2000 Proceedings of the Autumn Conference, Korean Statistical Society

A Bayesian Criterion for a Multiple test of Two Multivariate Normal Populations

Hea-Jung Kim [1] and Young Sook Son [2]

ABSTRACT

A Bayesian criterion is proposed for a multiple test of two independent multivariate normal populations. For a Bayesian test the fractional Bayes factor(FBF) of O'Hagan(1995) is used under the assumption of Jeffreys priors, noninformative improper proirs. In this test the FBF without the need of sampling minimal training samples is much simpler to use than the intrinsic Bayes facotr(IBF) of Berger and Pericchi(1996). Finally, a simulation study is performed to show the behaviors of the FBF.

Key Words : multivariate normal populations; Jeffreys prior; noninformative improper prior; fractional Bayes factor(FBF); intrinsic Bayes factor(IBF); posterior probability.

1. INTRODUCTION

Let Π_1 and Π_2 be two independent p-variate normal populations, where for $k = 1, 2$, $\Pi_k \sim N_p(\mu_k, \Sigma_k)$ with a $p \times 1$ unknown mean vector μ_k and a $p \times p$ unknown covariance matrix Σ_k. Suppose that we wish to do a multiple test composed of four hypotheses,

$$\begin{cases} H_0: & \mu_1 \neq \mu_2 \text{ and } \Sigma_1 \neq \Sigma_2 , \\ H_1: & \mu_1 = \mu_2 \text{ and } \Sigma_1 = \Sigma_2 , \\ H_2: & \mu_1 \neq \mu_2 \text{ and } \Sigma_1 = \Sigma_2 , \\ H_3: & \mu_1 = \mu_2 \text{ and } \Sigma_1 \neq \Sigma_2 . \end{cases} \tag{1.1}$$

In classical approach on the test problems of two independent normal populations there are various tests depending on whether two unknown means or covariances are equal or not. A multiple test is a merit of Bayesian test compared with the classical test of generally testing the null hypothesis versus the alternative hypothesis. A multiple test of (1.1) is to test two multivariate normal populations without any condition or previous information on the equality or the unequalness of two means or two covariances. Kim and Kim(2000) used the arithmetic intrinsic Bayes factor(IBF) of Berger and Pericchi(1996) for a multiple test of (1.1). But in case of using the IBF the number of minimal training samples to be sampled becomes tremendously big as the dimension of multivariate is bigger.

In this paper, we find a Bayesian criterion for a multiple test of two independent multivariate normal populations using the fractional Bayes factor(FBF) of O'Hagan(1995) without the need of sampling minimal training samples under the assumption of Jeffreys priors,

[1]Department of Statistics, Dongguk University, Seoul, 100-715, Korea.
[2]Department of Statistics, Chonnam National University, Kwangju, 500-757, Korea.

noninformative improper priors. In the next section, the FBF, the IBF, and the posterior probability of hypothesis are introduced. In section 3, we compute the FBF, and in section 4, a Bayesian criterion proposed in this paper is applied to some simulated data.

2. THE FRACTIONAL BAYES FACTOR

Suppose that we wish to test q hypotheses,

$$H_i : \boldsymbol{X} \sim f_i(X|\theta_i), \quad \theta_i \in \Theta_i,$$

for $i = 1, 2, \ldots, q$, with a random sample $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ of size n, where $f_i(X|\theta_i)$ is a probability density function, and θ_i and Θ_i are a parameter vector and a parameter space under the hypothesis H_i, respectively. We define the function of a random sample $\boldsymbol{X}$ and a constant b as follows, for $i = 1, 2, \ldots, q$, $j \neq i$,

$$B_{ji}(\boldsymbol{X}|b) = \frac{m_j(\boldsymbol{X}|b)}{m_i(\boldsymbol{X}|b)},$$

where

$$m_i(\boldsymbol{X}|b) = \int_{\Theta_i} \pi_i(\theta_i) L_i^b(\theta_i|\boldsymbol{X}) d\theta_i, \tag{2.1}$$

$\pi_i(\theta_i)$ is a prior distribution of θ_i, $L_i(\theta_i|\boldsymbol{X}) = \prod_{k=1}^n f_i(X_k|\theta_i)$ is a likelihood function, and b is a constant such that $0 < b \leq 1$.

The usual factor B_{ji} as a Bayesian tool to test the hypothesis H_j to the hypothesis H_i is defined by

$$B_{ji}(\boldsymbol{X}|b = 1) = \frac{m_j(\boldsymbol{X}|b = 1)}{m_i(\boldsymbol{X}|b = 1)},$$

where $m_i(\boldsymbol{X}|b = 1)$ is usually called a marginal or a predictive density of the hypothesis H_i.

The first step in a Bayesian inference is to choose the prior distributions of all the parameters in hypotheses or models. Default priors, most of which are typically noninformative improper, are objective priors that need not any subjective consideration. But the Bayes factor $B_{ji}(\boldsymbol{X}|b = 1)$ cannot be used because of arbitrary constants incorporated into the Bayes factor if priors are noninformative improper priors, $\pi_i^N(\theta_i)$ and $\pi_j^N(\theta_j)$, where throughout this paper the notation of superscript N implies the noninformative improper prior or its use, and have different dimensions in parameters. The fractional Bayes factor(FBF) of O'Hagan(1995) and the intrinsic Bayes factor(IBF) of Berger and Pericchi(1996, 1998) to overcome the problem due to the arbitrariness of noninformative improper priors are automatic and objective.

The idea of IBF is to use minimal training samples $\mathcal{X}_m = \{\boldsymbol{X}_m(l), l = 1, 2, \ldots, L\}$ to convert the improper prior to the proper posterior density. The minimal training sample implies the part of full sample with the minimal sample size to guarantee $0 < m_i^N(\boldsymbol{X}|b = 1) < \infty$ for all i. The IBF, $B_{ji}^I(l)$, given a minimal training sample, $\boldsymbol{X}_m(l)$, for some l is defined by

$$B_{ji}^I(l) = B_{ji}^N(\boldsymbol{X}|b = 1) \cdot B_{ij}^N(\boldsymbol{X}_m(l)|b = 1).$$

But, practically to prevent the IBF from depending on only one minimal training sample is used an arithmetic IBF(AIBF) as substituting an arithmetic mean of $B_{ij}^N(\boldsymbol{X}_m(l)|b = 1)$, $l = 1, 2, \ldots, L$, for $B_{ij}^N(\boldsymbol{X}_m(l)|b = 1)$ for some l, a geometric IBF(GIBF) as a geometric mean, or a median IBF(MIBF) as a median.

The FBF use a fraction b of each likelihood function to change noninformative improper priors into proper priors. The FBF is defined by

$$B_{ji}^F = B_{ji}^N(\boldsymbol{X}|b=1)\cdot B_{ij}^N(\boldsymbol{X}|b)\ . \tag{2.2}$$

Common choice of b is $b = m/n$, where m is the minimal sample size. Generally, for a Bayesian multiple test the posterior probabilities of hypotheses via the Bayes factors are useful. Under the assumption of prior probability p_i of the hypothesis H_i being true the posterior probability of H_i via the FBF is given by

$$P(H_i|\ \boldsymbol{X}) = \{\sum_{j=1}^{q}(p_j/p_i)\cdot B_{ji}^F\}^{-1},\quad i=1,2,\ldots,q. \tag{2.3}$$

3. COMPUTATION OF THE FBF

Let μ denote the common value of $\mu_1 = \mu_2$, and Σ the common value of $\Sigma_1 = \Sigma_2$. We use Jeffreys prior $\pi_i^N, i = 0,1,2,3,$ noninformative improper prior, for each hypothesis $H_i, i = 0,1,2,3,$ under the assumption of independence between a mean vector and covariance matrix as follows

$$\begin{cases} \pi_0^N(\mu_1,\mu_2,\Sigma_1,\Sigma_2) &= c_0\prod_{k=1}^{2}|\Sigma_k|^{-\frac{1}{2}(p+1)}, \quad \Sigma_1>0,\ \Sigma_2>0,\\ \pi_1^N(\mu,\Sigma) &= c_1|\Sigma|^{-\frac{1}{2}(p+1)}, \quad \Sigma>0,\\ \pi_2^N(\mu_1,\mu_2,\Sigma) &= c_2|\Sigma|^{-\frac{1}{2}(p+1)}, \quad \Sigma>0,\\ \pi_3^N(\mu,\Sigma_1,\Sigma_2) &= c_3\prod_{k=1}^{2}|\Sigma_k|^{-\frac{1}{2}(p+1)}, \quad \Sigma_1>0,\ \Sigma_2>0, \end{cases} \tag{3.1}$$

where c_i, $i = 0,1,2,3,$ is an undefined normalizing constant.

Let $\boldsymbol{X}_k = \{X_{k1}, X_{k2},\ldots,X_{kn_k}\}$ be a independent p-variate random sample of size n_k from Π_k with a distribution $N_p(\mu_k,\Sigma_k)$, $k = 1,2$. We use the following notation throughout this paper,

$$\begin{cases} n &= n_1+n_2\ ,\\ \boldsymbol{X} &= \{\boldsymbol{X}_1,\boldsymbol{X}_2\}\ ,\\ \bar{X}_k &= \dfrac{\sum_{j=1}^{n_k}X_{kj}}{n_k}\ ,\\ \bar{X} &= \dfrac{\sum_{k=1}^{2}n_k\bar{X}_k}{n}\ ,\\ V_k &= \sum_{j=1}^{n_k}(X_{kj}-\bar{X}_k)(X_{kj}-\bar{X}_k)'\ ,\\ V &= \sum_{k=1}^{2}\sum_{j=1}^{n_k}(X_{kj}-\bar{X})(X_{kj}-\bar{X})'. \end{cases}$$

The likelihood function $L_i(\cdot)$, $i = 0,1,2,3,$ under each hypothesis is given by

$$\begin{cases} L_0(\mu_1,\mu_2,\Sigma_1,\Sigma_2) &= \prod_{k=1}^{2}(2\pi)^{-\frac{n_kp}{2}}|\Sigma_k|^{-\frac{n_k}{2}}\exp\{-\frac{1}{2}tr[\Sigma_k^{-1}\Omega_k^*]\},\\ L_1(\mu,\Sigma) &= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\exp\{-\frac{1}{2}tr[\Sigma^{-1}\Omega]\},\\ L_2(\mu_1,\mu_2,\Sigma) &= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\prod_{k=1}^{2}\exp\{-\frac{1}{2}tr[\Sigma^{-1}\Omega_k^*]\},\\ L_3(\mu,\Sigma_1,\Sigma_2) &= \prod_{k=1}^{2}(2\pi)^{-\frac{n_kp}{2}}|\Sigma_k|^{-\frac{n_k}{2}}\exp\{-\frac{1}{2}tr[\Sigma_k^{-1}\Omega_k]\}, \end{cases} \tag{3.2}$$

where $\Omega = V + n(\mu - \bar{X})(\mu - \bar{X})'$, $\Omega_k = V_k + n_k(\mu - \bar{X}_k)(\mu - \bar{X}_k)'$, and $\Omega_k^* = V_k + n_k(\mu_k - \bar{X}_k)(\mu_k - \bar{X}_k)'$.

After using the kernel of multivariate normal density for the integration over a mean vector and the kernel of the inverted Wishart density for the integration over a covariance matrix the computation result of the function (2.2) with (3.1) and (3.2) is as follows

$$\begin{cases} m_0^N(\boldsymbol{X}|b) &= c_0(2\pi)^{-\frac{(bn-1)p}{2}} \prod_{k=1}^{2} \Delta_k(0)|V_k|^{-\frac{bn_k-1}{2}} n_k^{-\frac{p}{2}} , \\ m_1^N(\boldsymbol{X}|b) &= c_1(2\pi)^{-\frac{(bn-1)p}{2}} \Delta(1)|V|^{-\frac{bn-1}{2}} n^{-\frac{p}{2}} , \\ m_2^N(\boldsymbol{X}|b) &= c_2(2\pi)^{-\frac{(bn-2)p}{2}} \Delta(2)|V_1+V_2|^{-\frac{bn-2}{2}} \prod_{k=1}^{2} n_k^{-\frac{p}{2}} , \\ m_3^N(\boldsymbol{X}|b) &= c_3(2\pi)^{-\frac{bnp}{2}} \left\{ \prod_{k=1}^{2} \Delta_k(3)|V_k|^{-\frac{bn_k}{2}} \right\} \int_{-\infty}^{\infty} \prod_{k=1}^{2} \{1+W_k(\mu)\}^{-\frac{bn_k}{2}} \, d\mu , \end{cases} \tag{3.3}$$

where

$$\begin{cases} \Delta_k(0) &= 2^{\frac{p(bn_k-1)}{2}} b^{-\frac{bn_k p}{2}} \Gamma_p\left\{\frac{1}{2}(bn_k-1)\right\} , \\ \Delta(1) &= 2^{\frac{p(bn-1)}{2}} b^{-\frac{bnp}{2}} \Gamma_p\left\{\frac{1}{2}(bn-1)\right\} , \\ \Delta(2) &= 2^{\frac{p(bn-2)}{2}} b^{-\frac{bnp}{2}} \Gamma_p\left\{\frac{1}{2}(bn-2)\right\} , \\ \Delta_k(3) &= 2^{\frac{bn_k p}{2}} b^{-\frac{bn_k p}{2}} \Gamma_p\left\{\frac{1}{2}bn_k\right\} , \\ W_k(\mu) &= (\mu-\bar{X}_k)' S_k^{-1}(\mu-\bar{X}_k), \quad S_k = V_k/N_k , \\ \Gamma_p(t) &= \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma(t-\frac{i-1}{2}) . \end{cases}$$

The integration in $m_3^N(\boldsymbol{X}|b)$ of (3.3) is not analytically solved. This integration can be performed by the numerical integration procedure or Monte Carlo integration. In a simulation study of section 4, we estimate the integral function using the Monte Carlo integration method through the importance sampling. In $m_3^N(\boldsymbol{X}|b)$, we need to compute the following integral

$$M = \int_{-\infty}^{\infty} g(\mu) \, d\mu = \int_{-\infty}^{\infty} \frac{g(\mu)}{f_\mu(\mu)} \cdot f_\mu(\mu) \, d\mu = E_{f_\mu(\mu)}[\frac{g(\mu)}{f_\mu(\mu)}] ,$$

where $g(\mu) = \prod_{k=1}^{2} \{1+W_k(\mu)\}^{-\frac{bn_k}{2}}$, and $f_\mu(\mu)$ is an importance sampling density function. The Monte Carlo estimate of M is $\hat{M} = \sum_{j=1}^{G} \frac{g(\mu_j)}{f_\mu(\mu_j)}$, where μ_j, $j = 1, 2, \ldots, G$, is generated from the importance sampling density $f_\mu(\mu)$. It is well known that $Var(\hat{M})$ is small when $f_\mu(\mu) \propto |g(\mu)|$. The function $g(\mu)$ can be rewritten as

$$\begin{aligned} g(\mu) &= \prod_{k=1}^{2} \exp\left\{-\tfrac{1}{2}bn_k \cdot \ln\left(1+W_k(\mu)\right)\right\} \\ &\propto \exp\left\{-\tfrac{1}{2}(bn_1W_1(\mu)+bn_2W_2(\mu))\right\} \\ &\propto \exp\left\{-\tfrac{1}{2}(\mu-\mu_0)' J(\mu-\mu_0)\right\} , \end{aligned}$$

where the first proportional term is obtained using only the first term of Taylor series on $W_k(\mu) = 0$ of $\ln\{1+W_k(\mu)\}$. Thus the importance sampling density function of μ is $N_p(\mu_0, J^{-1})$, where $\mu_0 = K\bar{X}_1 + (I_p - K)\bar{X}_2$, $K = bn_1(S_1 J)^{-1}$, $J = b(n_1 S_1^{-1} + n_2 S_2^{-1})$, and I_p is a $p \times p$ identity matrix throughout this paper.

The size m of a minimal training sample equals to the condition that the marginal density $m_0^N(\boldsymbol{X}|b=1)$ of encompassing hypothesis H_0 is to be finite. If $n_k \le p$, then

$rank(V_k) \leq n_k - 1 \leq p - 1$. But $|V_k| = 0$, since a matrix V_k is a $p \times p$ matrix. Hence $n_k \geq p + 1$ for $k = 1, 2$. Thus the size of a minimal training sample is $m = 2(p+1)$. The conventional selection of a fraction, b, of likelihood function in the computation of FBF is $b = m/n = 2(p+1)/n$. For $k = 1, 2$ the sample size n_k must be restricted to $n_k \geq [p/b] + 1$, where $[\cdot]$ is a Gauss symbol, in order that the arguments of gamma functions in $m_i^N(\boldsymbol{X}|b)$, $i = 0, 1, 2, 3$, are to be positive.

Finally, the computation of the FBF, and the posterior probabilities of hypotheses via the FBF's are straightforward from (2.2) and (2.3).

4. A SIMULATION STUDY

We directly follow a simulation study of Kim and Kim(2000) originally based on the simulation scheme in Marks and Dunn(1974). All the experiments are performed for two independent p-variate ($p = 2, 4$) normal samples with sample size $n_1 = n_2 = 30$, 200 replications, and the importance samples of size 500. Let 0_p and 1_p be p-variate row vectors of zeroes and ones, respectively. Now we set $\mu_1 = 0$, $\mu_2 = [\tau(1+\sqrt{\lambda}) \ , 0_{p-1}]'$, where τ is called the measure of degree of separation of the two populations, $\Sigma_1 = I_p$, and $\Sigma_2 = \Lambda$, where Λ is a diagonal matrix with a vector, $[\lambda \cdot 1_{p/2}, 1_{p/2}]'$ of diagonal elements. Data with different choices, $\tau = 0, 2$ and $\lambda = 1, 4, 8$ of τ and λ are generated.

Table 4.1 shows the results of the averages and the standard deviations in parentheses of posterior probabilities for each hypothesis based on 200 replications. Also, though we prepared frequency plots on 200 replications of posterior probabilities for each hypothesis, we don't present here because of the limit of space.

Table 4.1 : The averages and the standard deviations in parentheses of posterior probabilities on 200 replications.

p	τ	λ	$P(M_0\|X)$	$P(M_1\|X)$	$P(M_2\|X)$	$P(M_3\|X)$
2	0	1	0.0044	0.7872	0.1256	0.0829
			(0.0084)	(0.1448)	(0.1061)	(0.1141)
		4	0.0425	0.1355	0.0272	0.7948
			(0.0453)	(0.2085)	(0.0894)	(0.2398)
		8	0.0721	0.0111	0.0012	0.9155
			(0.1213)	(0.0626)	(0.0070)	(0.1353)
	2	1	0.0411	0.0000	0.9589	0.0000
			(0.0689)	(0.0000)	(0.0689)	(0.0000)
		4	0.7425	0.0000	0.2575	0.0000
			(0.3064)	(0.0000)	(0.3064)	(0.0000)
		8	0.9867	0.0000	0.0133	0.0000
			(0.0535)	(0.0000)	(0.0535)	(0.0000)
4	0	1	0.0001	0.8685	0.1126	0.0188
			(0.0007)	(0.1999)	(0.1824)	(0.0997)
		4	0.0141	0.1691	0.0276	0.7891
			(0.0627)	(0.2574)	(0.1060)	(0.2963)
		8	0.0271	0.0005	0.0003	0.9721
			(0.0942)	(0.0035)	(0.0033)	(0.0953)
	2	1	0.0021	0.0000	0.9979	0.0000
			(0.0158)	(0.0000)	(0.0158)	(0.0000)
		4	0.6470	0.0000	0.3530	0.0000
			(0.3788)	(0.0000)	(0.3788)	(0.0000)
		8	0.9931	0.0000	0.0069	0.0000
			(0.0572)	(0.0000)	(0.0572)	(0.0000)

5. CONCLUDING REMARKS

We have proposed a Bayesian criterion for a multiple test of two independent multivariate normal populations. The test is performed by comparing with posterior probabilities of hypotheses via the FBF's under the assumption of Jefferys priors, noninformative improper priors. This multiple test doesn't require the prior knowledge or test on the equality or the unequalness of two means or two covariances, while the classical test requires that. Also, a Bayesian multiple test suggested in this paper can be flexibly applied to the classical tests of two independent multivariate normal populations. For example, the test of $H_0^{'} : \Sigma_1 = \Sigma_2$ versus $H_1^{'} : \Sigma_1 \neq \Sigma_2$ is to reject $H_0^{'}$ if $P(H_0|\boldsymbol{X}) + p(H_3|\boldsymbol{X}) > 0.5$. Then the Beherens-Fisher problem, the test of H_3 versus H_0, can be solved by comparing $P(H_0|\boldsymbol{X})$ with $P(H_3|\boldsymbol{X})$.

Concerning with the use of the IBF, the number of minimal training samples possible over the full sample is $L = \binom{n_1}{p+1} \cdot \binom{n_2}{p+1}$. For a example of $n_1 = n_2 = 30$, $L = 189,225$ when $p = 1$, $L = 16,483,600$ when $p = 2$, and $L = 20,308,000,000$ when $p = 4$. Also, importance sampling for each minimal training sample must be performed. So, the computation of posterior probabilities of hypotheses via the AIBF, GIBF, or MIBF which additionally need times for sorting is a job requiring tremendous computation times. While the FBF is very simple to use without the need of sampling minimal training samples. Also, wee can see that the results in this paper via the FBF confirm to our theoretical expectation for the test.

REFERENCES

Anderson, T. W. (1971), *An Introduction to Multivariate Statistical Analysis*, second edition, John Wiley and Sons, Inc.

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

Berger, J. O. and Pericchi, L. R.(1998), Accurate and Stable Bayesian Model Selection: The Median Intrinsic Bayes Factor, *Sankhya,* B, Vol.60, 1-18.

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

Kim, H. J. and Kim, S. W.(2000), Two-Smple Bayesian Test using Intrinsic Bayes Factors for Multivariate Normal Observations,*Communications in Statistics : Simulation and Computation.*

Marks, S. & Dunn, O.J.(1974) Discriminant funtions when covariance matrices are unequal, *Journal of the American Statistical Association,*, 69, pp. 555-559.

O'Hagan, A.(1995), Fractional Bayes Factors for Model Comparsion, *Journal of the Royal Statistical Socirty,* B, Vol.57, No.1, 99-138.