Journal of the Korean Statistical Society (2003), **32**: 4, pp 411–423

A BAYESIAN METHOD FOR FINDING MINIMUM GENERALIZED VARIANCE AMONG K MULTIVARIATE NORMAL POPULATIONS

HEA-JUNG KIM[1]

ABSTRACT

In this paper we develop a method for calculating a probability that a particular generalized variance is the smallest of all the K multivariate normal generalized variances. The method gives a way of comparing K multivariate populations in terms of their dispersion or spread, because the generalized variance is a scalar measure of the overall multivariate scatter. Fully parametric frequentist approach for the probability is intractable and thus a Bayesian method is pursued using a variant of weighted Monte Carlo (WMC) sampling based approach. Necessary theory involved in the method and computation is provided.

AMS 2000 subject classifications. Primary 62H12; Secondary 62F15.
Keywords. Multivarite normal population, generalized variances, vague prior, posterior probability, weighted Monte Carlo approach.

1. INTRODUCTION

The covariance matrix contains the variances of the variables and the covariances between pairs of variables and is a multifaceted picture of the overall variation in a multivariate normal population. Sometimes it is desirable to have a single numerical value for the overall multivariate scatter. One such measure is the generalized variance, defined as the determinant of the covariance matrix (see Press, 1982 and Rencher, 2002 for other measures).

In applications where variability of the multivariate population is of great practical importance, the generalized variance can be used to rank distinct groups and populations in order of their dispersion or spread. For example, a certain

Received December 2002; accepted August 2003.
[1]Department of Statistics, Dongguk University, Seoul 100-715, Korea (e-mail : kim3hj@dongguk.edu)

product, such as semiconductor, produced by a number of companies is characterized by a vector of p measurements. Although the same product is produced on the average, the companies can be distinguished on the basis of their associated generalized variances. Thus, the buyer can rank the supplying companies by ranking the normal populations using the generalized variance in an effort to minimize his chances of receiving a product that is sometimes of unacceptably poor quality. The usage of the generalized variance has been widely accepted by statisticians (Grizzle and Allen, 1969; Press, 1982; Rencher, 2002). However, due to complex distribution involved in inferencing the generalized variance, the analysis of it is yet to be seen in applied settings.

This paper, therefore, considers how inferences might be made about a possible differences in the variability of K multivariate normal populations. Particularly, our aim of this paper is to find minimum generalized variance among K multivariate normal populations. In comparing generalized variances of K multivariate normal populations, $N_p(\boldsymbol{\theta}_1, \boldsymbol{\Sigma}_1)$, ..., $N_p(\boldsymbol{\theta}_K, \boldsymbol{\Sigma}_K)$, there are many measures that could be used besides a linear contrasts $\log(|\boldsymbol{\Sigma}_j|) - \log(|\boldsymbol{\Sigma}_k|)$, for example, the ratios $|\boldsymbol{\Sigma}_j|/|\boldsymbol{\Sigma}_k|$, $|\boldsymbol{\Sigma}_j|^{1/2}/|\boldsymbol{\Sigma}_k|^{1/2}$ and other linear contrasts, $j, k = 1, \ldots, K$; $j \neq k$. Unlike the other measures, the ratios $|\boldsymbol{\Sigma}_j|/|\boldsymbol{\Sigma}_k|$, however, are easy to handle distributionally and computationally. Therefore, we focus on a Bayesian method for finding the minimum generalized variance in terms of the ratios. This is done by a sequence of transformations under joint posterior distribution of $\boldsymbol{\Lambda}_k$'s ($=\boldsymbol{\Sigma}_k^{-1}$'s), precision matrices, to put our problem into calculation of posterior probability for an event of the ratios. These transformations will be described in Section 3 based on the distributional results in Section 2. After the transformation, we are driven to do multidimensional integrations to evaluate the probability that can't be calculated analytical and numerical evaluation. As an alternative solution, we develop a weighted Monte Carlo (WMC) method to compute the probability. Section 4 gives complete description of the Monte Carlo algorithm, and Section 5 provides an illustrative example. Some concluding remarks concerning utility of the proposed method will be presented in Section 6.

2. PRELIMINARIES

In multivariate normal theory, Bayesian posterior distribution of a precision matrix comes out as an Wishart distribution. So that many Bayesian inferences for the multivariate normal covariance matrices (inverse of the precision matrices)

involve a function of Wishart matrices. The same case applies to our problem. To develop necessary theory involved in the problem of our interest, we shall need to make use of the following lemmas of some functions of Wishart matrices.

2.1. Functions of Wishart matrices

Suppose $\mathbf{S}_1, \ldots, \mathbf{S}_K$ are independently distributed as Wishart distributions with scale parameter $\mathbf{I}_p$ and degrees of freedom n_k. Each density function of $\mathbf{S}_k$ is

$$p(\mathbf{S}_k) = \left\{2^{n_k p/2}\pi^{p(p-1)/4}\Gamma_p\left(\frac{n_k}{2}\right)\right\}^{-1}|\mathbf{S}_k|^{(n_k-p-1)/2}\text{etr}\left\{-\frac{\mathbf{S}_k}{2}\right\}, \quad \mathbf{S}_k > 0$$

and we use the notation $\mathbf{S}_k \sim iid\, W_p(\mathbf{I}_p, n_k)$, $k = 1, \ldots, K$, where $\text{etr}\{\mathbf{C}\} = \exp(\text{tr}(\mathbf{C}))$ and

$$\Gamma_p(c) = \pi^{p(p-1)/2}\prod_{j=1}^{p}\Gamma\left(c - \frac{(j-1)}{2}\right),$$

the p-dimensional gamma function. For $\mathbf{S}_k \sim iid\ W_p(\mathbf{I}_p, n_k)$, let $\sum_{k=1}^{i+1}\mathbf{S}_k = \mathbf{T}_i'\mathbf{T}_i$ where $\mathbf{T}_i$ is an upper-triangular $p \times p$ matrix with positive diagonal matrix. If we make the transformations

$$\mathbf{R}_i = \mathbf{T}_i'^{-1}\left(\sum_{k=1}^{i}\mathbf{S}_k\right)\mathbf{T}_i^{-1} \tag{2.1}$$

in $\mathbf{S}_k$'s, $i = 1, \ldots, K-1$, we obtain the following lemmas.

LEMMA 2.1. *$\mathbf{R}_i$, $i = 1, \ldots, K-1$, are independently distributed as a multivariate Beta I distribution, $B_I(p; \sum_{k=1}^{i} n_k/2, n_{i+1}/2)$, with density given by*

$$f(\mathbf{R}_i) = \frac{|\mathbf{R}_i|^{(\sum_{k=1}^{i} n_k - p - 1)/2}|\mathbf{I}_p - \mathbf{R}_i|^{(n_{i+1}-p-1)/2}}{\beta_p(\sum_{k=1}^{i} n_k/2, n_{i+1}/2)}, \quad 0 < \mathbf{R}_i < \mathbf{I}_p, \tag{2.2}$$

where $\beta_p(a,b) = \Gamma_p(a)\Gamma_p(b)/\Gamma_p(a+b)$. Here $0 < \mathbf{R}_i < \mathbf{I}_p$ denotes that $\mathbf{R}_i$ and $\mathbf{I}_p - \mathbf{R}_i$ are symmetric positive definite matrices.

PROOF. Theorem 3.3.1 of Muirhead (1982) says that $\mathbf{R}_1 \sim B_I(p; n_1/2, n_2/2)$, and is independent of $\mathbf{T}_1$ and $\mathbf{S}_3$. So that $\mathbf{R}_1$ is independent of $\mathbf{R}_2$, a function of $\mathbf{T}_1$ and $\mathbf{S}_3$. Now $\mathbf{T}_1$ and $\mathbf{S}_3$ are independent and distributed as $W_p(\mathbf{I}_p, n_1 + n_2)$ and $W_p(\mathbf{I}_p, n_3)$. Thus, from the theorem, $\mathbf{R}_2 \sim B_I(p; (n_1+n_2)/2, n_3/2)$ and $\mathbf{R}_2$ is independent of $\mathbf{T}_2$ and $\mathbf{S}_4$. This implies that $\mathbf{R}_2$ independent of $\mathbf{R}_3$. Similar argument applies to prove the independence of all the $\mathbf{R}_i$'s. □

LEMMA 2.2 (Muirhead, 1982). *If* $\mathbf{R}_i \sim B_I(p; \sum_{k=1}^{i} n_k/2, n_{i+1}/2)$, *the joint density of the latent roots* $\lambda_1^i, \ldots, \lambda_p^i$ *of* $\mathbf{R}_i$ *is*

$$\pi^{p^2/2}\left\{\Gamma_p\left(\frac{p}{2}\right)\beta_p\left(\sum_{k=1}^{i}\frac{n_k}{2}, \frac{n_{i+1}}{2}\right)\right\}^{-1} \times \prod_{\ell}^{p}\left\{\left(\lambda_\ell^i\right)^{(\sum_{k=1}^{i} n_k - p - 1)/2}\left(1-\lambda_\ell^i\right)^{(n_{i+1}-p-1)/2}\right\}\prod_{u<v}^{p}\left(\lambda_u^i - \lambda_v^i\right),$$

where $1 > \lambda_1^i > \cdots > \lambda_p^i > 0$.

Above distributions lead to multivariate analogs of the Beta distribution. Some of these distributions arise naturally in various multivariate problems, *e.g.*, multivariate analysis of variance tests, sphericity test and multivariate slippage problems (see Box and Tiao, 1992; Huzurbazar and Butler, 1998).

3. BAYESIAN METHOD

3.1. The posterior distribution

Suppose $\mathbf{X}_1(k), \ldots, \mathbf{X}_{N_k}(k)$ are independent p-variate observations from $N_p(\boldsymbol{\theta}_k, \boldsymbol{\Lambda}_k^{-1})$, $k = 1, \ldots, K$, where $\boldsymbol{\Lambda}_k = \boldsymbol{\Sigma}_k^{-1}$ is the precision matrix. Let

$$\overline{\mathbf{X}}(k) = \sum_{j=1}^{N_k}\frac{\mathbf{X}_j(k)}{N_k} \text{ and } \mathbf{V}_k = \sum_{j=1}^{N_k}\left\{\mathbf{X}_j(k) - \overline{\mathbf{X}}(k)\right\}\left\{\mathbf{X}_j(k) - \overline{\mathbf{X}}(k)\right\}'.$$

Then the joint *pdf* of $\overline{\mathbf{X}}(k)$'s and $\mathbf{V}_k$'s is proportional to

$$\prod_{k=1}^{K}|\mathbf{V}_k|^{(N_k-p-2)/2}|\boldsymbol{\Lambda}_k|^{N_k/2}\text{etr}\left\{-\frac{1}{2}\boldsymbol{\Lambda}_k\left[\mathbf{V}_k + N_k\left\{\overline{\mathbf{X}}(k) - \boldsymbol{\theta}_k\right\}\left\{\overline{\mathbf{X}}(k) - \boldsymbol{\theta}_k\right\}'\right]\right\}.$$

To assure very little information is contributed to the analysis by a subjective prior density, we assume diffuse prior

$$p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_K) \propto \prod_{k=1}^{K}|\boldsymbol{\Lambda}_k|^{-(p+1)/2}.$$

The joint posterior density of the parameters is proportional to

$$\prod_{k=1}^{K}|\boldsymbol{\Lambda}_k|^{(N_k-p-1)/2}\text{etr}\left\{-\frac{1}{2}\boldsymbol{\Lambda}_k\left[\mathbf{V}_k + N_k\left\{\overline{\mathbf{X}}(k) - \boldsymbol{\theta}_k\right\}\left\{\overline{\mathbf{X}}(k) - \boldsymbol{\theta}_k\right\}'\right]\right\}. \quad (3.1)$$

Integrating (3.1) with respect $\boldsymbol{\theta}_k$'s, we have the marginal posterior distribution of $\boldsymbol{\Lambda}_k$:

$$\boldsymbol{\Lambda}_k \Big| \left(\overline{\mathbf{X}}(k), \mathbf{V}_k\right) \sim W_p\left(\mathbf{V}_k^{-1}, n_k\right), \quad n_k \geq p, \tag{3.2}$$

a Wishart distribution with scale parameter $\mathbf{V}_k^{-1}$ and degrees of freedom $n_k = N_k - 1$, $k = 1, \ldots, K$.

THEOREM 3.1. *Let* $\mathbf{S}_k = \mathbf{V}_k^{1/2} \boldsymbol{\Lambda}_k \mathbf{V}_k^{1/2}$, $k = 1, \ldots, K$. *Then* $\mathbf{R}_i$*'s defined by (2.1) are independent with* $\mathbf{R}_i \sim B_I(p; \sum_{k=1}^{i} n_k/2, n_{i+1}/2)$, $i = 1, \ldots, K-1$. *For each* i, *the joint posterior distribution of eigen values,* $\lambda_1^i, \ldots, \lambda_p^i$ *of* $\mathbf{R}_i$ *is*

$$\begin{aligned} p(\boldsymbol{\lambda}^i | Data) = c_i \prod_{\ell}^{p} \left(\lambda_\ell^i\right)^{(\sum_{k=1}^{i} n_k - p - 1)/2} \left(1 - \lambda_\ell^i\right)^{(n_{i+1} - p - 1)/2} \\ \times \prod_{u<v}^{p} \left(\lambda_u^i - \lambda_v^i\right) I(\boldsymbol{\lambda}^i), \end{aligned} \tag{3.3}$$

where $\boldsymbol{\lambda}^i = (\lambda_1^i, \ldots, \lambda_p^i)'$, $I(\boldsymbol{\lambda}^i) = I(1 > \lambda_1^i > \cdots > \lambda_p^i > 0)$, *an indicator function,* $c_i = \pi^{p^2/2} / \{\Gamma_p(p/2) B_p(\sum_{k=1}^{i} n_k/2, n_{i+1}/2)\}$, *and* $n_i = N_i - 1$.

PROOF. Note from (3.2) that $\mathbf{V}_k^{1/2} \boldsymbol{\Lambda}_k \mathbf{V}_k^{1/2} \sim W_p(\mathbf{I}_p, N_k - 1)$ are independent for $k = 1, \ldots, K$. Applying the distribution of $\mathbf{V}_k^{1/2} \boldsymbol{\Lambda}_k \mathbf{V}_k^{1/2}$ to Lemma 2.1 and Lemma 2.2, we have the result. □

COROLLARY 3.1. *For fixed* j, $j = 1, \ldots, K$, *the posterior expectation of a quantity of form* $h(\cap_{k=1, k\neq j}^{K} |\boldsymbol{\Sigma}_k| / |\boldsymbol{\Sigma}_j|)$ *is obtained from*

$$\int_{I(\boldsymbol{\lambda}^{K-1})} \cdots \int_{I(\boldsymbol{\lambda}^1)} \eta(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}) \prod_{i=1}^{K-1} p(\boldsymbol{\lambda}^i | Data)\, \partial\boldsymbol{\lambda}^1 \cdots \partial\boldsymbol{\lambda}^{K-1}, \tag{3.4}$$

where $h(\cap_{k=1, k\neq j}^{K} a_k)$ *denotes* $h(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_K)$ *and*

$$\eta(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}) = h\left(\bigcap_{k=1, k\neq j}^{K} \frac{|\mathbf{V}_k|}{|\mathbf{V}_j|} \frac{\prod_{u=1}^{p} \left(1 - \lambda_u^{j-1}\right) \prod_{i=1}^{k-1} \left(\prod_{u=1}^{p} \lambda_u^i\right)}{\prod_{u=1}^{p} \left(1 - \lambda_u^{k-1}\right) \prod_{i=1}^{j-1} \left(\prod_{u=1}^{p} \lambda_u^i\right)} \right).$$

PROOF. From the relation between $\mathbf{S}_k$'s and $\mathbf{R}_i$'s, we can see that

$$\frac{|\mathbf{S}_1|}{|\mathbf{S}_2|} = \frac{|\mathbf{R}_1|}{|\mathbf{I}_p - \mathbf{R}_1|}, \quad \frac{|\mathbf{S}_1|}{|\mathbf{S}_3|} = \frac{|\mathbf{R}_1||\mathbf{R}_2|}{|\mathbf{I}_p - \mathbf{R}_2|}, \ldots, \frac{|\mathbf{S}_1|}{|\mathbf{S}_K|} = \frac{\prod_{i=1}^{K-1} |\mathbf{R}_i|}{|\mathbf{I}_p - \mathbf{R}_{K-1}|}.$$

Since $|\mathbf{S}_k| = |\mathbf{V}_k||\mathbf{\Lambda}_k|$, the above equations yield

$$\frac{|\mathbf{\Sigma}_k|}{|\mathbf{\Sigma}_1|} = \frac{|\mathbf{\Lambda}_1|}{|\mathbf{\Lambda}_k|} = \frac{|\mathbf{V}_k|\prod_{i=1}^{k-1}|\mathbf{R}_i|}{|\mathbf{V}_1||\mathbf{I}_p - \mathbf{R}_{k-1}|} = \frac{|\mathbf{V}_k|}{|\mathbf{V}_1|}\frac{\prod_{i=1}^{k-1}\left(\prod_{u=1}^{p}\lambda_u^i\right)}{\prod_{u=1}^{p}\left(1-\lambda_u^{k-1}\right)} \tag{3.5}$$

for $k = 2, \ldots, K$. Thus, for $j \neq k$, (3.5) yields

$$\frac{|\mathbf{\Sigma}_k|}{|\mathbf{\Sigma}_j|} = \frac{|\mathbf{V}_k|}{|\mathbf{V}_j|}\frac{\prod_{u=1}^{p}\left(1-\lambda_u^{j-1}\right)\prod_{i=1}^{k-1}\left(\prod_{u=1}^{p}\lambda_u^i\right)}{\prod_{u=1}^{p}\left(1-\lambda_u^{k-1}\right)\prod_{i=1}^{j-1}\left(\prod_{u=1}^{p}\lambda_u^i\right)},$$

a function of $\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}$. Moreover, Theorem 3.1 says that the posterior distributions of $\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}$ are independent. These gives the result. □

3.2. Posterior probability of $\min|\mathbf{\Sigma}_k| = |\mathbf{\Sigma}_j|$

The distribution (3.4) enables us to obtain various integral-type posterior quantities of $|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j|$'s. Especially the posterior probability of a set $\cap_{k=1,k\neq j}^{K}$ $\{|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j| \in A_j\}$ is obtained from (3.4) if we set

$$h\Big(\bigcap_{k=1,k\neq j}^{K}|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j|\Big) = 1\Big[\bigcap_{k=1,k\neq j}^{K}\Big\{|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j| \in A_j\Big\}\Big],$$

where $1[\cdot]$ denotes the indicator function. When we take $A_j = \{|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j|;$ $|\mathbf{\Sigma}_k|/|\mathbf{\Sigma}_j| \geq 1,\ k, j = 1, \ldots, K;\ k \neq j\}$ in the indicator function, (3.4) gives the posterior probability of $\min|\mathbf{\Sigma}_k| = |\mathbf{\Sigma}_j|$. Therefore,

$$\begin{aligned} p_{\min}(j) &= p(\min|\mathbf{\Sigma}_k| = |\mathbf{\Sigma}_j| \,|\, \text{Data}) \\ &= \int_{I(\boldsymbol{\lambda}^{K-1})}\cdots\int_{I(\boldsymbol{\lambda}^1)}\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1})\prod_{i=1}^{K-1}p(\boldsymbol{\lambda}^i|\,\text{Data})\partial\boldsymbol{\lambda}^1\cdots\partial\boldsymbol{\lambda}^{K-1} \end{aligned} \tag{3.6}$$

where

$$\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}) = 1\Big[\bigcap_{k=1,k\neq j}^{K}\Big(\frac{|\mathbf{V}_k|}{|\mathbf{V}_j|}\frac{\prod_{u=1}^{p}\left(1-\lambda_u^{j-1}\right)\prod_{i=1}^{k-1}\left(\prod_{u=1}^{p}\lambda_u^i\right)}{\prod_{u=1}^{p}\left(1-\lambda_u^{k-1}\right)\prod_{i=1}^{j-1}\left(\prod_{u=1}^{p}\lambda_u^i\right)} \geq 1\Big)\Big].$$

An analytic evaluation of the probability is not available because the posterior distribution $\prod_{i=1}^{K-1}p(\boldsymbol{\lambda}^i|\,\text{Data})$ in (3.4) is complicated. In this regard, a Monte Carlo method, in particular, a variant of weighted Monte Carlo approach by Chen and Shao (1999) may naturally serve as an alternative solution for calculating the probability. The approach will be described in the next section.

4. A WEIGHTED MONTE CARLO METHOD

4.1. Importance sampling

Suppose that $\{\boldsymbol{\lambda}_{(t)}^i, t = 1, \ldots, m;\ i = 1, \ldots, K-1\}$ is a simulated sample from an appropriate importance function $g(\boldsymbol{\lambda}) = \prod_{i=1}^{K-1} g_i(\boldsymbol{\lambda}^i)$. Write the importance sampling weight as

$$w_{(t)} = \prod_{i=1}^{K-1} \frac{p(\boldsymbol{\lambda}_{(t)}^i | \text{ Data})}{g(\boldsymbol{\lambda}_{(t)}^i)}. \tag{4.1}$$

Then, the weighted Monte Carlo (WMC) estimator of the posterior probability (3.6) is given by

$$\widehat{p}_{\min}(j) = \frac{\sum_{t=1}^m w_{(t)} \eta_j^*(\boldsymbol{\lambda}_{(t)}^1, \ldots, \boldsymbol{\lambda}_{(t)}^{K-1})}{\sum_{\ell=1}^m w_{(\ell)}}, \tag{4.2}$$

where $\widehat{p}_{\min}(j) = \widehat{p}(\min |\boldsymbol{\Sigma}_k| = |\boldsymbol{\Sigma}_j| \mid \text{Data})$. Since

$$p_{\min}(j) = E_p\{\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1})\} = E_g\left\{\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}) \frac{\prod_{i=1}^{K-1} p(\boldsymbol{\lambda}^i | \text{ Data})}{\prod_{i=1}^{K-1} g_i(\boldsymbol{\lambda}^i)}\right\}$$

for the function $\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1})$ defined in (3.6), Geweke (1989) showed that

$$\widehat{p}_{\min}(j) \xrightarrow{a.s.} p_{\min}(j), \quad j = 1, \ldots, K, \tag{4.3}$$

as $m \to \infty$ where $\boldsymbol{\lambda}_{(1)}^i, \ldots, \boldsymbol{\lambda}_{(m)}^i$ are *iid* from $g_i(\boldsymbol{\lambda}^i)$. Notice that (4.2) gives a Monte Carlo estimate of the posterior probability using *iid* samples from the importance distribution in such a way that $\prod_{i=1}^{K-1} p(\boldsymbol{\lambda}^i | \text{ Data})$ and $\prod_{i=1}^{K-1} g_i(\boldsymbol{\lambda}^i)$ need only be known up to a constant of proportionality. The simulation standard error of $\widehat{p}_{\min}(j)$ is important, since it provides the magnitude of the simulation accuracy of the estimator $\widehat{p}_{\min}(j)$.

Let $U_t = w_{(t)} \eta_j^*(\boldsymbol{\lambda}_{(t)}^1, \ldots, \boldsymbol{\lambda}_{(t)}^{K-1})$, $S_t = w_{(t)}$ and $V_t = U_t - S_t E_p\{\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1})\}$. If the variances of U_t and V_t are both finite, then the standard error of $\widehat{p}_{\min}(j)$ can be calculated using the ratio estimate as in finite population sampling (Ripley, 1987, p. 158). The V_t's are *iid* random variables with zero mean and finite variance. The variance of V_t can be estimated by $\widehat{\sigma}_V^2 = m^{-1} \sum_{t=1}^m \widehat{V}_t^2$. The law of large numbers implies that $m^{-1/2} \sum_{t=1}^m V_t$ is asymptotically normal, $N(0, \sigma_V^2)$, and that

$$m^{1/2}\{\widehat{p}_{\min}(j) - p_{\min}(j)\} = \frac{m^{-1/2} \sum_{t=1}^m V_t}{m^{-1} \sum_{t=1}^m S_t} \longrightarrow N(0, \sigma^2),$$

where $\sigma^2 = \sigma_V^2/S^2$ and $S = E(S_t)$. It can be estimated by $\widehat{\sigma}^2 = \widehat{\sigma}_V^2/\widehat{S}^2$ with $\widehat{S} = m^{-1}\sum_{t=1}^{m} w_{(t)}$. The asymptotic variance of $\widehat{p}_{\min}(j)$ can be estimated by $m^{-1}\widehat{\sigma}^2$. In the calculation of $\widehat{\sigma}_V^2$, we can use $\widehat{p}_{\min}(j)$ for the unknown $E_p\{\eta_j^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1})\}$.

4.2. Choice of importance distribution

As one would expect, (4.2) will converge faster and generally be better behaved the closer the importance function is to the joint posterior distribution $\prod_{i=1}^{K-1} p(\boldsymbol{\lambda}^i|\,\text{Data})$. For this reason, choosing a good candidate is the main issue in applications of importance sampling.

Apart from $J(\boldsymbol{\lambda}^i) = \prod_{u<v}^{p}(\lambda_u^i - \lambda_v^i)$, the density in (3.3) has the shape of the joint density of order statistics from p independently and identically distributed Beta$((\sum_{k=1}^{i} n_k - p + 1)/2, (n_{i+1} - p + 1)/2)$ variates. This similarity in shape is exploited in developing the importance sampling scheme. Noticing from Theorem 3.1 that $\boldsymbol{\lambda}^i$'s, $i = 1, \ldots, K-1$, are independent, the most natural candidate for the importance distribution, say NIF, is the corresponding distribution of $K-1$ independent sets of ordered beta variates:

$$g(\boldsymbol{\lambda}) = \prod_{i=1}^{K-1} g_i(\boldsymbol{\lambda}^i) \propto \prod_{i=1}^{K-1}\prod_{u=1}^{p}\left\{(\lambda_u^i)^{(\sum_{k=1}^{i} n_k - p - 1)/2}\,(1-\lambda_u^i)^{(n_{i+1}-p-1)/2}\right\} I(\boldsymbol{\lambda}^i),$$

so that

$$p\big(\min|\boldsymbol{\Sigma}_k| = |\boldsymbol{\Sigma}_j|\;\big|\;\text{Data}\big) = \frac{\prod_{l=1}^{K-1} c_i}{c_0} E_g\left\{\eta^*(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{K-1}) \prod_{i=1}^{K-1} J(\boldsymbol{\lambda}^i)\right\},$$

where E_g is the expectation over $g(\lambda)$ and

$$c_0 = \prod_{i=1}^{K-1} p!\left\{\frac{\Gamma\big(\frac{1}{2}(\sum_{k=1}^{i+1} n_k - 2p + 2)\big)}{\Gamma\big(\frac{1}{2}(\sum_{k=1}^{i} n_k - p + 1)\big)\Gamma\big(\frac{1}{2}(n_{i+1} - p + 1)\big)}\right\}^p.$$

Huzurbazar and Butler (1998) discussed the inefficiency of the importance distributions. They showed that the inefficiency arose in factor $J(\boldsymbol{\lambda}^i) = \prod_{u<v}^{p}(\lambda_u^i - \lambda_v^i)$, and the factor tends to be larger when $\boldsymbol{\lambda}^i$ is more spread out over the range $(0, 1)$ and relatively evenly spaced. Then they suggested two modified importance functions of $\boldsymbol{\lambda}^i$ to improve the the efficiency of the importance sampling procedure. One modification reduces the degrees of freedom of the beta variates to increase their variance. They showed that a symmetric betas, Beta(df, df), with $1.5 \le df \le 2$ produced best result. The other entails sampling every r^{th} order

statistic from a sample of $(p-1)r+1$ beta variates with $r \geq 2$ (see Huzurbazar and Butler, 1998 for the modified importance function; MIF).

In this subsection we suggest another importance function denoted by SIF (suggested importance function):

$$g^*(\boldsymbol{\lambda}) = \prod_{i=1}^{K-1} g_i^*(\boldsymbol{\lambda}^i) \propto \prod_{i=1}^{K-1} \prod_{u=1}^{p} \left\{ \left(\lambda_u^i\right)^{\left(\sum_{k=1}^{i} n_k - p - 1\right)/2} \left(1-\lambda_u^i\right)^{(n_{i+1}-p-1)/2} \right\} I^*(\boldsymbol{\lambda}^i),$$

where $I^*(\boldsymbol{\lambda}^i) = I(1 > \lambda_1^i > \cdots > \lambda_p^i > 0)\, I(J(\boldsymbol{\lambda}^i) > \alpha)$, $i = 1, \ldots, K-1$. A new feature of SIF is that we can increase variance of the beta variates by controlling the value of α, $0 < \alpha < 1$, increasing the value of $J(\boldsymbol{\lambda}^i)$. This takes care of the assumption that $\mathbf{R}_i$'s are positive definite and non-intraclass structured symmetric matrices, *i.e.*, the constraint $J(\boldsymbol{\lambda}^i) > \alpha$ enables us to avoid near singularity and near intraclass structure problems in $\mathbf{R}_i$'s, so that it eliminates the space of $\boldsymbol{\lambda}^i$ from the importance sampling where $P(\boldsymbol{\lambda}^i | \text{ Data}) \approx 0$.

Using the inequality among arithmetic, geometric, and harmonic means (*cf.* Casella and Berger, 2002, p. 191), we see that

$$\prod_{u=1}^{p-1} \left(\frac{1}{p-u} \sum_{u<v}^{p} \frac{1}{\lambda_u^i - \lambda_v^i} \right)^{-(p-u)} \leq J(\boldsymbol{\lambda}^i) \leq \prod_{u=1}^{p-1} \left\{ \frac{\sum_{u<v}^{p} (\lambda_u^i - \lambda_v^i)}{p-u} \right\}^{p-u},$$

where the equality of the relation holds for a constant value of $\lambda_u^i - \lambda_v^i$ for all $u > v$. This relation gives a guideline for choosing a proper value of α. To guarantee that the constraint space should not be small, we may choose the value of α avoiding only the case where all the λ_u^i's are clustered in the vicinity of one point. One choice is $\alpha = (1/p)^{p(p-1)/2}$ and it is obtained from the above inequality by setting $\lambda_u^i - \lambda_v^i = 1/p$ for all $u > v$. Under SIF, the important sampling scheme is sampling order statistics from $K-1$ independent sequences of Beta(a_i, b_i) distributions, $i = 1, \ldots, K-1$, with additional constraint $J(\boldsymbol{\lambda}_{(t)}^i) > \alpha$. Here $a_i = (\sum_{k=1}^{i} n_k - p + 1)/2$, $b_i = (n_{i+1} - p + 1)/2$ and $\boldsymbol{\lambda}_{(t)}^i = (\lambda_{1(t)}^i, \ldots, \lambda_{p(t)}^i)'$. A simple accept/reject algorithm applies for generating the constrained Beta variates vector $\boldsymbol{\lambda}_{(t)}^i$. In this procedure, the weight function in (4.2) is the same as that obtained from using $g(\lambda)$. It reduces to

$$w_{(t)} = \prod_{i=1}^{K-1} J(\boldsymbol{\lambda}_{(t)}^i), \tag{4.4}$$

where $J(\boldsymbol{\lambda}_{(t)}^i) = \prod_{u<v}^{p} (\lambda_{u(t)}^i - \lambda_{v(t)}^i)$.

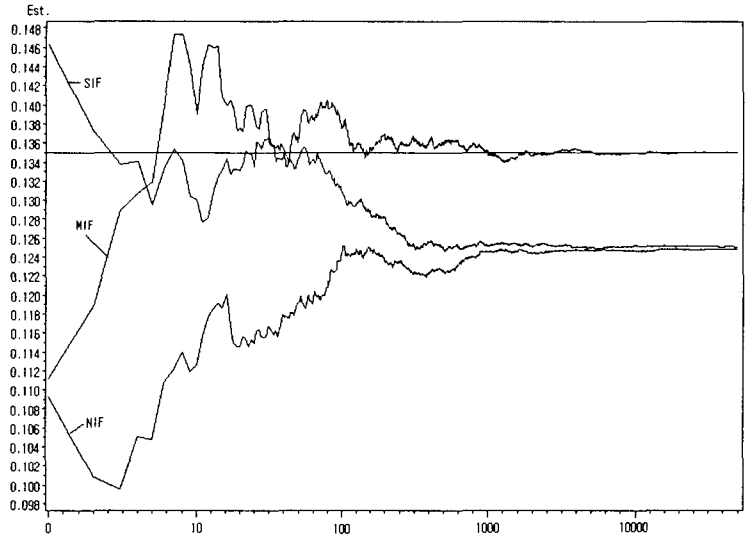

FIGURE 4.1 *Importance sampling using NIF, MIF and SIF. Horizontal axis shows number of importance samples on a log scale. Vertical axis shows three weighted MC estimates,* $|\widehat{\mathbf{R}}_1|$*'s. The horizontal line indicates the true value of* $|\mathbf{R}_1|$.

To give a simple comparison among the importance functions (NIF, MIF and SIF), we evaluated $|\mathbf{R}_1|$ using the weighted MC method, where $\mathbf{R}_1$ is a 3×3 matrix defined in Theorem 3.1. The posterior quantity to evaluate is

$$\int_{\boldsymbol{\lambda}^1} \prod_{u=1}^{3} \lambda_u^1 \, p(\boldsymbol{\lambda}^1 | \text{ Data}) d\boldsymbol{\lambda}^1,$$

where $\boldsymbol{\lambda}^1 = (\lambda_1^1, \lambda_2^1, \lambda_3^1)'$ is the vector of eigen values of $\mathbf{R}_1$. True value of $|\mathbf{R}_1|$ is 0.1351 and is obtained from a set of two independent samples generated from $N_3(\mathbf{0}, \mathbf{I_3})$ and $N_3(\mathbf{0}, 1.2\,\mathbf{I}_3)$. The size of each sample is $N = 100$. The results of weighted MC estimation using NIF, MIF, and SIF with $\alpha = (1/3)^3$ are shown in Figure 4.1. The horizontal axis shows the number (m) of importance samples on a log scale. From the figure we see that weighted MC estimates, $|\widehat{\mathbf{R}}_1|$'s, of the three samplers have settled down after about m=5,000 importance samples. From Figure 4.1, clearly SIF is better than the other importance functions. Weighted MC estimate using SIF sampler settled down closer to the true value than the others did. When $m = 5,000$ importance samples are used to estimate $|\mathbf{R}_1|$, the mean squared errors (MSE) of SIF, MIF, and NIF sampler estimates are 0.269×10^{-6}, 97.1×10^{-6} and 105.4×10^{-6}, respectively. Therefore, this simple

example illustrates that SIF is a proper and an appealing importance function.

5. ILLUSTRATIVE EXAMPLES

The goal of this section is to study the effectiveness of the posterior probability in comparing K generalized variances and to identify some situations where one would expect improvement in the Monte Carlo method.

5.1. A simulation study

To apply and illustrate our methodology, we consider the data generated from $K = 5$ multivariate normal populations. Since a linear transformation leaves the posterior probabilities $\widehat{p}_{\min}(j)$, $j = 1, \ldots, K$, invariant, there is no loss of generality in considering the case $\mathbf{\Pi}_j \sim N_p(\mathbf{0}, \mathbf{I} + \mathbf{D}_j)$, j^{th} population distribution. So that the first population has the minimum generalized variance. This canonical form is obtained via the transformation suggested by Dunn and Holloway (1967):

$$\mathbf{Y}_j = \mathbf{A}_j' \mathbf{\Sigma}_1^{-1/2} (\mathbf{Z}_j - \boldsymbol{\mu}_j),$$

where $\mathbf{A}_j$ is an orthogonal matrix such that $\mathbf{A}_j' \mathbf{\Sigma}_1^{-1/2} \mathbf{\Sigma}_j \mathbf{\Sigma}_1^{-1/2} \mathbf{A}_j = \mathbf{I} + \mathbf{D}_j$, a diagonal matrix. Each experiment generates N_j observation from the five multivariate normal populations with $\mathbf{D}_j = (j-1) \times \delta \mathbf{I}_p$, $j = 1, \ldots, K$. This experiment is replicated 200 times. For each experiment, with $m = 20,000$ importance sample size, the suggested posterior probability $\widehat{p}_{\min}(1)$ is calculated for each of the three importance functions (NIF, MIF, SIF) using a variety of parameters,

TABLE 5.1 *The mean and standard deviation (in parenthesis) of* $\widehat{p}_{\min}(1)$ *obtained from 200 replicates*

		$\delta = 0.1$			$\delta = 0.5$		
N	p	*NIF*	*MIF*	*SIF*	*NIF*	*MIF*	*SIF*
20	3	0.4099	0.4551	0.4592	0.8587	0.8704	0.9055
		(0.3206)	(0.3543)	(0.3126)	(0.2099)	(0.2223)	(0.1904)
	7	0.5976	0.5494	0.6103	0.9775	0.9701	0.9799
		(0.3963)	(0.4001)	(0.3933)	(0.1065)	(0.1413)	(0.0988)
100	3	0.6904	0.7280	0.7449	0.9974	0.9983	0.9997
		(0.3320)	(0.3221)	(0.2886)	(0.0240)	(0.0139)	(0.0018)
	7	0.8798	0.8531	0.8924	0.9999	1.0000	1.0000
		(0.2417)	(0.2966)	(0.2292)	(> 0.0001)	(> 0.0001)	(> 0.0001)

p, δ, N, and $\alpha = (1/p)^{p(p-1)/2}$ in the SIF, where $N_1 = \cdots = N_5 = N$.

The simulation results show that the suggested importance sampling procedure uniformly yields $\widehat{p}_{\min}(1)$ to have the highest posterior probability and it is tabulated in Table 5.1. In this regard, the procedure is a useful methodology that allows for finding a population having minimum generalized variance among K multivariate normal populations. The table also contrasts the procedures each using three different importance functions (NIF, MIF, SIF). In order to judge the efficiency of the procedures, the standard deviations of $\widehat{p}_{\min}(1)$ are compared. In all the examples we considered, importance distributions based on SIF scheme gave the best results.

5.2. A data example

To apply and illustrate our methodology, we consider the remote-sensing data on crops in SAS/STAT examples. The observations are grouped into five crops: clover, corn, cotton, soybeans, and sugar beets with sample sizes $N_1 = 11$, $N_2 = 7$, $N_3 = N_4 = N_5 = 6$, respectively. Four measures called X1–X4 make up the descriptive variables. Their generalized variances are $|\mathbf{V}_1| = 10^4 e^{23.64618}$, $|\mathbf{V}_2| = 6^4 e^{11.13472}$, $|\mathbf{V}_3| = 5^4 e^{13.23569}$, $|\mathbf{V}_4| = 5^4 e^{12.45263}$, $|\mathbf{V}_5| = 5^4 e^{17.76293}$. The importance distribution was chosen so that parameter α in the SIF was allowed to assume $\alpha = (1/p)^{p(p-1)/2} = (1/4)^6$. Under the SIF, the suggested procedure with $m = 20,000$ yielded the posterior probabilities listed in Table 5.2. For a reference, the table also contains the posterior probabilities obtained from using NIF. The table notes that, among the five crops, clover has the minimum generalized variance.

TABLE 5.2 *Estimated posterior probabilities and their standard errors*

	$\widehat{p}_{\min}(1)$	$\widehat{p}_{\min}(2)$	$\widehat{p}_{\min}(3)$	$\widehat{p}_{\min}(4)$	$\widehat{p}_{\min}(5)$
SIF	0.0000	0.8361	0.0457	0.1182	0.0000
	(0.0000)	(0.0584)	(0.0176)	(0.0301)	(0.0000)
NIF	0.0000	0.7312	0.1942	0.0746	0.0000
	(0.0000)	(0.0632)	(0.0596)	(0.0361)	(0.0000)

6. CONCLUDING REMARKS

In this paper we propose a weighted Monte Carlo method for estimating posterior probability of a particular population has minimum generalized variance

among K populations considered. The probability is a useful criterion for ranking the dispersion or spread of K multivariate normal populations. The methodology proposed in this paper can easily extended to inference integer-type posterior quantities of functions of eigenvalues. An immediate examples are multiple comparison of K generalized variances (see Kim, 2000 for the univariate case) and calculating Bayesian highest posterior density interval for "the proportion of variance explained" in principal component analysis. Although we have not pursued these in this paper, these are certainly worthy to be investigated.

REFERENCES

BOX, G. E. P. AND TIAO, G. C. (1992). *Bayesian Inference in Statistical Analysis*, Wiley, New York.

CASELLA, G. AND BERGER, R. L. (2002). *Statiatical Inference*, Duxbury, Pacific Grove.

CHEN, M.-H, AND SHAO, Q.-M. (1999). "Monte Carlo estimation of Bayesian credible and HPD intervals", *Journal of the Computational and Graphical Statistics*, **8**, 69–92.

DUNN, O. J. AND HOLLOWAY, L. N. (1967). "The robustness of Hotelling's T^{2}", *Journal of the American Statistical Association*, **62**, 124–136.

GEWEKE, J. (1989). "Bayesian inference in econometric models using Monte Carlo integration", *Econometrica*, **57**, 1317–1339.

GRIZZLE, J. E. AND ALLEN, D. M. (1969). "Analysis of growth and dose response curves", *Biometrics*, **25**, 357–381.

HUZURBAZAR, S. AND BUTLER, R. W. (1998). "Importance sampling for p-value computations in multivariate test", *Journal of the Computational and Graphical Statistics*, **7**, 342–355.

KIM, H. J. (2000). "Bayesian multiple comparisons for normal variances", *Journal of the Korean Statistical Society*, **29**, 155–168.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

PRESS, S. J. (1982). *Applied Multivariate Analysis*, 2nd ed., Krieger Publishing Co., Florida.

RENCHER, A. C. (2002). *Methods of Multivariate Analysis*, 2nd ed., Wiley, New York.

RIPLEY, B. D. (1987). *Stochastic Simulation*, Wiley, New York.