Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 6535 – 6542
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A test for the equality of covariance matrices when the dimension is large relative to the sample sizes

James R. Schott

Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL 32816-2370, USA

Received 28 September 2006; received in revised form 26 February 2007; accepted 2 March 2007
Available online 12 March 2007

Abstract

A simple statistic is proposed for testing the equality of the covariance matrices of several multivariate normal populations. The asymptotic null distribution of this statistic, as both the sample sizes and the number of variables go to infinity, is shown to be normal. Consequently, this test can be used when the number of variables is not small relative to the sample sizes and, in particular, even when the number of variables exceeds the sample sizes. The finite sample size performance of the normal approximation for this method is evaluated in a simulation study.

Keywords: Equal covariance matrices; High-dimensional data; Singular sample covariance matrix

1. Introduction

Analyses involving a large number of variables, m say, are becoming more prevalent in statistical applications. For instance, analyses of DNA microarray data typically have thousands of gene expressions but these are obtained on a group of individuals which often numbers much less than 100. For some examples, see Dudoit et al. (2002) and Ibrahim et al. (2002). Most of the statistical procedures currently in use are not well suited for this situation since they typically are based on asymptotic theory which has the sample sizes approaching infinity while the number of variables is fixed. Thus, many of these procedures will only be reliable when the sample sizes are substantially larger than m. A better approach in this high-dimensional data setting would be to use a procedure which is based on asymptotic theory which has both m and the sample sizes approaching infinity. Some examples of recent work on inference problems in this high-dimensional setting include Birke and Dette (2005), Ledoit and Wolf (2002), Fujikoshi (2004), Schott (2006), and Srivastava (2005).

In this paper, we consider tests for the equality of the covariance matrices of g m-dimensional multivariate normal populations. That is, if Σ_i denotes the $m \times m$ covariance matrix of the ith population, we wish to test the null hypothesis

$$\mathrm{H}_0 : \Sigma_1 = \cdots = \Sigma_g.$$

Suppose we have independent estimates, $S_1, \ldots, S_g$, of the covariance matrices $\Sigma_1, \ldots, \Sigma_g$ with $n_i S_i \sim W_m(\Sigma_i, n_i)$, that is, $n_i S_i$ has a Wishart distribution with n_i degrees of freedom and covariance matrix Σ_i. The modified likelihood ratio

E-mail address: jschott@ucf.edu.

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.03.004

test of H_0 is based on the statistic (see, for example, Muirhead, 1982, Section 8.2)

$$M = n \log |S| - \sum_{i=1}^{g} n_i \log |S_i|,$$

where $S = \sum_{i=1}^{g} n_i S_i / n$ and $n = \sum_{i=1}^{g} n_i$. In particular, if m is fixed, then the asymptotic null distribution of M, as $n_i \to \infty$ for $i = 1, \ldots, g$, is chi-squared with $v = (g-1)m(m+1)/2$ degrees of freedom. Since the sample covariance matrix S_i is singular if $n_i < m$, this likelihood ratio test is valid only if $n_i \geqslant m$ for $i = 1, \ldots, g$. An alternative test of H_0 having a wider range of applicability is based on the Wald statistic (Schott, 2001)

$$W = \frac{n}{2} \left\{ \sum_{i=1}^{g} \frac{n_i}{n} \operatorname{tr}(S_i S^{-1} S_i S^{-1}) - \sum_{i=1}^{g} \sum_{j=1}^{g} \frac{n_i n_j}{n^2} \operatorname{tr}(S_i S^{-1} S_j S^{-1}) \right\}.$$

This statistic has the same asymptotic null distribution as M and is valid as long as S is nonsingular, that is, as long as $n \geqslant m$.

The purpose of this paper is to develop a procedure for testing H_0 that performs well even when m is substantially larger than each of the n_i's. For situations in which m is very large, such as in the analysis of microarray data, it may be unrealistic to expect that H_0 holds. In this setting, the p-value from a test of H_0 may instead be used to quantify the closeness of the Σ_i's. A more realistic situation in which the procedure developed in this paper would be needed for a formal test of H_0 is when m is not particularly large, but the n_i's are very small. Such a situation is illustrated in Section 4.

We base our test on a fairly simple statistic computed from the sample covariance matrices, $S_1, \ldots, S_g$. It is shown that the asymptotic null distribution of this statistic, as $n_1, \ldots, n_g$, and m all approach infinity, is normal. In addition, some simulation results are obtained to assess the adequacy of the normal approximation and to compare the performance of this new test with that based on M and W.

2. A high-dimensional test of H_0

The statistic we propose for testing the equality of covariance matrices in high-dimensional settings is similar to some statistics that have been proposed for testing the equality of mean vectors in this same setting. For instance, Bai and Saranadasa (1996) based a test of the equality of two mean vectors on the sum of squared differences between the two sample mean vectors, that is, the statistic $(\bar{x}_1 - \bar{x}_2)'(\bar{x}_1 - \bar{x}_2)$. An extension to a test of the equality of g mean vectors was given in Schott (2007). In a similar fashion, we will base our test of H_0 on the sum of squared differences between elements of the sample covariance matrices. The expected value of this statistic is given by

$$\sum_{i<j} E[\operatorname{tr}\{(S_i - S_j)^2\}] = \sum_{i<j} (\operatorname{tr}\{(\Sigma_i - \Sigma_j)^2\} + n_i^{-1}[\operatorname{tr}(\Sigma_i^2) + \{\operatorname{tr}(\Sigma_i)\}^2] + n_j^{-1}[\operatorname{tr}(\Sigma_j^2) + \{\operatorname{tr}(\Sigma_j)\}^2]).$$

Since

$$E(\eta_i^{-1}[n_i(n_i-2)\operatorname{tr}(S_i^2) + n_i^2\{\operatorname{tr}(S_i)\}^2]) = \operatorname{tr}(\Sigma_i^2) + \{\operatorname{tr}(\Sigma_i)\}^2,$$

where $\eta_i = (n_i+2)(n_i-1)$, we find that if we define

$$\begin{aligned} t_{nm} &= \sum_{i<j} (\operatorname{tr}\{(S_i - S_j)^2\} - (n_i\eta_i)^{-1}[n_i(n_i-2)\operatorname{tr}(S_i^2) + n_i^2\{\operatorname{tr}(S_i)\}^2] \\ &\quad - (n_j\eta_j)^{-1}[n_j(n_j-2)\operatorname{tr}(S_j^2) + n_j^2\{\operatorname{tr}(S_j)\}^2]) \\ &= \sum_{i<j} [\{1-(n_i-2)/\eta_i\}\operatorname{tr}(S_i^2) + \{1-(n_j-2)/\eta_j\}\operatorname{tr}(S_j^2) - 2\operatorname{tr}(S_i S_j) \\ &\quad - n_i\eta_i^{-1}\{\operatorname{tr}(S_i)\}^2 - n_j\eta_j^{-1}\{\operatorname{tr}(S_j)\}^2], \end{aligned} \tag{1}$$

then

$$E(t_{nm}) = \sum_{i<j} \mathrm{tr}\{(\Sigma_i - \Sigma_j)^2\}.$$

Thus, the mean of t_{nm} is 0 if and only if H_0 holds. The variance of t_{nm} under H_0 can be expressed as

$$\sigma^2_{t_{nm}} = c_1\,\mathrm{tr}(\Sigma^4) + c_2\{\mathrm{tr}(\Sigma^2)\}^2,$$

where

$$c_1 = \sum_{i<j} \frac{4\{n_i n_j (n_i + n_j)^2 + (2n_i - n_j)(2n_j - n_i)(n_i + n_j) - 2(3n_i^2 + n_i n_j + 3n_j^2) + 8\}}{n_i n_j \eta_i \eta_j}$$
$$+ (g-1)(g-2)\sum_{i=1}^{g} \frac{4(n_i - 2)}{n_i \eta_i},$$

$$c_2 = \sum_{i<j} \frac{4\{n_i n_j (n_i + n_j)^2 + 3n_i n_j (n_i + n_j) - 2(2n_i^2 + n_i n_j + 2n_j^2) - 4(n_i + n_j) + 8\}}{n_i n_j \eta_i \eta_j}$$
$$+ (g-1)(g-2)\sum_{i=1}^{g} \frac{4}{\eta_i}.$$

The main result of this paper will establish the asymptotic normality of t_{nm} as $n_1, \ldots, n_g$ and m approach infinity. We can coordinate the movement of $n_1, \ldots, n_g$ and m to infinity by introducing a common index h and writing $n_1, \ldots, n_g$ and m as $n_{1h}, \ldots, n_{gh}$ and m_h where $h = 1, 2, \ldots$. We assume that for $i = 1, \ldots, g$, $\lim m_h/n_{ih} = b_i$ as $h \to \infty$, where $b_i \in [0, \infty)$ and $b_i > 0$ for at least one i. In a similar fashion, we could write the covariance matrices as $\Sigma_{1h}, \ldots, \Sigma_{gh}$ since they also depend on the index h due to the fact that they are $m_h \times m_h$ matrices. However, for notational convenience, the dependence of all these quantities on h will be suppressed throughout this paper. We will need the following conditions, regarding the limiting properties of the common covariance matrix Σ, to hold:

$$\lim m^{-1}\,\mathrm{tr}(\Sigma^i) = \gamma_i \in (0, \infty) \tag{2}$$

for $i = 1, \ldots, 8$. Note that under these conditions,

$$\lim \sigma^2_{t_{nm}} = \sum_{i<j} 4(b_i + b_j)^2\gamma_2^2 + (g-1)(g-2)\sum_{i=1}^{g} 4b_i^2\gamma_2^2 = \theta^2. \tag{3}$$

Theorem 1. *Suppose that the sample covariance matrices $S_1, \ldots, S_g$ have been computed from independent random samples from multivariate normal distributions with covariance matrices $\Sigma_1, \ldots, \Sigma_g$, respectively. If $\Sigma_1 = \cdots = \Sigma_g$ and the conditions in* (2) *hold, then t_{nm} converges in distribution to a normal random variable with mean* 0 *and variance θ^2.*

In order to use t_{nm} in practice, we will need to estimate θ^2 and this involves finding an estimator of γ_2. Now since under H_0, $nS \sim W_m(\Sigma, n)$, it follows that

$$E[\{\mathrm{tr}(S)\}^2] = 2n^{-1}\,\mathrm{tr}(\Sigma^2) + \{\mathrm{tr}(\Sigma)\}^2$$

and

$$E\{\mathrm{tr}(S^2)\} = n^{-1}(n+1)\mathrm{tr}(\Sigma^2) + n^{-1}\{\mathrm{tr}(\Sigma)\}^2,$$

from which we get $E(a) = \mathrm{tr}(\Sigma^2)$, where

$$a = n^2(n+2)^{-1}(n-1)^{-1}[\mathrm{tr}(S^2) - n^{-1}\{\mathrm{tr}(S)\}^2].$$

Thus, it follows from (2) that $\hat{\gamma}_2 = a/m$ is an asymptotically unbiased estimator of γ_2. Further,

$$E[\{\mathrm{tr}(S^2)\}^2] = n^{-3}(8n^2+20n+20)\mathrm{tr}(\Sigma^4) + n^{-3}(16n+16)\mathrm{tr}(\Sigma^3)\mathrm{tr}(\Sigma) + n^{-3}(n^3+2n^2+5n+4)\{\mathrm{tr}(\Sigma^2)\}^2 + n^{-3}(2n^2+2n+8)\mathrm{tr}(\Sigma^2)\{\mathrm{tr}(\Sigma)\}^2 + n^{-2}\{\mathrm{tr}(\Sigma)\}^4,$$

$$E[\mathrm{tr}(S^2)\{\mathrm{tr}(S)\}^2] = n^{-3}(24n+24)\mathrm{tr}(\Sigma^4) + n^{-3}(8n^2+8n+16)\mathrm{tr}(\Sigma^3)\mathrm{tr}(\Sigma) + n^{-3}(2n^2+2n+8)\{\mathrm{tr}(\Sigma^2)\}^2 + n^{-2}(n^2+n+10)\mathrm{tr}(\Sigma^2)\{\mathrm{tr}(\Sigma)\}^2 + n^{-1}\{\mathrm{tr}(\Sigma)\}^4,$$

$$E[\{\mathrm{tr}(S)\}^4] = 48n^{-3}\,\mathrm{tr}(\Sigma^4) + 32n^{-2}\,\mathrm{tr}(\Sigma^3)\mathrm{tr}(\Sigma) + 12n^{-2}\{\mathrm{tr}(\Sigma^2)\}^2 + 12n^{-1}\,\mathrm{tr}(\Sigma^2)\{\mathrm{tr}(\Sigma)\}^2 + \{\mathrm{tr}(\Sigma)\}^4,$$

and these lead to

$$\mathrm{var}(a) = (n+2)^{-2}(n-1)^{-2}[\{8n^3+\mathrm{o}(n^3)\}\mathrm{tr}(\Sigma^4) + \{4n^2+\mathrm{o}(n^2)\}\{\mathrm{tr}(\Sigma^2)\}^2].$$

Thus, again using the conditions in (2), we find that $\mathrm{var}(\hat{\gamma}_2)$ converges to 0, and so $\hat{\gamma}_2$ converges in probability to γ_2. In light of (3), a corresponding consistent estimator of θ^2 is given by

$$\hat{\theta}^2 = 4\left\{\sum_{i<j}\left(\frac{n_i+n_j}{n_in_j}\right)^2 + (g-1)(g-2)\sum_{i=1}^{g} n_i^{-2}\right\}a^2, \tag{4}$$

and a test of H_0 can be based on $t^*_{nm} = t_{nm}/\hat{\theta}$ since its asymptotic null distribution is the standard normal distribution. In particular, for a test with significance level α, we would reject H_0 if t^*_{nm} exceeds the $100(1-\alpha)$th quantile of the standard normal distribution.

3. Some simulation results

Some simulation results were obtained so as to assess the effectiveness of the asymptotic normal distribution in approximating the actual null distribution of t^*_{nm}. For simplicity, we generally restricted attention to the situation in which $n_1 = \cdots = n_g$. Both m and n_i varied over the values 4, 8, 16, 32, 64, and 128, and for each setting the significance level was estimated from 1000 simulations. The nominal significance level used was 0.05.

Table 1 has some results when $g = 2$ and the common covariance matrix, Σ, is I_m. For some of the settings with small values for m, the approximation yields inflated significance levels, but in none of the cases was the estimated significance level grossly different from 0.05. We also obtained some results, not tabulated here, when $n_2 = n_1/2$ and these were not substantially different than those given in Table 1. Thus, the information we get from the simulations reported in this section, does not appear to be heavily dependent on the equal sample size constraint that we have used. Table 2 has results, analogous to those given in Table 1, when Σ is block diagonal with each block matrix given by $0.5I_4 + 0.5 1_4 1_4'$, where 1_4 denotes the 4×1 vector which has each of its elements equal to 1. The significance levels in Table 2 are generally larger than those in Table 1 as it appears that the convergence to the standard normal distribution is somewhat slower when $\Sigma \neq I_m$. Some results when $g = 3$ and $\Sigma = I_m$ are given in Table 3. These significance levels

Table 1
Estimated significance levels for t^*_{nm} when $g = 2$ and $\Sigma = I_m$

m	n_i					
	4	8	16	32	64	128
4	0.054	0.069	0.068	0.089	0.071	0.084
8	0.033	0.064	0.055	0.055	0.077	0.065
16	0.040	0.054	0.054	0.051	0.071	0.067
32	0.037	0.050	0.041	0.057	0.047	0.057
64	0.044	0.050	0.043	0.060	0.051	0.055
128	0.044	0.040	0.045	0.054	0.046	0.053

Table 2
Estimated significance levels for t^*_{nm} when $g = 2$ and Σ has block-diagonal structure with each block matrix equal to $0.5I_4 + 0.5 1_4 1_4'$

m	n_i					
	4	8	16	32	64	128
4	0.060	0.106	0.084	0.086	0.100	0.096
8	0.059	0.072	0.078	0.089	0.092	0.090
16	0.047	0.051	0.066	0.072	0.077	0.093
32	0.044	0.046	0.056	0.072	0.063	0.078
64	0.040	0.051	0.055	0.054	0.047	0.067
128	0.042	0.046	0.048	0.052	0.058	0.069

Table 3
Estimated significance levels for t^*_{nm} when $g = 3$ and $\Sigma = I_m$

m	n_i					
	4	8	16	32	64	128
4	0.053	0.070	0.085	0.080	0.086	0.073
8	0.054	0.065	0.063	0.071	0.055	0.052
16	0.056	0.063	0.060	0.058	0.063	0.054
32	0.055	0.052	0.055	0.046	0.050	0.050
64	0.048	0.041	0.057	0.052	0.049	0.050
128	0.049	0.055	0.050	0.058	0.045	0.053

Table 4
Estimated significance levels for the likelihood ratio test when $g = 2$

m	n_i					
	4	8	16	32	64	128
4	0.085	0.052	0.041	0.050	0.051	0.061
8		0.187	0.049	0.054	0.044	0.051
16			0.515	0.045	0.061	0.050
32				0.925	0.055	0.039
64					1.0	0.061
128						1.0

are very similar to those in Table 1 so the performance of the approximation does not seem to be greatly affected by the value of g.

We also obtained simulation results for the test of H_0 based on the likelihood ratio statistic M as well as the Wald statistic W. Since the likelihood ratio procedure generally yielded significance levels closer to the nominal level than did the Wald procedure, we only report results for the likelihood ratio test. In the simulations, we used the F approximation to the null distribution of M as described in Box (1949). Some results when $g = 2$ are given in Table 4. The test yields inflated significance levels when $n_i = m$ and this becomes more pronounced as m increases. Clearly, when $n_i \geqslant 2m$, the asymptotic null distribution of M is well approximated by Box's F approximation. To get some idea of its performance when $m < n_i < 2m$, we obtained some additional results when $n_i = 128$. For instance, for $m = 72$, 80, 88, 96, 104, 112, and 120, we obtained 0.074, 0.083, 0.147, 0.323, 0.683, 0.984, and 1.000, respectively. These results suggest that it may not be appropriate to use the likelihood ratio test if m is much larger than $n_i/2$.

Additional simulations were conducted to estimate power. For $g = 2$, we used $\Sigma_1 = I_m$ while Σ_2 had block-diagonal structure with each block matrix given by diag(1, 1, 1, 2). The results for the test based on t^*_{nm} are given in Table 5. As expected, the power increases as n_i increases; the power is not substantially dependent on the value of m although

Table 5
Estimated power for t^*_{nm} when $g=2$, $\Sigma_1 = I_m$, and Σ_2 is block diagonal with each block matrix equal to diag(1, 1, 1, 2)

m	n_i					
	4	8	16	32	64	128
4	0.082	0.132	0.216	0.388	0.660	0.934
8	0.067	0.116	0.221	0.427	0.741	0.982
16	0.060	0.087	0.197	0.451	0.832	0.998
32	0.051	0.085	0.196	0.441	0.847	1.000
64	0.039	0.091	0.164	0.436	0.884	0.999
128	0.023	0.072	0.166	0.442	0.901	1.000

Table 6
Estimated power for the likelihood ratio test when $g=2$, $\Sigma_1 = I_m$, and Σ_2 is block diagonal with each block matrix equal to diag(1, 1, 1, 2)

m	n_i					
	4	8	16	32	64	128
4	0.099	0.046	0.064	0.115	0.269	0.677
8		0.217	0.050	0.117	0.324	0.778
16			0.527	0.099	0.300	0.851
32				0.957	0.261	0.824
64					1.000	0.756
128						1.000

it is approaching 1 at a slightly slower rate for small values of m. Corresponding power estimates for the likelihood ratio test are given in Table 6. The high power when $n_i = m$ is a direct consequence of the inflated significance levels observed in Table 4. Each power estimate when $n_i > m$ is smaller than the corresponding power estimate for t^*_{nm} in Table 5. This is partially due to the fact that the estimated significance levels for t^*_{nm} in Table 1 are generally slightly larger than those for M in Table 4. However, even when they have similar estimated significance levels, such as when $n_i = 2m$, t^*_{nm} yields substantially higher power than does M.

4. An example

We use some of the biochemical data given in Beerstecher et al. (1950) to illustrate the procedure developed in this paper. These data consist of 62 measurements on each of 12 individuals, 8 of whom were controls while the other 4 were alcoholics. We will restrict attention to one subset of the 62 variables, a set of 8 blood serum measurements, and test the hypothesis that the covariance matrix of these eight variables for alcoholics is the same as the corresponding covariance matrix for nonalcoholics. The likelihood ratio test cannot be applied here since the two sample covariance matrices are singular. Using (1) and (4), we find that $t^*_{nm} = t_{nm}/\hat{\theta} = 3.52$. Upon comparing this to the quantiles of the standard normal distribution, the hypothesis of equal covariance matrices is rejected at any reasonable significance level.

Appendix: Proof of Theorem 1

Since t_{nm} is unaffected by transformations of the form $P'S_iP$, where P is an orthogonal matrix, we may assume without loss of generality that the common covariance matrix Σ is diagonal. Let the (j,k)th element of S_i be denoted by $s_{jk,i}$ and note that it can be written as $s_{jk,i} = n_i^{-1}\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}z'_{j,i}z_{k,i}$, where σ_{jj} denotes the jth diagonal element of Σ and $z_{1,i},\ldots,z_{m,i}$ are independently and identically distributed as $N_{n_i}(0, I_{n_i})$. For $l = 1,\ldots,m$, let

$$X_{nl} = t_{nl} - t_{n,l-1} = \sum_{i<j} \tau_{ij,l},$$

where $\tau_{ij,l}=\rho_{1i,l}+\rho_{1j,l}-\rho_{2ij,l}-\rho_{3i,l}-\rho_{3j,l}$,

$$\rho_{1i,l}=\left\{\frac{1-(n_i-2)}{\eta_i}\right\}\left(2\sum_{h=1}^{l-1}s_{hl,i}^2+s_{ll,i}^2\right),$$

$$\rho_{2ij,l}=2\left(2\sum_{h=1}^{l-1}s_{hl,i}s_{hl,j}+s_{ll,i}s_{ll,j}\right),$$

$$\rho_{3i,l}=n_i\eta_i^{-1}\left(2\sum_{h=1}^{l-1}s_{hh,i}s_{ll,i}+s_{ll,i}^2\right),$$

and $t_{n0}=0$, so that $t_{nm}=\sum_{l=1}^{m}X_{nl}$. If we define the set $\mathscr{F}_{n,l-1}=\{z_{1,i},\ldots,z_{l-1,i},\ i=1,\ldots,g\}$, then when H_0 holds

$$E(s_{hl,i}^2|\mathscr{F}_{n,l-1})=\frac{\sigma_{hh}\sigma_{ll}}{n_i^2}z_{h,i}'z_{h,i},\qquad E(s_{ll,i}^2|\mathscr{F}_{n,l-1})=\frac{\sigma_{ll}^2}{n_i}(n_i+2),$$

$$E(s_{hl,i}s_{hl,j}|\mathscr{F}_{n,l-1})=0,\qquad E(s_{ll,i}s_{ll,j}|\mathscr{F}_{n,l-1})=\sigma_{ll}^2,$$

$$E(s_{hh,i}s_{ll,i}|\mathscr{F}_{n,l-1})=\frac{\sigma_{hh}\sigma_{ll}}{n_i}z_{h,i}'z_{h,i}.$$

Using these identities when computing the conditional expected value of $\tau_{ij,l}$, we find that $E(X_{nl}|\mathscr{F}_{n,l-1})=0$. Consequently, for each n, $\{t_{nl},\ l=1,\ldots,m\}$ is a martingale and $X_{n1},\ldots,X_{nm}$ are martingale differences. As a result, our theorem will follow from Corollary 3.1 of Hall and Heyde (1980, p. 58) if we can show that

$$\sum_{l}E(X_{nl}^2|\mathscr{F}_{n,l-1})\xrightarrow{p}\theta^2 \tag{5}$$

and

$$\sum_{l}E\{X_{nl}^2I(|X_{nl}|>\varepsilon)|\mathscr{F}_{n,l-1}\}\xrightarrow{p}0 \tag{6}$$

for all $\varepsilon>0$. Here $I(\cdot)$ denotes the indicator function. It is easily verified that $E(\tau_{ij,l}^2|\mathscr{F}_{n,l-1})\xrightarrow{p}4(b_i+b_j)^2\gamma_2^2$, $E(\tau_{ij,l}\tau_{ik,l}|\mathscr{F}_{n,l-1})\xrightarrow{p}4b_i^2\gamma_2^2$, and $E(\tau_{ij,l}\tau_{hk,l}|\mathscr{F}_{n,l-1})\xrightarrow{p}0$, from which (5) readily follows. The Lindeberg condition given in (6) can be established by showing that the stronger Liapounov condition

$$\sum_{l}E(X_{nl}^4|\mathscr{F}_{n,l-1})\xrightarrow{p}0 \tag{7}$$

holds. Now it is well known (see, for example, Chow and Teicher, 1978, p. 106) that for any random variables $Y_1,\ldots,Y_N$,

$$E\left|\sum_{i=1}^{N}Y_i\right|^p\leqslant N^{p-1}\sum_{i=1}^{N}E(|Y_i|^p)$$

if $p>1$. Using this result twice, we find that

$$E(X_{nl}^4)\leqslant\left\{\frac{g(g-1)}{2}\right\}^3\sum_{i<j}E(\tau_{ij,l}^4) \tag{8}$$

and

$$E(\tau_{ij,l}^4)\leqslant 5^3\{E(\rho_{1i,l}^{*4})+E(\rho_{1j,l}^{*4})+E(\rho_{2ij,l}^{*4})+E(\rho_{3i,l}^{*4})+E(\rho_{3j,l}^{*4})\}, \tag{9}$$

where we have been able to use the mean-corrected random variables $\rho^*_{1i,l} = \rho_{1i,l} - E(\rho_{1i,l})$, $\rho^*_{2ij,l} = \rho_{2ij,l} - E(\rho_{2ij,l})$, and $\rho^*_{3i,l} = \rho_{3i,l} - E(\rho_{3i,l})$ in (9) since $E(\tau_{ij,l}) = 0$. Straightforward, but tedious calculations reveal that

$$\begin{aligned} E(\rho^{*4}_{1i,l}) = {} & \sigma^4_{ll}[\mathrm{O}(n^{-4})\mathrm{tr}(\Sigma^4) + \mathrm{O}(n^{-5})\mathrm{tr}(\Sigma^3)\mathrm{tr}(\Sigma) + \mathrm{O}(n^{-5})\mathrm{tr}(\Sigma^2)\{\mathrm{tr}(\Sigma)\}^2 \\ & + \mathrm{O}(n^{-4})\{\mathrm{tr}(\Sigma^2)\}^2 + \mathrm{O}(n^{-6})\{\mathrm{tr}(\Sigma)\}^4] + \sigma^5_{ll}[\mathrm{O}(n^{-4})\mathrm{tr}(\Sigma^3) \\ & + \mathrm{O}(n^{-4})\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma) + \mathrm{O}(n^{-5})\{\mathrm{tr}(\Sigma)\}^3] + \sigma^6_{ll}[\mathrm{O}(n^{-3})\mathrm{tr}(\Sigma^2) \\ & + \mathrm{O}(n^{-4})\{\mathrm{tr}(\Sigma)\}^2] + \sigma^7_{ll}\mathrm{O}(n^{-3})\mathrm{tr}(\Sigma) + \sigma^8_{ll}\mathrm{O}(n^{-2}), \end{aligned}$$

$$E(\rho^{*4}_{2ij,l}) = \sigma^4_{ll}[\mathrm{O}(n^{-4})\mathrm{tr}(\Sigma^4) + \mathrm{O}(n^{-4})\{\mathrm{tr}(\Sigma^2)\}^2] + \sigma^6_{ll}\mathrm{O}(n^{-3})\mathrm{tr}(\Sigma^2) + \sigma^8_{ll}\mathrm{O}(n^{-2}),$$

$$\begin{aligned} E(\rho^{*4}_{3i,l}) = {} & \sigma^4_{ll}[\mathrm{O}(n^{-7})\mathrm{tr}(\Sigma^4) + \mathrm{O}(n^{-7})\mathrm{tr}(\Sigma^3)\mathrm{tr}(\Sigma) + \mathrm{O}(n^{-6})\mathrm{tr}(\Sigma^2)\{\mathrm{tr}(\Sigma)\}^2 \\ & + \mathrm{O}(n^{-6})\{\mathrm{tr}(\Sigma^2)\}^2 + \mathrm{O}(n^{-6})\{\mathrm{tr}(\Sigma)\}^4] + \sigma^5_{ll}[\mathrm{O}(n^{-7})\mathrm{tr}(\Sigma^3) \\ & + \mathrm{O}(n^{-5})\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma) + \mathrm{O}(n^{-6})\{\mathrm{tr}(\Sigma)\}^3] + \sigma^6_{ll}[\mathrm{O}(n^{-6})\mathrm{tr}(\Sigma^2) \\ & + \mathrm{O}(n^{-6})\{\mathrm{tr}(\Sigma)\}^2] + \sigma^7_{ll}\mathrm{O}(n^{-6})\mathrm{tr}(\Sigma) + \sigma^8_{ll}\mathrm{O}(n^{-6}), \end{aligned}$$

where Σ here denotes the $(l-1) \times (l-1)$ covariance matrix of the first $l-1$ variates. Using these equations along with (8), (9), and the conditions in (2), we find that

$$E\left\{\sum_l E(X^4_{nl}|\mathscr{F}_{n,l-1})\right\} = \sum_l E(X^4_{nl}) \to 0.$$

This guarantees that (7) holds and so the proof is complete. □

References

Bai, Z.D., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statist. Sinica 6, 311–329.
Beerstecher Jr., E., Sutton, H.E., Berry, H.K., Brown, W.D., Reed, J., Rich, G.B., Berry, L.J., Williams, R.J., 1950. Biochemical individuality. V. Explorations with respect to the metabolic patterns of compulsive drinkers. Arch. Biochem. 29, 27–40.
Birke, M., Dette, H., 2005. A note on testing the covariance matrix for large dimension. Stat. Probab. Lett. 74, 281–289.
Box, G.E.P., 1949. A general distribution theory for a class of likelihood criteria. Biometrika 36, 317–346.
Chow, S.C., Teicher, H., 1978. Probability Theory: Independence, Interchangeability, Martingales. Springer, New York.
Dudoit, S., Fridlyand, J., Speed, T.P., 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. J. Amer. Statist. Assoc. 97, 77–87.
Fujikoshi, Y., 2004. Multivariate analysis for the case when the dimension is large compared to the sample size. J. Korean Statist. Soc. 33, 1–24.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Applications. Academic Press, New York.
Ibrahim, J.G., Chen, M., Gray, R.J., 2002. Bayesian models for gene expression with DNA microarray data. J. Amer. Statist. Assoc. 97, 88–99.
Ledoit, O., Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Statist. 30, 1081–1102.
Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.
Schott, J.R., 2001. Some tests for the equality of covariance matrices. J. Statist. Plann. Inference 94, 25–36.
Schott, J.R., 2006. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. J. Multivariate Anal. 97, 827–843.
Schott, J.R., 2007. Some high dimensional tests for a one-way MANOVA. J. Mult. Anal., to appear.
Srivastava, M.S., 2005. Some tests concerning the covariance matrix in high dimensional data. J. Japan Statist. Soc. 35, 251–272.