MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Inference on Eigenvalues of Wishart Distribution Using Asymptotics with respect to the Dispersion of Population Eigenvalues

Yo SHEENA and Akimichi TAKEMURA

METR 2007–23 April 2007

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Inference on Eigenvalues of Wishart Distribution Using Asymptotics with respect to the Dispersion of Population Eigenvalues

Yo Sheena* and Akimichi Takemura†

April, 2007

Abstract

In this paper we derive some new and practical results on testing and interval estimation problems for the population eigenvalues of a Wishart matrix based on the asymptotic theory for block-wise infinite dispersion of the population eigenvalues. This new type of asymptotic theory has been developed by the present authors in Takemura and Sheena (2005) and Sheena and Takemura (2007a,b) and in these papers it was applied to point estimation problem of population covariance matrix in a decision theoretic framework. In this paper we apply it to some testing and interval estimation problems. We show that the approximation based on this type of asymptotics is generally much better than the traditional large-sample asymptotics for the problems.

Key words and phrases: eigenvalues of covariance matrix, Wishart distribution, test on eigenvalues, interval estimation of eigenvalues

1 Introduction

Let $\boldsymbol{S} = (s_{ij})$ be distributed according to Wishart distribution $\boldsymbol{W}_p(n, \boldsymbol{\Sigma})$, where p is the dimension, n is the degrees of freedom and $\boldsymbol{\Sigma}$ is the covariance matrix. Let $\lambda_1 \geq \ldots \geq \lambda_p > 0$ denote the eigenvalues of $\boldsymbol{\Sigma}$. In this paper we consider some testing and interval estimation problems for the eigenvalues of $\boldsymbol{\Sigma}$. Our aim is to give practical solutions to the problems based on the asymptotic theory for block-wise infinite dispersion of the population eigenvalues. In view of the intractability of the finite sample exact distribution of sample eigenvalues, usually the large sample asymptotic approximation is used. There exists an extensive literature on improving the first-order large sample approximation by an asymptotic expansion (see Siotani et al. (1985) for a comprehensive treatment). However for a moderate or small value of the sample size n, the large sample asymptotic theory often gives a poor approximation. In these cases asymptotic expansions tend to give an even larger error. On the other hand, we find that approximation

*Department of Economics, Shinshu University
†Graduate School of Information Science and Technology, University of Tokyo

based on the infinite dispersion of the population eigenvalues is more robust and gives a reasonable approximation even for a small sample size n.

The first problem we consider in this paper is testing the one-sided null hypothesis on the mth population eigenvalue

$$H_0^{(m)} : \lambda_m \geq \lambda_m^*. \tag{1}$$

For testing $H_0^{(m)}$ it is natural to consider a one-sided rejection region based on the mth sample eigenvalue of l_m of $\boldsymbol{S}$. We show that the least favorable distribution is given by $\lambda_m^* = \lambda_1 = \cdots = \lambda_m$ and $0 = \lambda_{m+1} = \cdots = \lambda_p$. This is exactly the situation covered by the asymptotic theory for block-wise infinite dispersion. Therefore it gives an explicit solution to the testing problem of $H_0^{(m)}$.

The second problem is the interval estimation for the largest population eigenvalue λ_1 in terms of the largest sample eigenvalue l_1 of $\boldsymbol{S}$. We will show that confidence interval based block-wise infinite dispersion gives much better coverage probability than the conventional large sample asymptotics.

The third problem is testing the hypothesis of equality of the several smallest eigenvalues: $\lambda_{m+1} = \cdots = \lambda_p$. This problem is important in determining the rank of the systematic part in a multivariate variance component model. We consider approximation to the null distribution of the likelihood ratio criterion under the block-wise infinite dispersion of population eigenvalues. Again this type of asymptotics gives much better approximation than the large sample asymptotics.

The organization of the paper is as follows. In Section 2 we set up notations for the paper and give some preliminary results on the asymptotic theory for block-wise infinite dispersion of the population eigenvalues. In Section 3 we study the above three problems, 1) One-sided test for a population eigenvalue in Section 3.1; 2) Interval estimation for extreme eigenvalues in Section 3.2; 3) Testing equality of the smallest eigenvalues in Section 3.3.

2 Asymptotic Distribution of Normalized Sample Eigenvalues

Denote the spectral decompositions of $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ by

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' \tag{2}$$

$$\boldsymbol{S} = \boldsymbol{G}\boldsymbol{L}\boldsymbol{G}', \tag{3}$$

where $\boldsymbol{G}, \boldsymbol{\Gamma} \in \mathcal{O}(p)$, the group of $p \times p$ orthogonal matrices, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $\boldsymbol{L} = \mathrm{diag}(l_1, \ldots, l_p)$ are diagonal matrices with the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p > 0$, $l_1 \geq \ldots \geq l_p > 0$ of $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$, respectively. We use the notations $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ and $\boldsymbol{l} = (l_1, \ldots, l_p)$ hereafter. We guarantee the uniqueness (almost surely) of the decomposition (3) by requiring that

$$\widetilde{\boldsymbol{G}} = (\widetilde{g}_{ij}) = \boldsymbol{\Gamma}'\boldsymbol{G} \tag{4}$$

has positive diagonal elements.

In Takemura and Sheena (2005) we considered what happens to appropriately normalized components of $\boldsymbol{S}$ if the population eigenvalues become infinitely dispersed, i.e.,

$$(\lambda_2/\lambda_1, \lambda_3/\lambda_2, \ldots, \lambda_p/\lambda_{p-1}) \to 0.$$

In Sheena and Takemura (2007b) we generalized the asymptotic result of Takemura and Sheena (2005) to the case when the population eigenvalues are block-wise infinitely dispersed.

Let the population eigenvalues be parameterized as follows;

$$\lambda_i = \begin{cases} \xi_i \alpha, & \text{if } i = 1, \ldots, m, \\ \xi_i \beta, & \text{if } i = m+1, \ldots, p, \end{cases} \tag{5}$$

where ξ_i's are fixed and "asymptotic parameter" α and β vary. When we say population eigenvalues are "(two-)block-wise infinitely dispersed", it means that

$$\beta/\alpha \to 0. \tag{6}$$

The above notation is used as a general notation including specific convergences (divergences) such as $(\alpha, \beta) \to (\infty, 1)$, $(\alpha, \beta) \to (1, 0)$ and so on. More precisely, the operation $\lim_{\beta/\alpha \to 0} f(\alpha, \beta)$ means $\lim_{i \to \infty} f(\alpha_i, \beta_i)$ with any specific sequences $\alpha_i, \beta_i,\ i = 1, 2, \ldots$, such that $\beta_i/\alpha_i \to 0$ as $i \to \infty$.

As Sheena and Takemura (2007b) indicates, appropriate normalization for the sample eigenvalues is given by

$$d_i = \begin{cases} l_i/\alpha, & \text{if } i = 1, \ldots, m, \\ l_i/\beta, & \text{if } i = m+1, \ldots, p, \end{cases} \tag{7}$$

while (4) also serves as an appropriate normalization for sample eigenvectors.

For the normalized population or sample eigenvalues, we use the following notations;

$$\begin{aligned} \boldsymbol{\xi}_1 &= (\xi_1, \ldots, \xi_m), & \boldsymbol{\xi}_2 &= (\xi_{m+1}, \ldots, \xi_p), \\ \boldsymbol{d}_1 &= (d_1, \ldots, d_m), & \boldsymbol{d}_2 &= (d_{m+1}, \ldots, d_p), \\ \boldsymbol{\Xi}_1 &= \mathrm{diag}(\xi_1, \ldots, \xi_m), & \boldsymbol{\Xi}_2 &= \mathrm{diag}(\xi_{m+1}, \ldots, \xi_p), \\ \boldsymbol{D}_1 &= \mathrm{diag}(d_1, \ldots, d_m), & \boldsymbol{D}_2 &= \mathrm{diag}(d_{m+1}, \ldots, d_p). \end{aligned}$$

Now we state the basic theorem on the asymptotic distributions of $\boldsymbol{d}_1, \boldsymbol{d}_2$.

Theorem 1 *Suppose that we have two independent Wishart distributions*

$$\widetilde{\boldsymbol{W}}_{11} \sim \boldsymbol{W}_m(n, \boldsymbol{\Xi}_1), \qquad \widetilde{\boldsymbol{W}}_{22} \sim \boldsymbol{W}_{p-m}(n-m, \boldsymbol{\Xi}_2)$$

and that their spectral decompositions are given by

$$\begin{aligned} \widetilde{\boldsymbol{W}}_{11} &= \widetilde{\boldsymbol{G}}_{11} \widetilde{\boldsymbol{D}}_1 \widetilde{\boldsymbol{G}}'_{11}, & \widetilde{\boldsymbol{D}}_1 &= \mathrm{diag}(\widetilde{d}_1, \ldots, \widetilde{d}_m), & \widetilde{\boldsymbol{d}}_1 &= (\widetilde{d}_1, \ldots, \widetilde{d}_m), \\ \widetilde{\boldsymbol{W}}_{22} &= \widetilde{\boldsymbol{G}}_{22} \widetilde{\boldsymbol{D}}_2 \widetilde{\boldsymbol{G}}'_{22}, & \widetilde{\boldsymbol{D}}_2 &= \mathrm{diag}(\widetilde{d}_{m+1}, \ldots, \widetilde{d}_p), & \widetilde{\boldsymbol{d}}_2 &= (\widetilde{d}_{m+1}, \ldots, \widetilde{d}_p), \end{aligned}$$

where $\widetilde{\boldsymbol{G}}_{11} \in \mathcal{O}(m)$, $\widetilde{\boldsymbol{G}}_{22} \in \mathcal{O}(p-m)$, $\widetilde{d}_1 \geq \cdots \geq \widetilde{d}_m$, $\widetilde{d}_{m+1} \geq \cdots \geq \widetilde{d}_p$. *Then as* $\beta/\alpha \to 0$,

$$\boldsymbol{d}_i \xrightarrow{d} \widetilde{\boldsymbol{d}}_i, \quad i = 1, 2.$$

Proof. Using Lemma 1 of Sheena and Takemura (2007b), we prove the convergence of the moment generating function. Let

$$x(\boldsymbol{G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta) = \exp\left(\alpha^{-1} \sum_{i=1}^{m} l_i \theta_i + \beta^{-1} \sum_{i=m+1}^{p} l_i \theta_i \right) = \exp\left(\sum_{i=1}^{p} d_i \theta_i \right),$$

where $|\theta_i| < 3^{-1} \min_j \xi_j^{-1}$, $\forall i$. Notice that (19) in Lemma 1 of Sheena and Takemura (2007b) is satisfied since

$$\begin{aligned} x(\boldsymbol{\Gamma G}, \boldsymbol{l}, \boldsymbol{\lambda}, \alpha, \beta) &\leq \exp\left(\alpha^{-1} \sum_{i=1}^{m} l_i |\theta_i| + \beta^{-1} \sum_{i=m+1}^{p} l_i |\theta_i| \right) \\ &\leq \exp\left(3^{-1}\alpha^{-1} \sum_{i=1}^{m} l_i \xi_i^{-1} + 3^{-1}\beta^{-1} \sum_{i=m+1}^{p} l_i \xi_i^{-1} \right) \\ &= \exp\left(3^{-1} \sum_{i=1}^{p} l_i \lambda_i^{-1} \right) \\ &\leq \exp\left(\operatorname{tr} 3^{-1} \boldsymbol{G L G}' \boldsymbol{\Lambda}^{-1} \right), \quad \forall \boldsymbol{G} \in \mathcal{O}(p),\ \forall \boldsymbol{l} \in \{\boldsymbol{l} | l_1 \geq \cdots \geq l_p \geq 0\}. \end{aligned}$$

For the last inequality, see e.g. Marshall and Olkin (1979) Ch.20.A.1. Since $x(\boldsymbol{d}, \boldsymbol{q}, \boldsymbol{\xi}, \alpha, \beta; \boldsymbol{\Gamma}, \boldsymbol{H}^{(\tau)}) = \exp\left(\sum_{i=1}^{p} d_i \theta_i\right)$, trivially we have

$$\bar{x}_{\boldsymbol{\Gamma}}(\boldsymbol{H}^{(\tau)} \boldsymbol{G}(\boldsymbol{q}_{11}, \boldsymbol{q}_{22}, \boldsymbol{0}), \boldsymbol{d}, \boldsymbol{Q}_{21}, \boldsymbol{\xi}) = \exp\left(\sum_{i=1}^{p} d_i \theta_i \right).$$

Therefore we have

$$\lim_{\beta/\alpha \to 0} E\left[\exp\left(\sum_{i=1}^{p} d_i \theta_i \right) \right] = E\left[\exp\left(\sum_{i=1}^{m} \widetilde{d}_i(\widetilde{\boldsymbol{W}}_{11}) \theta_i + \sum_{i=m+1}^{p} \widetilde{d}_i(\widetilde{\boldsymbol{W}}_{22}) \theta_i \right) \right].$$

∎

3 Inference on Population Eigenvalues

The asymptotic result in the previous section has possibly various applications for inference on the population eigenvalues. We give three inference problems as interesting applications.

3.1 One-sided Test for Population Eigenvalue

Consider the null hypothesis on the mth ($m = 1, \ldots, p$) population eigenvalue

$$H_0^{(m)} : \lambda_m \geq \lambda_m^*$$

against the alternative $H_1^{(m)} : \lambda_m < \lambda_m^*$. Need for testing H_0 arises in some practical cases, for example:

- In principal component analysis, $\lambda^* (= \lambda_1^* = \cdots = \lambda_p^*)$ may be a cut-off value and a test for $H_0^{(m)}$ is repeatedly carried out starting from $m = 1$ until $H_0^{(m)}$ is rejected. This is one of the methods for deciding the dimension of the principal components.

- Let $x_i (i = 1, \ldots, p)$ be the return of the ith asset in finance and $\boldsymbol{x} = (x_1, \ldots, x_p)$ is distributed as the p-dimensional normal distribution $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. $H_1^{(1)}$ is equivalent to the assertion $\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} < \lambda_1^*, \ \forall \boldsymbol{a} = (a_1, \ldots, a_p)$ such that $\|\boldsymbol{a}\| = 1$. If $H_0^{(1)}$ is rejected, then it means that the group of assets $\boldsymbol{x}$ is stable in view of volatility since any portfolio among the group is never beyond λ_1^* in its variance.

A natural rejection region in testing $H_0^{(m)}$ is given by $l_m \leq l_m^*(\gamma)$ for a given significance level γ. The following lemma and Theorem 1 give the critical point $l_m^*(\gamma)$.

Lemma 1 *For any positive c*

$$\sup_{H_0^{(m)}} P_{\boldsymbol{\Lambda}}(l_m \leq c) = \lim_{\beta \to 0} P_{\bar{\boldsymbol{\Lambda}}}(l_m \leq c),$$

where $\bar{\boldsymbol{\Lambda}} = \mathrm{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_p)$, $\bar{\lambda}_1 = \cdots = \bar{\lambda}_m = \lambda_m^*, \ \bar{\lambda}_{m+1} = \cdots = \bar{\lambda}_p = \beta$.

Proof. According to Theorem 1 of Anderson and Das Gupta (1964), $P_{\boldsymbol{\Lambda}}(l_m \leq c)$ is a monotonically decreasing function with respect to each $\lambda_i, \ (i = 1, \ldots, p)$, hence

$$P_{\boldsymbol{\Lambda}}(l_m \leq c) \leq P_{\bar{\boldsymbol{\Lambda}}}(l_m \leq c),$$

where $\beta = \lambda_p$. Furthermore $P_{\bar{\boldsymbol{\Lambda}}}(l_m \leq c)$ is monotonically increasing as β goes to zero. ∎

Because of the result of Theorem 1 with $\alpha = 1$, $\xi_i = \lambda_m^*, \ (i = 1, \ldots, m)$ and $\xi_i = 1, \ (i = m+1, \ldots, p)$,

$$\lim_{\beta \to 0} P_{\bar{\boldsymbol{\Lambda}}}(l_m \leq c) = P(\widetilde{l}_m \leq c),$$

where $\widetilde{l}_m$ is distributed as the smallest eigenvalues of $\boldsymbol{W}_m(n, \lambda_m^* \boldsymbol{I}_m)$. Therefore we have the following result.

Theorem 2 *For testing hypothesis $H_0^{(m)}$ against $H_1^{(m)}$, a test with significance level γ is given with the rejection region*

$$l_m \leq l_m^*(\gamma),$$

where $l_m^*(\gamma)$ is the lower $100\gamma\%$ point of the smallest eigenvalue of $\boldsymbol{W}_m(n, \lambda_m^* \boldsymbol{I}_m)$.

For analytic calculation of $l_m^*(\gamma)$, see Thompson (1962), Hanumara and Thompson (1968). In the case $m = 1$, which is practically the most important, it is given by $\lambda_1^* \chi_n^2(\gamma)$, where $\chi_n^2(\gamma)$ is the lower $100\gamma\%$ point of the χ^2 distribution with the degree of freedom n.

3.2 Interval Estimation of Extreme Eigenvalues

In this subsection we present a new way of constructing a confidence interval for the extreme population eigenvalues. Let $\lambda_1 \leq f_1(\boldsymbol{l})$ be a one-sided estimated interval with confidence level γ. For example, in the second case in Section 3.1, the maximum volatility in all possible portfolio among the assets $\boldsymbol{x}$ is estimated to be less than or equal to $f_1(\boldsymbol{l})$.

However if we use the exact finite distribution theory, it is not easy to find an appropriate $f_1(\boldsymbol{l})$ under a given γ even if we only consider an interval of the simplest form $\lambda_1 \leq c_1 l_1$ with some

constant c_1. (Note that $l_i/\lambda_i,\ (i=1,\ldots,p)$ is bounded in probability. See Lemma 1 of Takemura and Sheena (2005).) Therefore usually a large sample approximation is employed (e.g. Theorem 13.5.1. of Anderson (2003)):

$$\sqrt{n}\Big(\frac{l_i}{n}-\lambda_i\Big) \xrightarrow{d} N(0,2\lambda_i^2), \quad i=1,\ldots,p.$$

Let z_γ denote the upper 100γ percentile of the standard normal distribution. Since

$$\begin{aligned} &P\Big(\sqrt{\frac{n}{2}}\Big(\frac{l_1}{n\lambda_1}-1\Big)\geq z_\gamma\Big) \\ &= P\Big(l_1 \geq (\sqrt{2n}z_\gamma+n)\lambda_1\Big) \to \gamma \quad \text{as } n\to\infty, \end{aligned}$$

we have an approximate confidence interval

$$\lambda_1 \leq (\sqrt{2n}z_\gamma+n)^{-1}l_1, \tag{8}$$

with confidence level close to γ for sufficiently large n.

Now we propose an alternative approximation. Suppose $m=1$ in Theorem 1, then as β/α goes to zero,

$$d_1 \xrightarrow{d} \tilde{d}_1 = \widetilde{\boldsymbol{W}}_{11}.$$

Since $\widetilde{\boldsymbol{W}}_{11}/\xi_1 \sim \chi^2(n)$,

$$\frac{l_1}{\lambda_1} = \frac{d_1}{\xi_1} \xrightarrow{d} \chi^2(n).$$

as β/α goes to zero. From this asymptotics, we can make an approximate interval

$$\lambda_1 \leq (\chi^2_\gamma(n))^{-1}l_1, \tag{9}$$

where $\chi^2_\gamma(n)$ is the upper 100γ percentile of χ^2 distribution with the degree of freedom n. The interval (9) has approximately γ confidence level when β/α is sufficiently close to zero.

We are interested in how large n for (8) or how small β/α for (9) is required to get practically sufficient approximations. Because of difficulty in theoretical evaluations, we carried out a simulation study with the fixed parameters $p=3,\ m=1,\ \xi_1=\xi_2=\xi_3=1,\ \alpha=1$, while we select different n's (5, 10, 20, 50, 100, 500, 1000) and β's (1.0, 0.9, 0.8, 0.6, 0.5, 0.3, 0.1, 0.01, 0.001). For each case, 50000 Wishart random matrices are generated. We present the results in Table 1. The numbers under L1(U1) indicate the ratio of the largest sample eigenvalue which fall within the interval (8) with $\gamma=0.95(0.05)$, while those under U2(L2) show the similar ratio with respect to the interval (9). Numbers in bold indicate that they are within ± 0.01 deviation from the desired value, hence the approximation may be good enough for many practical purposes. We can summarize the result as follows;

1. In every case (9) gives better approximation than (8).

2. Since β is as large as 0.3, (9) already gives a good approximation. In that sense, the approximated interval (9) seems robust. When β is smaller or equal to 0.1, (9) works well even with small samples such as $n=5$ or 10, while (8) needs samples as large as 100 or 500.

Table 1: Approximated Interval Estimation

β	1				0.9				0.8			
	U1	U2	L1	L2	U1	U2	L1	L2	U1	U2	L1	L2
$n = 5$	.402	.322	1.00	1.00	.341	.267	1.00	1.00	.276	.209	1.00	1.00
$n = 10$	.416	.345	1.00	1.00	.327	.264	1.00	1.00	.247	.192	1.00	1.00
$n = 20$	.419	.361	1.00	1.00	.301	.250	1.00	1.00	.207	.165	1.00	1.00
$n = 50$	.416	.373	1.00	1.00	.256	.222	1.00	1.00	.153	.129	1.00	1.00
$n = 100$	.413	.380	1.00	1.00	.212	.189	1.00	1.00	.124	.109	.999	.999
$n = 500$	.405	.389	1.00	1.00	.118	.111	.999	.999	.080	.074	.984	.981
$n = 1000$	.407	.395	1.00	1.00	.097	.093	.995	.994	.070	.067	.972	.970

β	.6				.5				.3			
	U1	U2	L1	L2	U1	U2	L1	L2	U1	U2	L1	L2
$n = 5$	.167	.120	1.00	1.00	.133	.094	1.00	1.00	.095	.067	1.00	.998
$n = 10$	.139	.104	1.00	1.00	.111	.083	1.00	.999	.084	.063	.999	.989
$n = 20$	.112	.089	1.00	.998	.094	.073	.999	.995	.075	**.058**	.990	.974
$n = 50$	.089	.075	.995	.991	.080	.068	.988	.980	.069	**.058**	.974	.961
$n = 100$	.078	.068	.984	.978	.072	.064	.975	.968	.062	**.055**	.965	**.957**
$n = 500$	.064	**.059**	.963	**.959**	**.059**	**.055**	**.959**	**.955**	**.057**	**.053**	**.958**	**.954**
$n = 1000$	**.060**	**.057**	**.960**	**.958**	**.058**	**.055**	**.956**	**.953**	**.055**	**.053**	**.956**	**.954**

β	.1				.01				.001			
	U1	U2	L1	L2	U1	U2	L1	L2	U1	U2	L1	L2
$n = 5$	.075	**.055**	1.00	.975	.068	**.049**	1.00	**.953**	.071	**.052**	1.00	**.951**
$n = 10$	.072	**.055**	.992	**.959**	.068	**.052**	.989	**.951**	.065	**.048**	.988	**.950**
$n = 20$	.066	**.053**	.979	**.956**	.066	**.052**	.976	**.951**	.063	**.050**	.974	**.948**
$n = 50$	.061	**.051**	.966	**.953**	**.058**	**.048**	.964	**.950**	**.059**	**.050**	.965	**.950**
$n = 100$	**.057**	**.050**	.961	**.951**	**.057**	**.051**	.961	**.951**	**.056**	**.049**	**.960**	**.951**
$n = 500$	**.055**	**.051**	**.954**	**.950**	**.054**	**.050**	**.956**	**.952**	**.053**	**.049**	**.954**	**.950**
$n = 1000$	**.052**	**.050**	**.954**	**.951**	**.053**	**.050**	**.953**	**.949**	**.052**	**.049**	**.954**	**.951**

3. When β is from 0.5 to 0.6, both approximations need large samples such as 500 or 1000. If β is larger than 0.6, they need samples larger than 1000 for a good approximation.

Similarly we can make an interval estimation for the smallest eigenvalue, λ_p. Let $m = p - 1$ in Theorem 1, then β/α goes to zero,

$$d_p \xrightarrow{d} \tilde{d}_p = \widetilde{\boldsymbol{W}}_{22}.$$

Since $\widetilde{\boldsymbol{W}}_{22}/\xi_p \sim \chi^2(n-p+1)$,

$$\frac{l_p}{\lambda_p} = \frac{d_p}{\xi_p} \xrightarrow{d} \chi^2(n-p+1).$$

as β/α goes to zero. Using this fact, we can estimate λ_p to lie in the interval

$$\lambda_p \leq (\chi^2_\gamma(n-p+1))^{-1} l_p, \tag{10}$$

at approximately γ confidence level when β/α is sufficiently close to zero.

We now compare (8) and (9) more closely in view of the known results on asymptotic expansion of distribution of sample eigenvalues. Let

$$A_n = \sqrt{\frac{n}{2}}\left(\frac{l_1/\lambda_1}{n} - 1\right), \tag{11}$$

The asymptotic expansion of A_n up to the order $n^{1/2}$ is given by (see Sugiura (1973))

$$F_{A_n}(t) = \Phi(t) - \frac{\sqrt{2}\phi(t)}{3\sqrt{n}}\left((t^2-1) + \frac{1}{2}\sum_{i=2}^{p}\frac{\lambda_i}{\lambda_1 - \lambda_i}\right) + o(n^{-1/2}). \tag{12}$$

Now suppose $x_i,\ i = 1, \ldots, n$, are independently and identically distributed as $\chi^2(1)$ distribution. The normalized variable

$$\tilde{x}_i = \frac{1}{\sqrt{2}}(x_i - 1), \quad i = 1, \ldots, n$$

has zero mean and unit variance. Let

$$B_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{x}_i = \sqrt{\frac{n}{2}}\left(\frac{\sum_{i=1}^{n} x_i}{n} - 1\right). \tag{13}$$

The asymptotic expansion of B_n up to the order $n^{1/2}$ is given by

$$F_{B_n}(t) = \Phi(t) - \frac{\sqrt{2}\phi(t)}{3\sqrt{n}}(t^2-1) + o(n^{-1/2}). \tag{14}$$

Comparing (12) and (14), we notice that if $t^2 > 1$, then the absolute value of the second term in (14) is smaller than that of (12) by the margin

$$\frac{1}{2}\sum_{i=2}^{p}\frac{\lambda_i}{\lambda_1 - \lambda_i} \tag{15}$$

Since l_1/λ_1 is asymptotically distributed as $\chi^2(n)$ when the largest population eigenvalue is infinitely deviated from the others, l_1/λ_1 in (11) is similarly distributed as $\sum x_i$ in (13). In this case (15) vanishes and both expansions (12), (14) coincide. It is naturally conjectured that when the largest population eigenvalue λ_1 is positioned far away from the smaller eigenvalues, we can make an "easier" inference on λ_1. The fact that the term (15) shrinks in that situation supports this conjecture as well as our simulation results.

3.3 Testing Equality of the Smallest Eigenvalues

As in the introduction of Section 11.7.3 of Anderson (2003), the equality of the $p - m$ smallest population eigenvalues

$$\lambda_{m+1} = \cdots = \lambda_p \quad (\text{say } \sigma^2) \tag{16}$$

is equivalent to the covariance structure

$$\boldsymbol{\Sigma} = \boldsymbol{\Phi} + \sigma^2 \boldsymbol{I}_p,$$

where $\boldsymbol{\Phi}$, a positive semidefinite matrix with rank m, represents the variance-covariance matrix of a systematic part and $\sigma^2\boldsymbol{I}_p$ arises from measurement error. If hypothesis (16) is accepted, then it suggests that the systematic part might consist of m independent factors. Need for testing (16) also arises in principal component analysis when the dimension of principal components has to be decided. Once it is accepted and σ^2 is sufficiently small, which might require another hypothesis testing, we could ignore the last $p-m$ principal components.

The likelihood ratio statistic for testing (16) is given (see e.g. Theorem 9.6.1 of Muirhead (1982)) by

$$\boldsymbol{V} = \frac{\prod_{i=m+1}^{p} l_i}{\left(\sum_{i=m+1}^{p} l_i\right)^{p-m}}(p-m)^{p-m},$$

and the critical region is $\boldsymbol{V} \leq c(\gamma)$ for a given significance level γ .

In order to give the critical point $c(\gamma)$, we traditionally make use of the asymptotic convergence

$$-n\log\boldsymbol{V} \xrightarrow{d} \chi^2((p-m+2)(p-m-1)/2), \quad \text{as } n\to\infty. \tag{17}$$

Bartlett adjustment and further refinement on the asymptotic result are found in Section 9.6 of Muirhead (1982). From this convergence, the approximate critical point is given as

$$c(\gamma) = \exp\Big(-n^{-1}\chi^2_\gamma((p-m+2)(p-m-1)/2)\Big) \tag{18}$$

On the other hand we can approximate the critical point $c(\gamma)$ based on the asymptotic result in Theorem 1. We can expect that this approach yields good approximation since in testing hypothesis (16), we often encounter the situation where the eigenvalues $\lambda_{m+1},\ldots,\lambda_p$ are much smaller than the other eigenvalues. The hypothesis (16) with small σ^2 corresponds to the case $\xi_{m+1}=\cdots=\xi_p=1$ in Theorem 1. Consequently we can approximate the distribution of V in (17) by the distribution of

$$\widetilde{\boldsymbol{V}} = \frac{\prod_{i=m+1}^{p} d_i}{\left(\sum_{i=m+1}^{p} d_i\right)^{p-m}}(p-m)^{p-m},$$

where d_i $(i=m+1,\ldots,p)$ are the eigenvalues of Wishart matrix $\boldsymbol{W}_{p-m}(n-m,\boldsymbol{I}_{p-m})$. Even under the distribution $\boldsymbol{W}_{p-m}(n-m,\boldsymbol{I}_{p-m})$, it is not easy to derive analytical expressions for percentage points for $\widetilde{\boldsymbol{V}}$. For $p-m=2$, the distribution function is explicitly given (see 10.7.3. of Anderson (2003)) by

$$P(\widetilde{\boldsymbol{V}} \leq v) = v^{(n-m-1)/2}$$

which gives the critical point $c(\gamma)$ as

$$c(\gamma) = \gamma^{2/(n-m-1)}. \tag{19}$$

Generally a numerical calculation is needed for the exact evaluation of critical points. For this problem, refer to Consul (1967) and Pillai and Nagarsenker (1971).

We made a simulation for the comparison between the above two methods. Let $p=3,\ m=1$ and consider testing the hypothesis $\lambda_2=\lambda_3$. We examined the accuracy of the two critical points (18) and (19) with $\gamma=0.05$ and $\gamma=0.01$ by simulating the probability of $\boldsymbol{V}$ being smaller than these critical points when $\lambda_2=\lambda_3$, that is, the probability of the error of the first kind. We put

Table 2: Simulated Type 1 Error

β	1				.9				.8			
	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2
$n = 5$	.140	.041	.051	.008	.142	.041	.053	.008	.141	.042	.054	.008
$n = 10$	.063	.033	.015	.005	.064	.033	.016	.006	.067	.036	.017	.007
$n = 20$	.038	.026	.007	.004	.039	.028	.008	.005	.041	.030	.008	.005
$n = 50$	.025	.021	.004	.003	.027	.024	.004	.003	.032	.029	.006	.005
$n = 100$	.022	.020	.003	.003	.025	.023	.003	.003	.034	.031	.005	.005
$n = 500$	.016	.016	.002	.002	.029	.029	.005	.005	.043	.043	.008	.008
$n = 1000$	.017	.017	.002	.002	.035	.035	.006	.006	.048	.047	**.009**	**.009**

β	.6				.5				.3			
	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2
$n = 5$	.145	.044	.055	.008	.148	.044	.056	**.009**	.158	.047	.059	**.009**
$n = 10$	.072	.039	.019	.008	.076	.041	.021	.008	.086	.046	.023	**.009**
$n = 20$	**.051**	.036	**.011**	.006	.057	.042	.012	.005	.067	**.049**	.016	**.009**
$n = 50$	.046	.040	**.009**	.008	.053	.047	**.011**	**.009**	.056	**.050**	.012	**.010**
$n = 100$	**.050**	.047	**.010**	**.009**	**.050**	.047	**.010**	**.009**	.052	**.049**	**.010**	**.009**
$n = 500$	**.050**	**.050**	**.010**	**.009**	**.050**	**.050**	**.010**	**.010**	.052	**.051**	**.010**	**.010**
$n = 1000$	**.051**	**.050**	**.010**	**.010**	**.051**	**.050**	**.010**	**.010**	**.051**	**.051**	**.010**	**.010**

β	.1				.01				.001			
	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2	5%1	5%2	1%1	1%2
$n = 5$	.161	**.049**	.061	**.010**	.164	**.050**	.063	**.011**	.167	**.051**	.064	**.010**
$n = 10$	.091	**.051**	.026	**.011**	.092	**.051**	.025	**.009**	.090	**.050**	.024	**.010**
$n = 20$	.066	**.049**	.015	**.010**	.067	**.050**	.016	**.010**	.067	**.050**	.016	**.010**
$n = 50$	.057	**.051**	.012	**.010**	.057	**.051**	.012	**.010**	.055	.048	.012	**.010**
$n = 100$	.053	**.050**	**.011**	**.010**	.053	**.049**	**.011**	**.010**	.054	**.051**	**.011**	**.010**
$n = 500$	**.051**	**.051**	**.010**	**.010**	**.050**	**.049**	**.010**	**.010**	**.051**	**.050**	**.011**	**.010**
$n = 1000$	**.050**	**.050**	**.009**	**.009**	**.051**	**.051**	**.010**	**.010**	**.051**	**.050**	**.010**	**.010**

$\lambda_1 = 1,\ \lambda_2 = \beta,\ \lambda_3 = \beta$ and varied both β and n. Table 2 shows the result, where the labels 5(1)%1 and 5(1)%2 indicate that the numbers below correspond to the critical points (18) and (19) respectively with $\gamma = 0.05(0.01)$. Numbers in bold mean that they are within ± 0.001 deviation from the desired value.

We can summarize the result as follows;

1. If $\beta \geq 0.8$, both (18) and (19) need a large sample size. Especially when β is as large as 1.0 or 0.9, more than 1000 samples are required for a good approximation. There is no meaningful difference between both critical points.

2. If β equals 0.6 or 0.5, 50 (sometimes 20) samples are large enough to give a good approximation for both (18) and (19). There is no significant difference between both critical points.

3. If $\beta < 0.5$, (19) shows significantly better performance than (18). Even with such a small sample as 5, (19) gives very accurate approximations. The critical point (19) is robust in the sense that it already gives an excellent approximation when the smallest eigenvalues are 0.3 times as large as the largest eigenvalue.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed., Wiley, New Jersey.

[2] Anderson, T. W. and Das Gupta, S. (1964). A monotonicity property of the power functions of some tests of the equality of two covariance matrices. *Ann. Math. Statist.*, **35**, 1059-1063.

[3] Consul, P C. (1967). On the exact distributions of the criterion W for testing sphericity and in a p-variate normal distribution *Ann. Math. Statist.*, **38**, 1170-1174

[4] Hanumara, R. C. and Thompson, W. A. (1968). *Percentage points of the extreme roots of a Wishart matrix Biometrika*, **55**, 505-512.

[5] Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, California.

[6] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

[7] K. C. S. Pillai and Nagarsenker, B N. (1971). On the distribution of the sphericity test criterion in classical and complex normal populations having unknown covariance matrices. *Ann. Math. Statist.*, **42**, 764-767.

[8] Sheena, Y. and Takemura, A. (2007a). An asymptotic expansion of Wishart distribution when the population eigenvalues are infinitely dispersed. *Statistical Methodology*, **4**, 158-184.

[9] Sheena, Y. and Takemura, A. (2007b). Asymptotic distribution of Wishart matrix for block-wise dispersion of population eigenvalues. *Journal of Multivariate Analysis*, doi:10.1016/j.jmva.2007.04.001

[10] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Science Press, Columbus, Ohio.

[11] Sugiura, N. (1973) . Derivatives of the characteristic root of a symmetric or Hermitian matrix with two applications in multivariate analysis. *Comm. Statist.*, **1**, 393-417.

[12] Takemura, A. and Sheena, Y. (2005). Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues are infinitely dispersed and its application to minimax estimation of covariance matrix. *Journal of Multivariate Analysis*, **94**, 271-299.

[13] Thompson, W. A. (1962). Estimation of dispersion parameters. *J. Res. Natl. Bur. Standards Sec. B*, **66**, 161-164.