degrees of freedom. The corresponding P-values are $F(G(\boldsymbol{X})/2; 2n, 2n)$. The values of α and β when $n = 12, 20, 30$ are given in the second column of Table 1. If we observe $\overline{X}_1 = \overline{X}_2$, the data would support both H_0 and H_1 equally. The generalized likelihood ratio test would conclude H_0 and probabilities for both type I and type II errors are less than 0.05 as long as the sample size is 12 or more. Now let us apply Bayesian model selection for this problem. Suppose that we choose vague model probabilities, i.e., we have a 50% chance to select either H_0 or H_1. Also, μ_1 follows an informative prior such as the inverse gamma (1,1) prior under both H_0 and H_1. Then the posterior probabilities of H_0 and H_1 given $\boldsymbol{X}$ are

$$P(H_0|\boldsymbol{X}) = \frac{1}{1+B}, \quad \text{and} \quad P(H_1|\boldsymbol{X}) = \frac{B}{1+B},$$

respectively, where B is the Bayes factor, given by

$$B = \left[\frac{G(\boldsymbol{X})/2 + (n\overline{X}_2)^{-1} + 1}{G(\boldsymbol{X}) + (n\overline{X}_2)^{-1} + 1/2}\right]^{2n+1}.$$

The numerical values of the P-value, Bayes factor, and posterior probabilities for some $\overline{X}_1$ and $\overline{X}_2$ are also given in Table 1. Clearly, the Bayesian method gives a better solution.

Ideally, one would choose proper priors or informative priors in computing Bayes factors. However, limited information and time constraints often require the use of noninformative priors. In this paper, we use a Bayesian approach to test the problem given by (1) for the exponential distribution using noninformative priors. Since noninformative priors such as Jeffreys' (1961) priors or reference priors (Berger and Bernado, 1989, 1992) are typically improper so that such priors are only defined up to arbitrary constants which affects the values of Bayes factors. Many people have made efforts to compensate for that arbitrariness. See Geisser and Eddy (1979), Spiegelhalter and Smith (1982), and San Martini and Spezzaferri (1984) for related works.

Berger and Pericchi (1996b) introduced a new model selection criterion, called the Intrinsic Bayes factor (IBF) using a data-splitting idea, which would eliminate the arbitrariness of improper priors. This approach has shown to be quite useful. See Berger and Pericchi (1996a), Varshavsky (1996) and Lingham and Sivaganesan (1997).

The paper is arranged as follows. In Section 2, we review the concept of Bayes factors and intrinsic priors, and derive a general form of intrinsic priors for testing equal means against ordered means for k independent exponential distributions. Special cases when $k = 2$ and $k = 3$ are studied in detail. In Section 3, we give some numerical results along with real data analysis to illustrate theoretical results. Finally, few comments are given in Section 4.

2 Intrinsic Priors for Ordered Exponential Means

2.1 Preliminaries

Suppose that there are q different models, say $M_1, \ldots, M_q$, any of which would be possible for a statistical problem. If the model M_i holds, the data $\boldsymbol{X} = (X_1, \ldots, X_n)^t$ follow a parametric distribution with the density function $f_i(\boldsymbol{X}|\theta_i)$, where θ_i is a vector of unknown parameters. Let $\boldsymbol{\Theta}_i$ be the parameter space for θ_i. Based on the observations $\boldsymbol{X}$, one wants to select the correct model M_i among q possible models. Bayesian model selection proceeds by choosing a prior distribution $\pi_i(\theta_i)$ for θ_i under M_i, and a model probability $p(M_i)$ of M_i being true, for $i = 1, \ldots, q$. The posterior probability that M_i is true is then

$$P(M_i|\boldsymbol{X}) = \Big[\sum_{j=1}^{q} \frac{p(M_j)}{p(M_i)} B_{ji}\Big]^{-1}, \tag{2}$$

where B_{ji}, the Bayes factor of the model M_j to the model M_i, is defined by

$$B_{ji} = \frac{m_j(\boldsymbol{X})}{m_i(\boldsymbol{X})} = \frac{\int_{\boldsymbol{\Theta}_j} f_j(\boldsymbol{X}|\theta_j)\pi_j(\theta_j)d\theta_j}{\int_{\boldsymbol{\Theta}_i} f_i(\boldsymbol{X}|\theta_i)\pi_i(\theta_i)d\theta_i}, \tag{3}$$

where $m_i(\boldsymbol{X})$ is the marginal or predictive density of $\boldsymbol{X}$ under M_i. The posterior probabilities in (2) are used for selecting the most plausible model.

Let $\pi_i^N(\theta_i)$ be a noninformative prior for θ_i in model M_i. Then B_{ji}, the Bayes factor of M_j to M_i, could be defined by

$$B_{ji}^N = \frac{m_j^N(\boldsymbol{X})}{m_i^N(\boldsymbol{X})} = \frac{\int f_j(\boldsymbol{X}|\theta_j)\pi_j^N(\theta_j)d\theta_j}{\int f_i(\boldsymbol{X}|\theta_i)\pi_i^N(\theta_i)d\theta_i}. \tag{4}$$

A noninformative prior $\pi_i^N(\theta_i)$ is often improper, and is defined only up to an arbitrary constant c_i. Thus, B_{ji}^N is defined only up to (c_j/c_i), which is also arbitrary so that the Bayes factor is not well defined. To overcome the problem, one may use a part of the data as a so-called training sample, say $\boldsymbol{X}(l)$. The idea is to obtain the (intermediate) posterior $\pi_i^N(\theta_i|\boldsymbol{X}(l))$ then use $\pi_i^N(\theta_i|\boldsymbol{X}(l))$ as the prior to compute the Bayes factor for the $\boldsymbol{X}(-l)$, the remainder of the data. Consequently, the Bayes factor is as follows:

$$\begin{aligned} B_{ji}(l) &= \frac{\int_{\Theta_j} f_j(\boldsymbol{X}(-l)|\theta_j, \boldsymbol{X}(l))\pi_j^N(\theta_j|\boldsymbol{X}(l))d\theta_j}{\int_{\Theta_i} f_i(\boldsymbol{X}(-l)|\theta_i, \boldsymbol{X}(l))\pi_i^N(\theta_i|\boldsymbol{X}(l))d\theta_i} \\ &= B_{ji}^N \cdot B_{ij}^N(\boldsymbol{X}(l)), \end{aligned} \tag{5}$$

where for $h = i, j$,

$$m_h^N(\boldsymbol{X}(l)) = \int_{\Theta_h} f_h(\boldsymbol{X}(l)|\theta_h)\pi_h^N(\theta_h)d\theta_h.$$

In practice, $\boldsymbol{X}(l)$ is chosen to be a minimal training sample in the sense that the marginal $m_h^N(\boldsymbol{X}(l))$ is finite for all possible models, and no subset of $\boldsymbol{X}(l)$ gives finite marginals. Clearly, $B_{ji}(l)$ does not depends on (c_i, c_j). Furthermore, the Bayes factor defined in (5), depends on the choice of the minimal training sample. To avoid this dependence, Berger and Pericchi (1996b) suggested to take the average of $B_{ji}(l)$ over all $\boldsymbol{X}(l)$.

Definition 1 The arithmetic intrinsic Bayes factor (AI) of M_j to M_i is given by

$$B_{ji}^{AI} = \frac{1}{R}\sum_{l=1}^{R} B_{ji}(l) = B_{ji}^N \cdot \frac{1}{R}\sum_{l=1}^{R} B_{ij}^N(\boldsymbol{X}(l)), \tag{6}$$

where R is the number of all possible minimal training samples. Noticing that computation can be a problem if R is large, Berger and Pericchi (1996b) proposed the use of the following quantity.

Definition 2 The expected arithmetic intrinsic Bayes factor (EAI) of M_j to M_i is given by

$$B_{ji}^{EAI} = B_{ji}^N \cdot \frac{1}{R}\sum_{l=1}^{R} E_{\hat{\theta}_j}^{M_j}[B_{ij}^N(\boldsymbol{X}(l))], \tag{7}$$

where $\hat{\theta}_j = \hat{\theta}_{nj}$ is the MLE of θ_j. Alternatively, Berger and Pericchi (1996b) suggested finding a pair of proper priors such that the Bayes factor using the proper priors will be asymptotically equivalent to B^{AI}_{ji}. Such proper priors are called intrinsic priors if they exist. We need the following conditions to define intrinsic priors.

Condition 1 Under M_j, $\hat{\theta}_j \to \theta_j$, *a.s.* and $\hat{\theta}_i \to \psi_i(\theta_j)$, as $n \to \infty$.

Condition 2 Under M_i, $\hat{\theta}_i \to \theta_i$, *a.s.* and $\hat{\theta}_j \to \psi_j(\theta_i)$, as $n \to \infty$.

Here $\hat{\theta}_h$ is the MLE of θ_h under model M_h and ψ_h is a known function, for $h = i, j$, Normally we use

$$\psi_i(\theta_j) = \lim_{n\to\infty} E^{M_j}_{\theta_j}(\hat{\theta}_i). \tag{8}$$

Berger and Pericchi (1996b) showed that a pair of intrinsic priors (π_i^I, π_j^I) is a solution of the following system of functional equations:

$$\begin{cases} \dfrac{\pi_j^I(\theta_j)\pi_i^N(\psi_i(\theta_j))}{\pi_j^N(\theta_j)\pi_i^I(\psi_i(\theta_j))} = B_j^*(\theta_j), \\ \dfrac{\pi_j^I(\psi_j(\theta_i))\pi_i^N(\theta_i)}{\pi_j^N(\psi_j(\theta_i))\pi_i^I(\theta_i)} = B_i^*(\theta_i), \end{cases} \tag{9}$$

where for $h = i, j$,

$$B_h^*(\theta_h) = \lim_{R\to\infty} E^{M_h}_{\theta_h}[\frac{1}{R}\sum_{l=1}^{R} B^N_{ij}(\boldsymbol{X}(l))]. \tag{10}$$

The noninformative priors $\pi_i^N(\theta_i)$ and $\pi_j^N(\theta_j)$ are called starting priors. We note that solutions are not necessarily unique, nor necessarily proper. It is of interest to find proper intrinsic priors for given starting priors. Once we derive proper intrinsic priors, B^{AI}_{ji} can be replaced by the ordinary Bayes factors computed based on intrinsic priors.

2.2 Main Results

Suppose that we have independent observations $X_{ij} \sim Exp(\mu_i), i = 1, 2, \ldots, k;\ j = 1, 2 \ldots, n_i$. We want to test whether the k population means are equal or in ascending order. That is to say, to select between two competing models given by (1). Let

$$X_{i\cdot} = \sum_{j=1}^{n_i} X_{ij} \text{ and } \overline{X}_i = \frac{X_{i\cdot}}{n_i}.$$

Define the total sample size N by $N = \sum_{i=1}^{k} n_i$. Assume that there are k constants $a_i \in (0, 1)$ such that $a_1 + a_2 + \cdots + a_k = 1$ and for $i = 1, \ldots, k$,

$$\frac{n_i}{N} \to a_i \text{ as } N \to \infty. \tag{11}$$

Let $\mathbf{L}_k = \{\boldsymbol{\mu}_k = (\mu_1, \ldots, \mu_k) : 0 < \mu_1 \le \mu_2 \le \cdots \le \mu_k < \infty\}$. We use Jeffreys' priors as starting priors for both models M_1 and M_2. The reason to choose Jeffreys' prior under model M_1 is obvious. Under model M_2 and $k = 2$, Jeffreys' prior is both the reference prior and the matching prior when either parameter is of interest (cf. Ghosh and Sun, 1997). Analogously, we start from Jeffreys' prior for arbitrary k. Let μ be the common value of μ_i under M_1. Then

$$\pi_1^N(\mu) = \frac{1}{\mu},\ \mu > 0, \text{ and } \pi_2^N(\boldsymbol{\mu}_k) = \frac{1}{\mu_1 \cdots \mu_k},\ \boldsymbol{\mu}_k \in \mathbf{L}_k.$$

Recall that $B_{21}^N = m_2^N(\boldsymbol{X})/m_1^N(\boldsymbol{X})$, where

$$m_1^N(\boldsymbol{X}) = \int_0^\infty \frac{1}{\mu^{N+1}} \exp\left\{-\left[\frac{\sum_{i=1}^k X_{i\cdot}}{\mu}\right]\right\} d\mu = \frac{,\,(N)}{(\sum_{i=1}^k X_{i\cdot})^N},$$

and

$$m_2^N(\boldsymbol{X}) = \int_{\mathbf{L}_k} \frac{1}{\mu_1^{n_1+1}} \cdots \frac{1}{\mu_k^{n_k+1}} \exp\left\{-\sum_{i=1}^k \frac{X_{i\cdot}}{\mu_i}\right\} d\boldsymbol{\mu}_k.$$

A typical minimal training sample is $\boldsymbol{X}(l) = (X_{1h_1}, X_{2h_2}, \ldots, X_{kh_k})^t$. Then the marginal densities of $\boldsymbol{X}(l)$ are

$$\begin{aligned} m_1^N(\boldsymbol{X}(l)) &= \frac{,\,(k)}{(X_{1h_1} + \cdots + X_{kh_k})^k}; \\ m_2^N(\boldsymbol{X}(l)) &= \frac{1}{X_{1h_1}(X_{1h_1} + X_{2h_2}) \cdots (X_{1h_1} + X_{2h_2} + \cdots + X_{kh_k})}. \end{aligned}$$

Thus the Bayes factor based on the training sample $\boldsymbol{X}(l)$ is

$$B_{12}^N(\boldsymbol{X}(l)) = \frac{,\,(k) X_{1h_1}(X_{1h_1} + X_{2h_2}) \cdots (X_{1h_1} + \cdots + X_{k-1,h_{k-1}})}{(X_{1h_1} + \cdots + X_{kh_k})^{k-1}}. \tag{12}$$

Consequently, the AI Bayes factor and the EAI Bayes factors are

$$B_{21}^{AI} = B_{21}^N \cdot \frac{1}{n_1 \cdots n_k} \sum_{h_1=1}^{n_1} \cdots \sum_{h_k=1}^{n_k} B_{12}^N(\boldsymbol{X}(l)), \tag{13}$$

and

$$B_{21}^{EAI} = B_{21}^N \cdot E_{\hat{\theta}_2}^{M_2} B_{12}^N(\boldsymbol{X}(l)), \tag{14}$$

respectively. We need to find $\psi_1(\theta_2)$ and $\psi_2(\theta_1)$ in Conditions 1 and 2. Here $\theta_1 = \mu$ and $\theta_2 = \boldsymbol{\mu}_k$.

Fact 1 a) The MLE of μ under M_1 is given by

$$\hat{\mu} = N^{-1} \sum_{i=1}^k X_{i\cdot}. \tag{15}$$

b) The unrestricted MLE of μ_i is given by

$$\hat{\mu}_i^* = \overline{X}_i,\ i = 1, \ldots, k. \tag{16}$$

Proof: It is simple. $\square$

Let $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_1, \ldots, \hat{\mu}_k)^t$ be the restricted MLE of $\boldsymbol{\mu}_k$ under M_2. Note that $\hat{\boldsymbol{\mu}}_k$ can be computed by several algorithms. See Robertson et al. (1988).

Proposition 1 a) Under M_2, when $N \to \infty$, we have

$$\hat{\theta}_1 = \hat{\mu} \longrightarrow \psi_1(\boldsymbol{\mu}_k) \equiv \sum_{i=1}^{k} a_i \mu_i, \ a.s.,$$

where $\hat{\mu}$ is given by (15) and a_i is given by (11).

b) Under M_1, when $N \to \infty$, we have

$$\hat{\theta}_2 = \hat{\boldsymbol{\mu}}_k \longrightarrow \psi_2(\mu) \equiv (\mu, \ldots, \mu)^t, \ a.s..$$

Proof: For a) it is simple. For b), under M_2 we have the following inequality (see Robertson et al., 1988, p. 40),

$$\sum_{i=1}^{k} [\hat{\mu}_i - \mu_i]^2 \frac{n_i}{N} \leq \sum_{i=1}^{k} [\hat{\mu}_i^* - \mu_i]^2 \frac{n_i}{N}, \tag{17}$$

where $\hat{\mu}_i^*$ is given by (16). By the strong consistency of the unrestricted MLE of μ_i and the assumption (11), the right-hand side of (17) converges to zero as $N \to \infty$. Thus, the left-hand side of (17) also converges to zero. The result follows from the fact that under M_1, $\mu_i = \mu$ for each $i = 1, \ldots, k$. $\square$

Clearly $B_h^*(\theta_h)$ depends on k, the total number of populations. To distinguish the quantities $B_h^*(\theta_h)$ for different k, we write $B_{hk}^*(\theta_h) = B_h^*(\theta_h)$. From the definition (10), we see that

$$B_{hk}^*(\theta_h) = , (k) E_{\theta_h}^{M_h} \left[\frac{X_{11}(X_{11} + X_{21}) \cdots (X_{11} + \cdots + X_{k-1,1})}{(X_{11} + \cdots + X_{k1})^{k-1}} \right], \ h = 1, 2.$$

For any $k \geq 1$, define

$$\mathbf{A}_k = \{\mathbf{w}_k = (w_1, \ldots, w_k)^t : 0 < w_1, \ldots, w_k < 1\}. \tag{18}$$

Proposition 2 The quantities $B_{1k}^*(\theta_1)$ and $B_{2k}^*(\theta_2)$ are given by

$$B_{1k}^*(\mu) = , (k)^2 \int_{\mathbf{A}_{k-1}} \frac{w_1 w_2^3 \cdots w_{k-1}^{2k-3}}{(q_0 + q_1 + \cdots + q_{k-1})^k} d\mathbf{w}_{k-1} = \frac{, (k)}{2^{k-1}}, \tag{19}$$

and

$$B_{2k}^*(\boldsymbol{\mu}_k) = \frac{, (k)^2}{\mu_1 \cdots \mu_k} \int_{\mathbf{A}_{k-1}} \frac{w_1 w_2^3 \cdots w_{k-1}^{2k-3}}{[q_0/\mu_1 + \cdots + q_{k-1}/\mu_k]^k} d\mathbf{w}_{k-1}, \tag{20}$$

where

$$\begin{cases} q_0 = w_1 w_2 \cdots w_{k-1},\ q_1 = (1-w_1) w_2 \cdots w_{k-1}, \cdots, \\ q_{k-2} = (1-w_{k-2}) w_{k-1},\ q_{k-1} = 1 - w_{k-1}. \end{cases} \tag{21}$$

Proof: We first derive B^*_{2k}. The joint density of $(X_{11}, \ldots, X_{k1})$ is given by

$$f(x_{11}, \ldots, x_{k1}) = \Big(\prod_{i=1}^{k} \frac{1}{\mu_i}\Big) \exp\Big\{-(\frac{x_{11}}{\mu_1} + \cdots + \frac{x_{k1}}{\mu_k})\Big\},\ x_{i1} > 0.$$

Making the following transformations,

$$\begin{cases} W_1 = \frac{X_{11}}{(X_{11}+X_{21})}, \\ W_2 = \frac{(X_{11}+X_{21})}{(X_{11}+X_{21}+X_{31})}, \\ \cdots \\ W_{k-1} = \frac{(X_{11}+\cdots+X_{k-1,1})}{(X_{11}+\cdots+X_{k1})}, \\ W_k = X_{11} + \cdots + X_{k1}, \end{cases} \quad \text{or} \quad \begin{cases} X_{11} = W_1 W_2 \cdots W_k, \\ X_{21} = (1-W_1) W_2 \cdots W_k, \\ \cdots \\ X_{k-1,1} = (1-W_{k-2}) W_{k-1} W_k, \\ X_{k1} = (1-W_{k-1}) W_k. \end{cases} \tag{22}$$

We have

$$B^*_{2k}(\mu_1, \ldots, \mu_k) = \Gamma(k) E(W_1 W_2^2 \cdots W_{k-1}^{k-1}). \tag{23}$$

The Jacobian of this transformation is

$$\begin{aligned} |J| &= \left| \frac{\partial(x_{11}, x_{21}, \ldots, x_{k1})}{\partial(w_1, w_2, \ldots, w_k)} \right| \\ &= \begin{vmatrix} w_2 \cdots w_k & w_1 w_3 \cdots w_k & \cdots & w_1 \cdots w_{k-1} \\ -w_2 \cdots w_k & (1-w_1) w_3 \cdots w_k & \cdots & (1-w_1) w_2 \cdots w_{k-1} \\ 0 & -w_3 \cdots w_k & \cdots & (1-w_2) w_3 \cdots w_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 - w_{k-1} \end{vmatrix} \\ &= \begin{vmatrix} w_2 \cdots w_k & w_1 w_3 \cdots w_k & \cdots & w_1 \cdots w_{k-1} \\ 0 & w_3 \cdots w_k & \cdots & w_2 \cdots w_{k-1} \\ 0 & 0 & w_4 \cdots w_k & w_3 \cdots w_{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} \\ &= w_2 w_3^2 \cdots w_k^{k-1}. \end{aligned}$$

The joint density of $(W_1, \ldots, W_k)^t$ is then

$$f(w_1, \ldots, w_k) = \frac{w_2 w_3^2 \cdots w_k^{k-1}}{\mu_1 \cdots \mu_k} \exp\{-w_k(\frac{q_0}{\mu_1} + \cdots + \frac{q_{k-1}}{\mu_k})\},$$

where $\mathbf{w}_{k-1} = (w_1, \ldots, w_{k-1})^t \in \mathbf{A}_{k-1}$ and $w_k > 0$. Integrating out with respect to w_k, we get the following joint density of $(W_1, \ldots, W_{k-1})^t$

$$g(\mathbf{w}_{k-1}) = \frac{, (k)}{\mu_1 \cdots \mu_k} \frac{w_2 w_3^2 \cdots w_{k-1}^{k-2}}{[q_0/\mu_1 + \cdots + q_{k-1}/\mu_k]^k}, \ \mathbf{w}_{k-1} \in \mathbf{A}_{k-1},$$

where q_i's are given by (21). Hence equation (20) is established. Note that $\sum_{i=0}^{k-1} q_i = 1$ in (21). Since $\mu_i = \mu$ under M_1, equation (19) immediately follows from (20). This completes the proof. □

Denote $\mathbf{1}_k = (1, \ldots, 1)^t \in \mathbb{R}^k$. The system of equations (9) becomes

$$\begin{cases} \dfrac{\pi_2^I(\boldsymbol{\mu}_k)/(a_1\mu_1 + \cdots + a_k\mu_k)}{\pi_1^I(a_1\mu_1 + \cdots + a_k\mu_k)/(\mu_1 \cdots \mu_k)} = B_{2k}^*(\boldsymbol{\mu}_k), \ \boldsymbol{\mu}_k \in \mathbf{L}_k, \\ \dfrac{\pi_2^I(\mu \mathbf{1}_k)/\mu}{\pi_1^I(\mu)/\mu^k} = B_{1k}^*(\mu), \ \mu > 0. \end{cases} \tag{24}$$

Lemma 1 $B_{2k}^*(\boldsymbol{\mu}_k) \longrightarrow B_{1k}^*(\mu)$ as $\boldsymbol{\mu}_k \longrightarrow \mu \mathbf{1}_k$, where the limit $\boldsymbol{\mu}_k \longrightarrow \mu \mathbf{1}_k$ is taken within the region $\boldsymbol{\mu}_k \in \mathbf{L}_k$.

Proof: Since $W_1^1 W_2^2 \cdots W_{k-1}^{k-1}$ is bounded, it follows from (23) that $B_{2k}^* \to B_{1k}^*$ as $\boldsymbol{\mu}_k \longrightarrow \mu \mathbf{1}_k$. This completes the proof. □

Lemma 2 For any integer $l \geq 1$, and any constants $q_i \in (0,1)$ satisfying $q_0 + \cdots + q_l = 1$,

$$\int_{\mathbf{A}_l} \frac{1}{t_1^2 t_2^3 \cdots t_l^{l+1} [q_0/(\prod_{h=1}^l t_h) + q_1/(\prod_{h=2}^l t_h) + \cdots + q_{l-1}/t_l + q_l]^{l+1}} d\mathbf{t}_l$$
$$= \frac{1}{l! \ q_0(q_0 + q_1) \cdots (q_0 + q_1 + \cdots + q_l)}.$$

Proof: We use induction. For $l = 1$, we have

$$\int_{\mathbf{A}_1} \frac{1}{t_1^2 (q_0/t_1 + q_1)^2} dt_1 = \frac{1}{q_0(q_0 + q_1)} = \frac{1}{q_0}.$$

Assume that the result holds for $l - 1$. Now for l, we have

$$\begin{aligned} &\int_{\mathbf{A}_l} \frac{1}{t_1^2 t_2^3 \cdots t_l^{l+1} [q_0/(\prod_{h=1}^l t_h) + q_1/(\prod_{h=2}^l t_h) + \cdots + q_{l-1}/t_l + q_l]^{l+1}} d\mathbf{t}_l \\ = \ & \frac{1}{l q_l} \int_{\mathbf{A}_{l-1}} \frac{1}{t_1^2 t_2^3 \cdots t_{l-1}^l} \left\{ \frac{1}{[q_0/(\prod_{h=1}^{l-1} t_h) + q_1/(\prod_{h=2}^{l-1} t_h) + \cdots + q_{l-2}/t_{l-1} + q_{l-1}]^l} \right. \\ - \ & \left. \frac{1}{[q_0/(\prod_{h=1}^{l-1} t_h) + q_1/(\prod_{h=2}^{l-1} t_h) + \cdots + q_{l-2}/t_{l-1} + q_{l-1} + q_l]^l} \right\} d\mathbf{t}_{l-1}. \end{aligned} \tag{25}$$

Let $\xi = \sum_{i=0}^{l-1} q_i$. Then

$$\begin{aligned} &\int_{\mathbf{A}_{l-1}} \frac{1}{t_1^2 t_2^3 \cdots t_{l-1}^l [q_0/(\prod_{h=1}^{l-1} t_h) + q_1/(\prod_{h=2}^{l-1} t_h) + \cdots + q_{l-2}/t_{l-1} + q_{l-1}]^l} d\mathbf{t}_{l-1} \\ = \ & \frac{1}{\xi^l} \int_{\mathbf{A}_{l-1}} \frac{1}{t_1^2 t_2^3 \cdots t_{l-1}^l [(q_0/\xi)/(\prod_{h=1}^{l-1} t_h) + (q_1/\xi)/(\prod_{h=2}^{l-1} t_h) + \cdots + q_{l-1}/\xi]^l} d\mathbf{t}_{l-1}. \end{aligned}$$

Since $\sum_{i=0}^{l-1} q_i/\xi = 1$, it follows from the induction assumption for $l-1$ that the integral equals

$$\begin{aligned} & \frac{1}{\xi^l (l-1)!(q_0/\xi)(q_0/\xi + q_1/\xi)\cdots(q_0/\xi + \cdots + q_{l-1}/\xi)} \\ = \ & \frac{1}{(l-1)! q_0 (q_0+q_1)\cdots(q_0+\cdots+q_{l-1})}. \end{aligned}$$

Also, by the induction assumption, we have

$$\begin{aligned} & \int_{\mathbf{A}_{l-1}} \frac{1}{t_1^2 t_2^3 \cdots t_{l-1}^l [q_0/(\prod_{h=1}^{l-1} t_h) + q_1/(\prod_{h=2}^{l-1} t_h) + \cdots + q_{l-2}/t_{l-1} + q_{l-1} + q_l]^l} d\mathbf{t}_{l-1} \\ = \ & \frac{1}{(l-1)! q_0 (q_0+q_1)\cdots(q_0+\cdots+q_{l-2})(q_0+\cdots+q_{l-1}+q_l)}. \end{aligned}$$

Consequently, the left-hand side of (25) equals

$$\begin{aligned} & \frac{1}{lq_l}\left\{ \frac{1}{(l-1)! q_0 (q_0+q_1)\cdots(q_0+\cdots+q_{l-2})(q_0+\cdots+q_{l-1})} \right. \\ - \ & \left. \frac{1}{(l-1)! q_0 (q_0+q_1)\cdots(q_0+\cdots+q_{l-2})(q_0+\cdots+q_{l-1}+q_l)} \right\} \\ = \ & \frac{1}{l!\ q_0 (q_0+q_1)\cdots(q_0+\cdots+q_l)}. \end{aligned}$$

Hence, the result also holds for l, which completes the proof. □

Lemma 3 Define

$$t_1 = \frac{\mu_1}{\mu_2},\ t_2 = \frac{\mu_2}{\mu_3}, \ldots, t_{k-1} = \frac{\mu_{k-1}}{\mu_k},\ t_k = \mu_k. \tag{26}$$

Then B_{2k}^* depends only on $\mathbf{t}_{k-1} = (t_1, \ldots, t_{k-1})^t$.

Proof: It follows directly from (20). □

Theorem 1 For any proper density $g(\cdot)$ on $(0,\infty)$, the system of priors

$$\begin{cases} \pi_1^I(\mu) = g(\mu),\ 0 < \mu < \infty, \\ \pi_2^I(\boldsymbol{\mu}_k) = \dfrac{a_1\mu_1 + \cdots + a_k\mu_k}{\mu_1\cdots\mu_k} B_{2k}^*(\boldsymbol{\mu}_k)\pi_1^I(a_1\mu_1 + \cdots + a_k\mu_k),\ \boldsymbol{\mu}_k \in \mathbf{L}_k \end{cases} \tag{27}$$

is a solution of (24), where B_{2k}^* is given by (20). Furthermore, π_2^I is a proper density on $\mathbf{L}_k$.

Proof: From Lemma 1, we can see that (27) is a solution of (24). The Jacobian of the transformation from $\boldsymbol{\mu}_k$ to $\mathbf{t}_k$ in (26) is

$$|J| = \left|\frac{\partial(\mu_1,\ldots,\mu_k)}{\partial(t_1,\ldots,t_k)}\right| = t_2 t_3^2 \cdots t_k^{k-1}.$$

So,

$$\begin{aligned}
&\int_{\mathbf{A}_{k-1}} \int_0^\infty \pi_2^I(t_1, \ldots, t_k) dt_k d\mathbf{t}_{k-1} \\
= \quad &, (k)^2 \int_{\mathbf{A}_{k-1}} \Bigg\{ \frac{1}{t_1^2 t_2^3 \cdots t_{k-1}^k} \int_0^\infty \Big(a_1 \prod_{h=1}^{k-1} t_h + a_2 \prod_{h=2}^{k-1} t_h + \cdots + a_{k-1} t_{k-1} + a_k \Big) \\
&\pi_1^I \Big[t_k \Big(a_1 \prod_{h=1}^{k-1} t_h + a_2 \prod_{h=2}^{k-1} t_h + \cdots + a_{k-1} t_{k-1} + a_k \Big) \Big] dt_k \Bigg\} \\
&\Bigg\{ \int_{\mathbf{A}_{k-1}} \frac{w_1 w_2^3 \cdots w_{k-1}^{2k-3}}{[q_0/(\prod_{h=1}^{k-1} t_h) + q_1/(\prod_{h=2}^{k-1} t_h) + \cdots + q_{k-2}/t_{k-1} + q_{k-1}]^k} d\mathbf{w}_{k-1} \Bigg\} d\mathbf{t}_{k-1}, \qquad (28)
\end{aligned}$$

where $\mathbf{A}_{k-1}$ and the q_i's are defined by (18) and (21) respectively. Let $s = t_k(a_1 \prod_{h=1}^{k-1} t_h + a_2 \prod_{h=2}^{k-1} t_h + \cdots + a_k)$. Then $dt_k/ds = (a_1 \prod_{h=1}^{k-1} t_h + a_2 \prod_{h=2}^{k-1} t_h + \cdots + a_k)^{-1}$ and (28) equals

$$, (k)^2 \int_{\mathbf{A}_{k-1}} \int_{\mathbf{A}_{k-1}} \frac{w_1 w_2^3 \cdots w_{k-1}^{2k-3}}{t_1^2 t_2^3 \cdots t_{k-1}^k [q_0/(\prod_{h=1}^{k-1} t_h) + q_1/(\prod_{h=2}^{k-1} t_h) + \cdots + q_{k-1}]^k} d\mathbf{t}_{k-1} d\mathbf{w}_{k-1}. \qquad (29)$$

Notice that $q_0(q_0 + q_1) \cdots (q_0 + \cdots + q_{k-1}) = w_1 w_2^2 \cdots w_{k-1}^{k-1}$. ¿From Lemma 2, (29) becomes

$$, (k) \int_{\mathbf{A}_{k-1}} w_2 w_3^2 \cdots w_{k-1}^{k-2} d\mathbf{w}_{k-1} = 1.$$

This completes the proof. □

The following theorem explains the structure of the intrinsic prior $\pi_2^I(\boldsymbol{\mu}_k)$.

Theorem 2 a) The marginal intrinsic prior of $\mathbf{t}_{k-1}$ is

$$\pi_2^I(\mathbf{t}_{k-1}) = \frac{h_{k-1}(\mathbf{t}_{k-1})}{t_1 t_2 \cdots t_{k-1}}, \ \mathbf{t}_{k-1} \in \mathbf{A}_{k-1},$$

where

$$h_{k-1}(\mathbf{t}_{k-1}) = B_{2k}^* \Big(\prod_{h=1}^{k-1} t_h, \prod_{h=2}^{k-1} t_h, \ldots, t_{k-1}, 1 \Big), \ \mathbf{t}_{k-1} \in \mathbf{A}_{k-1}. \qquad (30)$$

b) The conditional intrinsic prior of t_k given $\mathbf{t}_{k-1}$ is

$$\pi_2^I(t_k | \mathbf{t}_{k-1}) \propto \pi_1^I(\xi t_k), \ t_k > 0,$$

where

$$\xi = a_1 \prod_{h=1}^{k-1} t_h + a_2 \prod_{h=2}^{k-1} t_h + \cdots + a_{k-1} t_{k-1} + a_k.$$

Proof: For part a), it follows from (27) that the joint intrinsic prior of $(\mathbf{t}_{k-1}^t, t_k)$ is

$$\pi_2^I(\mathbf{t}_{k-1}^t, t_k) \quad = \quad \frac{\xi}{t_1 \cdots t_{k-1}} B_{2k}^*\Big(\prod_{h=1}^k t_h, \prod_{h=2}^k t_h, \ldots, t_k\Big)\pi_1^I(\xi t_k). \tag{31}$$

Applying Lemma 3, the desired result follows from integrating equation (31) over t_k. The proof of part b) follows directly from part a). □

Corollary 1 When $g(t)$ is the probability density function of Inverse Gamma (λ, η), the pair of intrinsic priors is

$$\begin{cases} \pi_1^I(\mu) = \dfrac{\eta^\lambda}{\Gamma(\lambda)\mu^{\lambda+1}} e^{-\frac{\eta}{\mu}}, \ 0 < \mu < \infty, \\ \pi_2^I(\boldsymbol{\mu}_k) = \dfrac{\eta^\lambda \exp\{-\eta/(a_1\mu_1 + \cdots + a_k\mu_k)\}}{\Gamma(\lambda)(a_1\mu_1 + \cdots + a_k\mu_k)^\lambda \mu_1 \cdots \mu_k} B_{2k}^*(\boldsymbol{\mu}_k), \ \boldsymbol{\mu}_k \in \mathbf{L}_k. \end{cases} \tag{32}$$

2.3 Special cases when $k = 2$ and $k = 3$

We now derive the closed forms of $\pi_2^I(\mathbf{t}_{k-1})$ when $k = 2$ and $k = 3$.

Proposition 3 The quantities $h_1(t_1)$ and $h_2(t_1, t_2)$ are given by

a) $h_1(t_1) = B_{22}^*(t_1, 1), \ 0 < t_1 < 1,$

b) $h_2(t_1, t_2) = B_{23}^*(t_1t_2, t_2, 1), \ 0 < t_1, t_2 < 1,$

where

$$h_1(t_1) = \frac{t_1(-\log t_1 + t_1 - 1)}{(1-t_1)^2}, \ 0 < t_1 < 1, \tag{33}$$

and

$$\begin{aligned} h_2(t_1, t_2) \quad &= \quad 2t_1t_2\Big\{-\frac{t_1^2 t_2}{(1-t_1)(1-t_1t_2)^2} + \frac{1}{(1-t_2)(1-t_1t_2)^2} \\ &+ \quad \frac{t_2 \log t_2}{(1-t_1)^2(1-t_2)^2} - \frac{t_1^2 t_2(3 - 2t_1 - t_1t_2)\log(t_1t_2)}{(1-t_1)^2(1-t_1t_2)^3}\Big\}, \ 0 < t_1, t_2 < 1. \end{aligned} \tag{34}$$

Proof: From (20) the quantities B_{22}^* and B_{23}^* are

$$\begin{aligned} B_{22}^*(\mu_1, \mu_2) \quad &= \quad \mu_1\mu_2 \int_0^1 \frac{w_2}{[\mu_1 + (\mu_2 - \mu_1)w_2]^2} dw_2 = \frac{\mu_1\mu_2}{(\mu_2 - \mu_1)^2}[\log(\frac{\mu_2}{\mu_1}) + \frac{\mu_1}{\mu_2} - 1], \\ B_{23}^*(\mu_1, \mu_2, \mu_3) \quad &= \quad \frac{4}{\mu_1\mu_2\mu_3} \int_0^1 \int_0^1 \frac{w_1 w_2^3}{(\frac{w_1w_2}{\mu_1} + \frac{w_2(1-w_1)}{\mu_2} + \frac{1-w_2}{\mu_3})^3} dw_1 dw_2 \\ &= \quad 2\mu_1 \Big[\frac{\mu_1^2}{(\mu_1 - \mu_2)(\mu_1 - \mu_3)^2} - \frac{\mu_3^2}{(\mu_2 - \mu_3)(\mu_1 - \mu_3)^2} \\ &+ \quad \frac{\mu_2^3 \log(\mu_2/\mu_3)}{(\mu_1 - \mu_2)^2(\mu_2 - \mu_3)^2} + \frac{\mu_1^2(3\mu_2\mu_3 - \mu_1\mu_2 - 2\mu_1\mu_3)}{(\mu_1 - \mu_2)^2(\mu_1 - \mu_3)^3} \log(\frac{\mu_1}{\mu_3})\Big]. \end{aligned} \tag{35}$$

By (30) the desired results are established. □

Figure 1 is the plot of the marginal intrinsic prior density $\pi_2^I(t_1)$ of $t_1 = \mu_1/\mu_2$ when $k = 2$. Here, $\pi_2^I(t_1) = h_1(t_1)/t_1$, Note that $\pi_2^I(t_1)$ is monotonic decreasing, and goes to 0.5 when $t_1 \to 1$. Although $\pi_2^I(t_1)$ is unbounded at $t_1 = 0$, it is integrable. Figure 2 is the contour plot of the marginal intrinsic prior density of $t_1 = \mu_1/\mu_2$ and $\mu + 2\mu_3$ wen $k = 3$. Here $\pi_2^I(t_1, t_2) = h_2(t_1, t_2)/(t_1 t_2)$, which is unbounded as either t_1 or $t_2 \to 0$, but it is integrable.

For $k = 2$ and 3, with the pair of intrinsic priors given by Corollary 1, we compute the analytic forms of ordinary Bayes factors, which are denoted by $B_{21}^{I2}(\boldsymbol{X})$ and $B_{21}^{I3}(\boldsymbol{X})$ respectively.

Proposition 4 For a pair of intrinsic priors in (32) we have

$$B_{21}^{I2}(\boldsymbol{X}) = (X_1. + X_2. + \eta)^{\lambda+n_1+n_2} H_1(X_1., X_2., a_1, a_2), \tag{36}$$

where

$$H_1(X_1., X_2., a_1, a_2) = \int_0^1 \frac{t_1^{\lambda+n_2-1}(a_1t_1 + a_2)^{n_1+n_2} h_1(t_1)}{[(a_1t_1 + a_2)X_1. + t_1(a_1t_1 + a_2)X_2. + \eta t_1]^{\lambda+n_1+n_2}} dt_1,$$

where $h_1(\cdot)$ is defined by (33).

Proof: Under M_1, the marginal density of $\boldsymbol{X}$ is

$$m_1^{I2}(\boldsymbol{X}) = \frac{\eta^\lambda , (\lambda + n_1 + n_2)}{, (\lambda)(X_1. + X_2. + \eta)^{\lambda+n_1+n_2}}.$$

¿From (26), under M_2, $t_1 = \mu_1/\mu_2$ and $t_2 = \mu_2$. Then the likelihood function becomes

$$f(\boldsymbol{X}|t_1, t_2) = \frac{1}{t_1^{n_1} t_2^{n_1+n_2}} \exp\{-\frac{1}{t_2}[\frac{X_1.}{t_1} + X_2.]\},\ 0 < t_1 < 1,\ t_2 > 0.$$

¿From (31) the intrinsic prior $\pi_2^I(t_1, t_2)$ is given by

$$\pi_2^I(t_1, t_2) = \frac{\eta^\lambda h_1(t_1)}{, (\lambda) t_1 (a_1t_1 + a_2)^\lambda t_2^{\lambda+1}} \exp\{-\frac{\eta}{t_2(a_1t_1 + a_2)}\},\ 0 < t_1 < 1,\ t_2 > 0.$$

The marginal density of $\boldsymbol{X}$ is then

$$\begin{aligned} m_2^{I2}(\boldsymbol{X}) &= \int_0^1 \int_0^\infty \frac{\eta^\lambda}{, (\lambda)} \frac{h_1(t_1) t_2^{-(\lambda+n_1+n_2+1)}}{t_1^{n_1+1}(a_1t_1 + a_2)^\lambda} \exp\{-\frac{1}{t_2}[\frac{X_1.}{t_1} + X_2. + \frac{\eta}{a_1t_1 + a_2}]\} dt_2 dt_1 \\ &= \frac{\eta^\lambda , (\lambda + n_1 + n_2)}{, (\lambda)} \int_0^1 \frac{t_1^{\lambda+n_2-1}(a_1t_1 + a_2)^{n_1+n_2} h_1(t_1)}{[(a_1t_1 + a_2)X_1. + t_1(a_1t_1 + a_2)X_2. + \eta t_1]^{\lambda+n_1+n_2}} dt_1. \end{aligned}$$

Since $B_{21}^{I2}(\boldsymbol{X}) = m_2^{I2}(\boldsymbol{X})/m_1^{I2}(\boldsymbol{X})$, the result follows immediately. □

Proposition 5 For a pair of intrinsic priors in (32) we have

$$B_{21}^{I3}(\boldsymbol{X}) = (X_{1\cdot} + X_{2\cdot} + X_{3\cdot} + \eta)^{\lambda+N} H_3(X_{1\cdot}, X_{2\cdot}, X_{3\cdot}, a_1, a_2, a_3), \tag{37}$$

where

$$H_3(X_{1\cdot}, X_{2\cdot}, X_{3\cdot}, a_1, a_2, a_3) = \int_0^1 \int_0^1 \frac{t_1^{\lambda+n_2+n_3-1} t_2^{\lambda+n_3-1} h_2(t_1,t_2)(a_1t_1t_2 + a_2t_2 + a_3)^{-\lambda}}{[X_{1\cdot} + X_{2\cdot}t_1 + t_1t_2(X_{3\cdot} + \eta/(a_1t_1t_2 + a_2t_2 + a_3))]^{\lambda+N}} dt_1 dt_2,$$

where $N = n_1 + n_2 + n_3$ and h_2 is defined by (34).

Proof: Under M_1, the marginal density of $\boldsymbol{X}$ is

$$m_1^{I3}(\boldsymbol{X}) = \frac{\eta^\lambda , (\lambda + N)}{, (\lambda)(X_{1\cdot} + X_{2\cdot} + X_{3\cdot} + \eta)^{\lambda+N}}.$$

¿From (26), under M_2, $t_1 = \mu_1/\mu_2$, $t_2 = \mu_2/\mu_3$ and $t_3 = \mu_3$. Then the likelihood function becomes

$$f(\boldsymbol{X}|t_1,t_2,t_3) = \frac{1}{t_1^{n_1} t_2^{n_1+n_2} t_3^N} \exp\{-\frac{1}{t_3}[\frac{X_{1\cdot}}{t_1t_2} + \frac{X_{2\cdot}}{t_2} + X_{3\cdot}]\},\ 0 < t_1, t_2 < 1,\ t_3 > 0.$$

Again from (31) the intrinsic prior $\pi_2^I(t_1,t_2,t_3)$ is given by

$$\pi_2^I(t_1,t_2,t_3) = \frac{\eta^\lambda h_2(t_1,t_2)}{, (\lambda) t_1 t_2 t_3^{\lambda+1}(a_1t_1t_2 + a_2t_2 + a_3)^\lambda} \exp\{-\frac{\eta}{s(a_1t_1t_2 + a_2t_2 + a_3)}\},$$

for $0 < t_1, t_2 < 1,\ t_3 > 0$. The marginal density of $\boldsymbol{X}$ is then

$$\begin{aligned} m_2^{I3}(\boldsymbol{X}) &= \frac{\eta^\lambda}{, (\lambda)} \int_0^1 \int_0^1 \int_0^\infty \frac{h_2(t_1,t_2)}{t_1^{n_1+1} t_2^{n_1+n_2+1} t_3^{\lambda+N+1}(a_1t_1t_2 + a_2t_2 + a_3)^\lambda} \\ &\quad \exp\Big\{-\frac{1}{t_3}[\frac{X_{1\cdot}}{t_1t_2} + \frac{X_{2\cdot}}{t_2} + X_{3\cdot} + \frac{\eta}{a_1t_1t_2 + a_2t_2 + a_3}]\Big\} dt_3 dt_1 dt_2 \\ &= \frac{\eta^\lambda , (\lambda+N)}{, (\lambda)} \int_0^1 \int_0^1 \frac{t_1^{\lambda+n_2+n_3-1} t_2^{\lambda+n_3-1} h_2(t_1,t_2)(a_1t_1t_2 + a_2t_2 + a_3)^{-\lambda}}{\big[X_{1\cdot} + X_{2\cdot}t_1 + t_1t_2(X_{3\cdot} + \eta/(a_1t_1t_2 + a_2t_2 + a_3))\big]^{\lambda+N}} dt_1 dt_2. \end{aligned}$$

Since $B_{21}^{I3}(\boldsymbol{X}) = m_2^{I3}(\boldsymbol{X})/m_1^{I3}(\boldsymbol{X})$, the result follows immediately. □

3 Numerical Examples

Example 2 . Suppose that we want to select between the two models $M_1 : \mu_1 = \mu_2$ and $M_2 : \mu_1 \le \mu_2$. The commonly used F-test is the generalized likelihood ratio test. The P-value is $F(\overline{X}_1/\overline{X}_2; 2n_1, 2n_2)$. To illustrate the difference between the F-test and the Bayesian model selection under the intrinsic priors developed in Section 2, we examine the cases when $\overline{X}_1/\overline{X}_2 = 1, 2, 3$, and $n_1 = n_2 = 12, 20, 30$. The numerical values of the P-value for some choices of n_1 and n_2

are given in the column 3 of Table 2. The Bayes factors and the posterior probability of M_1 are computed for three choices of (λ, η). They are $(0.01, 0.01)$, $(1.0, 1.0)$ and $(10, 10)$. We see that the posterior probabilities tend to be bigger than P-values. For the cases when $\overline{X}_2/\overline{X}_1 = 2$ or 3, as the sample sizes become larger, the Bayes factors will select M_2. Furthermore, the Bayes factors are quite robust in terms of the change of the values (λ, η).

Example 3 . The following data, given by Proschan (1963), are time intervals of successive failures of the air conditioning system in Boeing 720 jet airplanes. We assume that the time between successive failures for each plane is independent and exponentially distributed.

plane 1	50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14
plane 2	102, 209, 14, 57, 54, 32, 67, 59, 134, 152, 27, 14, 230, 66, 61, 34
plane 3	90, 10, 60, 186, 61, 49, 14, 24, 56, 20, 79, 84, 44, 59, 29, 118, 25, 156, 310, 76, 26, 44, 23, 62, 130, 208, 70, 101, 208
plane 4	23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95
plane 5	97, 51, 11, 4, 141, 18, 142, 68, 77, 80, 1, 16, 106, 206, 82, 54, 31, 216, 46, 111, 39, 63, 18, 191, 18, 163, 24
plane 6	74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326

In Table 3, we provide the P-values, Bayes factors and the posterior probabilities $P^I(M_1|\boldsymbol{X})$ for testing equal means $(M_1 : \mu_1 = \mu_2 = \mu_3)$ against ascending ordered means $(M_2 : \mu_1 \leq \mu_2 \leq \mu_3)$ for failure times for the first 3 planes and the last 3 planes respectively. The P-value is computed based on asymptotic procedures using level probabilities (cf. Robertson, et al., 1988). For the first 3 sets of data, there is no strong evidence for supporting model M_2 in terms of both the P-value and the $P^I(M_1|\boldsymbol{X})$. Moreover, the ordinary Bayes factor B_{21}^{I3} computed by (37) with $\lambda = \eta = 1$ is very close to the AI Bayes factor B_{21}^{AI}. For the last 3 sets of data, there is a disagreement between the P-value and Bayes factors. When we just look at the sample means of each set of data, it seems that there is a strong evidence for supporting model M_2. However, we can see that three particular observations 502, 386 and 326 in plane 6 enlarge the sample mean $\overline{X}_3$, which makes the P-value very small. Meanwhile, Bayes factors give fairly reasonable answers. We notice that there is about a 9% difference between B_{21}^{AI} and B_{21}^{I3}, which is quite big. To compensate for this inaccuracy, we make some changes for observations in plane 6. When we change three observations $(502, 12, 21)$ to

$(418, 54, 63)$, the AI Bayes factor B_{21}^{AI} becomes 1.7622, which is almost equal to the intrinsic Bayes factor B_{21}^{I3}.

4 Comments

It has noticed that a P-value often does not agree with the posterior probability that the null hypothesis is correct. Delampady and Berger (1990) have showed that the lower bounds of Bayes factors and posterior probabilities in favor of null hypotheses are much larger than the corresponding P-values of the chi-squared goodness of fit test.

As we see from numerical results, P-values tend to reject the null hypothesis frequently. Furthermore, P-values are computed based only on sufficient statistics, which might be misleading for some cases. The average intrinsic Bayes factors are computed based on entire observations so that they give accurate interpretations and fairly steady answers.

ACKNOWLEDGEMENTS

Research supported in part by the US Federal Highway Administration, under subcontract E77690017-3Y from Los Alamos National Laboratory to the National Institute of Statistical Sciences.

References

[1] Berger, J. O. and Bernado, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors, *Journal of the American Statistical Association,* **84**, pp. 200-207

[2] Berger, J. O. and Bernardo, J. (1992). On the development of the reference priors, in *Bayesian Statistics 4,* (Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. eds), Oxford University Press: London, pp. 35-60.

[3] Berger, J. O., Brown, L. D. and Wolpert, R. L. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis, *The Annals of Statistics,* **22**, pp. 1787-1807.

[4] Berger, J. O. and Pericchi, L. (1996a). The Intrinsic Bayes factor for linear models (with discussion), in *Bayesian Analysis V*, J. M. Bernado, et al., (Eds.), pp. 25-44, Oxford University Press, Oxford.

[5] Berger, J. O. and Pericchi, L. (1996b). The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association,* **91**, pp. 109-122.

[6] Delampady, M. and Berger, J. O. (1990). Lower bounds on the Bayes factors for multinomial distribution, with application to chi-squared tests of fit, *The Annals of Statistics,* **18**, pp. 1295-1316.

[7] Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection, *Journal of the American Statistical Association,* **74**, pp. 153-160.

[8] Ghosh, M. and Sun, D. (1997). Recent developments of Bayesian inference for stress-strength models. To appear in the Indian Association for Productivity Quality and Reliability (IAPQR) volume titled *"Frontiers in Reliability"*.

[9] Jeffreys, H. (1961). *Theory of Probability.* London: Oxford University Press.

[10] Lingham, R. T. and Sivaganesan, S. (1997). Testing hypotheses about the power law process under failure truncation using intrinsic Bayes factors. To appear in *Annals of the Institute of Statistical Mathematics.*

[11] Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate, *Technometrics,* **5**, pp. 375-383.

[12] Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). Order restricted statistical inference, John Wiley & Sons Ltd.

[13] San Martini, A. and Spezzaferri, F. (1984). A predictive model selection criterion, *Journal of Royal Statistical Society, ser. B* **46**, pp. 296-303.

[14] Spiegelhalter, D. J. and Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information, *Journal of Royal Statistical Society, ser. B* **44**, pp. 377-387.

[15] Varshavsky, J. A. (1996). Intrinsic Bayes factors for model selection with autoregressive data, in *Bayesian Analysis V*, J. M. Bernado, et al., (Eds.), pp. 757-763, Oxford University Press, Oxford.

Table 1: Comparison of error probabilities for testing $H_0 : \mu_1/\mu_2 = 2$ versus $H_1 : \mu_1/\mu_2 = 0.5$.

n	$\alpha = \beta$	$(\overline{X}_1, \overline{X}_2)$	P-value	B	$P(H_0 \mid \boldsymbol{X})$	$P(H_1 \mid \boldsymbol{X})$
12	0.0480	$(1.0, 1.0)$	0.0480	1	0.5	0.5
		$(1.0, 1.5)$	0.0046	24.445	0.0393	0.9607
		$(1.0, 2.0)$	0.0006	216.39	0.0046	0.9954
20	0.0155	$(1.0, 1.0)$	0.0155	1	0.5	0.5
		$(1.0, 1.5)$	0.0004	207.33	0.0048	0.9952
		$(1.0, 2.0)$	0.000013	$4,999$	0.0002	0.9998
30	0.0041	$(1.0, 1.0)$	0.0041	1	0.5	0.5
		$(1.0, 1.5)$	0.000018	$3,332$	0.0003	0.9997
		$(1.0, 2.0)$	0.00000013	$666,666$	0.0000015	0.9999985

Table 2: P-values, Bayes factors, and $P(M_1 \mid \boldsymbol{X})$ for testing $M_1 : \mu_1 = \mu_2$ versus $M_2 : \mu_1 \leq \mu_2$.

			$(\lambda, \eta) = (.01, .01)$		$(\lambda, \eta) = (1.0, 1.0)$		$(\lambda, \eta) = (10, 10)$	
n	$(\overline{X}_1, \overline{X}_2)$	P-value	B_{21}	$P(M_1 \mid \boldsymbol{X})$	B_{21}	$P(M_1 \mid \boldsymbol{X})$	B_{21}	$P(M_1 \mid \boldsymbol{X})$
12	$(1.0, 1.0)$	0.5	0.23027	0.81283	0.22983	0.81312	0.22711	0.81492
	$(1.0, 2.0)$	0.04805	1.54115	0.39352	1.52474	0.39608	1.44197	0.40950
	$(1.0, 3.0)$	0.00465	10.5823	0.08634	10.4746	0.08715	10.3576	0.08805
20	$(1.0, 1.0)$	0.5	0.18258	0.84561	0.18244	0.84571	0.18146	0.84641
	$(1.0, 2.0)$	0.01549	3.19048	0.23864	3.17004	0.23981	3.05367	0.24669
	$(1.0, 3.0)$	0.000373	81.8325	0.01207	81.3373	0.01215	81.2479	0.01216
30	$(1.0, 1.0)$	0.5	0.15129	0.86859	0.15124	0.86863	0.15083	0.86894
	$(1.0, 2.0)$	0.004055	8.56653	0.10453	8.53005	0.10493	8.30891	0.10742
	$(1.0, 3.0)$	0.000018	1182.77	0.00084	1178.03	0.00085	1180.56	0.00085

Table 3: P-values, Bayes factors and $P^I(M_1 \mid \boldsymbol{X})$ for testing $M_1 : \mu_1 = \mu_2 = \mu_3$ versus $M_2 : \mu_1 \leq \mu_2 \leq \mu_3$ for airplane data.

$(n_1, n_2, n_3, \overline{X}_1, \overline{X}_2, \overline{X}_3)$	P-value	B_{21}^{AI}	B_{21}^{EAI}	B_{21}^{I3}	$P^I(M_1 \mid \boldsymbol{X})$
$(24, 16, 29, 64.13, 82.00, 83.52)$	0.2432	0.1054	0.1164	0.1041	0.9058
$(30, 27, 15, 59.60, 76.81, 121.27)$	0.0222	1.9303	1.9019	1.7621	0.3621

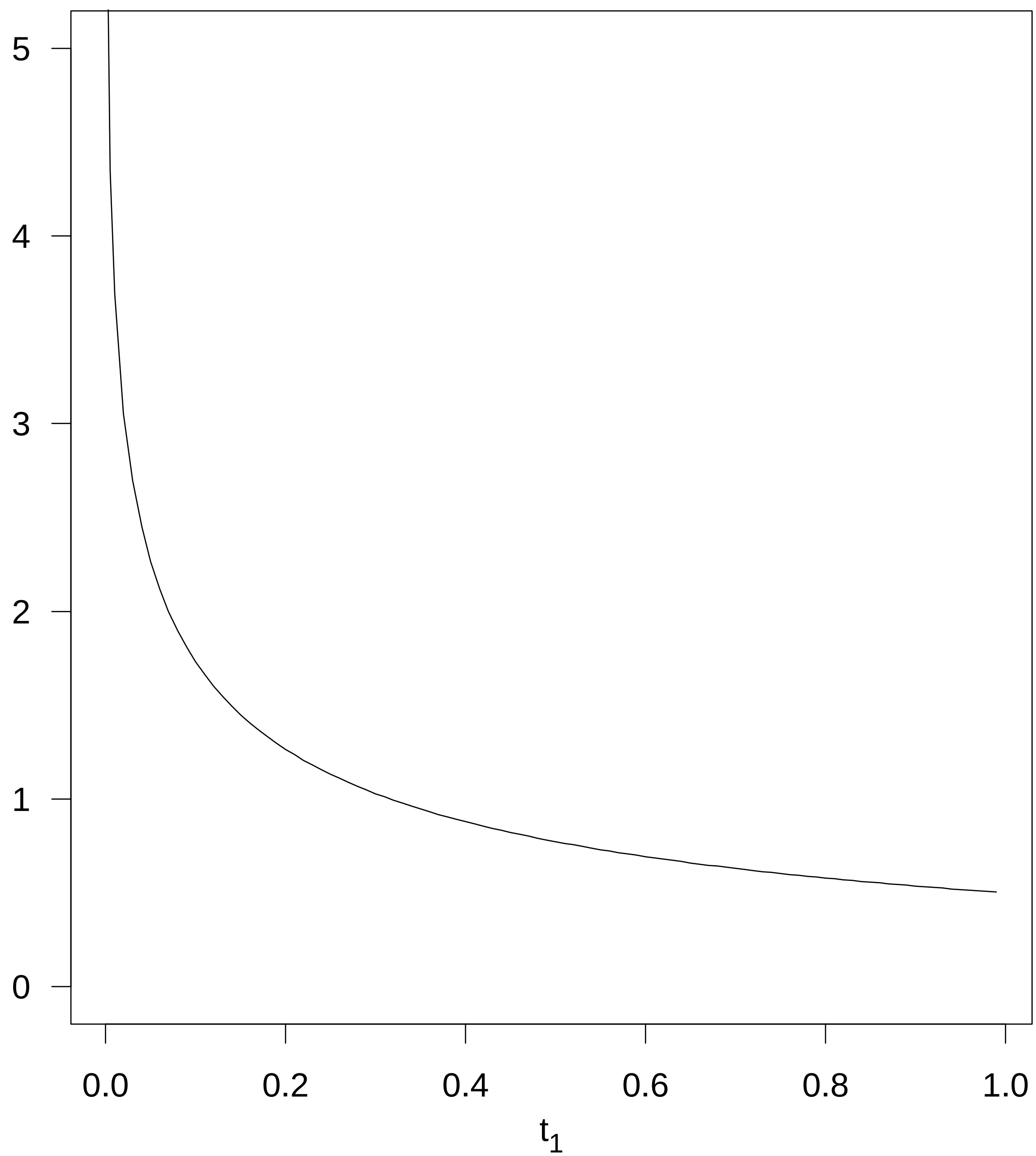

Figure 1. The marginal intrinsic prior density of $\pi_2^I(t_1)$.

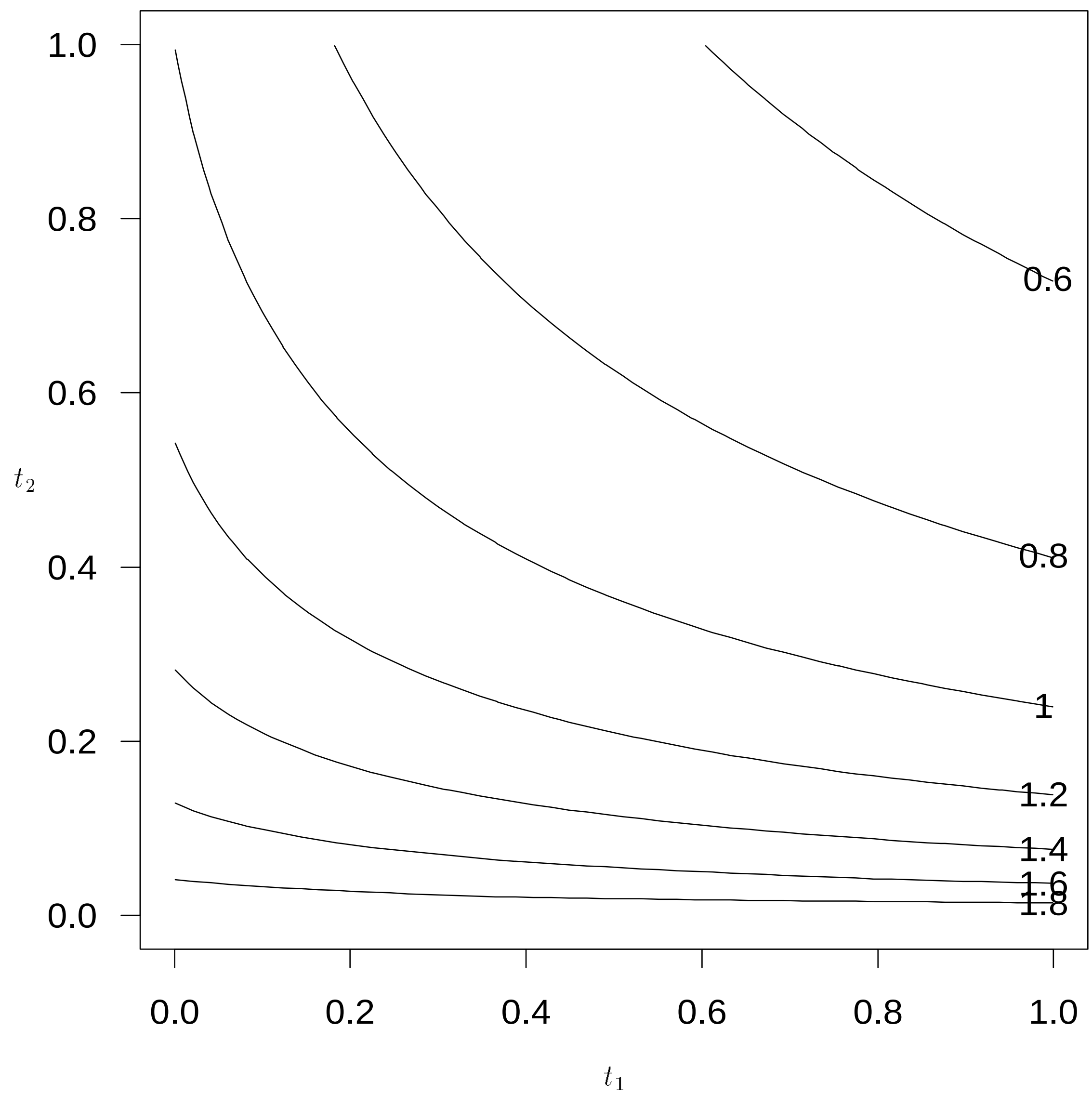

Figure 2. The marginal intrinsic prior density of $\pi_2^I(t_1, t_2)$.