Journal of the Korean Statistical Society (2000), **29**: 1, pp 81–94

Detecting the Influential Observation Using Intrinsic Bayes Factors †

Younshik Chung[1]

ABSTRACT

For the balanced variance component model, sometimes intraclass correlation coefficient is of interest. If there is little information about the parameter, then the reference prior(Berger and Bernardo, 1992) is widely used. Pettit nd Young(1990) considered a measure of the effect of a single observation on a logarithmic Bayes factor. However, under such a reference prior, the Bayes factor depends on the ratio of unspecified constants. In order to discard this problem, influence diagnostic measures using the intrinsic Bayes factor(Berger and Pericchi, 1996) is presented. Finally, one simulated dataset is provided which illustrates the methodology with appropriate simulation based computational formulas. In order to overcome the difficult Bayesian computation, MCMC methods, such as Gibbs sampler(Gelfand and Smith, 1990) and Metropolis algorithm, are employed.

Keywords: Conditional predictive ordinate; Gibbs sampler; Intraclass correlation coefficient; Intrinsic Bayes factor; Metropolis algorithm; Reference prior; Variance component model.

1. Introduction

Consider a balanced variance components model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \text{ for } i = 1, \dots, I; \; j = 1, \dots, J, \tag{1.1}$$

where μ is the mean effect, and the random effects α_i's are independent and identically distributed as $N(0, \sigma_\alpha^2)$. ϵ_{ij}'s are assumed to be independent and identically distributed as $N(0, \sigma^2)$. The α_i's are also assumed to be independent of

† This research was partially supported by the Matching Fund Programs of Research Institute for Basic Sciences, Pusan National University, Korea, 1999.

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

ϵ_{ij}'s. Further the parameters $(\mu, \sigma_\alpha^2, \sigma^2)$ are unknown. Define variance-covariance structure for the y_{ij}'s as

$$cov(y_{ij}, y_{i'j'}) = \begin{cases} 0, & i \neq i', \\ \rho\sigma^2 & i = i', j \neq j', \\ \sigma^2 & i = i', j = j'. \end{cases}$$

Thus σ^2 is the variance of each y_{ij}, and ρ is the correlation coefficient between y_{ij}'s within the same class. Usually we are interested in the inference concerning the variance ratio $\phi = J\sigma_\alpha^2/\sigma^2$ and the intraclass correlation coefficient $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma^2)$.

In this paper, we consider detecting the influential observation under default prior in a variance component model. Since our focus is fully Bayesian, the choice of priors is very important. Chaloner (1987) considered estimators of ρ using non-informative prior distributions which do not depend on the sample size. Palmer and Broemeling (1990) used the inverted gamma prior distribution for estimation of intraclass correlation coefficient ρ. In this case, we consider the reference prior proposed by Bernardo(1979) for the development of the noninformative prior for ρ. The refernce prior algorithm is now quite popular for the development of default priors for many interesting problems. Berger and Bernardo(1992) extend Bernardo(1979) algorithm to multi-parameter problem. Sun and Ye(1995) used Bayesian reference prior approach widely for estimating product of normal means. Ye(1994) considered the reference prior for the estimation of the ratio of variance for one-way random effects model. Following Berger and Bernardo(1992)'s algorithm, Chung and Dey(1998) obtained the reference priors for (ρ, μ, σ^2) which are of the form

$$\pi^r(\rho, \mu, \sigma^2) = \sigma^{-a}(1 + (J-1)\rho)^{-c}(1-\rho)^{-d}, \tag{1.2}$$

where a, c and d are some non-negative integers.

In this paper our objective is to consider a variance component model from a Bayesian perspective and devote model diagnostics under default prior specification for ρ using the intrinsic Bayes factor. The paper is organized as follows. Section 2 reviews the noninformative priors and develops the reference priors for different parameters. In Section 3, diagnostic measure using intrinsic Bayes factor is presented and computed using Gibbs sampler. Finally, in Section 4, we examine measuring the effect of observations on intrinsic Bayes factor to simulated data(Box and Tiao, 1973).

2. Reference Priors and Intrinsic Bayes Factor

2.1. Reference Priors

In Bayesian analysis, the choice of prior is very critical. Since there is no precise information about parameters, we use noninformative priors. Bernardo(1979) proposed the reference prior as a development of noninformative prior. The key feature was a possible dependence of the reference prior on specification of parameters of interest and nuisance parameters. Berger and Bernardo(1992) extended their algorithm to multi-parameter problems.

For our balanced variance components model, $\rho = \sigma^2/(\sigma_\alpha^2 + \sigma^2)$ is the parameter of interest. The reference prior distributions for different groups of ordering of (μ, σ^2) and (ρ, μ, σ^2) are found as follows.

Theorem 2.1. For the balanced variance components model in (1.1), if $\rho = \rho_0$ is given, then the reference prior distribution for any ordered group of (μ, σ^2) is

$$\pi_0(\mu, \sigma^2) \propto \sigma^{-2}. \tag{2.1}$$

Proof. Its proof is very similar to that of Chung and Dey (1998).

Theorem 2.2. (Chung and Dey, 1998) For the balanced variance component model, if $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma^2)$ is the parameter of interest, then the reference prior distributions for different groups of ordering of (μ, σ^2, ρ) are:

Group ordering	Reference prior
$\{\rho, \mu, \sigma^2\}, \{\rho, \sigma^2, \mu\}, \{(\rho, \sigma^2), \mu\}$	$\pi_1 \propto \sigma^{-2}(1 + (J-1)\rho)^{-1}(1-\rho)^{-1}$
$\{\rho, (\mu, \sigma^2)\}$	$\pi_2 \propto \sigma^{-3}(1 + (J-1)\rho)^{-1}(1-\rho)^{-1}$
$\{(\rho, \mu), \sigma^2\}$	$\pi_3 \propto \sigma^{-2}(1 + (J-1)\rho)^{-\frac{3}{2}}(1-\rho)^{-\frac{1}{2}}$
$\{(\rho, \mu, \sigma^2)\}$	$\pi_4 \propto \sigma^{-3}(1 + (J-1)\rho)^{-\frac{3}{2}}(1-\rho)^{-\frac{1}{2}}$.

Note that the Jeffreys's prior is same as the reference prior π_4 for (ρ, μ, σ^2).

2.2. Intrinsic Bayes Factor

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

$$B = \frac{p(H_0|Y)/p(H_1|Y)}{p(H_0)/p(H_1)} = \frac{m_0(Y)}{m_1(Y)}, \tag{2.2}$$

where $m_i(Y) = \int \pi_i(\theta_i) f_i(Y|\theta_i) d\theta_i$ is the marginal density of Y under model H_i and $\pi_i(\theta_i)$ is the prior density of θ_i under H_i for $i = 0, 1$. For instance, an improper prior for θ_i is written as $\pi_i^N(\theta_i) \propto g_i(\theta_i)$, where g_i is integrable function over θ_i-space. It could be expressed that

$$\pi_i^N(\theta_i) = c_i g_i(\theta_i), \quad i = 0, 1. \tag{2.3}$$

If $g_i(\theta_i)$ is not integrable , we can treat c_i as an unspecified constant. However, Bayes factor in favor of H_0, with respect to these priors,

$$B = \frac{c_0}{c_1} \frac{\int g_0(\theta_0) f_0(Y|\theta_0) d\theta_0}{\int g_1(\theta_1) f_1(Y|\theta_1) d\theta_1}, \tag{2.4}$$

dedends on the ratio of undefind constants c_0/c_1. Recently, to overcome this problem, various approaches have been advocated. See Aitkin(1991) and Berger and Pericchi(1996). Now we briefly review the IBF given by Berger and Pericchi (1996).

Let $Y_L = \{y(1), y(2), \ldots, y(L)\}$ denote the set of all minimal training samples, $y(l)$. That is, a training sample, $y(l)$, is called proper if $\int f_i(y(l)|\theta_i)\pi_i^N(\theta_i)d\theta_i < \infty$ and minimal if it is proper and no subset gives finite marginals. Berger and Pericchi (1996) introduced several intrinsic Bayes factors, including arithmetic intrinsic Bayes factor(AIBF) and geometric intrinsic Bayes factor(GIBF) defined respectively by

$$B_{01}^{AI} = \frac{1}{L}\sum_{l=1}^{L} B_{01}(y(l)) = B_{01}^N \frac{1}{L}\sum_{l=1}^{L} B_{10}^N(y(l)) \tag{2.5}$$

and

$$B_{01}^{GI} = \{\prod_{l=1}^{L} B_{01}(y(l))\}^{\frac{1}{L}} = B_{01}^N \{\prod_{l=1}^{L} B_{10}^N(y(l))\}^{\frac{1}{L}} \tag{2.6}$$

where

$$B_{01}^N = \frac{m_0^N(y)}{m_1^N(y)} = \frac{\int f_0(y|\theta_0)\pi_0^N(\theta_0)d\theta_0}{\int f_1(y|\theta_1)\pi_1^N(\theta_1)d\theta_1} \quad \text{and} \quad B_{10}^N(y(l)) = \frac{m_1^N(y(l))}{m_0^N(y(l))}.$$

3. Diagnostic measure using Intrinsic Bayes factor

Consider the following test for H_0(model M_0 is chosen) against H_1(model M_1 is chosen). Then in order to measure the effect on the Bayes factor of observation

d, Pettit and Young(1990) suggested the quantity k_d defined by

$$k_d = logBF_{01} - logBF_{01}^{(d)}, \tag{3.1}$$

where BF_{01} and $BF_{01}^{(d)}$ denote the Bayes factors with all data and with all data excluding observation d respectively. In different way,

$$k_d = \log \frac{m_0(y)}{m_1(y)} - \log \frac{m_0(y_{(d)})}{m_1(y_{(d)})} = \log \frac{m_0(y)}{m_0(y_{(d)})} - \log \frac{m_1(y)}{m_1(y_{(d)})}, \tag{3.2}$$

where y is the full data and $y_{(d)}$ is all the data excluding observation d and $m_i(y) = \int f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$ for $i = 0, 1$.

This k_d is expressed as the difference in the logarithms of the conditional predictive ordinates(CPO) for the two models which was discussed in Gelfand, Dey and Chang(1992). Also, Pettit(1990) mentioned that the CPO is a measure to detect surprising observations. Thus large values of $|k_d|$ indicate that such observation d has a large influence on the Bayes factor.

3.1 Diagnostic measures based on IBF

Let $y_L = \{y(1), y(2), \ldots, y(L)\}$ denote the set of all minimal training samples, $y(l)$, of all data y and let $y_{(d)M} = \{y_{(d)}(1), y_{(d)}(2), \ldots, y_{(d)}(M)\}$ the set of all minimal training samples, $y_{(d)}(m)$, of all data excluding observation d.

For notational convenience, let

$$B_{01(d)}^{AI} = \frac{1}{M}\sum_{m=1}^{M} B_{01}(y_{(d)}(m)) = B_{01(d)}^{N}\frac{1}{M}\sum_{m=1}^{M} B_{10}^{N}(y_{(d)}(m)) \tag{3.3}$$

and

$$B_{01(d)}^{GI} = \{\prod_{m=1}^{M} B_{01}(y_{(d)}(m))\}^{\frac{1}{M}} = B_{01(d)}^{N}\{\prod_{m=1}^{M} B_{10}^{N}(y_{(d)}(m))\}^{\frac{1}{M}}, \tag{3.4}$$

where

$$B_{01(d)}^{N} = \frac{m_0^N(y_{(d)})}{m_1^N(y_{(d)})} = \frac{\int f_0(y_{(d)}|\theta_0)\pi_0^N(\theta_0)d\theta_0}{\int f_1(y_{(d)}|\theta_1)\pi_1^N(\theta_1)d\theta_1} \quad \text{and} \quad B_{10}^{N}(y_{(d)}(m)) = \frac{m_1^N(y_{(d)}(m))}{m_0^N(y_{(d)}(m))}.$$

Following Pettit and Young(1990) and Dey(1996), we define the diagnostic measures K_d^{AI} and K_d^{GI} obtained by deleting dth observation as

$$K_d^{AI} = \log B_{01}^{AI} - \log B_{01(d)}^{AI} \tag{3.5}$$

and

$$K_d^{GI} = \log B_{01}^{GI} - \log B_{01(d)}^{GI}, \tag{3.6}$$

where B_{01}^{AI} and B_{01}^{GI} are in (2.6) and (2.7) respectively.

For interpretation of K_d, a negative K_d indicates less support for model M_0 from observation d and a positive K_d suggests that the dth observation favors M_0. To assess whether deletion of observation d changes our beliefs we have to compare K_d with $\log B_{01}^{AI}$ or $\log B_{01}^{GI}$ in (3.5) and (3.6) respectively. This can be done by plotting $\log B_{01}^{AI}$ and K_d.

Thus, a plot of K_d against the observation number can be used as a model-selection criterion. If we observe more positive K_d's than negative ones, the data support M_0 more than M_1; otherwise, the data support M_1 more. We can extend K_d to the case when we omit two or more observations and measure their joint effect. To avoid a combinatorial explosion, we should consider the order to delete observations. For simple possibility to decide the order, we can delete the observation with maximum value of $|K_d|$ first. Thus, we would have a sequence

$$K_d^{AI} = \log B_{01}^{AI} - \log B_{01(d)}^{AI}, \quad K_{de}^{AI} = \log B_{01(d)}^{AI} - \log B_{01(de)}^{AI} \tag{3.7}$$

and

$$K_d^{GI} = \log B_{01}^{GI} - \log B_{01(d)}^{GI}, \quad K_{de}^{GI} = \log B_{01(d)}^{GI} - \log B_{01(de)}^{GI}. \tag{3.8}$$

Here, $B_{01(de)}$ denotes the Bayes factor with all data excluding observations d and e.

3.2 Computation of the measures K_d^{AI} and K_d^{GI}

Suppose that we want to test $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$ where ρ_0 is a known specified number. In each case, the reference priors in Theorems 2.1 and 2.2 are assumed. That is, $\pi_0(\mu, \sigma^2) \propto \sigma^{-2}$ is assumed under H_0 and for H_1, $\pi(\mu, \sigma^2, \rho) \propto \sigma^{-a}(1 + (J-1)\rho)^{-c}(1-\rho)^{-d}$ is given where $d = 2 - c$ and $(a, c) = (2, 1), (2, 1.5), (3, 1)$ or $(3, 1.5)$.

For our model (1.1), the likelihood function of parameters (μ, σ^2, ρ) is given by

$$\begin{aligned} l(\mu, \sigma^2, \rho) &= (\sigma^2)^{-\frac{IJ}{2}} \left(\frac{1 + (J-1)\rho}{1-\rho}\right)^{-\frac{I}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\frac{IJ(1-\rho)}{1 + (J-1)\rho} (y_{\cdot\cdot} - \mu)^2 \right.\right. \\ &\left.\left. + \frac{J(1-\rho)}{1 + (J-1)\rho} \sum_{i=1}^{I} (y_{i\cdot} - y_{\cdot\cdot})^2 + \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - y_{i\cdot})^2 \right] \right\}, \end{aligned}$$

where $y_{i\cdot} = \sum_{j=1}^{J} y_{ij}/J$ and $y_{\cdot\cdot} = \sum_{i=1}^{I} y_{i\cdot}/I$.

Lemma 3.1. The marginal densities of Y under H_0 and H_1 are given respectively by

$$m_0^N(y) = K(1+(J-1)\rho_0)^{-\frac{I-1}{2}}(1-\rho_0)^{\frac{I-1}{2}}\left(\frac{1-\rho_0}{1+(J-1)\rho_0}W+1\right)^{-\frac{IJ-1}{2}} \quad (3.9)$$

and

$$\begin{aligned} m_1^N(y) &= \frac{1}{JW^p}\sqrt{\frac{\pi}{IJ}}2^{\frac{a+IJ-2}{2}}\Gamma\left(\frac{a+IJ-3}{2}\right) \\ &\times \left\{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-y_{i\cdot})^2\right\}^{-\frac{a+IJ-3}{2}} B_{q,p}\left(\frac{1}{1+W}\right), \end{aligned} \quad (3.10)$$

where $p = \frac{1}{2}(I+2c-3), q = \frac{1}{2}(I(J-1)+a-2c)$, $\quad B_{\alpha,\beta}(x) = \int_x^1 t^{\alpha-1}(1-t)^{\beta-1}dt$, $W = \frac{J\sum_{i=1}^{I}(y_{i\cdot}-y_{\cdot\cdot})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-y_{i\cdot})^2}$ and $K = \sqrt{\frac{\pi}{IJ}}2^{\frac{IJ}{2}}\Gamma(\frac{IJ-1}{2})\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-y_{i\cdot})^2\right\}^{-\frac{IJ-1}{2}}$.

Furthermore,

$$\begin{aligned} B_{01}^N = \frac{m_0^N(y)}{m_1^N(y)} &= \frac{JW^p\Gamma(\frac{IJ-1}{2})}{2^{a-2}\Gamma(\frac{a+IJ-3}{2})}\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-y_{i\cdot})^2\right\}^{\frac{a-2}{2}}\left(\frac{1-\rho_0}{1+(J-1)\rho_0}\right)^{\frac{I-1}{2}} \\ &\times \frac{1}{B_{q,p}(\frac{1}{W+1})(\frac{1-\rho_0}{1+(J-1)\rho_0}W+1)^{\frac{IJ-1}{2}}}. \end{aligned} \quad (3.11)$$

Proof. Its proof is straightforward.

From (3.9) and (3.10), MTS size is the minimal number of sample size such that both $m_0^N(y)$ and $m_1^N(y)$ are finite. That is, for $a = 2$ and 3, $\frac{a+IJ-3}{2} > 0$. Therefore, MTS size is (1,2) or (2,1). Let I_m and J_m denote the minimal training sample sizes of I and J, respectively. But it may be chosen for MTS size to be $I_m = J_m = 2$ in our model in order to combine the number of effects and the repeated observations.

For notational convenience, define $p = \frac{1}{2}(I+2c-3)$, $q = \frac{1}{2}[I(J-1)+a-2c]$, $p_m = \frac{1}{2}(I_m+2c-3)$, $q_m = \frac{1}{2}[I_m(J_m-1)+a-2c]$ and $W_m = \frac{J_m\sum_{i=1}^{I_m}(y_{i\cdot}-y_{\cdot\cdot})^2}{\sum_{i=1}^{I_m}\sum_{j=1}^{J_m}(y_{ij}-y_{i\cdot})^2}$.

Then, for the minimal training sample $y(l)$, the marginal densities of $Y(l)$ are obtained from (3.9) and (3.10) by replacing Y, W, I and J in (3.9) and (3.10) with $Y(l)$, W_m, I_m and J_m, respectively. Then the following collorary is obtained.

Collorary 3.1. For testing H_0 versus H_1 as above,

$$
\begin{aligned}
1/B_{10}^{N}(y(l)) &= \frac{J_m W_m^{p_m}\Gamma(\frac{I_m J_m - 1}{2})}{2^{a-2}\Gamma(\frac{a+I_m J_m-3}{2})}\left\{\sum_{i=1}^{I_m}\sum_{j=1}^{J_m}(y_{ij}-y_{i\cdot})^2\right\}^{\frac{a-2}{2}}\left(\frac{1-\rho_0}{1+(J_m-1)\rho_0}\right)^{\frac{I_m-1}{2}} \\
&\times \frac{1}{B_{q_m,p_m}(\frac{1}{W_m+1})(\frac{1-\rho_0}{1+(J_m-1)\rho_0}W_m+1)^{\frac{I_m J_m-1}{2}}}. \qquad (3.12)
\end{aligned}
$$

Therefore, B_{01}^{AI} and B_{01}^{GI} are directly computed from Lemma 3.1 and Collorary 3.1.

In order to compute $B_{01(d)}^{AI}$ and $B_{02(d)}^{GI}$, the following Lemma is needed.

Lemma 3.2. Under the same above hypothesis, $B_{01(d)}^{AI}$ and $B_{01(d)}^{GI}$ can be expressed as

$$
B_{01(d)}^{AI} = \pi_1(\rho_0|y_{(d)})E_{(d)}\left(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)}\right)\frac{1}{M}\sum_{m=1}^{M}\frac{1}{\pi_1(\rho_0|y_{(d)}(m))E_{(d)m}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})} \qquad (3.13)
$$

and

$$
B_{01(d)}^{GI} = \pi_1(\rho_0|y_{(d)})E_{(d)}\left(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)}\right)\left\{\prod_{m=1}^{M}\frac{1}{\pi_1(\rho_0|y_{(d)}(m))E_{(d)m}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})}\right\}^{\frac{1}{L}}, \qquad (3.14)
$$

where $E_{(d)}$ and $E_{(d)m}$ denote the expectations with respect to densities $\pi_1(\mu,\sigma^2|\rho_0,y_{(d)})$ and $\pi_1(\mu,\sigma^2|\rho_0,y_{(d)}(m))$, respectively.

Proof. Recall that $B_{01}(y(l)) = B_{01}^{N}B_{10}^{N}(y(l))$. By Verdinelli and Wasserman(1995),

$$
B_{01}^{N} = \pi_1(\rho_0|y)E\left(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)}\right) \text{ and } B_{10}^{N}(y(l)) = \left[\pi_1(\rho_0|y(l))E_l\left(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)}\right)\right]^{-1},
$$

where E and E_l denote the expectations with respect to densities $\pi_1(\mu,\sigma^2|\rho_0,y)$ and $\pi_1(\mu,\sigma^2|\rho_0,y(l))$, respectively, which completes the proof.

In particular, if $\pi_1(\mu,\sigma^2|\rho_0) = \pi_0(\mu,\sigma^2)$, then

$$
B_{01(d)}^{AI} = \frac{1}{M}\sum_{m=1}^{M}\frac{\pi_1(\rho_0|y_{(d)})}{\pi_1(\rho_0|y_{(d)}(m))} \qquad (3.15)
$$

and

$$
B_{01(d)}^{GI} = \pi_1(\rho_0|y_{(d)})\left\{\prod_{m=1}^{M}\frac{1}{\pi_1(\rho_0|y_{(d)}(m))}\right\}^{\frac{1}{M}}. \qquad (3.16)
$$

Finally we have to estimate $\pi_1(\rho \mid y_{(d)})$, $\pi_1(\rho \mid y_{(d)}(m))$, $E_{(d)}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})$ and $E_{(d)m}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})$ in (3.13) through (3.16). These four quantities are based on the full data excluding the observation d. Since we delete the data d from the balanced random effect model, in general we may consider the unbalanced model as follows;

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \text{ for } j = 1, \ldots, n_i \ ; i = 1, \ldots, I. \tag{3.17}$$

Let $N = \sum_{i=1}^{I} n_i$. Then under the prior $\pi_1(\rho, \mu, \sigma^2) \propto \sigma^{-a}(1 + (J-1)\rho)^{-c}(1-\rho)^{-d}$, the posterior joint density given all data excluding observation d is

$$\begin{aligned} \pi(\rho, \mu, \sigma^2|y_{(d)}) &\propto (\sigma^2)^{-\frac{N+a}{2}}(1+(J-1)\rho)^{-c}(1-\rho)^{\frac{I-2d}{2}} \prod_{i=1}^{I}(1+(n_i-1)\rho)^{-\frac{1}{2}} \\ &\quad \exp\{-\frac{1}{2\sigma^2}[\sum_{i=1}^{I} \frac{n_i(1-\rho)}{1+(n_i-1)\rho}(y_{i\cdot}-\mu)^2 + \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2]\} \\ &= l_{(d)}(\rho, \mu, \sigma^2)\pi_1(\rho, \mu, \sigma^2). \end{aligned} \tag{3.18}$$

Now, we can apply Gibbs sampler to compute $\pi_1(\rho_0|y_{(d)})$. The full conditional densities are given as follows:

$$\begin{aligned} [\mu|\rho, \sigma^2, y_{(d)}] &\propto exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{I}\frac{n_i(1-\rho)}{1+(n_i-1)\rho}(y_{i\cdot}-\mu)^2\} \\ &= N(\frac{\sum_{i=1}^{I}\frac{n_i y_{i\cdot}}{1+(n_i-1)\rho}}{\sum_{i=1}^{I}\frac{n_i}{1+(n_i-1)\rho}}, \frac{\sigma^2}{(1-\rho)\sum_{i=1}^{I}\frac{n_i}{1+(n_i-1)\rho}}), \end{aligned} \tag{3.19}$$

$$[\sigma^2|\mu, \rho, y_{(d)}] = IG(\frac{N+a-2}{2}, b), \tag{3.20}$$

and

$$\begin{aligned} [\rho|\mu, \sigma^2, y_{(d)}] &\propto (1+(J-1)\rho)^{-c}(1-\rho)^{\frac{I}{2}-d}\prod_{i=1}^{I}(1+(n_i-1)\rho)^{-\frac{1}{2}} \\ &\times \ exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{I}\frac{n_i(1-\rho)}{1+(n_i-1)\rho}(y_{i\cdot}-\mu)^2\}, \end{aligned} \tag{3.21}$$

where $b = 2[\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2+\sum_{i=1}^{I}\frac{n_i(1-\rho)}{1+(n_i-1)\rho}(y_{i\cdot}-\mu)^2]^{-1}$. Using Chen(1994) we propose the estimate of $\pi_1(\rho_0|y_{(d)})$ as follows:

$$\pi_1(\rho_0|y_{(d)}) \sim \frac{1}{G}\sum_{g=1}^{G} q(\rho^{(g)}|\mu^{(g)},\sigma^{2(g)})\frac{l_{(d)}(\rho_0,\mu^{(g)},\sigma^{2(g)})\pi_1(\rho_0,\mu^{(g)},\sigma^{2(g)})}{l_{(d)}(\rho^{(g)},\mu^{(g)},\sigma^{2(g)})\pi_1(\rho^{(g)},\mu^{(g)},\sigma^{2(g)})} \tag{3.22}$$

where $\{\rho^{(g)},\mu^{(g)},\sigma^{2(g)}\}$ is Gibbs output from the full conditional densities in (3.19) through (3.21) where $l_{(d)}(\rho,\mu,\sigma^2)$ and $\pi_1(\rho,\mu,\sigma^2)$ are given in (3.18). Choosing a good function of q can be quite difficult. In some cases a reasonable choice q is to use a normal density whose mean and variance are based on the sample mean and sample covariance of Gibbs output. Also we can estimate $\pi_1(\rho_0|y_{(d)}(m))$ by the similar method. Since the minimal training sample size of the unbalanced random effects model is exactly same as that of the balanced random effects model, we can put $I=2, J=2$ and $n_i=2$ in the equations (3.19), (3.20), and (3.21). For the notational convenience, we consider y_{11}, y_{12}, y_{21} and y_{22} as each minimal training sample. Then the full conditional distributions are given by

$$[\mu|\rho,\sigma^2,y_{(d)}(m)] = N(\frac{\sum_{i=1}^{2}\frac{2y_{i\cdot}}{1+\rho}}{\sum_{i=1}^{2}\frac{2}{1+\rho}}, \frac{\sigma^2}{(1-\rho)\sum_{i=1}^{2}\frac{2}{1+\rho}}), \tag{3.23}$$

$$[\sigma^2|\mu,\rho,y_{(d)}(m)] = IG(\frac{2+a}{2}, b), \tag{3.24}$$

and

$$\begin{aligned}[\rho|\mu,\sigma^2,y_{(d)}(m)] &\propto (1+\rho)^{-c-1}(1-\rho)^{c-1} \\ &\times exp\{-\frac{1}{\sigma^2}\sum_{i=1}^{2}\frac{(1-\rho)}{1+\rho}(y_{i\cdot}-\mu)^2\},\end{aligned} \tag{3.25}$$

where $b = 2[\sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij}-y_{i\cdot})^2+\sum_{i=1}^{2}\frac{2(1-\rho)}{1+\rho}(y_{i\cdot}-\mu)^2]^{-1}$. For simulating ρ from $[\rho|\mu,\sigma^2,y_{(d)}(m)]$ in (3.25), consider $[t|\mu,\sigma^2,y_{(d)}(m)] \propto t^{c-1}exp\{-\frac{\sum_{i=1}^{2}(y_{i\cdot}-\mu)^2}{\sigma^2}t\}$ with $t=(1-\rho)/(1+\rho)$ where $0<t<1$. Hence given μ,σ^2 and $y_{(d)}(m)$, the density of t is obtained as follows;

$$[t|\mu,\sigma^2,y_{(d)}(m)] = \frac{c}{1-\exp(-c)}t^{c-1}exp\{-\frac{\sum_{i=1}^{2}(y_{i\cdot}-\mu)^2}{\sigma^2}t\}. \tag{3.26}$$

Recall that from Theorem 2.2, we have

$$\pi(\mu, \sigma^2, \rho) \propto \sigma^{-a}(1 + (J-1)\rho)^{-c}(1-\rho)^{-d}$$

where $d = 2 - c$ and $(a, c) = (2, 1), (2, 1.5), (3, 1)$ or $(3, 1.5)$. Therefore, there are two choices of the values of c. That is, $c = 1$ or 1.5. For $c = 1$, since

$$[t|\mu, \sigma^2, y_{(d)}(m)] = \frac{1}{1 - \exp(-1)} exp\{-\frac{\sum_{i=1}^{2}(y_{i\cdot} - \mu)^2}{\sigma^2} t\},$$

we can use the inverse cumulative distribution function(CDF) method to simulate t from $[t|\mu, \sigma^2, y_{(d)}(m)]$. Finally, set $\rho = (1-t)/(1+t)$ where such ρ is considered as random variate from $[\rho|\mu, \sigma^2, y_{(d)}(m)]$. For $c = 1.5$, $[t|\mu, \sigma^2, y_{(d)}(m)] = Kt^{\frac{3}{2}-1}exp\{-kt\}$ where $K^{-1} = \int_0^1 t^{\frac{3}{2}-1}exp\{-kt\}dt$ and $k = \sum_{i=1}^{2}(y_{i\cdot} - \mu)^2/\sigma^2$. Let u be the random variate from uniform distribution on $(0, 1)$ and $G(t)$ and $F(t)$ the cumulative distribution functions of Gamma variable with parameters $\frac{3}{2}$ and k, and of randon variable t given $\mu, \sigma^2, y_{(d)}(m)$, in (3.26) respectively. That is, $G(t) = \int_0^t \frac{k^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} x^{\frac{3}{2}-1}\exp(-kx)dx$ and $F(t) = \int_0^t Kx^{\frac{3}{2}-1}exp\{-kx\}dx$. Thus, using inverse CDF method, $u = F(t) = K\Gamma(3/2)k^{-\frac{3}{2}}G(t)$ implies $t = G^{-1}(\frac{k^{\frac{3}{2}}u}{K\Gamma(\frac{3}{2})})$. Then such t is considered as a random variate from the conditional distribution $[t|\mu, \sigma^2, y_{(d)}(m)]$.

Let $\{\mu^{(g_1)}, \sigma^{2(g_1)}\}_{g_1=1}^{G_1}$ be the Gibbs output from $\pi(\mu, \sigma^2|\rho_0, y_{(d)})$. Then a Monte Carlo estimate of $E_{(d)}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})$ is

$$\hat{E}_{(d)} = \frac{1}{G_1}\sum_{g_1=1}^{G_1} \frac{\pi_0(\mu^{(g_1)}, \sigma^{2(g_1)})}{\pi_1(\mu^{(g_1)}, \sigma^{2(g_1)}|\rho_0)}. \tag{3.27}$$

Similar approximation is done for $E_{(d)m}(\frac{\pi_0(\mu,\sigma^2)}{\pi_1(\mu,\sigma^2|\rho_0)})$. Thus a Monte Carlo estimate of $E_{(d)m}$ can be expressed as

$$\hat{E}_{(d)m} = \frac{1}{G_2}\sum_{g_2=1}^{G_2} \frac{\pi_0(\mu^{(g_2)}, \sigma^{2(g_2)})}{\pi_1(\mu^{(g_2)}, \sigma^{2(g_2)}|\rho_0)}, \tag{3.28}$$

where $\{\mu^{(g_2)}, \sigma^{2(g_2)}\}_{g_2=1}^{G_2}$ is the Gibbs output from the posterior $\pi_1(\mu, \sigma^2|\rho_0, y_{(d)}(m))$.

4. Simulated Data

Now, a dataset (Table 4.1; Box and Tiao, 1973) is simulated as follows. The simulated dataset consists of six groups of five observations each. The errors

ϵ_{ij} were generated from Normal variates with mean 0 and variance $\sigma^2 = 4$, the effects α_i were generated from Normal variates with mean 0 and variance $\sigma_\alpha^2 = 2$, and the total effect μ was set to be equal to five.

Table 4.1. Generated Data I

Batch	1	2	3	4	5	6
Observations	7.298	5.220	0.110	2.212	0.282	1.722
	3.846	6.556	10.386	4.852	9.014	4.782
	2.434	0.608	13.434	7.092	4.458	8.106
	9.566	11.788	5.5104	9.288	9.446	0.758
	7.990	-0.892	8.166	4.980	7.198	3.758

Then we have $S_1^2 = 358.7014$ and $S_2^2 = 41.6816$ and so their corresponding mean squares are 14.95 and 8.33 in ANOVA table respectively and the estimates of σ^2 and σ_α^2 are $\hat{\sigma}^2 = 14.95$ and $\hat{\sigma}_\alpha^2 = -1.3219$ respectively. Thus the estimate of ρ is -0.13 which is negative but its true value of ρ is never negative. To overcome this problem, we apply our above mentioned Bayesian approach. To check the effect of observations, we will use the data in Table 4.1 except y_{13} which is 2.434. y_{13} is replaced by -3.434. That is, it is assumed that the observation y_{13} is generated from the model (1.1) with the same values of parameters with the total effect $\mu = 0$.

Figure 5.1 indicates the values of K_d^{AI} in (3.5) for testing $H_0 : \rho = 0.33$ against $H_1 : \rho \neq 0.33$, that is, $(a, c) = (2, 1)$ is chosen as prior of ρ in H_1. Note that all values of K_d^{AI} are positive except the value corresponding to the observation y_{13}. More carefully, the value corresponding to the observation y_{13} is very small. That is reasonable because y_{13} is far from all remaining data. It will be noticed that the data indeed support $H_0 : \rho = 0.33$.

$logB_{01}^{AI} = 1.0086$ is the reference line in Figure 4.1. Except y_{13}, the remaining values of K_{01}^{AI} are in the near neighborhood of reference point. Also, as Figure 4.1, there is a large effect on $B_{01}^{AI} = 10.02$ when we omit observation 3 (y_{13}) on the first column, $K_3^{AI} = -1.095$. We may regard the data as being $\rho = 0.33$ with one influential observation.

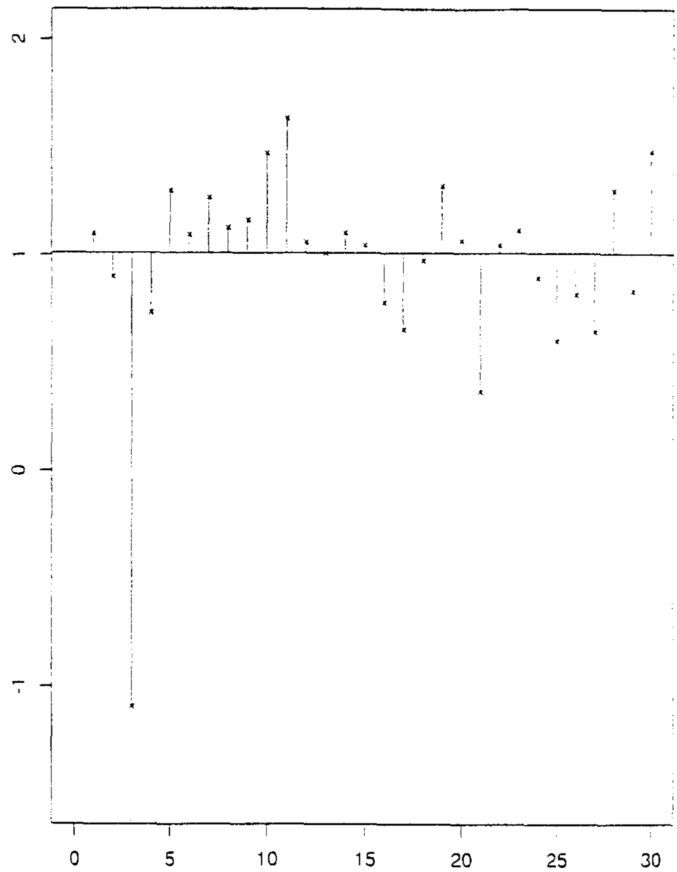

Figure 5.1. Values of K_d^{AI}

References

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

Berger, J. and J. Bernardo(1992). On the development of the reference prior method. In: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Eds., *Bayesian Statistics*, Vol 4. Oxford Univ. Press, London.35-60.

Berger, J.O. and Pericchi, L. R.(1996). The intrinsic Bayes factor for model selection and prediction, *Journal of the American statistical Association*, 91, 109-122.

Bernardo, J.(1979). Reference posterior distributions for Bayesian inference(with discussion). *J. Roy. Statis. Soc.*, B 41, 113-147.

Box, G. and G. Tiao(1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

Chaloner, K.(1987). A Bayesian approach to the estimation of variance component for the unbalanced one-way random model. *Technometrics*, 29, 323-337.

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

Chung, Y. and Dey, D.K. (1998). Bayesian approach to estimation of intraclass correlation using reference prior. *Communications in Statistics*, Vol 27,9, 2241-2255.

Dey, D.K.(1996). Discussion of intrinsic Bayes factor for linear model. *In: Bayesian Statistics 5*, (J. Bernardo, et. al, eds.), Oxford: Oxford University Press, 39-42.

Gelfand, A.E., Dey, D.K. and Chang, H.(1992). Model determination using predictive distributions with impletation via sampling-based methos. *In: Bayesian statistics 4*, (J. Bernardo, et. al, eds.), Oxford: Oxford University Press, 147-167.

Gelfand, A.E. and Smith, A. F. M.(1990) Sampling based approach to calculating marginal densities, *Journal of the American Statistical Association*, 85, 398-409.

Palmer, J.L. and L.D. Broemeling(1990) A comparison of Bayes and maximum likelihood estimation of the intraclass correlation coefficient. *Comm. Statist. Theory and Methods* 19, 953-975.

Pettit, L. I.(1990) The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society, Ser. B*, 52, 174-184.

Pettit, L. I. and Young, K. D. S. (1990) Measuring the effect of observations on Bayes factors. *Biometrika*, 77, 455-466.

Sun D. and Ye. K.(1995). The reference prior Bayesian analysis for normal mean products. *Journal of Americam Statistical Association*, 90, 589-597.

Ye, K.(1994). Bayesian reference prior analysis on the ration of variances for the balanced one-way random effect model. *Journal of statistical Planning Inference*, 41, 267-280.