Journal of the Korean Statistical Society (1999), **28***: 1, pp 73–92*

A Bayesian Test for Simple Tree Ordered Alternative using Intrinsic Priors

Seong W. Kim[1]

Abstract

In Bayesian model selection or testing problems, one cannot utilize standard or default noninformative priors, since these priors are typically improper and are defined only up to arbitrary constants. The resulting Bayes factors are 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 observatons to compute the Bayes factor. Surprisingly, such a Bayes factor can also be computed directly from the full sample by some proper priors, namely intrinsic priors. The present paper explains how to derive intrinsic priors for *simple tree ordered* exponential means. Some numerical results are also provided to support theoretical results and compare with classical methods.

Keywords: Intrinsic Bayes factor; Intrinsic priors; Jeffreys' priors; Noninformative priors; Order restricted maximum likelihood estimator

1. INTRODUCTION

In reliability theory or survival analysis, we often encounter the following testing problem,

$$\begin{aligned} M_1 &: \mu_0 = \mu_1 = \cdots = \mu_k, \text{ vs} \\ M_2 &: \max\{\mu_1, \ldots, \mu_k\} \le \mu_0, \end{aligned} \tag{1.1}$$

where μ_0 is the mean of the control group and μ_i, $i = 1, \ldots, k$ are the means of k treatment groups for certain distributions such as Exponential, Normal or Weibull. Robertson, Wright and Dykstra (1988) found the asymptotic distribution of the generalized likelihood ratio test statistic for M_1 against $M_2 - M_1$ for

[1]Visiting Instructor, Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, USA

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 most powerful test or generalized likelihood ratio test could be misleading, even when the sample sizes are large. Berger, Brown and Wolpert (1994) showed that the test for a simple hypothesis against a simple alternative on the normal means based on Neyman-Pearson Lemma rejects the null hypothesis systematically while the Bayes factor is just 1. Sun and Kim (1997) showed that the generalized likelihood ratio test for comparing two exponential means rejects the null hypothesis automatically while the Bayes factor is also just 1.

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 (1.1) for the exponential distribution using noninformative priors.

Noninformative priors such as Jeffreys' priors (1966) or reference priors (Berger and Bernado, 1989,1992) are typically improper, and thus are only defined up to arbitrary constants, which affects the values of Bayes factors. Berger and Pericchi (1996b) introduced a new model selection criterion, called the intrinsic Bayes factor (IBF) using the *training sample* method, which would remove the arbitrariness of improper priors. The training sample idea has been informally used. See Geisser and Eddy (1979), Spiegelhalter and Smith (1982), San Martini and Spezzaferri (1984), and Gelfand and Dey (1992) for related work. There has been some work using the IBF criterion. Berger and Pericchi (1996a) discussed some problems in linear models with several different error distributions. Varshavsky (1996) made use of the IBF for a stationary autoregressive process. Lingham and Sivaganesan (1997) conducted a test for the shape parameter of the power law process. Sun and Kim (1997) derived several intrinsic priors for testing equal means against the ascending ordered exponential means. Kim and Sun (1997) analyzed comparisons of two exponential means using the encompassing model.

The present article is organized as follows. In Section 2, we review the concept of Bayes factors and intrinsic priors. In Section 3, we derive a general form of intrinsic priors for testing equal means against ordered means for $k+1$ independent exponential distributions. Special cases when $k=1$ and $k=2$ are examined in detail. In Section 4, we provide some numerical results along with real data analysis to illustrate theoretical results. We finish this paper with concluding remarks in Section 5.

2. PRELIMINARIES

Suppose we want to select a model among q different models $M_1, \ldots, M_q$. If the model M_i holds, the data $\mathbf{X} = (X_1, \ldots, X_n)$ follow a parametric distribution with the density $f_i(\mathbf{X}|\theta_i)$, where θ_i is a vector of unknown parameters. Let Θ_i be the parameter space for θ_i. Based on observations $\mathbf{X}$, one wants to select the correct model M_i among q possible models. Let $\pi_i(\theta_i)$ be the prior distribution for θ_i under M_i, and let $p(M_i)$ be the prior model probability of M_i being true, for $i = 1, \ldots, q$. Then the posterior probability that M_i is true is

$$P(M_i|\mathbf{X}) = \Big[\sum_{j=1}^{q} \frac{p(M_j)}{p(M_i)} B_{ji}\Big]^{-1}, \tag{2.1}$$

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

$$B_{ji} = \frac{m_j(\mathbf{X})}{m_i(\mathbf{X})} = \frac{\int_{\Theta_j} f_j(\mathbf{X}|\theta_j)\pi_j(\theta_j)d\theta_j}{\int_{\Theta_i} f_i(\mathbf{X}|\theta_i)\pi_i(\theta_i)d\theta_i}, \tag{2.2}$$

where $m_i(\mathbf{X})$ is the marginal or predictive density of $\mathbf{X}$ under M_i. The posterior probabilities in (2.1) are used for selecting the most plausible model. If we use a noninformative prior $\pi_i^N(\theta_i)$, (2.2) becomes

$$B_{ji}^N = \frac{m_j^N(\mathbf{X})}{m_i^N(\mathbf{X})} = \frac{\int_{\Theta_j} f_j(\mathbf{X}|\theta_j)\pi_j^N(\theta_j)d\theta_j}{\int_{\Theta_i} f_i(\mathbf{X}|\theta_i)\pi_i^N(\theta_i)d\theta_i}. \tag{2.3}$$

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. One remedy for removing this arbitrariness is to use a part of data as a so-called training sample. Let $\mathbf{X}(l)$ be a training sample and let $\mathbf{X}(-l)$ be the remainder of the data. First, compute the (intermediate) posterior $\pi_i^N(\theta_i|\mathbf{X}(l))$, then compute the Bayes factors with the $\mathbf{X}(-l)$ as data, using $\pi_i^N(\theta_i|\mathbf{X}(l))$ as the prior. Consequently, the Bayes factor is as follows:

$$\begin{aligned} B_{ji}(l) &= \frac{\int_{\Theta_j} f_j(\mathbf{X}(-l)|\theta_j, \mathbf{X}(l))\pi_j^N(\theta_j|\mathbf{X}(l))d\theta_j}{\int_{\Theta_i} f_i(\mathbf{X}(-l)|\theta_i, \mathbf{X}(l))\pi_i^N(\theta_i|\mathbf{X}(l))d\theta_i} \\ &= B_{ji}^N \cdot B_{ij}^N(\mathbf{X}(l)), \end{aligned} \tag{2.4}$$

where for $h = i, j$,

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

In practice, $\mathbf{X}(l)$ is chosen to be minimal in the sense that the marginal $m_h^N(\mathbf{X}(l))$ is finite for all possible models, and no subset of $\mathbf{X}(l)$ gives finite marginals. Note that in (2.4) $B_{ji}(l)$ does not depend on arbitrary constants and thus is well defined. Furthermore, the Bayes factor defined by (2.4) depends on the choice of the minimal training sample. To avoid this dependence, Berger and Pericchi (1996b) suggested taking the average of $B_{ji}(l)$ over all $\mathbf{X}(l)$.

Definition 2.1. *The arithmetic intrinsic (AI) Bayes factor of M_j to M_i is given by*

$$B_{ji}^{AI} = \frac{1}{R}\sum_{l=1}^{R} B_{ji}(l) = B_{ji}^{N} \cdot \frac{1}{R}\sum_{l=1}^{R} B_{ij}^{N}(\mathbf{X}(l)), \tag{2.5}$$

where R is the number of all possible minimal training samples.

Define the correction factor, CFA_{ij} by

$$CFA_{ij} = \frac{1}{R}\sum_{l=1}^{R} B_{ij}^{N}(\mathbf{X}(l)). \tag{2.6}$$

By virtue of (2.5) and (2.6) we have

$$B_{ji}^{AI} = B_{ji}^{N} \cdot CFA_{ij}. \tag{2.7}$$

Noticing that computation can be lengthy if R is large, Berger and Pericchi (1996b) proposed the use of the following quantity as an approximation of (2.7).

Definition 2.2. *The expected arithmetic intrinsic (EAI) Bayes factor of M_j to M_i is given by*

$$B_{ji}^{EAI} = B_{ji}^{N} \cdot \frac{1}{R}\sum_{l=1}^{R} E_{\hat{\theta}_j}^{M_j}[B_{ij}^{N}(\mathbf{X}(l))], \tag{2.8}$$

where $\hat{\theta}_j = \hat{\theta}_{nj}$ is the MLE of θ_j based on n observations.

Alternatively, Berger and Pericchi (1996b) suggested finding a pair of proper priors such that the Bayes factor using proper priors will be asymptotically equivalent to B_{ji}^{AI}. Such priors, if they exist, are called intrinsic priors. 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

$$\psi_i(\theta_j) = \lim_{n\to\infty} E_{\theta_j}^{M_j}(\hat{\theta}_i). \tag{2.9}$$

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:

$$\begin{cases} \dfrac{\pi_j^I(\theta_j)\pi_i^N(\psi_i(\theta_j))}{\pi_j^N(\theta_j)\pi_i^I(\psi_i(\theta_j))} &= B_j^*(\theta_j), \\ \dfrac{\pi_j^I(\psi_j(\theta_i))\pi_i^N(\theta_i)}{\pi_j^N(\psi_j(\theta_i))\pi_i^I(\theta_i)} &= B_i^*(\theta_i), \end{cases} \tag{2.10}$$

where for $h = i,\ j$,

$$B_h^*(\theta_h) = \lim_{R\to\infty} E_{\theta_h}^{M_h}[\frac{1}{R}\sum_{l=1}^{R} B_{ij}^N(\mathbf{X}(l))]. \tag{2.11}$$

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_{ji}^{AI} can be replaced by the ordinary Bayes factors computed based on intrinsic priors.

3. MAIN RESULTS

3.1. Results for arbitrary k

Suppose that we have independent observations $X_{ij} \sim Exp(\mu_i), i = 0, 1, \ldots, k$; $j = 1, 2, \ldots, n_i$. We want to select between two competing models given by (1.1). 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 $N = \sum_{i=0}^{k} n_i$. Assume that there are $k+1$ constants $a_i \in (0,1)$ such that $a_0 + a_1 + \cdots + a_k = 1$ and for $i = 0, \ldots, k$,

$$\frac{n_i}{N} \to a_i \text{ as } N \to \infty. \tag{3.1}$$

Let $\mathbf{L}_k = \{\boldsymbol{\mu}_k = (\mu_0, \ldots, \mu_k) : 0 < \max(\mu_1, \ldots, \mu_k) \le \mu_0 < \infty\}$. 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 with $k = 1$, 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

$$\pi_1^N(\mu) = \frac{1}{\mu}, \ \mu > 0, \ \text{and} \ \pi_2^N(\boldsymbol{\mu}_k) = \frac{1}{\mu_0 \cdots \mu_k}, \ \boldsymbol{\mu}_k \in \mathbf{L}_k.$$

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

$$m_1^N(\mathbf{X}) = \int_0^\infty \frac{1}{\mu^{N+1}} \exp\left\{-\left[\frac{\sum_{i=0}^k X_{i\cdot}}{\mu}\right]\right\} d\mu = \frac{\Gamma(N)}{(\sum_{i=0}^k X_{i\cdot})^N},$$

and

$$m_2^N(\mathbf{X}) = \int_{\mathbf{L}_k} \frac{1}{\mu_0^{n_0+1}} \cdots \frac{1}{\mu_k^{n_k+1}} \exp\left\{-\sum_{i=0}^k \frac{X_{i\cdot}}{\mu_i}\right\} d\boldsymbol{\mu}_k.$$

A typical minimal training sample is $\mathbf{X}(l) = (X_{0h_0}, X_{1h_1}, \ldots, X_{kh_k})$. Then the marginal densities of $\mathbf{X}(l)$ are

$$\begin{aligned} m_1^N(\mathbf{X}(l)) &= \frac{\Gamma(k+1)}{(X_{0h_0} + \cdots + X_{kh_k})^{k+1}}; \\ m_2^N(\mathbf{X}(l)) &= \frac{1}{X_{1h_1} \cdots X_{kh_k}(X_{0h_0} + \cdots + X_{kh_k})}. \end{aligned}$$

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

$$B_{12}^N(\mathbf{X}(l)) = \frac{\Gamma(k+1) X_{1h_1} \cdots X_{kh_k}}{(X_{0h_0} + \cdots + X_{kh_k})^k}. \tag{3.2}$$

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

$$\begin{aligned} B_{21}^{AI} &= B_{21}^N \cdot CFA_{12} \\ &= B_{21}^N \cdot \frac{1}{n_0 \cdots n_k} \sum_{h_0=1}^{n_0} \cdots \sum_{h_k=1}^{n_k} B_{12}^N(\mathbf{X}(l)), \end{aligned} \tag{3.3}$$

and

$$B_{21}^{EAI} = B_{21}^N \cdot E_{\hat{\theta}_2}^{M_2} B_{12}^N(\mathbf{X}(l)), \tag{3.4}$$

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 3.1. a) The MLE of μ under M_1 is given by

$$\hat{\mu} = N^{-1} \sum_{i=0}^{k} X_{i\cdot\cdot}. \tag{3.5}$$

b) The unrestricted MLE of μ_i is given by

$$\hat{\mu}_i^* = \overline{X}_i, \; i = 0, \ldots, k. \tag{3.6}$$

Let $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_0, \ldots, \hat{\mu}_k)$ 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). For example, when $k = 1$,

$$\begin{aligned} \hat{\boldsymbol{\mu}}_1 &= (\hat{\mu}_0, \hat{\mu}_1) \\ &= (\overline{X}_0, \overline{X}_1)\mathbf{1}(\overline{X}_0 \geq \overline{X}_1) + (\hat{\mu}, \hat{\mu})\mathbf{1}(\overline{X}_0 < \overline{X}_1), \end{aligned} \tag{3.7}$$

and when $k = 2$,

$$\begin{aligned} \hat{\boldsymbol{\mu}}_2 &= (\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2) \\ &= (\overline{X}_0, \overline{X}_1, \overline{X}_2)\mathbf{1}(\overline{X}_1 \leq \overline{X}_2 \leq \overline{X}_0) + (\overline{X}_0, \overline{X}_1, \overline{X}_2)\mathbf{1}(\overline{X}_2 \leq \overline{X}_1 \leq \overline{X}_0) \\ &+ (\frac{X_{0\cdot} + X_{2\cdot}}{n_0 + n_2}, \overline{X}_1, \frac{X_{0\cdot} + X_{2\cdot}}{n_0 + n_2})\mathbf{1}(\overline{X}_1 \leq \overline{X}_0 < \overline{X}_2) \\ &+ (\frac{X_{0\cdot} + X_{1\cdot}}{n_0 + n_1}, \frac{X_{0\cdot} + X_{1\cdot}}{n_0 + n_1}, \overline{X}_2)\mathbf{1}(\overline{X}_2 \leq \overline{X}_0 < \overline{X}_1) \\ &+ (\hat{\mu}, \hat{\mu}, \hat{\mu})\mathbf{1}(\overline{X}_0 < \overline{X}_1 \leq \overline{X}_2) + (\hat{\mu}, \hat{\mu}, \hat{\mu})\mathbf{1}(\overline{X}_0 < \overline{X}_2 \leq \overline{X}_1), \end{aligned} \tag{3.8}$$

where $\hat{\mu}$ is given by (3.5).

Proposition 3.1. *a) Under M_2, when $N \to \infty$, we have*

$$\hat{\theta}_1 = \hat{\mu} \longrightarrow \psi_1(\boldsymbol{\mu}_k) \equiv \sum_{i=0}^{k} a_i \mu_i \; a.s.,$$

where $\hat{\mu}$ is given by (3.5) and a_i is given by (3.1).
b) Under M_1, when $N \to \infty$, we have

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

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

$$\sum_{i=0}^{k} [\hat{\mu}_i - \mu_i]^2 \frac{n_i}{N} \leq \sum_{i=0}^{k} [\hat{\mu}_i^* - \mu_i]^2 \frac{n_i}{N}, \tag{3.9}$$

where $\hat{\mu}_i^*$ is given by (3.6). By the strong consistency of the unrestricted MLE of μ_i and the assumption (3.1), the right-hand side of (3.9) converges to zero as $N \to \infty$. Thus, the left-hand side of (3.9) also converges to zero. The result follows from the fact that under M_1, $\mu_i = \mu$ for each $i = 0, \ldots, k$. $\square$

Clearly $B_h^*(\theta_h)$ depends on k, the total number of treatment groups. To distinguish the quantities $B_h^*(\theta_h)$ for different k, we write $B_{hk}^*(\theta_h) = B_h^*(\theta_h)$. From the definition (2.11), we see that

$$B_{hk}^*(\theta_h) = \Gamma(k+1) E_{\theta_h}^{M_h}\left[\frac{X_{11}\cdots X_{k1}}{(X_{01}+\cdots+X_{k1})^k}\right],\ h = 1, 2.$$

For any $k \geq 1$, define

$$\mathbf{A}_k = \{\mathbf{w}_k = (w_1, \ldots, w_k) : 0 < w_1, \ldots, w_k < 1\}. \tag{3.10}$$

Proposition 3.2. *The quantities $B_{1k}^*(\theta_1)$ and $B_{2k}^*(\theta_2)$ are given by*

$$B_{1k}^*(\mu) = \Gamma(k+1)^2 \int_{\mathbf{A}_k} \frac{w_1\cdots w_k}{(r_0+r_1+\cdots+r_k)^{k+1}} d\mathbf{w}_k = \frac{\Gamma(k+1)^2}{2^k}, \tag{3.11}$$

and

$$B_{2k}^*(\boldsymbol{\mu}_k) = \frac{\Gamma(k+1)^2}{\mu_0\cdots\mu_k} \int_{\mathbf{A}_k} \frac{w_1\cdots w_k}{[r_0/\mu_0+\cdots+r_k/\mu_k]^{k+1}} d\mathbf{w}_k, \tag{3.12}$$

where

$$\begin{cases} r_0 = 1 - w_1 - w_2 - \cdots - w_k, \\ r_1 = w_1,\ r_2 = w_2, \ldots, r_k = w_k. \end{cases} \tag{3.13}$$

Proof: We first derive B_{2k}^*. The joint density of $(X_{01}, \ldots, X_{k1})$ is given by

$$f(x_{01}, \ldots, x_{k1}) = \left(\prod_{i=0}^{k} \frac{1}{\mu_i}\right) \exp\left\{-\left(\frac{x_{01}}{\mu_0}+\cdots+\frac{x_{k1}}{\mu_k}\right)\right\},\ x_{i1} > 0,\ 0 \leq i \leq k.$$

Making the following transformations,

$$\begin{cases} W_1 = \frac{X_{11}}{X_{01}+\cdots X_{k1}}, \\ W_2 = \frac{X_{21}}{X_{01}+\cdots X_{k1}}, \\ \cdots \\ W_k = \frac{X_{k1}}{X_{01}+\cdots X_{k1}}, \\ W_{k+1} = X_{01}+\cdots+X_{k1}, \end{cases} \quad \text{or} \quad \begin{cases} X_{01} = W_{k+1}(1 - W_1 - \cdots - W_k), \\ X_{11} = W_1 W_{k+1}, \\ \cdots \\ X_{k-1,1} = W_{k-1} W_{k+1}, \\ X_{k1} = W_k W_{k+1}, \end{cases} \tag{3.14}$$

we have

$$B^*_{2k}(\mu_1,\ldots,\mu_k)=\Gamma(k+1)E(W_1W_2\cdots W_k). \tag{3.15}$$

The Jacobian of this transformation is

$$\begin{aligned}
|J| &= \left|\frac{\partial(x_{01},x_{11},\ldots,x_{k1})}{\partial(w_1,w_2,\ldots,w_{k+1})}\right| \\
&= \begin{vmatrix} -w_{k+1} & -w_{k+1} & \cdots & 1-w_1-\cdots w_k \\ w_{k+1} & 0 & \cdots & w_1 \\ 0 & w_{k+1} & \cdots & w_2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_k \end{vmatrix} \\
&= \begin{vmatrix} -w_{k+1} & -w_{k+1} & \cdots & 1-w_1-\cdots w_k \\ 0 & -w_{k+1} & \cdots & 1-w_1-\cdots w_k \\ 0 & 0 & -w_{k+1} & 1-w_1-\cdots w_k \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} \\
&= w_{k+1}^k.
\end{aligned}$$

The joint density of $(W_1,\ldots,W_{k+1})$ is then

$$f(w_1,\ldots,w_{k+1})=\frac{w_{k+1}^k}{\mu_0\cdots\mu_k}\exp\{-w_{k+1}(\frac{r_0}{\mu_0}+\cdots+\frac{r_k}{\mu_k})\},$$

where $\mathbf{w}_k\in\mathbf{A}_k$ and $w_{k+1}>0$. Integrating with respect to w_{k+1}, we get the following joint density of $(W_1,\ldots,W_k)$

$$g(\mathbf{w}_k)=\frac{\Gamma(k+1)}{\mu_0\cdots\mu_k[r_0/\mu_0+\cdots r_k/\mu_k]^{k+1}},\ \mathbf{w}_k\in\mathbf{A}_k,$$

where the r_i's are defined by (3.13). Hence equation (3.12) is established. Note that $\sum_{i=0}^k r_i=1$ in (3.13). Since $\mu_i=\mu$ under M_1, equation (3.11) immediately follows from (3.12). This completes the proof. □

The system of equations (2.10) becomes

$$\begin{cases} \dfrac{\pi_2^I(\boldsymbol{\mu}_k)/(a_0\mu_0+\cdots+a_k\mu_k)}{\pi_1^I(a_0\mu_0+\cdots+a_k\mu_k)/(\mu_0\cdots\mu_k)}=B^*_{2k}(\boldsymbol{\mu}_k),\ \boldsymbol{\mu}_k\in\mathbf{L}_k, \\ \dfrac{\pi_2^I(\mu\mathbf{1}_k)/\mu}{\pi_1^I(\mu)/\mu^{k+1}}=B^*_{1k}(\mu),\ \mu>0, \end{cases} \tag{3.16}$$

where $\mathbf{1}_k=(1,\ldots,1)\in\mathbb{R}^{k+1}$.

Lemma 3.1. *$B^*_{2k}(\boldsymbol{\mu}_k) \longrightarrow B^*_{1k}(\mu)$ 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 $W_1 \cdots W_k$ is bounded, it follows from (3.15) that $B^*_{2k} \to B^*_{1k}$ as $\boldsymbol{\mu}_k \longrightarrow \mu\mathbf{1}_k$. This completes the proof. □

Lemma 3.2. *For any integer $l \geq 1$, and any constants $r_i \in (0,1)$ satisfying $r_0 + \cdots + r_l = 1$,*

$$\int_{\mathbf{A}_l} \frac{1}{s_1^2 \cdots s_l^2 (r_0 + r_1/s_1 + \cdots + r_l/s_l)^{l+1}} d\mathbf{s}_l = \frac{1}{l! r_1 \cdots r_l (r_0 + r_1 + \cdots + r_l)}.$$

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

$$\int_{A_1} \frac{1}{s_1^2 (r_0 + r_1/s_1)^2} ds_1 = \frac{1}{r_1(r_0 + r_1)} = \frac{1}{r_1}.$$

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

$$\int_{\mathbf{A}_l} \frac{1}{s_1^2 \cdots s_l^2 (r_0 + r_1/s_1 + \cdots + r_l/s_l)^{l+1}} d\mathbf{s}_l$$

$$= \frac{1}{l r_l} \int_{\mathbf{A}_{l-1}} \frac{1}{s_1^2 \cdots s_{l-1}^2 (r_0 + r_1/s_1 + \cdots + r_{l-1}/s_{l-1} + r_l)^l} d\mathbf{s}_{l-1}. \quad (3.17)$$

By the induction assumption, (3.17) becomes

$$\frac{1}{l r_l} \frac{1}{(l-1)! r_1 \cdots r_{l-1}(r_0 + \cdots + r_l)} = \frac{1}{l! r_1 \cdots r_l (r_0 + \cdots + r_l)}.$$

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

Lemma 3.3. *Define*

$$s_0 = \mu_0, \ s_1 = \frac{\mu_1}{\mu_0}, \ s_2 = \frac{\mu_2}{\mu_0}, \ldots, s_k = \frac{\mu_k}{\mu_0}. \quad (3.18)$$

Then B^*_{2k} depends only on $\mathbf{s}_k = (s_1, \ldots, s_k)$.

Proof: It follows directly from (3.12). □

Theorem 3.1. *For any proper density $g(\cdot)$ on $(0, \infty)$, the set of intrinsic priors*

$$\begin{cases} \pi_1^I(\mu) = g(\mu), \ 0 < \mu < \infty, \\ \pi_2^I(\boldsymbol{\mu}_k) = \dfrac{a_0\mu_0 + \cdots + a_k\mu_k}{\mu_0 \cdots \mu_k} B^*_{2k}(\boldsymbol{\mu}_k) \pi_1^I(a_0\mu_0 + \cdots + a_k\mu_k), \ \boldsymbol{\mu}_k \in \mathbf{L}_k \end{cases} \quad (3.19)$$

is a solution of (3.16), where B^*_{2k} is given by (3.12). Furthermore, π_2^I is a proper density on $\mathbf{L}_k$.

Proof: From Lemma 3.1, we can see that (3.19) is a solution of (3.16). The Jacobian of the transformation from $\boldsymbol{\mu}_k$ to $(\mathbf{s}_k, s_0)$ in (3.18) is

$$|J| = \left|\frac{\partial(\mu_0, \ldots, \mu_k)}{\partial(s_0, \ldots, s_k)}\right| = s_0^k.$$

So,

$$\begin{aligned}
&\int_{\mathbf{A}_k}\int_0^\infty \pi_2^I(s_0, \ldots, s_k)ds_0 d\mathbf{s}_k \\
= \;& \Gamma(k+1)^2 \int_{\mathbf{A}_k}\left\{\frac{1}{s_1^2\cdots s_k^2}\int_0^\infty (a_0 + a_1 s_1 + \cdots a_k s_k)\pi_1^I\Big[s_0(a_1 s_1 + \cdots \right.\\
&\left. + a_k s_k)\Big]ds_0\right\}\left\{\int_{\mathbf{A}_k}\frac{w_1\cdots w_k}{[r_0 + r_1/s_1 + \cdots + r_k/s_k]^{k+1}}d\mathbf{w}_k\right\}d\mathbf{s}_k, \qquad (3.20)
\end{aligned}$$

where $\mathbf{A}_k$ and the r_i's are defined by (3.10) and (3.13), respectively. Let $\tau = s_0(a_0 + a_1 s_1 + \cdots + a_k s_k)$. Then $ds_0/d\tau = (a_0 + a_1 s_1 + \cdots + a_k s_k)^{-1}$, so (3.20) equals

$$\Gamma(k+1)^2 \int_{\mathbf{A}_k}\int_{\mathbf{A}_k}\frac{w_1\cdots w_k}{s_1^2\cdots s_k^2[r_0 + r_1/s_1 + \cdots + r_k/s_k]^{k+1}}d\mathbf{s}_k d\mathbf{w}_k. \qquad (3.21)$$

From Lemma 3.2, (3.21) reduces to $\Gamma(k+1)$. This completes the proof. □

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

Theorem 3.2. *a) The marginal intrinsic prior of* $\mathbf{s}_k$ *is*

$$\pi_2^I(\mathbf{s}_k) \propto \frac{h_k(\mathbf{s}_k)}{s_1\cdots s_k}, \;\; \mathbf{s}_k \in \mathbf{A}_k,$$

where

$$h_k(\mathbf{s}_k) = B_{2k}^*(1, s_1, s_2, \ldots, s_k), \;\; \mathbf{s}_k \in \mathbf{A}_k, \qquad (3.22)$$

and the normalizing constant is $1/\Gamma(k+1)$.
b) The conditional prior of s_0 *given* $\mathbf{s}_k$ *is*

$$\pi_2^I(s_0|\mathbf{s}_k) \propto \pi_1^I(\lambda s_0), \;\; s_0 > 0,$$

where

$$\lambda = a_0 + a_1 s_1 + \cdots + a_k s_k.$$

Proof: For part a), it follows from (3.19) that the joint intrinsic prior of $(\mathbf{s}_k, s_0)$ is

$$\pi_2^I(\mathbf{s}_k, s_0) = \frac{\lambda}{s_1 \cdots s_k} B_{2k}^*(s_0, s_0 s_1, \ldots, s_0 s_k)\pi_1^I(\lambda s_0). \tag{3.23}$$

Applying Lemma 3.3, the desired result follows from integrating equation (3.23) over s_0. The proof of part b) follows directly from part a). □

Remark 3.1. Clearly, from Theorem 3.1 we can see that there are infinitely many proper sets of intrinsic priors. However, if we choose the conjugate prior for model M_1, that is, the Inverse Gamma prior, the computation should be tractable for both M_1 and M_2.

Corollary 3.1. *When $g(t)$ is the probability density function of Inverse Gamma (ξ, ν), the pair of intrinsic priors is*

$$\begin{cases} \pi_1^I(\mu) = \dfrac{\nu^\xi}{\Gamma(\xi)\mu^{\xi+1}} e^{-\frac{\nu}{\mu}}, \ 0 < \mu < \infty, \\ \pi_2^I(\boldsymbol{\mu}_k) = \dfrac{\nu^\xi \exp\{-\nu/(a_0\mu_1 + \cdots + a_k\mu_k)\}}{\Gamma(\xi)(a_0\mu_1 + \cdots + a_k\mu_k)^\xi \mu_0 \cdots \mu_k} B_{2k}^*(\boldsymbol{\mu}_k), \ \boldsymbol{\mu}_k \in \mathbf{L}_k. \end{cases} \tag{3.24}$$

Once we choose a special class of intrinsic priors given by Corollary 3.1, there could be a sensitivity problem in terms of hyperparameters (ξ, ν). Thus, we suggest another set of intrinsic priors which does not depend on hyperparameters.

Theorem 3.3. *Set $\pi_1^I(\mu) = \pi_1^N(\mu)$. Write $\pi_2^I(\boldsymbol{\mu}_k) = \pi_2^I(\mu_1, \ldots, \mu_k|\mu_0)\pi_2^I(\mu_0)$. Then*

$$\begin{cases} \pi_1^I(\mu) = \pi_1^N(\mu), \ 0 < \mu < \infty, \\ \pi_2^I(\mu_0) = \pi_1^N(\mu), \ 0 < \mu_0 < \infty, \\ \pi_2^I(\mu_1, \ldots, \mu_k|\mu_0) = \dfrac{B_{2k}^*(\boldsymbol{\mu}_k)}{\mu_1 \cdots \mu_k}, \ \boldsymbol{\mu}_k \in \mathbf{L}_k \end{cases} \tag{3.25}$$

is a solution of (3.16). Furthermore, $\pi_2^I(\mu_1, \ldots, \mu_k|\mu_0)$ is a proper density on $\mathbf{L}_k$.

Proof: It readily follows by an identical argument as in the proof of Theorem 3.1. □

Remark 3.2. Obviously, the intrinsic priors $\pi_1^I(\mu)$ and $\pi_2^I(\mu_0)$ are improper. However, when we compute Bayes factors with these intrinsic priors, the unspecified arbitrary constant in $\pi_1^N(\mu)$ is cancelled out. Thus, the resulting Bayes factor is well defined.

3.2. Special cases when $k = 1$ and $k = 2$

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

Proposition 3.3. *The quantities $h_1(s_1)$ and $h_2(s_1, s_2)$ are given respectively by*

a) $h_1(s_1) = B_{21}^*(1, s_1),\ 0 < s_1 < 1,$

b) $h_2(s_1, s_2) = B_{22}^*(1, s_1, s_2),\ 0 < s_1, s_2 < 1,$

where

$$h_1(s_1) = \frac{s_1(-\log s_1 + s_1 - 1)}{(1 - s_1)^2},\ 0 < s_1 < 1, \tag{3.26}$$

and

$$h_2(s_1, s_2) = \frac{2s_2}{(1 - s_1)(1 - s_2)^2}\left\{\frac{s_1 \log(s_1 + s_2 - s_1 s_2)}{1 - s_1} - \frac{s_2}{s_1 + s_2 - s_1 s_2} + 1\right\},$$
$$0 < s_1, s_2 < 1. \tag{3.27}$$

Proof: From (3.12) the quantities B_{21}^* and B_{22}^* are

$$\begin{aligned}
B_{21}^*(\mu_0, \mu_1) &= \mu_0\mu_1 \int_0^1 \frac{w_1}{[\mu_1 + (\mu_0 - \mu_1)w_1]^2} dw_1 \\
&= \frac{\mu_0\mu_1}{(\mu_0 - \mu_1)^2}[\log(\frac{\mu_0}{\mu_1}) + \frac{\mu_1}{\mu_0} - 1], \\
B_{22}^*(\mu_0, \mu_1, \mu_2) &= \frac{4}{\mu_0\mu_1\mu_2} \int_0^1 \int_0^1 \frac{w_1 w_2}{(\frac{1 - w_1 - w_2}{\mu_0} + \frac{w_1}{\mu_1} + \frac{w_2}{\mu_2})^3} dw_1 dw_2 \\
&= \frac{2\mu_0^2\mu_2}{(\mu_0 - \mu_1)(\mu_0 - \mu_2)^2}\left[\frac{\mu_1 \log(\mu_0\mu_1 + \mu_0\mu_2 - \mu_1\mu_2)}{\mu_0 - \mu_1}\right. \\
&\quad \left. - \frac{\mu_0\mu_2}{\mu_0\mu_1 + \mu_0\mu_2 - \mu_1\mu_2} + \frac{\mu_0 - \mu_1 - 2\mu_1 \log \mu_0}{\mu_0 - \mu_1}\right].
\end{aligned} \tag{3.28}$$

By (3.22) the desired results are established. □

Figure 3.1 is the plot of the marginal intrinsic prior density $\pi_2^I(s_1)$ of $s_1 = \mu_1/\mu_0$ when $k = 1$. Here, $\pi_2^I(s_1) = h_1(s_1)/s_1$. Note that $\pi_2^I(s_1)$ is monotonic decreasing, and converges to 0.5 when $s_1 \to 1$. Although $\pi_2^I(s_1)$ is unbounded at $s_1 = 0$, it is integrable. Figure 3.2 is the perspective plot of the marginal intrinsic prior density of $s_1 = \mu_1/\mu_0$ and $s_2 = \mu_2/\mu_0$ when $k = 2$. Here $\pi_2^I(s_1, s_2) = h_2(s_1, s_2)/(s_1 s_2)$, which is unbounded as $s_1 \to 0$ and $s_2 \to 0$, but it is integrable.

For $k = 1$ and 2, with the pair of intrinsic priors given by Corollary 3.1, we derive the closed forms of the ordinary Bayes factors, which are denoted by

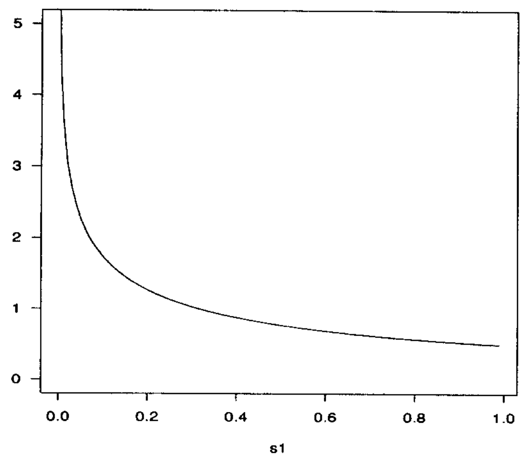

Figure 3.1: The marginal intrinsic prior density of $\pi_2^I(s_1)$.

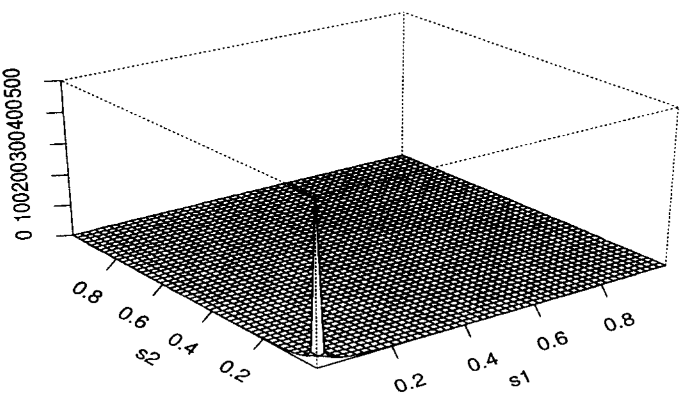

Figure 3.2: The marginal intrinsic prior density of $\pi_2^I(s_1, s_2)$.

$B_{21}^{I1}(\mathbf{X})$ and $B_{21}^{I2}(\mathbf{X})$ respectively. We also compute the ordinary Bayes factors with the pair of intrinsic priors in Theorem 3.3. These are denoted by $B_{21}^{I1*}(\mathbf{X})$ and $B_{21}^{I2*}(\mathbf{X})$

Proposition 3.4. *a1) For a pair of intrinsic priors in (3.24) we have*

$$B_{21}^{I1}(\mathbf{X}) = (X_{0\cdot} + X_{1\cdot} + \nu)^{\xi+n_0+n_1} H_1(X_{0\cdot}, X_{1\cdot}, a_0, a_1), \tag{3.29}$$

where

$$H_1(X_{0\cdot}, X_{1\cdot}, a_0, a_1) = \int_0^1 \frac{s_1^{\xi+n_0-1}(a_0 + a_1 s_1)^{n_0+n_1} h_1(s_1)}{[s_1(a_0 + a_1 s_1)X_{0\cdot} + (a_0 + a_1 s_1)X_{1\cdot} + \nu s_1]^{\xi+n_0+n_1}} ds_1.$$

a1′) For a pair of intrinsic priors in (3.25) we have

$$B_{21}^{I1*}(\mathbf{X}) = (X_{0\cdot} + X_{1\cdot})^{n_0+n_1} H_1^*(X_{0\cdot}, X_{1\cdot}), \tag{3.30}$$

where

$$H_1^*(X_{0\cdot}, X_{1\cdot}) = \int_0^1 \frac{s_1^{n_0-1} h_1(s_1)}{(s_1 X_{0\cdot} + X_{1\cdot})^{n_0+n_1}} ds_1,$$

where $h_1(\cdot)$ is defined by (3.26).
b1) Similarly, for $k = 2$, $B_{21}^{I2}(\mathbf{X})$ is given by

$$B_{21}^{I2}(\mathbf{X}) = \frac{1}{2}(X_{0\cdot} + X_{1\cdot} + X_{2\cdot} + \nu)^{\xi+N} H_2(X_{0\cdot}, X_{1\cdot}, X_{2\cdot}, a_0, a_1, a_2), \tag{3.31}$$

where

$$H_2(X_{0\cdot}, X_{1\cdot}, X_{2\cdot}, a_0, a_1, a_2) = \int_0^1\int_0^1 \frac{s_1^{\xi+n_0+n_2-1} s_2^{\xi+n_0+n_1-1} h_2(s_1, s_2)(a_0 + a_1 s_1 + a_2 s_2)^{-\xi}}{[X_{1\cdot} s_2 + X_{2\cdot} s_1 + s_1 s_2(X_{0\cdot} + \nu/(a_0 + a_1 s_1 + a_2 s_2))]^{\xi+N}} ds_1 ds_2.$$

b1′) For a pair of intrinsic priors in (3.25) we have

$$B_{21}^{I2*}(\mathbf{X}) = \frac{1}{2}(X_{0\cdot} + X_{1\cdot} + X_{2\cdot})^N H_2^*(X_{0\cdot}, X_{1\cdot}, X_{2\cdot}), \tag{3.32}$$

where

$$H_2^*(X_{0\cdot}, X_{1\cdot}, X_{2\cdot}) = \int_0^1\int_0^1 \frac{s_1^{n_0+n_2-1} s_2^{n_0+n_1-1} h_2(s_1, s_2)}{[X_{1\cdot} s_2 + X_{2\cdot} s_1 + s_1 s_2 X_{0\cdot}]^N} ds_1 ds_2,$$

with $N = n_0 + n_1 + n_2$ and h_2 defined by (3.27).

Proof: It is straightforward. □

4. NUMERICAL EXAMPLES

Example 4.1. Suppose that we want to select between two models $M_1 : \mu_0 = \mu_1$ and $M_2 : \mu_1 \leq \mu_0$. The P-value is $F(\overline{X}_1/\overline{X}_0; 2n_1, 2n_0)$ based on the generalized likelihood ratio test, where $F(\cdot; 2n_1, 2n_0)$ is the cdf (cumulative distribution function) of an F distribution with $2n_1$ and $2n_0$ degrees of freedom. To illustrate the difference between the frequentist method and the Bayesian model selection procedure under the intrinsic priors developed in Section 3, we examine the cases when $\overline{X}_0/\overline{X}_1 = 1, 2, 3$, and $n_0 = n_1 = 12, 20, 30$. The P-values for some choices of n_0 and n_1 are given in the column 3 of Table 4.1. The Bayes factors and the posterior probability of M_1 being true are computed for three choices of (ξ, ν) assuming equal prior model probabilities. They are $(0.01, 0.01)$, $(1.0, 1.0)$ and $(10, 10)$. We see that the posterior probabilities are bigger than P-values. For the cases when $\overline{X}_0/\overline{X}_1 = 2, 3$, as the sample sizes become larger, the Bayes factors will select M_2. Moreover, the Bayes factors are quite robust in terms of the change of the values (ξ, ν).

Table 4.1: P-values, Bayes factors, and $P(M_1|\mathbf{X})$ for testing $M_1 : \mu_0 = \mu_1$ versus $M_2 : \mu_1 \leq \mu_0$.

			$(\xi,\nu) = (.01,.01)$		$(\xi,\nu) = (1.0,1.0)$		$(\xi,\nu) = (10,10)$	
n	$(\overline{X}_1, \overline{X}_0)$	P-value	B_{21}	$P(M_1\|\mathbf{X})$	B_{21}	$P(M_1\|\mathbf{X})$	B_{21}	$P(M_1\|\mathbf{X})$
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

Example 4.2. 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 0	74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326
plane 1	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 2	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

Let μ_0, μ_1, and μ_2 denote the means of successive failures for plane 0,1,2 respectively. In Table 4.2, we provide Bayes factors and the posterior probabilities $P^I(M_1|\mathbf{X})$ for testing equal means ($M_1 : \mu_0 = \mu_1 = \mu_2$) against the simple tree ordered means ($M_2 : \max\{\mu_1, \mu_2\} \leq \mu_0$) for failure times of three planes. The P-value is 0.0347 which is computed based on asymptotic procedures by a χ^2 test using level probabilities (cf. Robertson, et al., 1988). The Bayes factors for the full sample were computed by (3.31) with six choices of (ξ, ν). They are (1,1), (1,10), (1,20), (1,30), (1,33), and (1,40). The computation here requires two dimensional numerical integration, which is done by IMSL routines. As a comparison, we also compute the Bayes factor using (3.32), which is denoted by B_{21}^{I2*} in Table 4.2. 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 0 enlarge the sample mean $\overline{X}_0$, which makes the P-value very small. Meanwhile, Bayes factors give fairly reasonable answers. We notice that the EAI Bayes factor does not give a good approximation. However, the ordinary Bayes factor B_{21}^{I2} computed by (3.31) with $\xi = 1$ and $\nu = 33$ is very close to the AI Bayes factor B_{21}^{AI}.

Table 4.2: Bayes factors and $P^I(M_1|\mathbf{X})$ for testing $M_1 : \mu_0 = \mu_1 = \mu_2$ versus $M_2 : \max\{\mu_1, \mu_2\} \leq \mu_0$ for airplane data. Here $(n_0, n_1, n_2) = (15, 30, 27)$, and $(\overline{X}_0, \overline{X}_1, \overline{X}_2) = (121.27, 59.60, 76.81)$

	Bayes factor	$P^I(M_1\|\mathbf{X})$
AI	0.8707	0.5345
EAI	2.7238	0.2685
$(\xi,\nu) = (1,1)$	1.3108	0.4327
$(\xi,\nu) = (1,5)$	1.2463	0.4452
$(\xi,\nu) = (1,10)$	1.1701	0.4608
$(\xi,\nu) = (1,20)$	1.0317	0.4922
$(\xi,\nu) = (1,30)$	0.9099	0.5236
$(\xi,\nu) = (1,33)$	0.8762	0.5330
$(\xi,\nu) = (1,40)$	0.8026	0.5547
B_{21}^{I2*}	1.3529	0.4250

5. CONCLUDING REMARKS

There has been a considerable amount of literature on the controversy between a P-value and a Bayes factor. It has been noticed that, more often than not, a P-value does not agree with the posterior probability that the null hypothesis is correct. Delampady and Berger (1990) have shown that the lower bounds of posterior probabilities in favor of null hypotheses are much larger than the corresponding P-values. Meng (1994) also found some discrepancy between the P-value and the Bayes factor by introducing posterior predictive P-values. See Kass and Raftery (1995) for more details.

The IBF methodology provides fully authentic Bayes factors in the sense of dealing only with default or standard noninformative priors. It is well defined and seems to be reasonably close to actual Bayes factors. The IBFs can be easily applied to nonnested as well as to nested model selection problems. They can also be applied in general when the samples come from any distribution.

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

The author gratefully acknowledges very constructive comments of the editors and referees. The author would like to thank Professor Youngsook Son, for her useful comments and discussions.

REFERENCES

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

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.

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.

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.

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.

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.

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

Gelfand, A. E. and Dey, D. K. (1992). "Bayesian model choice: Asymptotics and exact calculations," Technical report, Department of Statistics, University of Connecticut.

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"*.

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

Kass, R. E. and Raftery, A. E. (1995). "Bayes factors," *Journal of the American Statistical Association,* **90**, pp. 773-795.

Kim, S. W. and Sun, D. (1997). "Intrinsic priors for model selection using an encompassing model," *Lifetime Data Analysis*, under revision.

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

Meng, X. (1994). "Posterior predictive p-values," *The Annals of Statistics,* **22**, pp. 1142-1160.

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

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

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

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.

Sun, D. and Kim, S. W. (1997). "Intrinsic priors for testing ordered exponential means," submitted.

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.