Journal of the Korean
Statistical Society
Vol. 27, No. 1, 1998

A Bayes Criterion for Testing Homogeneity of Two Multivariate Normal Covariances

Hea-Jung Kim [1]

ABSTRACT

A Bayes criterion for testing the equality of covariance matrices of two multivariate normal distributions is proposed and studied. Development of the criterion involves calculation of Bayes factor using the imaginary sample method introduced by Spiegelhalter and Smith (1982). The criterion is designed to develop a Bayesian test criterion, so that it provides an alternative test criterion to those based upon asymptotic sampling theory (such as Box's M test criterion). For the constructed criterion, numerical studies demonstrate routine application and give comparisons with the traditional test criteria.

Key Words : Homogeneity of covariance matrices; Bayes test criterion; Bayes factor; Imaginary training sample method; Box's M test.

[1]Department of Statistics, Dongguk University, Seoul 100-715, Korea

1. INTRODUCTION

The comparison of two rival models on the basis of Bayes factor was pioneered by Jeffreys (1935) and has been used by many authors (see Kass and Raftery (1995) for a list of references). Suppose we have data D(say, training sample), assumed to have arisen under one of two alternative models M_1 and M_2 having probability densities $p(D|\theta_i, M_i)$, under $M_i, i = 1, 2$, where parameter vectors are unknown and are of dimension k_i. Then, in general, the data produce the Bayes factor for M_1 against M_2 defined by

$$B_{12} = \frac{p(D|M_1)}{p(D|M_2)} = \frac{\int p(D|\theta_1, M_1)\pi(\theta_1|M_1)d\theta_1}{\int p(D|\theta_2, M_2)\pi(\theta_2|M_2)d\theta_2}. \quad (1)$$

where the densities $p(D|M_i)$ are obtained by integrating over the parameter space, so that

$$p(D|M_i) = \int p(D|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i, \quad (2)$$

where $p(D|M_i)$ is called marginal likelihood or predictive density of D under M_i. The Bayes factor denotes the ratio of the posterior odds of M_i to its prior odds, regardless of the value of the prior odds. Thus B_{12} can be viewed as the weighted likelihood ratio of M_1 to M_2 and hence can be solely in terms of comparative support of the data for the two models(cf. Kass and Raftery, 1995).

Computing B_{12} in equation (1) requires specification of the priors, $\pi(\theta_i|M_i)$ $i = 1, 2$. In model comparison, most Bayesians today prefer to use noninformative priors that are typically improper. Unfortunately, for the most model selection problems, one can not use standard improper noninformative priors; such priors are defined only up to a constant multiple, and hence the Bayes factor B_{12} is itself a multiple of this arbitrary constant. Thus, it is typically not possible to utilize standard noninformative (or default) prior distributions for the model comparison. A common solution to this problem is to construct a default Bayes factor which uses part of the data as a sub-training sample to eliminate the constant. Formal developments of the idea can be found in work of Gelfand, Dey, and Chang (1992) and Berger and Pericchi (1996). However, as pointed by Berger and Pericchi (1996), the training sample approach is clearly impractical if the complete data set itself is rather small, or if the data are driven from a highly structured situation. Another solution to the arbitrary constant problem, which does not involve the above problems attached to the training sample approach, is the imaginary training sample method of Spiegelhalter and Smith (1982).

In Section 2, we shall derive a Bayesian criterion (Bayes factor) for testing the equality of two multivariate normal covariance matrices by means of the imaginary training sample method. Section 3 examines the performance of the suggested criterion and notes some merits over the usual test (Box's M test). Finally, Section 4 includes some concluding remarks.

2. TEST CRITERION

Suppose we have two multivariate normal populations Π_1 and Π_2 each specified by a model M_i, $i = 1, 2$, where M_i defines the distribution of each population distribution $\Pi_k \sim N_p(\mu_k, \Sigma_k)$, $k = 1, 2$, where parameters are unknown. Let our interest of model comparison be homogeneity (or heterogeneity) of the covariance matrices between two multivariate normal populations. Then the model specification becomes

$$M_1 : \Sigma_1 = \Sigma_2 = \Sigma \text{ versus } M_2 : \Sigma_1 \neq \Sigma_2. \tag{3}$$

Let $X_1(k), X_2(k), \ldots, X_{N_k}(k)$ denote independent p variate sample of size N_k from Π_k with distribution $N_k(\mu_k, \Sigma_k)$, $k = 1, 2$, and let denote the two independent samples as D. If we define $\bar{X}(k) = \sum_{j=1}^{N_k} X_j(k)/N_k$, and $V_k = \sum_{j=1}^{N_k}(X_j(k) - \bar{X}(k))(X_j(k) - \bar{X}(k))'$. Then the data D is to have arisen under M_2 according to probability density given by

$$P(D|\mu_1, \mu_2, \Sigma_1, \Sigma_2, M_2) = \prod_{k=1}^{2} (2\pi)^{-N_k p/2} |\Sigma_k|^{-N_k/2} \exp\{-1/2tr[\Sigma_k^{-1}\Omega_k]\}, \tag{4}$$

where $\Omega_k = V_k + N_k(\mu_k - \bar{X}(k))(\mu_k - \bar{X}(k))'$. Setting $\Sigma_1 = \Sigma_2 = \Sigma$ in equation (4), we get the joint probability density conditionally on M_1.

Since our interest focuses primarily on a statement concerning to relative probability that D comes from one or another of the model, and not about of making probability statement about where a parameter lies, we shall use a particular convenient prior density to reflect a noninformative information about the unknown parameters. In this paper, we shall be concerned with the case where both priors have Jeffreys' priors that can be written by

$$\pi(\mu_1, \mu_2, \Sigma|M_1) = c_1|\Sigma|^{-(p+1)/2} \tag{5}$$

and

$$\pi(\mu_1, \mu_2, \Sigma_1, \Sigma_2|M_2) = c_2 \prod_{k=1}^{2} |\Sigma_k|^{-(p+1)/2},$$

where $c_i,\ i=1,2$, are arbitrary constants.

Lemma 1. Under the improper priors (5), the marginal likelihoods (or predictive densities) conditional on M_1 and M_2, respectively, are

$$p(D|M_1) = c_1(2\pi)^{-(N-2)p/2}\Delta|\sum_{k=1}^{2} V_k|^{-(N-2)/2}\prod_{k=1}^{2} N_k^{-p/2}, \quad (6)$$

$$p(D|M_2) = c_2\prod_{k=1}^{2}\left((2\pi)^{-(N_k-1)p/2}N_k^{-p/2}\Delta_k|V_k|^{-(N_k-1)/2}\right), \quad (7)$$

where $\Delta = 2^{p(N-2)/2}\prod_{j=1}^{p}\Gamma_p\{(N-2)/2\}$, $N=\sum_{k=1}^{2}N_k$ and $\Delta_k = 2^{p(N_k-1)/2}$ $\prod_{j=1}^{p}\Gamma_p\{(N_k-1)/2\}$, where $\Gamma_p(\theta) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma\left(\theta-\frac{j-1}{2}\right)$ denotes the p-dimensional gamma function.

Proof. From the definition of the marginal density (2) and the likelihood (4),

$$\begin{aligned} p(D|M_1) &= \int p(D|\mu_1,\mu_2,\Sigma,M_1)\pi(\mu_1,\mu_2,\Sigma|M_1)\prod_{k=1}^{2}d\mu_k d\Sigma \quad (8)\\ &= c_1\int(2\pi)^{-Np/2}|\Sigma|^{-(N+p+1)/2}\exp\{-\frac{1}{2}tr\Sigma^{-1}[\sum_{k=1}^{2}V_k\\ &+ \sum_{k=1}^{2}N_k(\mu_k-\bar{X}(k))(\mu_k-\bar{X}(k))']\}\prod_{k=1}^{2}d\mu_k d\Sigma. \end{aligned}$$

We obtain the marginal likelihood $p(D|M_1)$ in the following way. Integrate (8) with respect to μ_k's using multivariate normal distribution. This gives

$$c_1(2\pi)^{-(N-2)p/2}\prod_{k=1}^{2}(N_k)^{-p/2}|\Sigma|^{-(N+p-1)/2}\exp\left\{-\frac{1}{2}tr[\Sigma^{-1}\sum_{k=1}^{2}V_k]\right\}, \quad (9)$$

and hence the desired marginal likelihood (6) can be found by integrating with respect to Σ, using the inverted Wishart normalizing constant. Similar integrations with respect to μ_k and $\Sigma_k,\ k=1,2$ for the expression below yield (7).

$$\begin{aligned} p(D|M_2) &= \int p(D|\mu_1,\mu_2,\Sigma_1,\Sigma_2,M_2)\pi(\mu_1,\mu_2,\Sigma_1,\Sigma_2|M_2)\prod_{k=1}^{2}d\mu_k\prod_{k=1}^{2}d\Sigma_k\\ &= c_2\int\prod_{k=1}^{2}(2\pi)^{-N_kp/2}|\Sigma_k|^{-(N_k+p+1)/2}\exp\{-\frac{1}{2}tr\Sigma_k^{-1}[V_k\\ &+ N_k(\mu_k-\bar{X}(k))(\mu_k-\bar{X}(k))']\}\prod_{k=1}^{2}d\mu_k\prod_{k=1}^{2}d\Sigma_k. \quad (10) \end{aligned}$$

These marginal likelihoods, (6) and (7), are clearly indeterminate, due to the presence of the undefined constants, c_1 and c_2. Therefore, Bayes factor for comparing M_1 with M_2 is itself a multiple of these arbitrary constants such that

$$B_{12} = \frac{P(D|M_1)}{P(D|M_2)} = \frac{c_1}{c_2} \frac{\Gamma_p\{(N-2)/2\} |\sum_{k=1}^2 V_k|^{-(N-2)/2}}{\prod_{k=1}^2 \Gamma_p\{(N_k-1)/2\} |V_k|^{-(N_k-1)/2}}. \tag{11}$$

A solution for this problem is to construct a default Bayes factor by means of imaginary training sample. This method, so called the imaginary training sample method, for assigning some value to c_1/c_2 has been proposed by Spiegelhalter and Smith (1982). The basic idea, a variation on a theme of Good (1947), is to image that a imaginary training sample data set is available which is defined as follows.

Definition 1. (Spiegelhalter and Smith 1982). A data set is called the imaginary training data set if it is available, which

(i) involves the smallest possible sample size permitting a comparison of M_1 and M_2;

(ii) provides maximum possible support for M_1, so that it may yield the Bayes factor $B_{12} \approx 1$.

Using the definition, we can obtain the imaginary training sample to eliminate the indeterminacy of the Bayes factor in (11). Lemma 1 and the condition (i) of Definition 1 require that a minimal size of the imaginary training sample is $p+1$ for each population $\Pi_k \sim N_p(\mu_k, \Sigma_k)$ defined by M_2, $k = 1, 2$, because we need at least $p+1$ observations in order to be able to estimate Σ_k and μ_k. Since M_1 is nested in M_2, we see that the imaginary training sample of size $p+1$ for each population is minimal sample size for the comparison of M_1 and M_2. With the imaginary training sample of total sample size $2(p+1)$, the Bayes factor (11) will be evaluated as

$$\frac{P(D|M_1)}{P(D|M_2)} = \frac{c_1}{c_2} \frac{\Gamma_p\{p\} |\sum_{k=1}^2 V_k|^{-p}}{(\Gamma_p\{p/2\})^2 \prod_{k=1}^2 |V_k|^{-p/2}}. \tag{12}$$

Suppose we denote Λ by

$$\frac{|\sum_{k=1}^2 V_k|^{-p}}{\prod_{k=1}^2 |V_k|^{-p/2}} 2^{p^2}.$$

Then it is easy to check that Λ is 1/2 times M test statistic for testing the null hypothesis that M_1 is true (cf. Rencher 1995). Perlman (1980) has shown that the test based on Λ is unbiased. This test statistic is suggested by Box (1949); an alternative of the likelihood ratio criterion for testing the null hypothesis that M_1 is true. Thus the Bayes factor (12) can be expressed in terms of Λ, so that

$$\frac{P(D|M_1)}{P(D|M_2)} = \frac{c_1}{c_2}\frac{\Gamma_p\{p\}}{(\Gamma_p\{p/2\})^2}2^{-p^2}\Lambda. \tag{13}$$

Since the statistic Λ takes a value $0 \leq \Lambda \leq 1$ (cf. Anderson 1984, p.419), the condition (ii) of Definition 1 leads to the value of Λ to be one achieving maximum support to M_1. Furthermore, if we set the Bayes factor (12) (arising from this imaginary experiment) equal to one, we can immediately deduce, from Definition 1, that

$$\frac{c_1}{c_2} = \frac{(\Gamma_p\{p/2\})^2}{\Gamma_p\{p\}}2^{p^2}. \tag{14}$$

Lemma 2. The imaginary training sample method yields a default Bayes factor, B^I_{12}, for comparing M_1 with M_2 given by

$$B^I_{12} = 2^{p^2}\frac{(\prod_{j=1}^p \Gamma\{(p-j+1)/2\})^2}{\prod_{j=1}^p \Gamma\{(2p-j+1)/2\}}\frac{\prod_{j=1}^p \Gamma\{(N-j-1)/2\}|\sum_{k=1}^2 V_k|^{-(N-2)/2}}{\prod_{k=1}^2(\prod_{j=1}^p \Gamma\{(N_k-j)/2\})|V_k|^{-(N_k-1)/2}}. \tag{15}$$

Proof. Substitute c_1/c_2 of (11) with the right hand side of the equality (14), and express the multivariate gamma functions in terms of corresponding gamma functions using $\Gamma_p(\theta) = \prod_{j=1}^p \Gamma\{\theta-(j-1)/2\}$, then we have the result.

This yields the following Bayes criterion for testing M_1.

Proposition 1. If the posterior probability of M_1, $P(M_1|D)$, is larger than 1/2, then we choose M_1 as a model best supported by the data, D. Otherwise, we choose M_2, where

$$P(M_1|D) = \frac{\pi_1 B^I_{12}}{\pi_2 + \pi_1 B^I_{12}}. \tag{16}$$

Here π_i denotes the prior probability of M_i, $i = 1, 2$.

Proof. Under the prior probabilities, Bayes theorem gives $P(M_1|D) = P(D|M_1)\pi_1/(\sum_{i=1}^k P(D|M_i)\pi_i)$. Since $B^I_{12} = P(D|M_1)/P(D|M_2)$, expressing $P(M_1|D)$ in terms of B^I_{12}, we have the result.

When $P(M_1|D) = 1/2$, we may randomly choose one of the two model. Note that when $\pi_1 = \pi_2 = 1/2$, $P(M_1|D) > 1/2$ is equivalent to $B_{12}^I > 1$. It is easy to check that, in the scalar case ($p = 1$), the Bayes factor B_{12}^I reduces to a function of the two sample F-test statistic

$$B_{12}^I = \frac{2\pi\Gamma\{n/2\}}{\Gamma\{n_1/2\}\Gamma\{n_2/2\}} \frac{n_1^{n_1/2} n_2^{n_2/2} F^{n_1/2}}{(n_1 F + n_2)^{n/2}}, \tag{17}$$

where $F = s_1^2/s_2^2$, s_1^2 and s_2^2 are the usual unbiased estimators of σ_1^2 and σ_2^2 (the two normal population variances) and $n_k = N_k - 1$, $k = 1, 2$, $n = n_1 + n_2$.

3. NUMERICAL STUDY

The interest of this section is the relationship between the p value (or observed significance level) and the suggested Bayesian measures of evidence that M_1 is true, B_{12}^I and $P(M_1|D)$, under the assumption that the prior probability of M_1 is $\pi_1 = 1/2$.

Together with the posterior probability $P(M_1|D)$ in Proposition 1, the Bayes factor is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another. Jeffreys (1961, Appendix B) suggested interpreting B_{12}^I by the following "order of magnitude" (see Kass and Raftery (1995), for the other "order of magnitude").

$B_{12}^I > 1$	evidence supports M_1;
$1 > B_{12}^I > 10^{-1/2}$	very slight evidence against M_1;
$10^{-1/2} > B_{12}^I > 10^{-1}$	moderate evidence against M_1;
$10^{-1} > B_{12}^I > 10^{-2}$	strong to very strong evidence against M_1;
$10^{-2} > B_{12}^I$	decisive evidence against M_1.

3.1. Comparison with Classical Tests

In this subsection, we shall summarize the p-values of standard tests and corresponding values of $P(M_1|D)$ and the default Bayes factor obtained from (15), (16) and (17). The standard tests considered here are the usual F-test and Box's M-test for the following Case 1 and Case 2, respectively.

Case 1: Univariate Normal Case; test for equality of covariances.

Table 1 gives B_{12}^I and $P(M_1|D)$ for various p-values of the standard F-statistic and $\nu = n_1 = n_2$ (B_{12}^I and $P(M_1|D)$ being chosen from (17) to correspond to the indicated p-value). As having been well known, the conflict between p-values and $P(M_1|D)$ (and B_{12}^I) is noted and is marked with asterisk. If $\nu = 60$ and $F = 1.67$, one can classically reject the modelM_1 at significance level $p = .05$, although $P(M_1|D) = .5787$ which would actually indicate that the evidence favors M_1. For practical examples and discussions of this conflict see Jeffreys (1961) or Berger and Sellke (1987).

Table 1. $P(M_1|D)$ **and** B_{12}^I **equivalents of** F **in (17), where** p**-value** $= P(F_{\nu,\nu} > F)$.

		ν					
p-value	values	5	10	20	30	60	120
.10	$P(M_1\|D)$	.3763	.4826	.5821	.6329	.7155	.7810
	B_{12}^I	.6033	.9327	1.3932	1.7243	2.5155	3.5663
	F	5.05	2.98	2.12	1.87	1.53	1.35
.05	$P(M_1\|D)$	.2449	.3394	.4379	.4938	.5787*	.6700*
	B_{12}^I	.3244	.5138	.7790	.9756	1.3738*	2.0312*
	F	7.15	3.72	2.46	2.07	1.67	1.43
.01	$P(M_1\|D)$	.0671	.1065	.1555	.1885	.2562	.3201
	B_{12}^I	.0719	.1192	.1842	.2322	.3445	.4710
	F	14.9	5.85	3.32	2.63	1.96	1.61

Case 2: Multivariate Normal Case; test for equality of covariance matrices.

Table 2 give p-values of the standard Box's M-test statistic corresponding to critical values of the Bayes factor (15) for various values of $N_1 = N_2$. It has shown by Box (1949) that a close approximation to the distribution of Λ under M_1 is given by

$$P(-a\Lambda \le t) = P(\chi^2(f) \le t) + b\{P(\chi^2(f+4) \le t) - P(\chi^2(f) \le t)\} + O((N-2)^{-3}), \quad (18)$$

where

$$a = 1 - \left(\sum_{i=1}^{2} \frac{1}{(N_i - 1)} - \frac{1}{N-2}\right) \frac{2p^2 + 3p - 1}{6(p+1)},$$

$$b = \frac{p(p+1)}{48a^2}\left\{(p-1)(p+2)\left(\sum_{i=1}^{2}\frac{1}{(N_1-1)^2}-\frac{1}{(N-m)^2}\right)-6(1-a)^2\right\},$$

$f=(p+1)p/2$, and $t=-a\times(\Lambda$ statistic value).

The correspondence between B^I_{12}, $P(M_1|D)$ and p-value of Box M-test statistic may be roughly summarized as follows. For moderate values of total sample size, critical values $10^{-1/2}$, 10^{-1}, and 10^{-2} of the Bayes factor correspond to less than p-value=.04 of the M-test statistic. In all cases, for large experiments evidence at a very high significance level is required for the Bayes factor to favor strongly the more complex hypothesis. The phenomenon is related to the "Lindley paradox", discussed in detail in Lee (1988). It is easy to check that Table 1 also reveals this phenomenon. In general, the Bayes factors provide an automatic assessment of significance, taking into account the size and structure of the experiment. For very small experiments, the use of the imaginary training sample approach has not been fully justified (cf. Spiegelhalter and Smith 1982), and hence we drop the case in this numerical study.

Table 2. p-values of Box M-Test for $N_1 = N_2 = N^*$.

		B^I_{12}						B^I_{12}			
p	N^*	1	$10^{-1/2}$	0.1	0.01	p	N^*	1	$10^{-1/2}$	0.1	0.01
2	10	.128	.052	.021	.003	4	10	.395	.268	.174	.066
	15	.063	.024	.009	.001		15	.087	.048	.026	.007
	20	.039	.024	.005	.000		20	.025	.013	.006	.001
	30	.019	.007	.002	.000		30	.004	.001	.000	.000
3	10	.201	.108	.056	.014	5	10	.609	.490	.284	.159
	15	.062	.029	.013	.002		15	.159	.103	.065	.023
	20	.026	.011	.005	.000		20	.034	.019	.001	.003
	30	.008	.003	.001	.000		30	.002	.001	.001	.000

3.2. Simulation Study

The goal of this subsection is to study the performance of the suggested test criterion in Proposition 1 by means of simulation study. Based upon the simulated data, the posterior probability $P(M_1|D)$ is compared with the corresponding p-value of Box's M test under the default prior values $\pi_1 = \pi_2 = 1/2$.

Our SAS/IML program generated sample of size N_k, say D_k, from each population Π_k, $k=1,2$, calculated $P(M_1|D)$ and p-value of Box's M test, where $D=(D_1,D_2)$. Each experiment consisted with 100 replications under the following distributional assumptions. The population conditional distributions, in this study, were defined as follows. Since a linear transformation leaves the criterion and Box's M test statistic invariant, there is no loss of generality in considering the case $\Pi_1 \sim N_p(0, I)$ and $\Pi_2 \sim N_p(\delta, I + \alpha\Delta)$.

Table 3. Simulation Result with $N_1 = N_2 = 30$ and m=1.

	p=2				p=4			
	d=2		d=4		d=2		d=4	
α	$Pr(M_1\|D)$	p-value	$Pr(M_1\|D)$	p-value	$Pr(M_1\|D)$	p-value	$Pr(M_1\|D)$	p-value
.00	.920	.568	.920	.568	.999	.511	.999	.511
	(.146)	(.286)	(.146)	(.286)	(.004)	(.236)	(.004)	(.236)
.01	.922	.586	.926	.538	.999	.509	.999	.473
	(.144)	(.286)	(.137)	(.291)	(.003)	(.262)	(.001)	(.264)
.05	.926	.553	.759	.168	.999	.489	.968	.137
	(.138)	(.289)	(.291)	(.199)	(.001)	(.026)	(.135)	(.203)
.1	.923	.508	.261	.009	.999	.442	.538	.005
	(.141)	(.293)	(.328)	(.024)	(.001)	(.266)	(.432)	(.019)
.2	.887	.370	.000	.000	.997	.311	.004	.000
	(.181)	(.284)	(.003)	(.000)	(.018)	(.269)	(.029)	(.000)
.3	.811	.222	.000	.000	.982	.187	.000	.000
	(.257)	(.234)	(.000)	(.000)	(.099)	(.234)	(.000)	(.000)
.4	.697	.121	.000	.000	.947	.097	.000	.000
	(.318)	(.162)	(.000)	(.000)	(.180)	(.168)	(.000)	(.000)
.5	.551	.056	.000	.000	.874	.043	.000	.000
	(.355)	(.097)	(.000)	(.000)	(.273)	(.102)	(.000)	(.000)
.6	.397	.023	.000	.000	.715	.016	.000	.000
	(.356)	(.051)	(.000)	(.000)	(.373)	(.050)	(.000)	(.000)
.7	.261	.009	.000	.000	.538	.005	.000	.000
	(.318)	(.024)	(.000)	(.000)	(.432)	(.019)	(.000)	(.000)
.8	.158	.003	.000	.000	.371	.001	.000	.000
	(.261)	(.010)	(.000)	(.000)	(.430)	(.005)	(.000)	(.000)
.9	.088	.001	.000	.000	.257	.000	.000	.000
	(.194)	(.003)	(.000)	(.000)	(.387)	(.001)	(.000)	(.000)
.95	.063	.000	.000	.000	.201	.000	.000	.000
	(.160)	(.002)	(.000)	(.000)	(.345)	(.000)	(.000)	(.000)
.99	.048	.000	.000	.000	.159	.000	.000	.000
	(.134)	(.001)	(.000)	(.000)	(.315)	(.000)	(.000)	(.000)
1.0	.044	.000	.000	.000	.150	.000	.000	.000
	(.127)	(.001)	(.000)	(.000)	(.308)	(.000)	(.000)	(.000)

Here, α ($0 \leq \alpha \leq 1$) is a mixing proportion designed to define the degree of homo/heterogeneity of the two covariance matrices associated with the distributions of Π_1 and Π_2. Thus, for given Δ, $\alpha = 0$ and $\alpha = 1$ denote the homogeneity and the most heterogeneity between the two covariance matrices, respectively. This canonical form is obtained from the transformation suggested by Dunn and Holloway(1967);

$$Y = A'\Sigma_1^{-1/2}(Z - \mu_1), \tag{19}$$

where A is an orthogonal matrix such that $A'\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}A = I + \alpha\Delta$, a diagonal matrix. Therefore, for given α, parameters involved in this study are confined to δ and Δ combinations:

$$\Delta \ = \ \mathrm{diag}(d, \ldots\ldots, d), \ \ \delta = (m(1 + d^{1/2}), 0, \ldots, 0)'.$$

In selecting values of δ, we introduced a parameter m as a measure of the degree of separation of two populations to ensure a particular distance between the populations for any choices of d and δ. m is defined as the Euclidean distance from the mean of Π_1 to the best linear discriminant hyperplane of the two populations(for details see Marks and Dunn(1974)). Thus parameters that are varied in this study include distance between the populations (m), covariance matrices (Δ), number of dimensions (p), and mixing proportion (α). Table 3, summarizing the 100 replications of the simulation with $N_1 = N_2 = 30$ and m=1, present the average of the posterior probabilities, $Pr(M_1 \mid D)$'s and p-values. Their standard deviations are given in the parentheses. The simulation results with other values of N_1, N_2 and m revealed the same implications as Table 3, and hence we eliminated them from the presentation.

It is noted from the table that the posterior probability of M_1 obtained from (16), is consistent with the probability value of Box's M test. Moreover, the table conveys the same informations as those in Table 2.

4. CONCLUDING REMARKS

We have suggested a development of the imaginary training sample method introduced by Spiegelhalter and Smith (1982). The development is pertaining to the comparison of homo/heteroscedasticity of the multivariate normal covariances. The appeal of the method is that it provides a simple method for evaluating a value for the arbitrary constant attached to the Bayes factor (using improper prior distributions) to coming at a Bayes test criterion. It is seen that the Bayes factor so obtained is expressed as a function of classical test criterion. Using numerical studies, we compare the suggested test criterion with p-value of a classical test criterion. The study notes that the criterion generally gives more conservative critical value than the classical test does. Thus, this study can be taken as an another illustration of the result by Berger and Sellke (1987).

Suggested Bayes factor for the comparison of homo/heteroscedasticity of several multivariate normal covariances can be applied for incorporating model uncertainty for many multivariate techniques such as discriminant analysis and MANOVA. A study pertaining to these applications are not unimportant. It is left as a future study of interest.

REFERENCES

(1) Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis,* John Wiley and Sons.

(2) 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.

(3) Berger, J. O. and Sellke, T. (1987). "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence," *Journal of the American Statistical Association,* 82, 112-122.

(4) Box, G. E. P. (1949). "A General Distribution Theory for a Class of Likelihood Criteria, *Biometrika,* 36, 317-346.

(5) Dunn, O. J. and Holloway, L. N. (1967). "Robustness of Hotelling's T^2," *Journal of the American Statistical Association,* 62, 124-136.

(6) Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distribution with implementations via sampling based methods (with discussion), *Bayesian Statistics IV,* Eds. by Bernardo, J. M. et al., Oxford University Press.

(7) Good, I. J. (1947). *Probability and the Weighing of Evidence,* Haffner, New York.

(8) Jeffreys, H. (1935). "Some Tests of Significance, Treated by the Theory of Probability," *Proceedings of the Cambridge Philosophy Society,* 31, 203-222.

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

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

(11) Lee, P. M. (1988). *Bayesian Statistics: An Introduction*, Wiley, New York.

(12) Marks, S. and Dunn, O. J. (1974). "Discriminant functions when covariance matrices are unequal," *Journal of the American Statistical Association,* 69, 555-559.

(13) Perlman, M. D. (1980). "Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations," *Annals of Statistics,* 8, 247-263.

(14) Press, S. J. (1982). *Applied Multivariate Analysis,* Krieger Publishing Co., Florida.

(15) Rencher, A. C. (1995). *Methods of Multivariate Analysis,* John Wiley and Sons Co., New York.

(16) Smith, A. F. M. and Spiegelhalter, D. J. (1980). "Bayes factors and choice criteria for linear models," *Journal of the Royal Statistical Society,* B, 42, 213-220.

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