Outlier Detection with Fractional Bayes Factor

Sang-Jeen Lee [1] and Choon-Il Park [2]

ABSTRACT

In this paper, we modify the fractional Bayes factor (FBF; O'Hagan, 1995) with the generalized Savage-Dickey density ratio (Verdinelli and Wasserman, 1995) to overcome the unknown constant problem in Bayes factor from the improper priors. This modified FBF are applied to detecting outlier in random effect model with a mean-shift model. Finally, we have a simulation with a hypthetical data set including an outlier and analyze a real data set

KEYWORDS: Fraction Bayes factor; Mean-shift model; Outlier detection; Savage-Dickey density ratio; Important sampling method.

1. Introduction

Recently, the Bayesian approaches as well as frequentists have had an effort in the method of outlier detection Their methods are classified as two main procedures, according to using alternative model for outliers or not.

The methods not using alternative model are proceeded with the predictive distribution as in Geisser (1985) or posterior distribution as in Johnson and Geisser (1983), Chaloner and Brant (1988) and Guttman and Pena (1993).

We use mainly the mean-shift model or the variance-inflation model as an alternative model for outliers. Let $\mathbf{y}$ be a observation vector from $N(\mu, \sigma^2)$. The mean-shift model is that a spurious observation is distributed as $N(\mu + m, \sigma^2)$. If m is not equal to 0, the corresponding observation is decided as an outlier. Guttman (1973) applied the mean-shift model to a linear model. The variance-inflation model is that an observation y_i be from $\mathrm{N}(\mu, b_i\sigma^2)$. The observation, y_i with $b_i >> 1$, is treated as an outlier (Box and Tiao, 1968).

[1] Lecturer, Dept. of Applied Mathematics, Korea Maritime University.
[2] Professor, Dept. of Applied Mathematics, Korea Maritime University.

Sharples (1990) showed how variance inflation can be incorporated easily into general hierarchical models, retaining tractability of analysis.

In this paper, we will apply the mean-shift model to random effect model. Let $\mathbf{Y} = (y_{ij})_{I\times J}$ be a data matrix from the model,

$$y_{ij} = \mu + e_i + \epsilon_{ij}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \tag{1.1}$$

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. But the fear exists that one observations, y_{ks}, may come from

$$y_{ks} = \mu + m_{ks} + e_k + \epsilon_{ks}, \quad m_{ks} \neq 0, \tag{1.2}$$

where m_{ks} is the mean-shift parameter for the spurious observation y_{ks}. If $m_{ks} = 0$, the observation, y_{ks}, is not an outlier. But If $m_{ks} \neq 0$, y_{ks} is an outlier.

Bayes factor is used mainly in Bayesian test. When improper priors are used, Bayes factor includes the unspecified constants. To overcome this problem, O'Hagan (1995) proposed the fractional Bayes factor (FBF) in favor of H_0,

$$B_r(\mathbf{Y}) = \frac{q_0(r, \mathbf{Y})}{q_1(r, \mathbf{Y})}, \tag{1.3}$$

where for $i = 0, 1$, $q_i(r, \mathbf{Y}) = \frac{\int \pi_i(\boldsymbol{\theta}_i) f_i(\mathbf{Y}|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i}{\int \pi_i(\boldsymbol{\theta}_i) f_i^r(\mathbf{Y}|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i}$ and r is the coefficient for fractional approximation.

Nevetheless, these measures have some computational difficulties, yet. Hence, we will compute FBF with following concept. Dickey (1971, 1976) showed that if the Dickey's condition, $\pi_1^N(\xi|\omega) = \pi_0^N(\xi)$, is satisfied then Bayes factor is expressed as $B_{01} = \pi_1^N(\omega_0|\mathbf{Y})/\pi_1^N(\omega_0)$ in a simple hypothesis test, which is called *Savage-Dickey density ratio*. Verdinelli and Wasserman (1995) generalized it for the case unsatisfying Dickey's condition.

Lemma 1.1. (Verdinelli and Wasserman, 1995) If we assume that $0 < \pi_1^N(\omega_0|\mathbf{Y}), \pi_1^N(\omega_0, \xi) < \infty$ for almost all ξ, then the Bayes factor in favor of $H_0 : \omega = \omega_0$ is computed as

$$B_{01} = \frac{\pi_1^N(\omega_0|\mathbf{Y})}{\pi_1^N(\omega_0)} \cdot E^{\pi_1^N(\xi|\omega_0,\mathbf{Y})}[\frac{\pi_0^N(\xi)}{\pi_1^N(\xi|\omega_0)}], \tag{1.4}$$

where $E^{\pi_1^N(\xi|\omega_0,\mathbf{Y})}$ denotes the expectation with respect to $\pi_1^N(\xi|\omega_0,\mathbf{Y})$, which is called the generalized Savage-Dickey density ratio. If Dickey's condition is satisfied, the expectation term of the equation in (1.4) is disappeared.

In Section 2, we will compose Bayesian framework of random effects model for outlier detection. We will modify and compute FBF for outlier detection with the generalized Savage-Dickey density ratio concept in Section 3. In Section 4, illustrative examples will be given and proposed methods will be performed.

2. Bayesian Formulation in Random Effects Model

Consider an observation matrix, $\mathbf{Y} = \{y_{ij}, i = 1,\ldots,I, j = 1,\ldots,J\}$, from the model in (1.1) and (1.2). To decide whether y_{ks} is an outlier or not, we want to test the hypothesis H_0 : there is no outlier in $\mathbf{Y}$ versus $H_1 : y_{ks}$ is an outlier. This test is same with testing hypotheses, for each k and s,

$$H_0 : m_{ks} = 0 \ \text{ versus } \ H_1 : m_{ks} \neq 0. \tag{2.1}$$

For the convenience, we define the variance ratio as $\phi = J\sigma_e^2/\sigma^2$ and hence the parameter vector is $\boldsymbol{\theta} = (\mu, \sigma^2, \phi)$. Then, the likelihood function under H_0 is given by

$$L_0(\mu,\sigma^2,\phi) \propto \sigma^{-IJ}(1+\phi)^{-I/2}\exp\{-\frac{1}{2\sigma^2}(\frac{S_1^2 + IJ(\bar{y}_{..}-\mu)^2}{1+\phi} + S_2^2)\}, \tag{2.2}$$

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

¿From now for notational convenience, let $m_{ks} = m$. Next, we find the likelihood function under H_1. The data $\mathbf{Y}$ including $y_{ks} - m$ in stead of y_{ks},

$$\{y_{ij}, (i,j) \neq (k,s),\ y_{ks} - m\}, \tag{2.3}$$

can be treated as data set without outlier under H_1. Let $\bar{y}_{m..}$, $\bar{y}_{mk.}$, S_{m1}^2 and S_{m2}^2 be $\bar{y}_{..}$, $\bar{y}_{k.}$, S_1^2 and S_2^2 computed with data in (2.3) in stead of $\mathbf{Y}$, respectively. Then, $\bar{y}_{m..} = \bar{y}_{..} - m/IJ$, $\bar{y}_{mk.} = \bar{y}_{k.} - m/J$, $S_{m1}^2 = S_1^2 - 2(\bar{y}_{k.} - \bar{y}_{..})m + \frac{I-1}{IJ}m^2$, and $S_{m2}^2 = S_2^2 - 2(y_{ks} - \bar{y}_{k.})m + \frac{J-1}{J}m^2$. Hence, the likelihood function under

H_1 is

$$\begin{aligned} L_1(\mu, m, \sigma^2, \phi) &\propto \sigma^{-IJ}(1+\phi)^{-I/2}\exp[-\frac{1}{2\sigma^2}\{\frac{S_{m1}^2 + IJ(\bar{y}_{m..}-\mu)^2}{1+\phi} + S_{m2}^2\}] \\ &\propto L_0(\mu, \sigma^2, \phi)\cdot\exp[-\frac{1}{2\sigma^2}\{(\frac{IJ-I+1}{IJ(1+\phi)} - 2\frac{I-1}{IJ})m^2 \\ &\quad -2(\frac{1}{1+\phi}(y_{ks}-\mu) + \frac{\phi}{1+\phi}(\bar{y}_{k.}-\bar{y}_{..}))m\}]. \end{aligned} \quad (2.4)$$

Assume that we have no priori information about the parameters and hence the noninformative priors should be used for null and alternative. ¿From Tiao and Tan (1966) and Box and Tiao (1973), we can have the prior density for null hypothesis as follows;

$$\pi_0^N(\mu, \sigma^2, \phi) = \sigma^{-2}(1+\phi)^{-1}. \quad (2.5)$$

Since mean-shift parameter m is assumed to be independent of $\{\mu, \sigma^2, \phi\}$, the prior under H_1 can be found as

$$\pi_1^N(\mu, m, \sigma^2, \phi) = \pi_1^N(m)\cdot\pi_1^N(\mu, \sigma^2, \phi), \quad (2.6)$$

where $\pi_1^N(m) = 1$ and $\pi_1^N(\mu, \sigma^2, \phi) = \pi_0^N(\mu, \sigma^2, \phi)$ and hence $\pi_1^N(\mu, \sigma^2, \phi|m) = \pi_0^N(\mu, \sigma^2, \phi)$, *i.e.* the Dickey's condition in Lemma 1.1 is satisfied.

We doubt an observation, y_{ks}, in $\mathbf{Y}$ being an outlier. The FBF can be used for overcoming the unknown constant problem of Bayes factor with improper priors. The noninformative priors (2.5) and (2.6) are assumed for null and alternative hypotheses, respectivley. Then, the FBF in favor of H_0 in (2.1)

$$B_r(\mathbf{Y}) = \frac{q_0(r, \mathbf{Y})}{q_1(r, \mathbf{Y})}, \quad (2.7)$$

where for $i = 0, 1$,

$$q_0(r, \mathbf{Y}) = \frac{\int\int\int \pi_0^N(\mu, \sigma^2, \phi)L_0(\mu, \sigma^2, \phi; \mathbf{Y})d\mu d\sigma^2 d\phi}{\int\int\int \pi_0^N(\mu, \sigma^2, \phi)L_0^r(\mu, \sigma^2, \phi; \mathbf{Y})d\mu d\sigma^2 d\phi},$$

and

$$q_1(r, \mathbf{Y}) = \frac{\int\int\int\int \pi_1^N(\mu, m, \sigma^2, \phi)L_1(\mu, m, \sigma^2, \phi; \mathbf{Y})d\mu dm d\sigma^2 d\phi}{\int\int\int\int \pi_1^N(\mu, m, \sigma^2, \phi)L_1^r(\mu, m, \sigma^2, \phi; \mathbf{Y})d\mu dm d\sigma^2 d\phi}.$$

In the equation (2.7), the unknown constants in the Bayes factor from improper prior densities are canceled. But there remain the computational burden and the integration problem about m in FBF. Hence, we will modify FBF with the generalized Savage-Dickey density ratio and integrate with the importance sampling method.

3. Computation of FBF

For the notational convenience, let $\boldsymbol{\theta} = (m, \boldsymbol{\xi})$ with $\boldsymbol{\xi} = (\mu, \sigma^2, \phi)$. Let $m_0 = 0$. We wish to test the null hypothesis $H_0 : m = m_0$ versus alternative $H_1 : m \neq m_0$ with prior distributions $\pi_0^N(\boldsymbol{\xi})$ and $\pi_1^N(m, \boldsymbol{\xi})$ under H_0 and H_1, respectively. The Bayes factor is computed as the ratio of marginal densities for data under H_0 and H_1. It is often difficult to find its analytic form. For these cases, we compute approximately the marginal densities, such as Laplace's approximation (Tierney and Kadane, 1986; Kass, Tierney, and Kadane, 1990), Bartlett adjustment method (DiCiccio and Stern, 1994) and bridge sampling method (Meng and Wong, 1996).

In this section, FBF of O'Hagan (1995) is modified with the Savage-Dickey density ratio in Lemma 1.1 to reduce their computational burden as follows. Let r be the approximation coefficient for FBF. See O'Hagan (1995) for more detail of r.

Lemma 3.1. For an appropriate r, the FBF is expressed as

$$B_{01}^r = B_{01}^N \cdot B_{r10}^N, \tag{3.1}$$

where B_{01}^N and B_{r10}^N are the Bayes factors using likelihood $f(\mathbf{Y}|m, \boldsymbol{\xi})$ and $f^r(\mathbf{Y}|m, \boldsymbol{\xi})$ with the given improper priors, respectively.

Proof. Let $m_i(\mathbf{Y})$ and $m_{i,r}(\mathbf{Y})$ be the marginal densities of data with likelihood functions $f(\mathbf{Y}|m, \boldsymbol{\xi})$ and $f^r(\mathbf{Y}|m, \boldsymbol{\xi})$ under H_i, $i = 0, 1$. Then,

$$\begin{aligned} B_{01}^r &= \frac{q_0(r, \mathbf{Y})}{q_1(r, \mathbf{Y})} \\ &= \frac{\int \pi_0^N(\boldsymbol{\xi}) f(\mathbf{Y}|m_0, \boldsymbol{\xi}) d\boldsymbol{\xi}}{\int \pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0, \boldsymbol{\xi}) d\boldsymbol{\xi}} \Big/ \frac{\int\int \pi_1^N(m, \boldsymbol{\xi}) f(\mathbf{Y}|m, \boldsymbol{\xi}) dm d\boldsymbol{\xi}}{\int\int \pi_1^N(m, \boldsymbol{\xi}) f^r(\mathbf{Y}|m, \boldsymbol{\xi}) dm d\boldsymbol{\xi}} \\ &= \frac{\int \pi_0^N(\boldsymbol{\xi}) f(\mathbf{Y}|m_0, \boldsymbol{\xi}) d\boldsymbol{\xi}}{\int\int \pi_1^N(m, \boldsymbol{\xi}) f(\mathbf{Y}|m, \boldsymbol{\xi}) dm d\boldsymbol{\xi}} \cdot \frac{\int\int \pi_1^N(m, \boldsymbol{\xi}) f^r(\mathbf{Y}|m, \boldsymbol{\xi}) dm d\boldsymbol{\xi}}{\int \pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0, \boldsymbol{\xi}) d\boldsymbol{\xi}} \end{aligned}$$

$$= \frac{m_0(\mathbf{Y})}{m_1(\mathbf{Y})} \cdot \frac{m_{1,r}(\mathbf{Y})}{m_{0,r}(\mathbf{Y})} = B_{01}^N \cdot B_{r10}^N.$$

In the above notations , the marginal densities $m_{0,r}(\mathbf{Y}) = m_0(\mathbf{Y})$ and $m_{1,r}(\mathbf{Y}) = m_1(\mathbf{Y})$ and hence $B_{10}^N = B_{r10}^N$ with $r = 1$. Let $c(r, m, \boldsymbol{\xi}) = \int f^r(\mathbf{Y}|m, \boldsymbol{\xi})d\mathbf{Y}$ and $c(r, \boldsymbol{\xi}) = \int f^r(\mathbf{Y}|m_0, \boldsymbol{\xi})d\mathbf{Y}$. The marginal densities can be expressed as $m_{1,r}(\mathbf{Y}) = \int\int \frac{\pi_1^N(m,\boldsymbol{\xi})}{c(r,m,\boldsymbol{\xi})} \cdot c(r, m, \boldsymbol{\xi}) f^r(\mathbf{Y}|m, \boldsymbol{\xi})\, dm d\boldsymbol{\xi}$ and $m_{0,r}(\mathbf{Y}) = \int \frac{\pi_0^N(\boldsymbol{\xi})}{c(r,\boldsymbol{\xi})} \cdot c(r, \boldsymbol{\xi}) f^r(\mathbf{Y}|m_0,\ \boldsymbol{\xi})d\boldsymbol{\xi}$. Then, $c(r, m, \boldsymbol{\xi}) \times f^r(\mathbf{Y}|m, \boldsymbol{\xi})$ and $c(r, \boldsymbol{\xi}) \times f^r(\mathbf{Y}|m_0, \boldsymbol{\xi})$ can be considered as the marginal densities for $\mathbf{Y}$ and $\frac{\pi_1^N(m,\boldsymbol{\xi})}{c(r,m,\boldsymbol{\xi})}$ and $\frac{\pi_0^N(\boldsymbol{\xi})}{c(r,\boldsymbol{\xi})}$ can be regarded as prior densities for $(m, \boldsymbol{\xi})$ and $\boldsymbol{\xi}$, respectively. Hence, we can treat B_{r10}^N as a Bayes factor. Since B_{01}^N is also a Bayes factor for a simple test, we can apply the Savage-Dickey density ratio concept to these two Bayes factors. But in practice, the computations of $c_1(r, m, \boldsymbol{\xi})$ and $c(r, \boldsymbol{\xi})$ are not needed. Hence, the SDFBF is found as in the following theorem.

Theorem 3.2. With an appropriate r, the FBF in favor of $H_0 : m = m_0$ can be computed as

$$B_{01}^r = \frac{\pi_1^N(m_0|\mathbf{Y})}{\pi_{1,r}^N(m_0|\mathbf{Y})} \cdot \frac{E^{\pi_1^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(\boldsymbol{\xi}|m_0)]}{E^{\pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(\boldsymbol{\xi}|m_0)]}, \tag{3.2}$$

where $\pi_{1,r}^N(\boldsymbol{\xi}|m_0, \mathbf{Y}) = \pi_{1,r}^N(m_0, \boldsymbol{\xi}|\mathbf{Y})/\pi_{1,r}^N(m_0|\mathbf{Y})$, $\pi_{1,r}^N(m_0|\mathbf{Y}) = \int \pi_{1,r}^N(m_0, \boldsymbol{\xi}|\mathbf{Y})\, d\boldsymbol{\xi}$, $\pi_{1,r}^N(m_0, \boldsymbol{\xi}|\mathbf{Y}) = \pi_1^N(m, \boldsymbol{\xi}) f^r(\mathbf{Y}|m, \boldsymbol{\xi})/m_{1,r}(\mathbf{Y})$. and $E^{g(\boldsymbol{\xi})}(\cdot)$ denotes the expectation with respect to the density $g(\boldsymbol{\xi})$.

Proof. By Lemma 3.1, FBF is the product of two Bayes factors in (3.1), which are expressed by Lemma 1.1 as follows;

$$B_{01}^N = \frac{\pi_1^N(m_0|\mathbf{Y})}{\pi_1^N(m_0)} \cdot E[\frac{\pi_0^N(\boldsymbol{\xi})}{\pi_1^N(\boldsymbol{\xi}|m_0)}]$$

and

$$\begin{aligned} B_{r01}^N &= 1/B_{r10}^N = \frac{m_{0,r}(\mathbf{Y})}{m_{1,r}(\mathbf{Y})} = \frac{\int \pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0, \boldsymbol{\xi})d\boldsymbol{\xi}}{m_{1,r}(\mathbf{Y})} \\ &= \pi_{1,r}^N(m_0|\mathbf{Y}) \cdot \int \frac{\pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0, \boldsymbol{\xi})\pi_{1,r}^N(\boldsymbol{\xi}|m_0, \mathbf{Y})}{m_{1,r}(\mathbf{Y})\pi_{1,r}^N(m_0|\mathbf{Y})\pi_{1,r}^N(\boldsymbol{\xi}|m_0, \mathbf{Y})} d\boldsymbol{\xi} \end{aligned}$$

Since $\pi_{1,r}^N(m_0|\mathbf{Y}) = \pi_{1,r}^N(m_0,\boldsymbol{\xi}|\mathbf{Y})/\pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})$,

$$
\begin{aligned}
B_{r01} &= \pi_{1,r}^N(m_0|\mathbf{Y}) \cdot \int \frac{\pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0,\boldsymbol{\xi}) \pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}{m_{1,r}(\mathbf{Y}) \pi_{1,r}^N(m_0,\boldsymbol{\xi}|\mathbf{Y})} d\boldsymbol{\xi} \\
&= \pi_{1,r}^N(m_0|\mathbf{Y}) \cdot \int \frac{\pi_0^N(\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0,\boldsymbol{\xi}) \pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}{m_{1,r}(\mathbf{Y}) \frac{\pi_1^N(m_0,\boldsymbol{\xi}) f^r(\mathbf{Y}|m_0,\boldsymbol{\xi})}{m_{1,r}(\mathbf{Y})}} d\boldsymbol{\xi} \\
&= \pi_{1,r}^N(m_0|\mathbf{Y}) \cdot \int \frac{\pi_0^N(\boldsymbol{\xi}) \pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}{\pi_1^N(m_0,\boldsymbol{\xi})} d\boldsymbol{\xi} \\
&= \frac{\pi_{1,r}^N(m_0|\mathbf{Y})}{\pi_1^N(m_0)} \cdot \int \frac{\pi_0^N(\boldsymbol{\xi})}{\pi_1^N(\boldsymbol{\xi}|m_0)} \pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y}) d\boldsymbol{\xi} \\
&= \frac{\pi_{1,r}^N(m_0|\mathbf{Y})}{\pi_1^N(m_0)} \cdot E^{\pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(\boldsymbol{\xi}|m_0)].
\end{aligned}
$$

Since $B_{01}^r = B_{01}^N / B_{r01}^N$ from Lemma 3.1, the proof is completed.

Note that this form of FBF is called *Savage-Dickey FBF* (SDFBF). By multiplying an independent quantity $\pi_1^N(m_0)$ to both numerator and denominator of the expectation terms in (3.2), we can replace $\pi_1^N(\boldsymbol{\xi}|m_0)$ by $\pi_1^N(m_0,\boldsymbol{\xi})$ in (3.2),

$$
\frac{E^{\pi_1^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(\boldsymbol{\xi}|m_0)]}{E^{\pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(\boldsymbol{\xi}|m_0)]} = \frac{E^{\pi_1^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(m_0,\boldsymbol{\xi})]}{E^{\pi_{1,r}^N(\boldsymbol{\xi}|m_0,\mathbf{Y})}[\pi_0^N(\boldsymbol{\xi})/\pi_1^N(m_0,\boldsymbol{\xi})]}. \tag{3.3}
$$

This computing approach is convenient in practice. Especially, the FBF is reduced under the condition of Dickey as follows;

Remark 3.3. If the Dickey's condition is satisfied, the SDFBF in (3.2) is reduced to the form that expectation terms disappear.

$$
B_{01}^r = \pi_1^N(m_0|\mathbf{Y})/\pi_{1,r}^N(m_0|\mathbf{Y}). \tag{3.4}
$$

Now, we compute SDFBF for the outlier detection in model (2.1). First, we should find the marginal posterior density of m.

Lemma 3.4. The prior densities in (2.5) and (2.6) are assumed under H_0 and H_1, respectively. We use $\{L_1(\mu,m,\sigma^2,\phi)\}^r$ of equation in (2.4) as a likelihood function. Then, the marginal posterior density for m is

$$
\pi_{1,r}(m|\mathbf{Y}) = C_r \beta_{p_r,q_r}\left(\frac{W_m}{W_m+1}\right)(S_{m2}^2)^{-(p_r+q_r)} W_m^{-p_r}, \quad 0 < r \le 1 \tag{3.5}
$$

where

$$C_r^{-1} = \int_{-\infty}^{\infty} \beta_{p_r,q_r}\left(\frac{W_m}{W_m+1}\right) W_m^{-p_r} S_{m2}^{2(-p_r-q_r)} dm, \quad 0 < r \leq 1 \tag{3.6}$$

S_{m1}^2 and S_{m2}^2 are defind in (2.4), $W_m = S_{m1}^2/S_{m2}^2$, $p_r = (rI-1)/2$, $q_r = rI(J-1)/2$ and $\beta_{i,j}(x) = \int_0^x t^{i-1}(1-t)^{j-1}dt$, $0 < r \leq 1$.

Proof. The joint posterior density for (μ, m, σ^2, ϕ) is

$$\begin{aligned} \pi_{1,r}(\mu, m, \sigma^2, \phi | y_{..}, s_1^2, s_2^2) \;\; &\propto \;\; \sigma^{-(rIJ+2)}(1+\phi)^{-(rI+2)/2} \\ &\quad \cdot \exp\left[-\frac{r}{2\sigma^2}\left\{\frac{S_{m1}^2 + IJ(y_{m..}-\mu)^2}{1+\phi} + S_{m2}^2\right\}\right]. \end{aligned}$$

So the marginal posterior density for m is computed as

$$\begin{aligned} \pi_{1,r}(m|\mathbf{Y}) \;\; &\propto \;\; \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \pi_{1,r}(\mu, m, \sigma^2, \phi|\mathbf{Y}) d\mu d\sigma^2 d\phi \\ &\propto \;\; \int_0^\infty \int_0^\infty \sigma^{-(rIJ+2)}(1+\phi)^{-(rI+2)/2} \exp\left\{-\frac{r}{2\sigma^2}\left(S_{m2}^2 + \frac{S_{m1}^2}{1+\phi}\right)\right\} \\ &\quad \cdot \left\{\frac{2\pi}{rIJ}\sigma^2(1+\phi)\right\}^{1/2} d\sigma^2 d\phi \\ &\propto \;\; \int_0^\infty (1+\phi)^{-(rI+1)/2} \int (\sigma^2)^{-(rIJ+1)/2} \exp\left\{-\frac{r}{2\sigma^2}\left(S_{m2}^2 + \frac{S_{m1}^2}{1+\phi}\right)\right\} \\ &\quad d\sigma^2 d\phi \\ &\propto \;\; \int_0^\infty (1+\phi)^{-(rI+1)/2}\left(S_{m2}^2 + \frac{S_{m1}^2}{1+\phi}\right)^{-(rIJ-1)/2} d\phi. \end{aligned}$$

With transformation $Z_m = W_m/(W_m+1+\phi)$, we have

$$\begin{aligned} \pi_{1,r}(m|\mathbf{Y}) \;\; &\propto \;\; \int_0^\infty (1+\phi)^{-(rI+1)/2}\left(1+\frac{W_m}{1+\phi}\right)^{-(rIJ-1)/2} S_{m2}^{-(rIJ-1)} d\phi \\ &\propto \;\; (S_{m2}^2)^{-(p_r+q_r)} W_m^{-p_r} \int_0^{\frac{W_m}{W_m+1}} Z_m^{p_r-1}(1-Z_m)^{q_r-1} dZ_m \\ &= \;\; (S_{m2}^2)^{-(p_r+q_r)} W_m^{-p_r} \beta_{p_r,q_r}\left(\frac{W_m}{W_m+1}\right). \end{aligned}$$

We compute the SDFBF in (3.2) for our Bayesian outlier detection approach with appropriate r.

Theorem 3.5. The SDFBF in favor of H_0 in (2.1) is written as

$$B_{01}^r(Y) = \frac{C_1}{C_r} \frac{\beta_{p,q}\left(\frac{W}{W+1}\right)}{\beta_{p_r,q_r}\left(\frac{W}{W+1}\right)} W^{-p+p_r} S_2^{2(p_r-p+q_r-q)}, \tag{3.7}$$

where the normalized constants C_r^{-1} is defined in (3.6), p, q, p_r, q_r, W_m, S_{m1}^2, S_{m2}^2, and $\beta_{i,j}(x)$ are defined in (2.4) and (3.6).

Proof. Since the Dickey's condition is satisfied, the marginal posterior density for m with likelihood, $\{L_1(\mu, m, \sigma^2, \phi)\}^r$, was found in the Lemma 3.4 and $\pi_1(m|\mathbf{Y}) = \pi_{1,r}(m|\mathbf{Y})$ with $r = 1$, the proof is complete by Remark 3.3.

The normalizing constants C_r in (3.6) are impossible to find analytically. So the sampling based computing methods are needed to estimate them, such as importance sampling method. Let

$$g_r(m) = \beta_{p_r,q_r}(\frac{W_m}{W_m+1})W_m^{-p_r}(S_{m2}^2)^{-(p_r+q_r)}$$

The quantity C_r can be found by computing

$$C_r^{-1} = \int_{-\infty}^{\infty} g_r(m)dm = \int_{-\infty}^{\infty} \frac{g_r(m)}{I(m)} I(m)dm = E\{\frac{g_r(m)}{I(m)}\},$$

where $I(m)$ is a certain probability density with same support of $g_r(m)$. After we generate $\{m^{(1)}, \ldots, m^{(G)}\}$ from the distribution with density $I(m)$, we can estimate the value of C_r^{-1} by Monte Carlo method,

$$\hat{C}_r^{-1} = \frac{1}{G}\sum_{g=1}^{G} \frac{g_r(m^{(g)})}{I(m^{(g)})}. \qquad (3.8)$$

The choice of $I(m)$ is very important. We suggest to use the normal density with mean $\bar{m}$ and variance s_m^2 as an importance function $I(m)$, where $\bar{m}$ and s_m^2 are the sample mean and sample variance of the sample generated from $g_r(m)$ with sampling method, such as the Metropolice algorithm.

4. Illustrative Examples

We perform simulations of SDIBF and SDFBF for detecting outliers explained in previous sections with a generated data including an outlier. This data set is generated from the balanced random effect model in (1.1) with mean $\mu = 5$, $\sigma_e^2 = 6$ and $\sigma^2 = 8$. But the observation y_{52} is generated from the model in (1.1) with mean $\mu = 0$ and same σ_e^2 and σ^2, that is, it is generated from the model in (1.2) with $m_{52} = -5$, which is set out in Table 4.1. As a real data, Dyestuff data in Table 4.2 is also analyzed with the proposed outlier detection procedure. The procedures are executed for every observation in data. For the computation of normalizing constant C_r, we use the importance sampling method.

The values of SDFBF for outlier detection in (3.7) in the generated data set are computed and the results are listed in Table 4.3 The mean-shifted observation y_{52} has the smallest value. Hence, we can conclude that out suggested method is certified as a good procedure. However, the data in the 5-th batch are far from center. ¿From this, we can know that the 5-th batch random fact e_5 is a little far from center. This is coincided with the factor that the values in the 5-th column of data in Table 4.1 are smaller than values in other columns.

Table 4.1. Generated data including an outlier

Batch	1	2	3	4	5	6
obs. 1	7.8925	-0.0030	10.1009	13.6895	0.5623	5.3777
obs. 2	12.6125	7.0934	5.0114	10.6080	-8.4583	10.1637
obs. 3	4.3213	12.4114	7.9833	9.6563	0.7844	3.4680
obs. 4	13.1566	7.8590	11.1319	11.2744	5.6431	5.4790
obs. 5	12.8839	9.0184	6.7217	3.8906	5.1731	7.5221

Table 4.2. Dyestuff data(Yield of Dyestuff of standard color)

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

In the real data, Dyestuff data, the values of SDFBF for all observations are listed in Table 4.4. All observations are able to be decided as no outliers. Since all values in the first column of Table 4.4 are similar to 1 but less than 1, the first batch random factor e_1 may be determined as farther than other batches.

Table 4.3. SDFBF values for outlier in generated data

s k	1	2	3	4	5	6
1	1.0673	0.6704	0.9292	0.8395	0.9109	1.0651
2	0.8869	1.0207	1.1297	0.9524	0.5180	0.8811
3	0.7738	0.8223	1.0103	0.9887	0.9223	0.8619
4	0.8672	0.9907	0.8910	0.9273	0.9014	1.0610
5	0.8770	0.9463	1.0602	0.7651	0.9191	0.9805

Table 4.4. SDFBF values for outlier in Dyestuff data

s k	1	2	3	4	5	6
1	0.9915	0.9970	0.9900	1.0152	0.9966	0.9913
2	1.0178	0.9933	1.0012	1.0165	0.9879	1.0075
3	1.0178	1.0095	0.9875	0.9779	1.0166	1.0088
4	0.9978	0.9921	1.0112	1.0102	0.9867	1.0013
5	0.9829	0.1008	0.9987	0.9903	0.9892	1.0100

5. Concluding Remarks

A Bayesian outlier detection method was suggested with the mean-shift model by FBF in random effect model. Since the FBF are impossible to find analytically with improper priors, we modified it as SDFBF. The SDFBF were used to test whether m is zero or not. The performance of the approache is decided as good in simulation study.

In the computation of normalizing constant C_r, Meng and Wong (1996)'s method are needed for their better estimators. However, the computational difficulty with very long CPU time was an obstacle. Hence, we should study more efficient algorithm. It should also be studied the simultaneous detection method for two or more outliers.

References

[1] Box, G. E. P. and Tiao, G. C. (1968), A Bayesian Approach to Some Outlier Problems, *Biometrika,* **55,** 119-129.

[2] Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis,* Addison-Wesley Publishing Co., U.S.A.

[3] Chaloner, K. and Brant, R. (1988), A Bayesian Approach to Outlier Detection and Residual Analysis, *Biometrika,* **75,** 651-659.

[4] Chung, Younshik and Lee, Sangjeen (1998), Computing Fractional Bayes Factor Using Generalized Savage-Dickey Density Ratio A Bayesian Approach to Outlier Detection and Residual Analysis, *Journal of Korean Statistical Society,* **27,** 387-396.

[5] Dickey, J. (1971), The Weighted Likelihood Ratio Linear Hypotheses on Normal Location Parameters, *the Annals of Mathematical Statistics,* **42,** 204-223.

[6] Dickey, J. (1976), Approximate Posterior Distributions, *Journal of the American Statistical Association,* **71,** 680-689.

[7] Geisser, S. (1985), On the Predicting of Observables: a Selective Update, in *Bayesian Statistics 2*, Ed. Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M., 203-230, Amsterdam: North Holland.

[8] Geisser, S. (1987), Influential Observations, Diagnostics and Discordancy Tests, *Journal of Applied Statistics,* **14,** 133-142.

[9] Guttman, I. (1973), Care and Handling of Univariate or Multivariate Outliers in Detecting Spuriosity - A Bayesian Approach, *Technometrics,* **15,** 4, 723-738.

[10] Guttman, I., Dutter, R. and Freeman, P. R. (1978), Care and Handling of Univariate Outliers in the General Linear Model to Detect Spuriosity - A Bayesian Approach, *Technometrics,* **20,** 2, 187-193.

[11] Guttman, I. and Pena, D. (1993), A Bayesian Look at Diagnostics in the Univariate Linear Model, *Statistical Sinica,* **3**, 367-390.

[12] Johnson, W. and Geisser, S. (1983), A Predictive View of the Detection and Characterization of Influential Observations in Regression Analysis, *Journal of the American Statistical Association,* **78,** 137-144.

[13] Meng, X. L. and Wong, W. H. (1993), Simulating Ratios of Normalizing Constants Via a Simple Identity : A Theoretical Exploration, Technical Report 365, University of Chicago, Dept. of Statistics.

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

[15] Tiao, G. C. and Guttman, I. (1967), Analysis of Outliers with Adjusted Residuals, *Technometrics,* **9,** 541-568.

[16] 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.