Journal of the Korean Statistical Society (1998), **27**: *4, pp 385–396*

Computing Fractional Bayes Factor Using the Generalized Savage-Dickey Density Ratio†

Younshik Chung and Sangjeen Lee[1]

Abstract

A computing method of fractional Bayes factor (FBF) for a point null hypothesis is explained. We propose alternative form of FBF that is the product of density ratio and a quantity using the generalized Savage-Dickey density ratio method. When it is difficult to compute the alternative form of FBF analytically, each term of the proposed form can be estimated by MCMC method. Finally, two examples are given.

Keywords: Fractional Bayes factor; Generalized Savage-Dickey density ratio; Gibbs sampling; Testing hypothesis

1. INTRODUCTION

The Bayesian approach to testing hypothesis was developed by Jeffreys (1961) as a major part of his program for scientific inference. Mainly Bayes factor has been used for the Bayesian approach of model selection or hypothesis testing. But it has a drawback, that is, it is hard to approach with improper prior.

Consider a statistical model with data $\mathbf{Y}$ and its parameter vector θ. Suppose that we wish to test the null hypothesis H_0 versus alternative H_1, according to a probability density $f_0(\mathbf{Y}|\theta_0)$ or $f_1(\mathbf{Y}|\theta_1)$ respectively. Given priori probabilities $p(H_0)$ and $p(H_1) = 1 - p(H_0)$, the data $\mathbf{Y}$ produces posterior probabilities $p(H_0|\mathbf{Y})$ and $p(H_1|\mathbf{Y})$. The Bayes factor, $B(\mathbf{Y})$, in favor of H_0 is

$$B(\mathbf{Y}) = \frac{p(H_0|\mathbf{Y})/p(H_1|\mathbf{Y})}{p(H_0)/p(H_1)} = \frac{m_0(\mathbf{Y})}{m_1(\mathbf{Y})}, \tag{1.1}$$

where $m_i(\mathbf{Y}) = \int \pi_i(\theta_i) f_i(\mathbf{Y}|\theta_i) d\theta_i$ is the marginal density of $\mathbf{Y}$ under model H_i and $\pi_i(\theta_i)$ is the prior density of θ_i under H_i, for $i = 0, 1$. Bayes factors

†The research was supported (in part) by the Matching Fund Programs of Research Institute for Basic Sciences, Pusan National University, Korea, 1998.

[1]Department of Statistics, Pusan National University, Pusan, 609-735 Korea.

typically depend rather strongly on the prior distributions much more so than in estimations. For instance, as the sample size grows, the influence of the prior distribution disappears in estimation, but does not in hypothesis testing or model selction. Especially, when improper priors are used, it is difficult to compare hypothesis or models by Bayes factor.

An improper prior for θ_i is usually written as $\pi_i(\theta_i) \propto g_i(\theta_i)$, where g_i is a integrable function over θ_i-space, for example, a location noninformative prior is given by $\pi_i(\theta_i) \propto 1$. That is, it could be expressed that

$$\pi_i(\theta_i) = c_i g_i(\theta_i), \;\; i = 0, 1.$$

If there is no normalizing constant, we can treat c_i as an unspecified constant. However, Bayes factor in favor of H_0, with respect to these priors,

$$B(\mathbf{Y}) = c_0/c_1 \frac{\int g_0(\theta_0) f_0(\mathbf{Y}|\theta_0) d\theta_0}{\int g_1(\theta_1) f_1(\mathbf{Y}|\theta_1) d\theta_1},$$

depends on the ratio c_0/c_1. Recently, to overcome this problem, various approaches have been advocated. Aitkin (1991) proposed the posterior Bayes factor which is used proper posterior distribution instead of improper prior distribution. In this approach, data are doubly used. As an alternative, the concept of the partial Bayes factor is thought. Berger and Pericchi (1996) provided intrinsic Bayes factor with the (imaginary) training sample idea of Smith and Spiegelhalter (1980) and Spiegelhalter and Smith (1982). Fractional Bayes factor (FBF) method is proposed by O'Hagan (1995). The FBF in favor of H_0,

$$B_r(\mathbf{Y}) = \frac{q_0(r, \mathbf{Y})}{q_1(r, \mathbf{Y})}, \tag{1.2}$$

is simple and has practical merits, where for $i = 0, 1$, $q_i(r, \mathbf{Y}) = \frac{\int \pi_i(\theta_i) f_i(\mathbf{Y}|\theta_i) d\theta_i}{\int \pi_i(\theta_i) f_i^r(\mathbf{Y}|\theta_i) d\theta_i}$ and for the choice of r, refer to O'Hagan (1995). In particular, the possible choices of r are m/n, $n^{-1}\log n$ or $n^{-\frac{1}{2}}$ where n and m are the original sample size and minimal training sample size which is defined in Berger an Pericchi (1996), respectively.

In section 2, we review Verdinelli and Wasserman's (1995) method for computing Bayes factor using the generalized Savage - Dickey density ratio method. Then we suggest a computing method for the fractional Bayes factor (FBF) to test a point null hypothesis. In section 3, the suggested computing method will be applied to variance ratio, ϕ, of random effects model and the mean parameter of truncated normal distribution involving censored data.

2. COMPUTING FRACTIONAL BAYES FACTOR

Consider a parameter $\theta = (\omega, \xi) \in \Theta = \Omega \times \Xi$ and suppose that we wish to test the null hypothesis $H_0 : \omega = \omega_0$ versus alternative $H_1 : \omega \neq \omega_0$. Suppose that $\pi_0(\xi)$ is the prior density for ξ under H_0 and $\pi_1(\omega, \xi)$ is the prior density for (ω, ξ) under H_1. Let r be the approximation coefficient of fractional Bayes factor explained in the equation (1.2). The FBF, $B_r(\mathbf{Y})$, in favor of H_0 is expressed as

$$\begin{aligned} B_r(\mathbf{Y}) &= \frac{q_0(r, \mathbf{Y})}{q_1(r, \mathbf{Y})} \\ &= \frac{\frac{\int \pi_0(\xi) f(\mathbf{Y}|\omega_0, \xi) d\xi}{\int \pi_0(\xi) f^r(\mathbf{Y}|\omega_0, \xi) d\xi}}{\frac{\int\int \pi_1(\omega, \xi) f(\mathbf{Y}|\omega, \xi) d\omega d\xi}{\int\int \pi_1(\omega, \xi) f^r(\mathbf{Y}|\omega, \xi) d\omega d\xi}}. \end{aligned} \tag{2.1}$$

Then we can consider $f^r(\mathbf{Y}|\omega_0, \xi)$ and $f^r(\mathbf{Y}|\omega, \xi)$ as the sampling distribution densities, by normalizing them. That is, if let $1/c_0(r, \xi) = \int f^r(\mathbf{Y}|\omega_0, \xi)\ d\mathbf{Y}$ and $1/c_1(r, \omega, \xi) = \int f^r(\mathbf{Y}|\omega, \xi) d\mathbf{Y}$, then $c_0(r, \xi) \times f^r(\mathbf{Y}|\omega_0, \xi)$ and $c_1(r, \omega, \xi)$ $\times f^r(\mathbf{Y}|\omega, \xi)$ are densities for $\mathbf{Y}$ given (ω_0, ξ) and (ω, ξ), respectively. Thus the equation (2.1) can be written

$$\begin{aligned} B_r(\mathbf{Y}) &= \frac{\int \pi_0(\xi) f(\mathbf{Y}|\omega_0, \xi) d\xi}{\int\int \pi_1(\omega, \xi) f(\mathbf{Y}|\omega, \xi) d\omega d\xi} \cdot \frac{\int\int \pi_1(\omega, \xi) f^r(\mathbf{Y}|\omega, \xi) d\omega d\xi}{\int \pi_0(\xi) f^r(\mathbf{Y}|\omega_0, \xi) d\xi} \\ &= \frac{m_0(\mathbf{Y})}{m_1(\mathbf{Y})} \cdot \frac{m_{1,r}(\mathbf{Y})}{m_{0,r}(\mathbf{Y})} \\ &= BF_{01}(\mathbf{Y}) \cdot BF_{10}^r(\mathbf{Y}), \end{aligned} \tag{2.2}$$

where $m_{0,r}(\mathbf{Y}) = \int \pi_0(\xi) f^r(\mathbf{Y}|\omega_0, \xi) d\xi$, $\quad m_{1,r}(\mathbf{Y}) = \int\int \pi_1(\omega, \xi) f^r(\mathbf{Y}|\omega, \xi) d\omega d\xi$, $BF_{01}(\mathbf{Y}) = \frac{m_0(\mathbf{Y})}{m_1(\mathbf{Y})}$, and $BF_{10}^r(\mathbf{Y}) = \frac{m_{1,r}(\mathbf{Y})}{m_{0,r}(\mathbf{Y})}$.

Then $m_{0,r}(\mathbf{Y})$ and $m_{1,r}(\mathbf{Y})$ are not the marginal densities since there are no normalizing constant terms of $f^r(\mathbf{Y}|\omega_0, \xi)$ and $f^r(\mathbf{Y}|\omega, \xi)$, respectively. But $m_{0,r}(\mathbf{Y})$ can be regarded as the marginal density of $\mathbf{Y}$ with the sampling density $c_0(r, \xi) \times f^r(\mathbf{Y}|\omega_0, \xi)$ and prior $\frac{\pi_0(\xi)}{c_0(r, \xi)}$. Also, $m_{1,r}(\mathbf{Y})$ will be considered by the same way. Thus, the FBF in (2.2) seems to be the product of two Bayes factors. Therefore, we can compute the FBF in (2.2) by computing each Bayes factor and then producting them. Dickey (1971) showed that if $\pi_1(\xi|\omega) = \pi_0(\xi)$, then Bayes factor is $\pi_1(\omega_0|\mathbf{Y})/\pi_1(\omega_0)$, where $\pi_1(\omega|\mathbf{Y})$ and $\pi_1(\omega)$ are the marginal posterior density and the marginal prior density under H_1, respectively. Dickey (1971) attributed this formula to Savage and called the expression Savage's density ratio.

If Dickey's condition is not satisfied, we can apply to Verdinelli and Wasserman's approach, which is called the generalized Savage-Dickey density ratio metho d.

Lemma 2.1 (Verdinelli and Wasserman, 1995) *Assume that* $0 < \pi_1(\omega_0|\mathbf{Y})$, $\pi_1(\omega_0, \xi) < \infty$ *for almost all* ξ, *then*

$$BF_{01}(\mathbf{Y}) = \frac{\pi_1(\omega_0|\mathbf{Y})}{\pi_1(\omega_0)} \cdot E[\frac{\pi_0(\xi)}{\pi_1(\xi|\omega_0)}] = \pi_1(\omega_0|\mathbf{Y}) \cdot E[\frac{\pi_0(\xi)}{\pi_1(\xi, \omega_0)}]$$

assuming that the expectation is finite with respect to $\pi_1(\xi|\omega_0, \mathbf{Y})$.

Proof: see Verdinelli and Wasserman (1995). □

For the notational convenience, we will represent various notations as follows;

$$\pi_{1,r}(\omega, \xi|\mathbf{Y}) = \pi_1(\omega, \xi)\frac{f^r(\mathbf{Y}|\omega, \xi)}{m_{1,r}(\mathbf{Y})} = \frac{\pi_1(\omega, \xi)f^r(\mathbf{Y}|\omega, \xi)}{\int\int \pi_1(\omega, \xi)f^r(\mathbf{Y}|\omega, \xi)d\omega d\xi}$$

$$\pi_{1,r}(\omega_0|\mathbf{Y}) = \int \pi_{1,r}(\omega_0, \xi|\mathbf{Y})d\xi, \text{ and } \pi_{1,r}(\xi|\omega_0, \mathbf{Y}) = \frac{\pi_{1,r}(\omega_0, \xi|\mathbf{Y})}{\pi_{1,r}(\omega_0|\mathbf{Y})}.$$

Note that $\pi_{1,r}(\omega, \xi|\mathbf{Y})$ can be regarded as posterior density of ω and ξ given $\mathbf{Y}$, because $\pi_1(\omega, \xi)f^r(\mathbf{Y}|\omega, \xi)$ can be expressed as $\frac{\pi_1(\omega,\xi)}{c_1(r,\omega,\xi)} \cdot c_1(r, \omega, \xi)f^r(\mathbf{Y}|\omega, \xi)$ which are the product of prior and likelihood where $c_1(r, \omega, \xi)$ is defined as before. But, in practice the computation of $c_1(r, \omega, \xi)$ is not needed.

By applying Lemma 2.1 to the FBF in (2.2), we have the following theorem.

Theorem 2.1. *Assume that the following densities and expectations are finite. The fractional Bayes factor in (2.1) is*

$$B_r(\mathbf{Y}) = \frac{\pi_1(\omega_0|\mathbf{Y})}{\pi_{1,r}(\omega_0|\mathbf{Y})} \cdot \frac{E^{\pi_1(\xi|\omega_0,\mathbf{Y})}[\frac{\pi_0(\xi)}{\pi_1(\xi|\omega_0)}]}{E^{\pi_{1,r}(\xi|\omega_0,\mathbf{Y})}[\frac{\pi_0(\xi)}{\pi_1(\xi|\omega_0)}]}. \tag{2.3}$$

Proof: By Lemma 2.1, we can write that

$$BF_{01}(\mathbf{Y}) = \frac{\pi_1(\omega_0|\mathbf{Y})}{\pi_1(\omega_0)} \cdot E[\frac{\pi_0(\xi)}{\pi_1(\xi|\omega_0)}]$$

and

$$\begin{aligned} 1/BF_{10}^r(\mathbf{Y}) &= \frac{m_{0,r}(\mathbf{Y})}{m_{1,r}(\mathbf{Y})} = \frac{\int \pi_0(\xi) f^r(\mathbf{Y}|\omega_0,\xi)d\xi}{m_{1,r}(\mathbf{Y})} \\ &= \int \frac{\pi_0(\xi) f^r(\mathbf{Y}|\omega_0,\xi)}{m_{1,r}(\mathbf{Y})\pi_{1,r}(\omega_0|\mathbf{Y})} d\xi \cdot \pi_{1,r}(\omega_0|\mathbf{Y}). \end{aligned}$$

Since $\pi_{1,r}(\omega_0|\mathbf{Y}) = \pi_{1,r}(\omega_0,\xi|\mathbf{Y})/\pi_{1,r}(\xi|\omega_0,\mathbf{Y})$,

$$\begin{aligned} 1/BF_{10}^r(\mathbf{Y}) &= \pi_{1,r}(\omega_0|\mathbf{Y}) \cdot \int \frac{\pi_0(\xi) f^r(\mathbf{Y}|\omega_0,\xi)\pi_{1,r}(\xi|\omega_0,\xi)}{m_{1,r}(\mathbf{Y})\pi_{1,r}(\omega_0,\xi|\mathbf{Y})} d\xi \\ &= \pi_{1,r}(\omega_0|\mathbf{Y}) \cdot \int \frac{\pi_0(\xi)\pi_{1,r}(\xi|\omega_0,\mathbf{Y})}{\pi_1(\omega_0,\xi)} d\xi \\ &= \frac{\pi_{1,r}(\omega_0|\mathbf{Y})}{\pi_{1,r}(\omega_0)} \cdot E^{\pi_{1,r}(\xi|\omega_0,\mathbf{Y})}[\pi_0(\xi)/\pi_1(\xi|\omega_0)]. \end{aligned}$$

Therefore, it follows from (2.2) that

$$\begin{aligned} B_r(\mathbf{Y}) &= BF_{01}(\mathbf{Y}) \cdot BF_{10}^r(\mathbf{Y}) \\ &= \frac{\pi_1(\omega_0|\mathbf{Y})}{\pi_{1,r}(\omega_0|\mathbf{Y})} \cdot \frac{E^{\pi_1(\xi|\omega_0,\mathbf{Y})}[\pi_0(\xi)/\pi_1(\xi|\omega_0)]}{E^{\pi_{1,r}(\xi|\omega_0,\mathbf{Y})}[\pi_0(\xi)/\pi_1(\xi|\omega_0)]}. \end{aligned}$$

□

Remark 2.1: For a special case, if $\pi_1(\xi|\omega_0) = \pi_0(\xi)$, then the equation (2.3) is simplified to

$$B_r(\mathbf{Y}) = \frac{\pi_1(\omega_0|\mathbf{Y})}{\pi_{1,r}(\omega_0|\mathbf{Y})}. \tag{2.4}$$

If it is not easy to compute $B_r(\mathbf{Y})$ in (2.3) analytically, we can use the Gibbs sampler (Gelfand and Smith, 1990) which is recently powerful method to overcome the difficulty of Bayesian computation. Let $\{(\omega_{r1},\xi_{r1}),\ldots,(\omega_{rG},\xi_{rG})\}$ is a samples from the posterior $\pi_{1,r}(\omega,\xi|\mathbf{Y})$ by MCMC method. First we consider the computations of $\pi_1(\omega_0|\mathbf{Y})$ and $\pi_{1,r}(\omega|\mathbf{Y})$. Since the numerical estimations of $\pi_1(\omega_0|\mathbf{Y})$ and $\pi_{1,r}(\omega|\mathbf{Y})$ are similar and if $r = 1$ $\pi_{1,r}(\omega|\mathbf{Y}) = \pi_1(\omega_0|\mathbf{Y})$, only the case for $\pi_{1,r}(\omega_0|\mathbf{Y})$ will be presented. If $\pi_{1,r}(\omega|\xi,\mathbf{Y})$ is in the closed form, then the marginal posterior density for ω evaluated at ω_0 is estimated using Rao-Blackwellizing estimator of Gelfand and Smith (1990) as follow;

$$\hat{\pi}_{1,r}(\omega_0|\mathbf{Y}) = \frac{1}{G}\sum_{i=1}^{G} \pi_{1,r}(\omega_0|\mathbf{Y},\xi_{ri}).$$

If $\pi_{1,r}(\omega|\xi, \mathbf{Y})$ is not in the closed form, we use the another method due to Chen (1994) as follows;

$$\hat{\pi}_{1,r}(\omega_0|\mathbf{Y}) = \frac{1}{G}\sum_{i=1}^{G} q(\omega_{ri}|\xi_{ri})\frac{g(\omega_0, \xi_{ri}|\mathbf{Y})}{g(\omega_{ir}, \xi_{ri}|\mathbf{Y})},$$

where $g(\omega, \xi|\mathbf{Y}) = f^r(\mathbf{Y}|\omega, \xi)\pi_1(\omega, \xi)$ and $q(\omega, \xi)$ is an arbitrary probability density function and so $q(\omega|\xi)$ is its conditional density . Chen (1994) proposed that in some cases a reasonable choice for q is to use a normal density whose mean and covariance are based on the sample mean and sample covariance of $\{(\omega_{r1}, \xi_{r1}), (\omega_{r2}, \xi_{r2}), \ldots, (\omega_{rG}, \xi_{rG})\}$.

Next, in order to estimate $C_r = E^{\pi_{1,r}(\xi|\omega_0,\mathbf{Y})}\frac{\pi_0(\xi)}{\pi_1(\omega_0,\xi)}$, we draw a sample $\{\xi_r^{(1)}, \ldots, \xi_r^{(G)}\}$ from $\pi_{1,r}(\xi|\omega_0, \mathbf{Y})$. Then, we estimate C_r by

$$\hat{C}_r = \frac{1}{G}\sum_{i=1}^{G}\frac{\pi_0(\xi_r^{(i)})}{\pi_1(\omega_0, \xi_r^{(i)})}.$$

Also, similarly we can estimate $C_1 = E^{\pi_1(\xi|\omega_0,\mathbf{Y})}\frac{\pi_0(\xi)}{\pi_1(\omega_0,\xi)}$. Therefore, we could have the estimate for $B_r(\mathbf{Y})$, $\hat{B}_r(\mathbf{Y}) = \frac{\hat{\pi}_1(\omega_0|\mathbf{Y})}{\hat{\pi}_{1,r}(\omega_0|\mathbf{Y})} \cdot \frac{\hat{C}_1}{\hat{C}_r}$.

3. SOME APPLICATIONS

In this section, we consider the two examples which use the forms of FBF in (2.3) and (2.4)

3.1. Random effects model

Consider a one-way balanced random effect model:

$$y_{ij} = \mu + e_i + \epsilon_{ij}, \text{ for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J \tag{3.1}$$

where μ is the mean of y_{ij}, and e_i and ϵ_{ij} are independent normal variables with 0 means and variances σ_e^2 and σ^2, respectively. Let $\phi = J\frac{\sigma_e^2}{\sigma^2}$, which is the variance ratio, of interest in various fields. For notational convenience, let $\theta = (\phi, \mu, \sigma^2)$ and $\mathbf{Y} = (y_{ij})_{I\times J}$.

We are interested in the ratio of variances, ϕ, of model (3.1). This pararameter in the random effect model has been of interest for a long time in various fields. We want to test the hypothesis $H_0 : \phi = \phi_0$ versus $H_1 : \phi \neq \phi_0$ with proposed method.

For model (3.1), the likelihood function of parameter $\theta = (\phi, \mu, \sigma^2)$ is given by

$$L(\phi, \mu, \sigma^2) \propto \sigma^{-IJ}(1+\phi)^{-I/2} \exp\{-\frac{1}{2\sigma^2}(\frac{S_1^2 + IJ(y_{..} - \mu)^2}{1+\phi} + S_2^2)\}$$

where $y_{i.} = \sum_j y_{ij}/J$, $y_{..} = \sum_i \sum_j y_{ij}/IJ$, $S_1^2 = J\sum_i (y_{i.} - y_{..})^2$ and $S_2^2 = \sum_i \sum_j (y_{ij} - y_{i.})^2$.

Theorem 3.1. *In the random effects model (3.1), Jeffreys' priors are assumed in both hypotheses. Then, the FBF in favor of H_0 is*

$$B_r(\mathbf{Y}) = \frac{\beta_{p_r,q_r}(\frac{W}{1+W})}{\beta_{p,q}(\frac{W}{1+W})}(\frac{1+\phi_0}{W})^{q-q_r}(1+\frac{1+\phi_0}{W})^{p_r-p+q_r-q}, \tag{3.2}$$

where $p = I/2$, $q = I(J-1)/2$, $p_r = rI/2$, $q_r = rI(J-1)/2$, and $W = S_1^2/S_2^2$, and $\beta_{i,j}(x) = \int_0^x t^{i-1}(1-t)^{j-1}dt$ is the incomplete beta function of (i,j) evaluated at x.

Proof: Recall that Jeffreys' priors for H_0 and H_1 are $\pi_0(\mu, \sigma^2) = \sigma^{-3}$ and $\pi_1(\phi, \mu, \sigma^2) = \sigma^{-3}(1+\phi)^{-3/2}$, respectively. This situation is the special case of our computing method and we can use the FBF in (2.4). Thus we want to find the marginal posterior density for ϕ. Since the joint posterior density under H_1 is

$$\begin{aligned}\pi_{1,r}(\phi, \mu, \sigma^2|\bar{y}, s_1^2, s_2^2) &\propto \sigma^{-(rIJ+3)}(1+\phi)^{-(rI+3)/2} \\ &\quad \exp\{-\frac{r}{2\sigma^2}(\frac{S_1^2 + IJ(y_{..} - \mu)^2}{1+\phi} + S_2^2)\},\end{aligned}$$

the marginal posterior density for ϕ is

$$\begin{aligned}\pi_{1,r}(\phi|\mathbf{Y}) &= \int\int \pi_{1,r}(\phi, \mu, \sigma^2|\bar{y}, s_1^2, s_2^2)d\mu d\sigma^2 \\ &\propto \int \sigma^{-(rIJ+3)}(1+\phi)^{-(rI+3)/2} \exp\{-\frac{r}{2\sigma^2}(S_2^2 + \frac{S_1^2}{1+\phi})\} \\ &\quad \{\frac{2\pi}{rIJ}\sigma^2(1+\phi)\}^{1/2}d\sigma^2 \\ &\propto (1+\phi)^{-(rI+2)/2}\int (\sigma^2)^{-(rIJ+2)/2} \exp\{-\frac{r}{2\sigma^2}(S_2^2 + \frac{S_1^2}{1+\phi})\}d\sigma^2 \\ &\propto (1+\phi)^{-(rI+2)/2}(1+\frac{W}{1+\phi})^{-rIJ/2}.\end{aligned}$$

To find its normalizing constant, we have to get the value of integral,

$$\int_0^\infty (1+\phi)^{-(rI+2)/2}(1+\frac{W}{1+\phi})^{-rIJ/2}d\phi. \tag{3.3}$$

To do this, let $Z = W/(W+1+\phi)$. Then the integral (3.3) can be obtained as follows;

$$\begin{aligned}
&\int_0^{\frac{W}{W+1}} (\frac{W}{Z} - W)^{-(rI+2)/2+rIJ/2}(\frac{W}{Z})^{-rIJ/2}\frac{W}{Z^2}dZ \\
&= W^{-rI/2}\cdot\int_0^{\frac{W}{W+1}} Z^{rI/2-1}(1-Z)^{rI(J-1)/2-1}dZ \\
&= W^{-p_r}\cdot\beta_{p_r,q_r}(\frac{W}{W+1}).
\end{aligned}$$

Hence,

$$\pi_{1,r}(\phi|\mathbf{Y}) = \frac{1}{\beta_{p_r,q_r}(\frac{W}{W+1})}\frac{(1+\phi)^{q_r-1}}{W^{q_r}}(1+\frac{1+\phi}{W})^{-p_r-q_r}. \tag{3.4}$$

And $\pi_1(\phi|\mathbf{Y})$ is the special case of $\pi_{1,r}(\phi|\mathbf{Y})$ with $r = 1$. Since $B_r(\mathbf{Y}) = \pi_{1,r}(\phi_0|\mathbf{Y})/\pi_1(\phi_0|\mathbf{Y})$, the proof is complete. □

Now, we apply this approach to a real data, Box and Tiao (1973)'s Dyestuff data, which is set out in Table 3.1. The object of the experiment was to learn to what extent batch to batch variation in a certain raw material was responsible for variation in the final product yield with Jeffreys prior. Five samples from each of six randomly chosen batches of raw material were taken. From data, S_1^2 = 56,357.5, S_2^2 =58,830.0 and $W = 0.95797$. So, $MS_1^2 = S_1^2/\nu_1 = 11,271.50$ and $MS_2^2 = S_2^2/\nu_2 = 2,451.25$ where $\nu_1 = 5$ and $\nu_2 = 24$ are between and within batch degrees of freedom, respectively. Hence, $\hat{\sigma}^2 = MS_2^2 = 2451.25$ and $\hat{\sigma}_e^2$ $= (MS_1^2 - MS_2^2)/\nu_1 = 1764.05$ and so $\hat{\phi} = J\frac{\hat{\sigma}_e^2}{\hat{\sigma}^2} = 3.6$ which is the maximum likelihood estimator and near its posterior mode. $r = n^{-\frac{1}{2}} = 1/\sqrt{30}$ is used for the approximation coefficient of FBF. The value of FBF for testing the hypothesis $H_0 : \phi = 3.6$ is 2.2805. Therefore, we can say that this non-Bayesian estimator for ϕ is decided as a good estimator in Bayesian viewpoint.

Table 3.1: Dyestuff Data

Batch	1	2	3	4	5	6
Observations	1545	1540	1595	1445	1595	1520
	1440	1555	1550	1440	1630	1455
	1440	1490	1605	1595	1515	1450
	1520	1560	1510	1465	1635	1480
	1580	1495	1560	1545	1625	1445

3.2. The truncated normal data involving censored data

Assume that $x_1, \ldots, x_n$ be a random sample from normal distribution with mean μ and variance σ^2 and each x_j be in A_j for $j = 1, \cdots, n$. Let the full data index set $I = \{1, \ldots, n\}$ be divided to the uncensored data index set $I_u = \{i_1, \ldots, i_{n_u}\}$ and the censored data index set $I_c = \{j_1, \ldots, j_{n_c}\}$, that is, $n = n_u + n_c$. Suppose that $x_{i_1} \in A_{i_1}, \ldots, x_{i_{n_u}} \in A_{i_{n_u}}$ are only observed for $i_1, \ldots, i_{n_u} \in I_u$, and $x_{j_1} \in A_{j_1}, \ldots, x_{j_{n_c}} \in A_{j_{n_c}}$ are censored for $j_1, \ldots, j_{n_c} \in I_c$. We want to test the hypothesis $H_0 : \mu = \mu_0$ where μ_0 is a known value. In this situation, following on Jeffreys (1961), we take $\pi_0(\sigma) \propto \sigma^{-1}$ under H_0 and $\pi_1(\mu, \sigma) = \pi_1(\mu|\sigma)\pi_1(\sigma)$ where $\pi_1(\mu|\sigma) = N(\mu_0, \sigma^2)$ and $\pi_1(\sigma) = \pi_0(\sigma)$ under H_1. The true values of censored data $x_{j_1}, \ldots, x_{j_{n_c}}$ are treated as parameters. So, the marginal posterior density $\pi_1(\mu|x_1, \cdots, x_n)$ can be estimated using the method of Gelfand and Smith (1990).

To overcome the unknown constant problem from improper priors, we use the FBF in section 2. Especially, the form of FBF in (2.3) is needed because $\pi_1(\sigma|\mu_0) \neq \pi_0(\sigma)$. At first, for the case $f(x_1, \cdots, x_n|\mu, \sigma)$ we draw

$$\begin{aligned} \mu|x_1, \cdots, x_n, \sigma &\sim N\left(\frac{n\bar{x} + \mu_0}{n+1}, \frac{\sigma^2}{n+1}\right), \\ \sigma|x_1, \cdots, x_n, \mu &\sim \frac{S_1^{1/2}}{\chi_{n+1}} \end{aligned}$$

and

$$x_j|\mu, \sigma \sim N_{A_j}(\mu, \sigma^2), \;\; j \in I_c$$

where $S_1 = (n-1)s^2 + n(\bar{x}-\mu)^2 + (\mu-\mu_0)^2$, $s^2 = \sum_j (x_j - \bar{x})^2/(n-1)$, $\bar{x} = \sum_j x_j/n$. $N_B(\mu, \sigma^2)$ is a normal truncated to the set B, and I_c is defined as before.

Next, for the case $f^r(x_1,\cdots,x_n|\mu,\sigma)$ we draw

$$\begin{aligned}\mu|x_1,\cdots,x_n,\sigma &\sim N(\frac{rn\bar{x}+\mu_0}{rn+1},\frac{\sigma^2}{rn+1}),\\ \sigma|x_1,\cdots,x_n,\mu &\sim \frac{S_r^{1/2}}{\chi_{rn+1}}\end{aligned}$$

and

$$x_j|\mu,\sigma \sim N_{A_j}(\mu,\sigma^2/r),\ \ j\in I_c$$

where $S_r = r(n-1)s^2 + rn(\bar{x}-\mu)^2 + (\mu-\mu_0)^2$.

After these Gibbs sampling processes are performed N times iteratively, we have two samples, $\{\mu^{(g)},\sigma^{(g)},x_{j_1}^{(g)},\ldots,x_{j_{n_c}}^{(g)},g=1,\ldots,N\}$ and $\{\mu^{r(g)},\sigma^{r(g)},x_{j_1}^{r(g)},\ldots,x_{j_{n_c}}^{r(g)},\ g=1,\ \ldots,\ N\}$. Let $\bar{x}^{(g)}$ and $\bar{x}^{r(g)}$ be the means of full data including g-th generated values for censored data $\{x_j^{(g)},j\in I_c\}$ and $\{x_j^{r(g)},j\in I_c\}$, respectively. From these samples, we estimate that

$$\hat{\pi}_1(\mu_0|x_1\in A_1,\cdots,x_n\in A_n)=\frac{1}{N}\sum_{g=1}^{N}N(\frac{n\bar{x}^{(g)}+\mu_0}{n+1},\frac{(\sigma^{(g)})^2}{n+1})(\mu_0)$$

and

$$\hat{\pi}_{1,r}(\mu_0|x_1\in A_1,\cdots,x_n\in A_n)=\frac{1}{N}\sum_{g=1}^{N}N(\frac{rn\bar{x}^{r(g)}+\mu_0}{rn+1},\frac{(\sigma^{r(g)})^2}{rn+1})(\mu_0)$$

where the notation $N(a,b)(c)$ means the normal density value at c with parameters of mean a and variance b.

Now, to compute both of expectation terms of (2.3) we need two samples $\{\tilde{\sigma}_{1,1},\cdots,\tilde{\sigma}_{1,N}\}$ and $\{\tilde{\sigma}_{r,1},\cdots,\tilde{\sigma}_{r,N}\}$ from $\pi_1(\sigma|\mu_0,x_1\in A_1,\cdots,x_n\in A_n)$ and $\pi_{1,r}(\sigma|\mu_0,x_1\in A_1,\cdots,x_n\in A_n)$, respectively. By proceeding as before with μ fixed at μ_0, we can get them. Since $\pi_0(\sigma)/\pi_1(\mu_0,\sigma)=1/\pi(\mu_0|\sigma)=\sqrt{2\pi}\sigma$, $C_1=\sqrt{2\pi}E_1(\sigma|\mu_0,X_1\in A_1,\cdots,X_n\in A_n)$ and $C_r=\sqrt{2\pi}E_r(\sigma|\mu_0,X_1\in A_1,\cdots,X_n\in A_n)$, where $E_1(\cdot)$ and $E_r(\cdot)$ are the expectations with respect to the distributions $\pi_1(\sigma|\mu_0,x_1,\cdots,x_n)$ and $\pi_{1,r}(\sigma|\mu_0,x_1,\cdots,x_n)$, respectively.

Hence, they are estimated by

$$\hat{C}_1=\sqrt{2\pi}\sum_{g=1}^{N}\tilde{\sigma}_{1,g}/N\ \text{ and }\ \hat{C}_r=\sqrt{2\pi}\sum_{g=1}^{N}\tilde{\sigma}_{r,g}/N.$$

Therefore, we estimate $B_r(\mathbf{x_1}, \cdots, \mathbf{x_n})$ as

$$\hat{B}_r(\mathbf{x_1}, \cdots, \mathbf{x_n}) = \frac{\hat{\pi}_1(\mu_0 | x_1 \in A_1, \cdots, x_n \in A_n)}{\hat{\pi}_{1,r}(\mu_0 | x_1 \in A_1, \cdots, x_n \in A_n)} \cdot \frac{\hat{C}_1}{\hat{C}_r}. \tag{3.5}$$

We work numerically with data from an experiment described by Sampford and Taylor (1959) involving 17 pairs of rats that were litter-mates. One of each pair was injected with vitamin B_{12}; the other was a control. The differences between the $\log_{10}$ survival times in minites are $< -.25, -.18, -.83, -.57, -.49, -.12, -.11, -.05, -.04, -.03, -.11, .14, .30, .33, .43, .45$ and $> .30$. The first and the last observations are censored. We are interested in testing whether $\mu = \mu_0 = 0$. Since Berger and Pericchi (1996)'s minimal training sample size is 1, we choose $r = 1/17$. To certify the convergence of each drawn sample, we use the checking method proposed by Gelman and Rubin (1992). The value of FBF in (3.5) for favoring $H_0 : \mu = \mu_0 = 0$ is computed as 1.4945. Therefore, we can conclude that the injection of vitamin B_{12} is not effective, since the mean of differences is decided as zero.

REFERENCES

Aitkin, M. (1991). "Posterior Bayes factor (with discussion)," *Journal of the Royal Statistical Society,* Ser. **B 53,** 111-142.

Berger, J. O. and Pericchi, L. R. (1996). "The Intrinsic Bayes Factor for Model Selection and Prediction," *Journal of the American Statistical Association,* **91,** No. 433, 109-122.

Chen, M.H. (1994). "Importance-weighted marginal Bayesian posterior density estimator," *Journal of the American Statistical Association,* 89. 818-824.

Dickey, J. (1971). "The Weighted Likelihood Ratio Linear Hypotheses on Normal Location Parameters," *The Annuals of Statistics,* **42,** 204-223.

Gelfand, A. and Smith, A. F. M. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* **85,** 398-409.

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulation (with disscussion)," *Statistical Science,* **7,** 457-511.

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

Sampford, M.R. and Taylor, J. (1959). "Censored Observations in Randomized Block Experiments," *Journal of the Royal Statistical Society*, Ser.**B**, **21,** 214-237.

Smith, A. F. M. and Spiegelhalter, D. J. (1980). "Bayes Factors and Choice Criteria for Linear Models," *Journal of the Royal Statistical Society,* Ser. **B**, **42,** 213-220.

Spiegelhalter, D. J. and Smith, A. F. M. (1982). "Bayes Factors for Liner and Log-Linear Models with Vague Prior Information," *Journal of the Royal Statistical Society,* Ser. **B**, **44,** 377-387.

O'Hagan, A. (1995). "Fractional Bayes Factors for Model Comparison," *Journal of the Royal Statistical Society,* Ser. **B**, **57,** 99-138.

Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes Factors Using a Generalization of Savage-Dickey Density Ratio," *Journal of the American Statistical Association,* **90,** 614-618.