mathematics

Article

A Two-Sample Test of High Dimensional Means Based on Posterior Bayes Factor

Yuanyuan Jiang [1] and Xingzhong Xu [1,2,*]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; 3120160590@bit.edu.cn
[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081, China
* Correspondence: xuxz@bit.edu.cn

Abstract: In classical statistics, the primary test statistic is the likelihood ratio. However, for high dimensional data, the likelihood ratio test is no longer effective and sometimes does not work altogether. By replacing the maximum likelihood with the integral of the likelihood, the Bayes factor is obtained. The posterior Bayes factor is the ratio of the integrals of the likelihood function with respect to the posterior. In this paper, we investigate the performance of the posterior Bayes factor in high dimensional hypothesis testing through the problem of testing the equality of two multivariate normal mean vectors. The asymptotic normality of the linear function of the logarithm of the posterior Bayes factor is established. Then we construct a test with an asymptotically nominal significance level. The asymptotic power of the test is also derived. Simulation results and an application example are presented, which show good performance of the test. Hence, taking the posterior Bayes factor as a statistic in high dimensional hypothesis testing is a reasonable methodology.

Keywords: high dimension; mean test; posterior Bayes factor; asymptotic normality

MSC: 62H15

check for updates

Citation: Jiang, Y.; Xu, X. A Two-Sample Test of High Dimensional Means Based on Posterior Bayes Factor. *Mathematics* **2022**, *10*, 1741. https://doi.org/10.3390/math10101741

Academic Editor: María Purificación Galindo Villardón

Received: 5 April 2022
Accepted: 17 May 2022
Published: 19 May 2022

1. Introduction

The likelihood ratio is the primary test statistic in hypothesis testing owing to its dominating power. However, for high dimensional data, the likelihood ratio statistic is sometimes undefined. For example, the likelihood function of a multivariate normal distribution is unbounded when the dimension of data is greater than the sample size. Even if the likelihood ratio is well-defined, its performance is unsatisfactory when the dimension is proportionally "close to" the sample size [1]. Therefore, when the dimension is large relative to the sample size, that is the so-called "large p small n" situation; how to choose a test statistic plays a key role in statistical inference.

In this article, we try to use the posterior Bayes factor to be a test statistic for high dimensional data, applying it to equality testing of two multivariate normal mean vectors. The classical likelihood ratio test statistic is the ratio of the maximum values of likelihoods, whereas the Bayes factor is the ratio of the integrated likelihoods. We chose the posterior Bayes factor rather than the prior Bayes factor because when the dimension is fixed, the former is less affected by the variations of the prior. This paper aims to investigate the ability of the posterior Bayes factor as a test statistic. As a result, a simple prior is taken for the parameter.

In multivariate analysis, testing the equality of two means is a fundamental problem. The classical procedure for this problem is the famous Hotelling T^2 test in [2], which is based on Mahalanobis distance between the sample mean vectors weighted by the inverse sample covariance matrix. Hotelling's T^2 test is the most powerful invariant test when the dimension is fixed and much smaller than the total sample size [3], but it is unsatisfactory when the dimension is large relative to the sample size [1]. However, in recent decades, hypothesis testing viable for high dimensional data is increasingly demanded in many application areas such as genomics, finance, medicine, and so on. An important work [1] modifies Hotelling's T^2 statistic in a high dimensional setting by removing the inverse of the sample covariance from the Hotelling formulation. Some new test statistics for the mean vector are introduced by replacing the sample covariance with its diagonal in [4–6]. In [7], a statistic is constructed by retaining the cross-product terms in work [1]. In the sequel, ref. [8] standardizes each component of $\bar{X} - \bar{Y}$ in [7] by the corresponding variance estimation and proposes a scale-invariant test. The test statistics introduced above are called "sum-of-squares type statistics" (see [9]) and attempt to get around the ill-formed sample covariance matrix. Another major approach called "projecting data" transforms high dimensional data into low dimensional data with random projection so that traditional tests can be applied. See, for example [10–12]. By maximizing an average signal to noise ratio, ref. [13] finds the optimal projection subspace and proposes a new test procedure based on it. Besides the two main approaches mentioned above, ref. [14] studies the rates of convergence for the high-dimensional mean and proposes tests based on the sample mean. A new test based on random subspaces is proposed by [15]. A generalized component test is presented in [16], whose statistic is the average of the squared t-statistics for all the component testing problems. A method using a multiple hypothesis test based on the maximum of standardized partial sums of logarithmic p-values statistic is introduced in [17]. More works about testing the mean vectors are presented in [18–20].

Few articles develop tests for the means of two samples with Bayesian machineries in high dimensional settings. A Bayes factor-based testing procedure is developed by [12]. However, the statistic is still constructed with lower dimensional random projections of the high dimensional data vectors because Bayes factors based on Jeffrey's prior involve inversion of the ill-formed sample covariance matrices, as in the classical Hotelling T^2 test statistic in a "large p small n" setting. The approach of random projection cannot be applied when the difference of two mean vectors is dense. However, whether the difference of two mean vectors is dense or sparse is not known in applications. Aitkin [21] proposed the posterior Bayes factor, which is the ratio of the posterior means of the likelihood under each model rather than the usual prior means. Suppose two models M_0 and M_1 for common data x are considered, under which the likelihood function is $L_i(\boldsymbol{\theta}_j)$, where $\boldsymbol{\theta}_j$ is the parameter of dimension p_j and belongs to the parameter space Θ_j, $j \in \{0,1\}$. Specifying prior $\pi_j(\boldsymbol{\theta}_j)$ to $\boldsymbol{\theta}_j$, $j \in \{0,1\}$, then the posterior Bayes factor in favor of the model M_1, denoted by PBF_{10}, is defined as

$$\mathrm{PBF}_{10} = \frac{\bar{L}_1}{\bar{L}_0}, \tag{1}$$

where $\bar{L}_j$ is the posterior mean:

$$\bar{L}_j = \int_{\Theta_j} L_j(\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j|x,y)\,\mathrm{d}\boldsymbol{\theta}_j, \quad j \in \{0,1\},$$

and $\pi_j(\boldsymbol{\theta}_j|x,y)$ is the posterior density of $\boldsymbol{\theta}_j$:

$$\pi_j(\boldsymbol{\theta}_j|x,y) = \frac{L_j(\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)\,\mathrm{d}\boldsymbol{\theta}_j}{\int_{\Theta_j} L_j(\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j)\,\mathrm{d}\boldsymbol{\theta}_j}, \quad j \in \{0,1\}. \tag{2}$$

Unlike the Bayes factor, which is highly dependent on the prior and may be very sensitive to variations in the prior, the posterior Bayes factor reduces this sensitivity to the prior. Specifically, when model M_0 is a regular submodel of M_1, the logarithm of the posterior Bayes factor under model M_1 has

$$2\ln \mathrm{PBF}_{10} \xrightarrow{d} -v\ln 2 + \chi^2(v),$$

where "$\xrightarrow{d}$" means the convergence in distribution, and $v = p_1 - p_0$ and $\chi^2(v)$ denotes a Chi-square distribution with v degrees of freedom. The asymptotic distribution of the logarithm of the posterior Bayes factor is independent of the prior distribution, which further illustrates that the posterior Bayes factor is insensitive to the prior.

Inspired by [21], we consider testing the equality of two high dimensional means with the posterior Bayes factor. With an appropriate prior, the posterior Bayes factor no longer suffers the impediment of the inversion of ill-formed matrices. Additionally, compared with the approach in [12], which proposed a test based on the Bayes factor with random projections, the posterior Bayes factor can be applied for both dense and sparse cases. In this paper, a non-informative prior also works for the location parameters, while an inverse Wishart prior is taken for the covariance matrix. We establish the asymptotic normality of the logarithm of the posterior Bayes factor under the null hypothesis and derive the asymptotic power of the test. Simulation studies are carried out to investigate the performance of the proposed test. The numerical results show that the power of our test outperforms the competitors in most cases.

The rest of this article is organized as follows. In Section 2, we derive the posterior Bayes factor for testing the equality of two mean vectors in the "large p small n" setting. The asymptotic null distribution of the posterior Bayes factor and the local power function of the test are also presented. Simulation results are given in Section 3. We apply the proposed test to a real dataset in Section 4. Section 5 concludes the paper. Technical proofs and the code for performing the simulation studies are deferred to Appendices A and B.

2. Test Based on Posterior Bayes Factor

This section tries to construct the test based on the posterior Bayes factor. Let $X = (X_1, \ldots, X_{n_1})$ and $Y = (Y_1, \ldots, Y_{n_2})$ be iid samples from p-dimensional multivariate normal distributions $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively, where μ_1 and μ_2 are $p \times 1$ vectors, and Σ is a positively definite $p \times p$ matrix. The goal is to test the hypotheses

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2. \tag{3}$$

In order to test Hypotheses (3) by the posterior Bayes factor, we specify the priors for the parameters μ_1, μ_2 and Σ under both the null and alternative hypotheses as

$$\pi_0(\mu) = 1, \quad \Sigma \sim W_p^{-1}(m_0, V^{-1}), \tag{4}$$

and

$$\pi_1(\mu_1) = \pi_1(\mu_2) = 1, \quad \Sigma \sim W_p^{-1}(m_1, V^{-1}), \tag{5}$$

respectively, where μ is the common mean vector under the null hypothesis, and $W_p^{-1}(m_j, V^{-1})$ is the inverse Wishart distribution with real degrees of freedom m_j and a positive definite matrix V^{-1}, $j \in \{0, 1\}$.

The reasons for choosing the above priors are as follows.

1. When no knowledge about the prior is available, a non-informative prior is suggested. A usual one is Jeffrey's prior. As a result, for the parameters μ_1, μ_2, and the common parameter μ under the null hypothesis, we choose Jeffrey's prior, i.e., Lebesgue measure.
2. For the covariance matrix Σ, the posterior distribution with Jeffrey's prior does not exist when $p > n - 2$, where $n = n_1 + n_2$. Therefore, we take the inverse Wishart distribution, which is a conjugate for a normal covariance matrix.
3. This paper aims to investigate whether the test with the posterior Bayes factor statistic in high dimensional settings performs better than the existing methods. If the results turn out to be as expected, the posterior Bayes factor could be suggested to be the test statistic for high dimensional datasets. Hence, we will take simple priors. Furthermore, we take $V = k\mathrm{I}_p$ in the priors for the covariance matrices with small k so that the variation of the Σ is large.

The joint densities of X and Y under the null and alternative hypotheses are

$$(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\left[\sum_{i=1}^{n_1}(x_i-\mu)(x_i-\mu)^T+\sum_{i=j}^{n_2}(y_j-\mu)(y_j-\mu)^T\right]\right\}$$

and

$$(2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}\left[\sum_{i=1}^{n_1}(x_i-\mu_1)(x_i-\mu_1)^T+\sum_{i=j}^{n_2}(y_j-\mu_2)(y_j-\mu_2)^T\right]\right\},$$

respectively. Then the posterior mean $\bar{L}_1$ under H_1 can be calculated as

$$\bar{L}_1=\int_{\Theta_1}L_1(\boldsymbol{\theta}_1)\left[\frac{L_1(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)}{\int_{\Theta_1}L_1(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}\right]\mathrm{d}\boldsymbol{\theta}_1=\frac{\int_{\Theta_1}L_1^2(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int_{\Theta_1}L_1(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1},$$

where

$$\begin{aligned}&\int_{\Theta_1}L_1(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1\\=&(2\pi)^{-\frac{(n-2)p}{2}}n_1^{-\frac{p}{2}}n_2^{-\frac{p}{2}}\frac{2^{\frac{(m_1+n-2)p}{2}}|V^{-1}+S_1+S_2|^{-\frac{m_1+n-2}{2}}\Gamma_p(\frac{m_1+n-2}{2})}{2^{\frac{m_1p}{2}}|V|^{\frac{m_1}{2}}\Gamma_p(\frac{m_1}{2})},\end{aligned}$$

and $\Gamma_p(\cdot)$ denotes the multivariate gamma function, that is,

$$\Gamma_p(a)=\pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma[a+(1-j)/2].$$

$$S_1=\sum_{i=1}^{n_1}(X_i-\bar{X})(X_i-\bar{X})^T,\quad S_2=\sum_{j=1}^{n_2}(Y_j-\bar{Y})(Y_j-\bar{Y})^T$$

with $\bar{X}=\sum_{i=1}^{n_1}X_i/n_1$, $\bar{Y}=\sum_{j=1}^{n_2}Y_j/n_2$, and

$$\begin{aligned}&\int_{\Theta_1}L_1^2(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_1)\,\mathrm{d}\boldsymbol{\theta}_1\\=&(2\pi)^{-\frac{(2n-2)p}{2}}(2n_1)^{-\frac{p}{2}}(2n_2)^{-\frac{p}{2}}\frac{2^{\frac{(m_1+2n-2)p}{2}}|V^{-1}+2S_1+2S_2|^{-\frac{m_1+2n-2}{2}}\Gamma_p(\frac{m_1+2n-2}{2})}{2^{\frac{m_1p}{2}}|V|^{\frac{m_1}{2}}\Gamma_p(\frac{m_1}{2})}.\end{aligned}$$

The posterior mean $\bar{L}_0$ under H_0 can be calculated as

$$\bar{L}_0=\frac{\int_{\Theta_0}L_0^2(\boldsymbol{\theta}_0)\pi_0(\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{\theta}_0}{\int_{\Theta_0}L_0(\boldsymbol{\theta}_0)\pi_0(\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{\theta}_0},$$

where

$$\begin{aligned}&\int_{\Theta_0}L_0(\boldsymbol{\theta}_0)\pi_0(\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{\theta}_0\\=&(2\pi)^{-\frac{(n-1)p}{2}}n^{-\frac{p}{2}}\frac{2^{\frac{(m_0+n-1)p}{2}}|V^{-1}+S_1+S_2+\frac{n_1n_2}{n}(\bar{X}-\bar{Y})(\bar{X}-\bar{Y})^T|^{-\frac{m_0+n-1}{2}}\Gamma_p(\frac{m_0+n-1}{2})}{2^{\frac{m_0p}{2}}|V|^{\frac{m_0}{2}}\Gamma_p(\frac{m_0}{2})}\end{aligned}$$

and

$$
\begin{aligned}
&\int_{\Theta_0} L_0^2(\boldsymbol{\theta}_0)\pi_0(\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{\theta}_0 \\
=&(2\pi)^{-\frac{(2n-1)p}{2}}(2n)^{-\frac{p}{2}} \\
&\times \frac{2^{\frac{(m_0+2n-1)p}{2}}|V^{-1}+2S_1+2S_2+4\frac{n_1n_2}{n}(\bar{X}-\bar{Y})(\bar{X}-\bar{Y})^T|^{-\frac{m_0+2n-1}{2}}\Gamma_p(\frac{m_0+2n-1}{2})}{2^{\frac{m_0p}{2}}|V|^{\frac{m_0}{2}}\Gamma_p(\frac{m_0}{2})}.
\end{aligned}
$$

For simplicity, we specify $m_0 = m+1$ and $m_1 = m+2$. Then the posterior Bayes factor in favor of H_1 against H_0 in (3) denoted by PB admits an expression as

$$
\mathrm{PB}(X,Y) = (\frac{1}{2})^{p/2}\frac{\left[1+2\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+2(S_1+S_2))^{-1}(\bar{X}-\bar{Y})\right]^{\frac{m+2n}{2}}}{\left[1+\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+S_1+S_2)^{-1}(\bar{X}-\bar{Y})\right]^{\frac{m+n}{2}}}. \tag{6}
$$

Multiplying the logarithm of the posterior Bayes factor by 2, we have

$$
\begin{aligned}
2\ln\mathrm{PB}(X,Y) =& -p\ln 2 + (m+2n)\ln\Big[1+2\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+2(S_1+S_2))^{-1}(\bar{X}-\bar{Y})\Big] \\
&-(m+n)\ln\Big[1+\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+S_1+S_2)^{-1}(\bar{X}-\bar{Y})\Big].
\end{aligned} \tag{7}
$$

Now we want to determine a critical value c_α, which makes the test given by the rejection region

$$
\{(x,y): 2\ln\mathrm{PB}(x,y) \geq c_\alpha\}
$$

have a significance level α. Since the distribution of $2\ln\mathrm{PB}(X,Y)$ under the null hypothesis is unknown, the critical value c_α is determined by means of the asymptotic distribution of it. In order to obtain its asymptotic distribution, Taylor series expansion of the logarithm function $\ln(1+x)$ in (7) around 0 is carried out, which can be summarized as

$$
2\ln\mathrm{PB}(X,Y) = -p\ln 2 + (m+2n)\left[A_1 - \frac{A_1^2}{2(1+A_1^*)^2}\right] - (m+n)\left[A_2 - \frac{A_2^2}{2(1+A_2^*)^2}\right], \tag{8}
$$

where

$$
A_1 = 2\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+2(S_1+S_2))^{-1}(\bar{X}-\bar{Y}),
$$

$$
A_2 = \frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+S_1+S_2)^{-1}(\bar{X}-\bar{Y}),
$$

$A_1^* \in (0, A_1)$ and $A_2^* \in (0, A_2)$.

In (8), the quadratic form in $\bar{X}-\bar{Y}$ is

$$
(m+2n)A_1 - (m+n)A_2 = \frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T B(\bar{X}-\bar{Y}),
$$

where

$$
B = 2(m+2n)(V^{-1}+2S_1+2S_2)^{-1} - (m+n)(V^{-1}+S_1+S_2)^{-1}.
$$

Denotes the spectral decomposition of $\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}$ by G^TAG, where $A = \mathrm{diag}(a_1,\ldots,a_p)$. Let $\xi = (\xi_1,\xi_2,\ldots,\xi_p)^T = \sqrt{\frac{n_1n_2}{n}}G\Sigma^{-\frac{1}{2}}(\bar{X}-\bar{Y})$. Then we have

$$
\begin{aligned}
\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T B(\bar{X}-\bar{Y}) &= \xi^T G\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}G^T\xi \\
&= \xi^T GG^TAGG^T\xi \\
&= \sum_{i=1}^p a_i\xi_i^2.
\end{aligned} \tag{9}
$$

When the null hypothesis is true,

$$\sqrt{\frac{n_1 n_2}{n}}(\bar{X} - \bar{Y}) \sim N_p(\mathbf{0}_p, \Sigma),$$

$\xi \sim N_p(\mathbf{0}_p, \mathrm{I}_p)$. The asymptotic distribution of the above formulation can be derived with the following Lemma.

Lemma 1 ([22])**.** *Let $\zeta_{n,i}$, $i \in \{1, \ldots, n\}$, $n = 1, 2, \ldots$, be iid s-dimensional random vectors with mean zero, covariance matrix M and finite fourth moment. For $n = 1, 2, \ldots$, let $\{a_{n,i}\}_{i=1}^{n}$ be real random variables which are independent of $\{\zeta_{n,i}\}_{i=1}^{n}$ and satisfy*

$$\frac{\max_{1 \le i \le n} a_{n,i}^2}{\sum_{i=1}^{n} a_{n,i}^2} \xrightarrow{P} 0. \tag{10}$$

Then

$$\frac{\sum_{i=1}^{n} a_{n,i} \zeta_{n,i}}{\sqrt{\sum_{i=1}^{n} a_{n,i}^2}} \xrightarrow{d} N_s(\mathbf{0}_s, \mathrm{M}).$$

We take $\zeta_{n,i} = \xi_i^2 - 1$, such that $\mathrm{E}\zeta_{n,i} = 0$, $i \in \{1, \ldots, p\}$. From Lemma 1, (9) needs to be normalized by

$$\sqrt{2\sum_{i=1}^{n} a_{n,i}^2} = \sqrt{2\mathrm{tr}A^2} = \sqrt{2\mathrm{tr}(\Sigma B)^2}$$

because $\mathrm{Var}(\zeta_{n,i}) = 2$, $i \in \{1, \ldots, p\}$. To ensure equality,

$$\sum_{i=1}^{p} a_i = \mathrm{tr}A = \mathrm{tr}(B\Sigma)$$

is added to the right side of the equality. By now, we have

$$2\ln \mathrm{PB} + p \ln 2 = \sum_{i=1}^{p} a_i(\xi_i^2 - 1) + \mathrm{tr}(B\Sigma) - (m + 2n)\frac{A_1^2}{2(1 + A_1^*)^2} + (m + n)\frac{A_2^2}{2(1 + A_2^*)^2}.$$

As a result,

$$\begin{aligned}
&\frac{2\ln \mathrm{PB} + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)}}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \\
=&\frac{\sum_{i=1}^{p} a_i(\xi_i^2 - 1)}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} + \frac{\mathrm{tr}(B\Sigma) - \widehat{\mathrm{tr}(B\Sigma)}}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \\
&- (m + 2n)\frac{A_1^2}{2(1 + A_1^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}} + (m + n)\frac{A_2^2}{2(1 + A_2^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}},
\end{aligned} \tag{11}$$

where $\widehat{\mathrm{tr}(B\Sigma)}$ is the estimator of $\mathrm{tr}(B\Sigma)$. We take

$$\widehat{\mathrm{tr}(B\Sigma)} = \mathrm{tr}(BS_n),$$

where $S_n = (S_1 + S_2)/(n - 2)$.

We shall next prove that the first item on the right side of (11) converges in distribution to $N(0,1)$ and the remaining items converge in probability to 0. If a ratio consistent estimator of $\mathrm{tr}(B\Sigma)^2$ is obtained, a test with a level of asymptotical significance α can be constructed by (11). To this end, some usual assumptions are made as follows:

$$\frac{p}{n} \to c \in (0,\infty), \quad \frac{n_1}{n} \to \tau \in (0,1) \quad as\ n \to \infty, \quad and \quad m = O(n). \tag{12}$$

$$\frac{\lambda_1(\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}} \to 0, \tag{13}$$

where $\lambda_1(A)$ is the largest eigenvalue of a matrix A . Let $\delta = \mu_1 - \mu_2$. We also assume

$$\frac{n_1 n_2}{n}\frac{\delta^T \Sigma \delta}{\mathrm{tr}\Sigma^2} \to 0, \tag{14}$$

and

$$\frac{\delta^T \delta}{\mathrm{tr}\Sigma^2} = O(1). \tag{15}$$

Carefully choosing $k = \varepsilon_n / [np\lambda_1(S_1 + S_2)]$, where $\varepsilon_n \to 0$, we ensure that

$$\frac{m+2n}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\frac{A_1^2}{2(1+A_1^*)^2} \xrightarrow{P} 0 \quad and \quad \frac{m+n}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\frac{A_2^2}{2(1+A_2^*)^2} \xrightarrow{P} 0 \tag{16}$$

under condition (12), (13) and (15). See Appendix A for the proof. For the estimator $\mathrm{tr}(BS_n)$, the following theorem shows its property.

Lemma 2. *If conditions (12) and (13) are true, the estimator* $\widehat{\mathrm{tr}(B\Sigma)}$ *satisfies*

$$\frac{\mathrm{tr}(BS_n) - \mathrm{tr}(B\Sigma)}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0. \tag{17}$$

Combining (11) with (16) and (17), we have

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\left[2\ln \mathrm{PB} + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)}\right] - \frac{\sum_{i=1}^p a_i(\xi_i^2 - 1)}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0. \tag{18}$$

By now, the asymptotic distributions of the linear function of the logarithm of the posterior Bayes factor can be derived.

Theorem 1. *Under the conditions in (12) and (13), the posterior Bayes factor* PB *has properties as follows.*

1. *Under the null hypothesis,*

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)}\right] \xrightarrow{d} N(0,1). \tag{19}$$

2. *Under the local alternative in (14) and condition (15),*

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)} - \frac{n_1 n_2}{n}\delta^T B\delta\right] \xrightarrow{d} N(0,1). \tag{20}$$

In order to formulate a test procedure based on Lemma 2, the estimator of $\mathrm{tr}(B\Sigma)^2$ is demanded. The following ratio-consistent estimator for $\mathrm{tr}\Sigma^2$ is proposed in [1]:

$$\widehat{\mathrm{tr}\Sigma^2} = \frac{(n-2)^2}{n(n-1)}\left[\mathrm{tr}S_n^2 - \frac{1}{n-2}(\mathrm{tr}S_n)^2\right].$$

Inspired by it, we propose the following estimator for $\mathrm{tr}(B\Sigma)^2$:

$$\widehat{tr(B\Sigma)^2} = \mathrm{tr}(BS_n)^2 - \frac{1}{n-2}[\mathrm{tr}(BS_n)]^2. \tag{21}$$

The following theorem shows the property of $\widehat{\mathrm{tr}(B\Sigma)^2}$.

Theorem 2. *Under the conditions in (12) and (13), the estimator $\widehat{\mathrm{tr}(B\Sigma)^2}$ is a ratio-consistent estimator of $\mathrm{tr}(B\Sigma)^2$, which means*

$$\frac{\widehat{\mathrm{tr}(B\Sigma)^2}}{\mathrm{tr}(B\Sigma)^2} \xrightarrow{P} 1 \quad \text{as } p \text{ and } n \to \infty. \tag{22}$$

By Theorems 1 and 2, we obtain a test statistic for (3),

$$T_{\mathrm{PB}} = \frac{1}{\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}}}[2\ln(\mathrm{PB}) + p\ln 2 - \mathrm{tr}(BS_n)],$$

which is asymptotically distributed as $N(0,1)$ when the null hypothesis is true. Then the rejection region of the test with approximate significance level α is

$$\left\{2\ln(\mathrm{PB}) \geq z_{1-\alpha}\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}} - p\ln 2 + \mathrm{tr}(BS_n)\right\},$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$.

Previous results allow us to investigate the asymptotic power of the proposed test. By Theorem 1, the following conclusion is obtained.

Corollary 1. *Under the conditions in (12), (13), the local alternative (14) and condition (15), the power of the posterior Bayes factor-based test is*

$$\beta_{PBF}(\delta) - \mathrm{E}_{\Sigma}\left[\Phi\left(\frac{\frac{n_1 n_2}{n}\delta^T B\delta}{\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}}} - z_{1-\alpha}\right)\right] \to 0, \tag{23}$$

where "E_{Σ}" means the expectation about random variance S_n.

3. Simulation

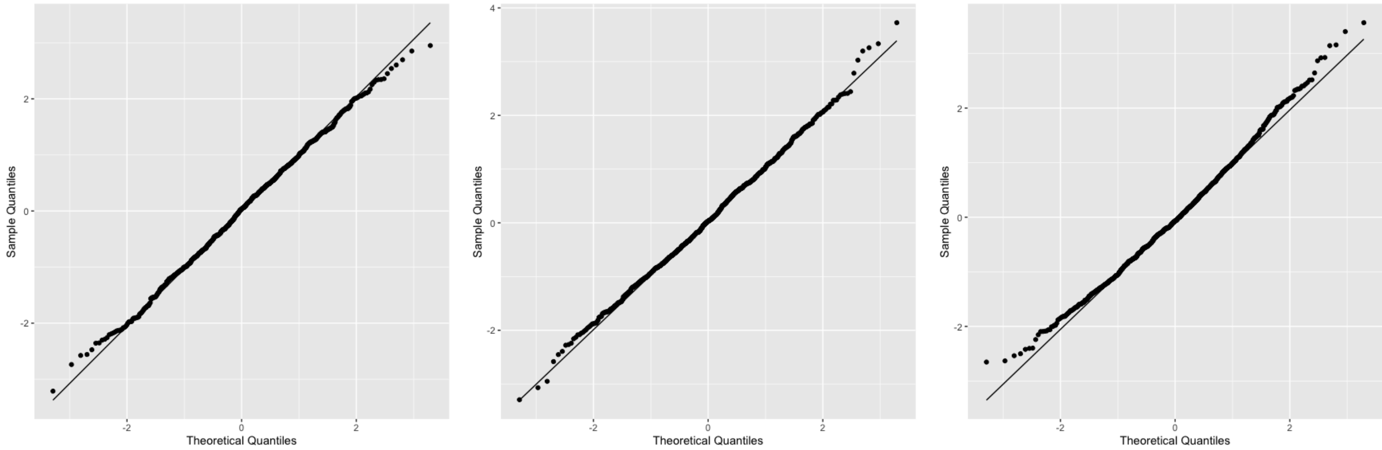
In this section, we conduct simulation studies using R language to evaluate the performance of the posterior Bayes factor-based test for various scenarios. The significance level is set to $\alpha = 0.05$ in all the simulations; $p = 1000$ and the sample sizes are $n_1 = n_2 = 70$. The data $X_1, \ldots, X_{n1}$ and $Y_1, \ldots, Y_{n2}$ are generated from multivariate normal distributions $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively.

We consider the following choices for $\Sigma = ((\sigma_{i,j}))$.

1. $\Sigma_1 = \mathrm{I}_p$ is the identity matrix.
2. Σ_2 is a covariance matrix with $\sigma_{i,j} = 0.4^{|i-j|}$.
3. Σ_3 is block diagonal matrix, with block $B_{25\times 25}$ in which the diagonal entries are 1 and the off-diagonal entries are 0.15.

Σ_1 is for independent cases, while Σ_2 and Σ_3 are for dependent cases.

Theorem 1 shows that T_{PB} is a linear function of the logarithm of the posterior Bayes factor, which is asymptotically distributed as $N(0,1)$. Q–Q plots are presented in Figure 1 to reveal the asymptotic behavior of T_{PB} for $\mu_1 = \mu_2 = \mathbf{0}_{p\times 1}$ and different choices of Σ. We can see that points in Figure 1a–c are closely aligned along the identity line, indicating that the distributions of T_{PB} with different Σ are close to $N(0,1)$.



(a) $\Sigma = \Sigma_1$ **(b)** $\Sigma = \Sigma_2$ **(c)** $\Sigma = \Sigma_3$

Figure 1. Quantile–Quantile plot of asymptotic distribution for T_{PBF} under the null hypothesis $\mu_1 = \mu_2$ against $N(0,1)$ for different Σ based on 1000 independently generated T_{PBF} with $n_1 = n_2 = 70$, $p = 1000$.

We also compare the empirical significance levels and powers of the proposed test with several other tests, including not only tests based on the sum-of-squares-type statistics in [4], referred to as SD, and [7], referred to as CQ, but also a Bayes factor-based test which relies on two random projection approaches as in [12], referred to as RMPBT$_1$ and RMPBT$_2$. In this section, the test we proposed is denoted as PB. The results of SD, CQ and RMPBT are cited from [12].

As in [12], we consider two possible alternatives, as follows. Without loss of generality, we shall always take $\mu_2 = \mathbf{0}_{p\times 1}$ in the simulations. The proportion of entries of the vector $\delta = \mu_1 - \mu_2$ that are exactly zero is denoted by p_0.

1. Simulate $\mu_1 \sim N_p(\mathbf{1}, \mathrm{I}_p)$, set p_0 randomly selected elements to 0, and scale μ_1 so that $\delta^T \Sigma^{-1} \delta = 2$.
2. Simulate $\mu_1 \sim N_p(\mathbf{1}, \mathrm{I}_p)$, set p_0 randomly selected elements to 0, and scale μ_1 so that $\frac{||\delta||^2}{\sqrt{\mathrm{tr}\Sigma^2}} = 0.1$.

We take $p_0 = 0.5, 0.75, 0.80, 0.95, 0.975, 1$. Note that the case $p_0 = 1$ corresponds to the null hypothesis and the power becomes the empirical level. A larger p_0 corresponds to a more sparse alternative, while a smaller p_0 corresponds to a denser one.

For the PB test, we take $m = 2p$ and $\varepsilon_n = 1/\ln(n)$. The numerical results are calculated from 1000 replications and summarized in Tables 1–4. Table 1 compares the empirical sizes of the tests. In general, the test PB performs best in maintaining the significance level. It can be seen that the estimated sizes of PB are reasonably close to the nominal level 0.05. Tests RMPBT and SD show lower empirical levels than the nominal one, whereas test CQ is a little higher.

Tables 2–4 compare the powers of the tests. Covariance matrix Σ in Tables 2–4 are Σ_1, Σ_2 and Σ_3, respectively. Table 2 shows that our test PB substantively outperforms the other three tests for both dense and sparse alternatives. This implies that our method provides the most powerful test compared with the approaches of [4,7,12] for independent cases. In Table 3, the test PB performs better than its competitors in most cases. In Table 4, PB also performs better than the competitors with dense alternatives. Finally, from Tables 3 and 4, either the prior or the posterior Bayes factor-based tests are better than others. For the dense alternative, the PB test is more powerful than RMPBT.

Table 1. Empirical sizes based on 1000 replications with $\alpha = 0.05$, $n_1 = n_2 = 70$ and $p = 1000$. RMPBT is the approach of [12], SD is the approach of [4] and CQ is the approach of [7].

	PB	$\mathbf{RMPBT_1}$	$\mathbf{RMPBT_2}$	SD	CQ
Σ_1	0.049	0.031	0.030	0.040	0.063
Σ_2	0.052	0.038	0.035	0.037	0.049
Σ_3	0.060	0.060	0.040	0.045	0.063

Table 2. Power analysis of 4 tests assuming the true covariance matrix is $\Sigma = \Sigma_1$; $n_1 = n_2 = 70$ and $p = 1000$. RMPBT is the approach of [12], SD is the approach of [4] and CQ is the approach of [7].

	p_0	PB	$\mathbf{RMPBT_1}$	$\mathbf{RMPBT_2}$	SD	CQ
Alternative 1	0.975	**0.470**	0.332	0.309	0.384	0.450
	0.950	**0.478**	0.388	0.339	0.423	0.474
	0.800	**0.482**	0.337	0.304	0.389	0.448
	0.750	**0.482**	0.348	0.294	0.401	0.470
	0.500	**0.485**	0.372	0.343	0.422	0.473
Alternative 2	0.975	**0.764**	0.685	0.612	0.722	0.761
	0.950	**0.797**	0.694	0.612	0.741	0.775
	0.800	**0.785**	0.660	0.581	0.717	0.762
	0.750	**0.806**	0.695	0.616	0.756	0.789
	0.500	**0.786**	0.677	0.588	0.727	0.767

Table 3. Power analysis of 4 tests assuming the true covariance matrix is $\Sigma = \Sigma_2$. $n_1 = n_2 = 70$ and $p = 1000$; RMPBT is the approach of [12], SD is the approach of [4] and CQ is the approach of [7].

	p_0	PB	$\mathbf{RMPBT_1}$	$\mathbf{RMPBT_2}$	SD	CQ
Alternative 1	0.975	**0.269**	0.259	0.243	0.219	0.266
	0.950	**0.277**	0.249	0.232	0.209	0.258
	0.800	**0.282**	0.261	0.222	0.221	0.270
	0.750	**0.299**	0.264	0.236	0.242	0.284
	0.500	**0.336**	0.303	0.265	0.268	0.326
Alternative 2	0.975	0.783	**0.791**	0.738	0.722	0.768
	0.950	0.780	**0.786**	0.734	0.718	0.766
	0.800	**0.794**	0.755	0.699	0.700	0.756
	0.750	**0.792**	0.772	0.722	0.730	0.785
	0.500	**0.789**	0.753	0.686	0.720	0.766

Table 4. Power analysis of 4 tests assuming the true covariance matrix is $\Sigma = \Sigma_3$; $n_1 = n_2 = 70$ and $p = 1000$. RMPBT is the approach of [12], SD is the approach of [4] and CQ is the approach of [7].

	p_0	PB	$\mathbf{RMPBT_1}$	$\mathbf{RMPBT_2}$	SD	CQ
Alternative 1	0.975	0.296	**0.315**	0.278	0.245	0.294
	0.950	0.303	**0.335**	0.307	0.270	0.311
	0.800	0.332	**0.348**	0.318	0.285	0.343
	0.750	**0.357**	0.327	0.294	0.278	0.331
	0.500	**0.422**	0.414	0.379	0.353	0.401
Alternative 2	0.975	0.785	**0.836**	0.776	0.716	0.755
	0.950	0.801	**0.827**	0.776	0.730	0.782
	0.800	0.795	**0.796**	0.734	0.728	0.775
	0.750	**0.793**	0.790	0.727	0.718	0.764
	0.500	**0.778**	0.774	0.717	0.720	0.761

4. An Application Example

To further explore the practical utility of the posterior Bayes factor-based test, we analyze a real dataset about the small round blue cell tumors (SRBCTs), which is available at https://file.biolab.si/biolab/supp/bi-cancer/projections/info/SRBCT.html, accessed on 1 March 2022.

The SRBCTs are four different childhood tumors including Ewing's family of tumors (EWS), neuroblastoma (NB), non-Hodgkin lymphoma (BL) and rhabdomyosarcoma (RMS). Our interest is in examining the equality of means of the genes between the EWS and the RMS tumor groups. The dataset contains 29 examples of EWS and 25 examples of RMS with 2038 genes. The observed test statistic of PB is $\mathrm{T}_{PB} = 14.19842$ with p-value ≈ 0, indicating a serious deviation from the null hypothesis.

5. Conclusions

In this article, we explore the potential for the posterior Bayes factor to be a statistic for testing the mean equality of two high dimensional populations. A closed form of the posterior Bayes factor is obtained with simple priors for the model parameters. Asymptotic normality of the posterior Bayes factor is established, and the corresponding test is constructed. Numerical studies and a real-life example show the superiority of the test. Therefore, we recommend the posterior Bayes factor as a test statistic for hypothesis testing in high dimensional settings.

Author Contributions: Conceptualization, X.X.; methodology, X.X.; software, Y.J.; validation, Y.J. and X.X.; formal analysis, Y.J. and X.X.; writing—original draft preparation, Y.J.; writing—review and editing, Y.J.; visualization, Y.J.; supervision, X.X.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant number 11471035.

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the anonymous referee for the insightful suggestions and comments, which significantly improve the quality and the exposition of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Proof

The Proof of (16).

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} = \frac{4(m+2n)\left[\frac{n_1 n_2}{n}(\bar{X}-\bar{Y})^T(V^{-1}+2(S_1+S_2))^{-1}(\bar{X}-\bar{Y})\right]^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}$$

Substituting $k\mathrm{I}_p$ for V in B,

$$\begin{aligned} B =& 2(m+2n)(\frac{1}{k}\mathrm{I}_p + 2(S_1+S_2))^{-1} - (m+n)(\frac{1}{k}\mathrm{I}_p + S_1 + S_2)^{-1} \\ =& k\Big[2(m+2n)(\mathrm{I}_p + 2k(S_1+S_2))^{-1} - (m+n)(\mathrm{I}_p + k(S_1+S_2))^{-1}\Big] \\ =& k(m+3n) \times \frac{\big[2(m+2n)(\mathrm{I}_p + 2k(S_1+S_2))^{-1} - (m+n)(\mathrm{I}_p + k(S_1+S_2))^{-1}\big]}{m+3n}. \end{aligned}$$

Let

$$C = \frac{1}{m+3n}\Big[2(m+2n)(\mathrm{I}_p + 2k(S_1+S_2))^{-1} - (m+n)(\mathrm{I}_p + k(S_1+S_2))^{-1}\Big],$$

then

$$B = k(m+3n)C, \tag{A1}$$

$$\begin{aligned}\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} &= \frac{4(m+2n)k^2\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\mathrm{I}_p+2k(S_1+S_2))^{-1}(\bar{X}-\bar{Y})\right]^2}{(m+3n)k\sqrt{2\mathrm{tr}((C\Sigma)^2)}} \\ &\le \frac{4k(m+2n)\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\right]^2}{\sqrt{2}(m+3n)\sqrt{\mathrm{tr}((C\Sigma)^2)}}\end{aligned} \tag{A2}$$

$$\begin{aligned}\frac{\mathrm{tr}(C\Sigma)^2}{\mathrm{tr}\Sigma^2} &= \frac{\mathrm{tr}(\Sigma + (C-\mathrm{I}_p)\Sigma)^2}{\mathrm{tr}\Sigma^2} \\ &= \frac{\mathrm{tr}\Sigma^2 + 2\mathrm{tr}[(C-\mathrm{I}_p)\Sigma^2] + \mathrm{tr}[(C-\mathrm{I}_p)\Sigma]^2}{\mathrm{tr}\Sigma^2} \\ &= 1 + \frac{2\mathrm{tr}[(C-\mathrm{I}_p)\Sigma^2]}{\mathrm{tr}\Sigma^2} + \frac{\mathrm{tr}[(C-\mathrm{I}_p)\Sigma]^2}{\mathrm{tr}\Sigma^2}\end{aligned}$$

$$\begin{aligned}C \ge &\frac{1}{m+3n}\Big[(m+2n)(\mathrm{I}_p+k(S_1+S_2))^{-1} - (m+n)(\mathrm{I}_p+k(S_1+S_2))^{-1}\Big] \\ = &\frac{n}{m+3n}(\mathrm{I}_p+k(S_1+S_2))^{-1} \\ > &0,\end{aligned}$$

and

$$\begin{aligned}C &\le \frac{1}{m+3n}\Big[2(m+2n)(\mathrm{I}_p+k(S_1+S_2))^{-1} - (m+n)(\mathrm{I}_p+k(S_1+S_2))^{-1}\Big] \\ &= (\mathrm{I}_p+k(S_1+S_2))^{-1} \\ &\le \mathrm{I}_p.\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{I}_p - C = &\frac{1}{m+3n}\Big\{2(m+2n)[\mathrm{I}_p - (\mathrm{I}_p+2k(S_1+S_2))^{-1}] - (m+n)[\mathrm{I}_p - (\mathrm{I}_p+k(S_1+S_2))^{-1}]\Big\} \\ = &\frac{1}{m+3n}\Big\{2(m+2n)2k(S_1+S_2)(\mathrm{I}_p+2k(S_1+S_2))^{-1} - (m+n)k(S_1+S_2)(\mathrm{I}_p+k(S_1+S_2))^{-1}\Big\} \\ \le &\frac{1}{m+3n}\Big[(3m+7n_1+7n_2)k(S_1+S_2)(\mathrm{I}_p+k(S_1+S_2))^{-1}\Big] \\ \le &3k(S_1+S_2)(\mathrm{I}_p+k(S_1+S_2))^{-1}.\end{aligned}$$

Because $k = \varepsilon_n/[np\lambda_1(S_1+S_2)]$, we have $k(S_1+S_2) \le \varepsilon_n/(np)\mathrm{I}_p$. Hence,

$$0 \le \mathrm{I}_p - C \le \frac{3\varepsilon_n}{np}\mathrm{I}_p. \tag{A3}$$

It follows that

$$\left|\frac{\mathrm{tr}[(C-\mathrm{I}_p)\Sigma^2]}{\mathrm{tr}\Sigma^2}\right| \le \frac{\frac{3\varepsilon_n}{np}\mathrm{tr}\Sigma^2}{\mathrm{tr}\Sigma^2} = \frac{3\varepsilon_n}{np} \to 0,$$

and

$$\frac{\mathrm{tr}[(C-\mathrm{I}_p)\Sigma]^2}{\mathrm{tr}\Sigma^2} \le \frac{9\varepsilon_n^2}{n^2p^2} \to 0.$$

Consequently,

$$\frac{\text{tr}(C\Sigma)^2}{\text{tr}\Sigma^2} \xrightarrow{P} 1. \tag{A4}$$

Therefore,

$$\frac{(m+2n)A_1^2}{\sqrt{2\text{tr}(B\Sigma)^2}} \le \frac{4k(m+2n)\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\right]^2 O_p(1)}{\sqrt{2}(m+3n)\sqrt{\text{tr}\Sigma^2}} = \frac{4(m+2n)\varepsilon_n\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\right]^2 O_p(1)}{\sqrt{2}np(m+3n)(n-2)\sqrt{\text{tr}\Sigma^2}\lambda_1(S_n)}. \tag{A5}$$

S_n can be written as

$$S_n = \frac{\sum_{i=1}^{n-2}\Sigma^{1/2}z_iz_i^T\Sigma^{1/2}}{n-2},$$

where z_i, $i \in (1, \dots, n-2)$ are independently distributed according to the normal distribution $N_p(\mathbf{0}_p, \mathrm{I}_p)$. It follows that

$$\text{Var}(\text{tr}S_n) = \frac{\sum_{i=1}^{n-2}\text{Var}(z_i^T\Sigma z_i)}{(n-2)^2}.$$

By elementary calculation, we have

$$\text{E}(z_i^T\Sigma z_i)^2 = 2\text{tr}\Sigma^2 + (\text{tr}\Sigma)^2.$$

Since

$$\text{E}[z_i^T\Sigma z_i] = \text{tr}\Sigma,$$

it follows that,

$$\text{Var}(\text{tr}S_n) = \frac{2\text{tr}\Sigma^2}{n-2}.$$

Hence,

$$\text{tr}S_n = \text{tr}\Sigma\left\{1 + O_p\left[\frac{\sqrt{2\text{tr}\Sigma^2}}{\sqrt{n-2}\text{tr}\Sigma}\right]\right\}. \tag{A6}$$

By (A6), we have

$$\frac{1}{\lambda_1(S_n)} \le \frac{n-2}{\text{tr}S_n} = \frac{n-2}{\text{tr}\Sigma\left\{1 + O_p\left[\frac{\sqrt{2\text{tr}\Sigma^2}}{\sqrt{n-2}\text{tr}\Sigma}\right]\right\}}.$$

Because

$$\frac{\sqrt{2\text{tr}\Sigma^2}}{\sqrt{n-2}\text{tr}\Sigma} = o(1),$$

then (A5) becomes

$$\frac{(m+2n)A_1^2}{\sqrt{2\text{tr}(B\Sigma)^2}} \le \frac{4(m+2n)\varepsilon_n\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\right]^2 O_p(1)}{\sqrt{2}np(m+3n)\sqrt{\text{tr}\Sigma^2}\text{tr}\Sigma\{1+o_p(1)\}}. \tag{A7}$$

Under the null hypothesis,

$$\sqrt{\frac{n_1n_2}{n}}(\bar{X}-\bar{Y}) \sim N_p(\mathbf{0}_p, \Sigma).$$

By elementary calculation, we have

$$\text{E}\left[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\right]^2 = 2\text{tr}\Sigma^2 + (\text{tr}\Sigma)^2.$$

Therefore,

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \leq \frac{4(m+2n)\varepsilon_n\left(2\sqrt{\mathrm{tr}\Sigma^2}/\mathrm{tr}\Sigma+\mathrm{tr}\Sigma/\sqrt{\mathrm{tr}\Sigma^2}\right)O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}}.$$

Because

$$1 \leq \frac{\mathrm{tr}\Sigma}{\sqrt{\mathrm{tr}\Sigma^2}} \leq \sqrt{p}, \tag{A8}$$

we can obtain

$$\frac{4(m+2n)\varepsilon_n\left(2\sqrt{\mathrm{tr}\Sigma^2}/\mathrm{tr}\Sigma+\mathrm{tr}\Sigma/\sqrt{\mathrm{tr}\Sigma^2}\right)O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}} \leq \frac{4(m+2n)\varepsilon_n(2+\sqrt{p})O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}} \xrightarrow{P} 0 \tag{A9}$$

as $\varepsilon_n \xrightarrow{P} 0$.

We can conclude

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0.$$

Because $A_1^* \in (0, A_1)$, we have

$$\frac{(m+2n)A_1^2}{2(1+A_1^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0.$$

Similarly, we can prove that

$$\frac{(m+n)A_2^2}{2(1+A_2^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0.$$

Under the alternative hypothesis,

$$\mathrm{E}\Big[\frac{n_1n_2}{n}(\bar{X}-\bar{Y})^T(\bar{X}-\bar{Y})\Big]^2 = 2\mathrm{tr}\Sigma^2+(\mathrm{tr}\Sigma)^2+4\frac{n_1n_2}{n}\delta^T\Sigma\delta+(\frac{n_1n_2}{n}\delta^T\delta)^2+\frac{2n_1n_2}{n}(\mathrm{tr}\Sigma)\delta^T\delta.$$

(A7) becomes

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \leq \frac{4(m+2n)\varepsilon_n\Big[2\mathrm{tr}\Sigma^2+(\mathrm{tr}\Sigma)^2+4\frac{n_1n_2}{n}\delta^T\Sigma\delta+(\frac{n_1n_2}{n}\delta^T\delta)^2+\frac{2n_1n_2}{n}(\mathrm{tr}\Sigma)\delta^T\delta\Big]O_p(1)}{\sqrt{2}np(m+3n)\sqrt{\mathrm{tr}\Sigma^2}\mathrm{tr}\Sigma\{1+o_p(1)\}} \tag{A10}$$

By (A9),

$$\frac{4(m+2n)\varepsilon_n\Big[2\mathrm{tr}\Sigma^2+(\mathrm{tr}\Sigma)^2+4\frac{n_1n_2}{n}\delta^T\Sigma\delta+(\frac{n_1n_2}{n}\delta^T\delta)^2+\frac{2n_1n_2}{n}(\mathrm{tr}\Sigma)\delta^T\delta\Big]O_p(1)}{\sqrt{2}np(m+3n)\sqrt{\mathrm{tr}\Sigma^2}\mathrm{tr}\Sigma\{1+o_p(1)\}}$$
$$-\frac{4(m+2n)\varepsilon_n\Big[4\frac{n_1n_2}{n}\frac{\delta^T\Sigma\delta}{\sqrt{\mathrm{tr}\Sigma^2}(\mathrm{tr}\Sigma)}+\frac{(\frac{n_1n_2}{n}\delta^T\delta)^2}{\sqrt{\mathrm{tr}\Sigma^2}(\mathrm{tr}\Sigma)}+\frac{\frac{2n_1n_2}{n}\delta^T\delta}{\sqrt{\mathrm{tr}\Sigma^2}}\Big]O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}} \xrightarrow{P} 0$$

Together with (A8) and (A10) becomes

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \leq \frac{4(m+2n)\varepsilon_n\Big[4\frac{n_1n_2}{n}\frac{\delta^T\Sigma\delta}{\mathrm{tr}\Sigma^2}+(\frac{\frac{n_1n_2}{n}\delta^T\delta}{\sqrt{\mathrm{tr}\Sigma^2}})^2+2\frac{\frac{n_1n_2}{n}\delta^T\delta}{\sqrt{\mathrm{tr}\Sigma^2}}\Big]O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}}$$

Under the assumption (14),

$$\frac{(m+2n)A_1^2}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \leq \frac{4(m+2n)\varepsilon_n\left[\left(\frac{\frac{n_1n_2}{n}\delta^T\delta}{\sqrt{\mathrm{tr}\Sigma^2}}\right)^2 + 2\frac{\frac{n_1n_2}{n}\delta^T\delta}{\sqrt{\mathrm{tr}\Sigma^2}}\right]O_p(1)}{\sqrt{2}np(m+3n)\{1+o_p(1)\}}$$

Additionally, with assumption (15),

$$\frac{(m+2n)A_1^2}{2(1+A_1^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0$$

holds as $\varepsilon_n \xrightarrow{P} 0$. Similarly,

$$\frac{(m+n)A_2^2}{2(1+A_2^*)^2\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0.$$

□

Proof of Lemma 2. By (A1) and (A4), we need only show that

$$\frac{\mathrm{tr}(CS_n) - \mathrm{tr}(C\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}} \xrightarrow{P} 0.$$

$$\begin{aligned}
\frac{\mathrm{tr}(CS_n) - \mathrm{tr}(C\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}} &= \frac{\mathrm{tr}(S_n - \Sigma + (C - \mathrm{I}_p)(S_n - \Sigma))}{\sqrt{\mathrm{tr}\Sigma^2}} \\
&= \frac{O_p\left[\sqrt{\frac{2\mathrm{tr}\Sigma^2}{n-2}}\right] + \mathrm{tr}((C-\mathrm{I}_p)(S_n - \Sigma))}{\sqrt{\mathrm{tr}\Sigma^2}} \\
&= \frac{1}{\sqrt{n-2}}O_p(1) + \frac{\mathrm{tr}((C-\mathrm{I}_p)(S_n-\Sigma))}{\sqrt{\mathrm{tr}\Sigma^2}},
\end{aligned}$$

where the second equality follows from (A6).

By Cauchy–Schwarz inequality and (A3),

$$\begin{aligned}
\frac{\left|\mathrm{tr}((C-\mathrm{I}_p)(S_n-\Sigma))\right|}{\sqrt{\mathrm{tr}\Sigma^2}} &\leq \frac{\sqrt{\mathrm{tr}(C-\mathrm{I}_p)^2}\sqrt{\mathrm{tr}(S_n-\Sigma)^2}}{\sqrt{\mathrm{tr}\Sigma^2}} \\
&\leq \frac{\frac{3\varepsilon_n}{np}(n-2)\sqrt{p}\sqrt{\mathrm{tr}(S_n-\Sigma)^2}}{\sqrt{\mathrm{tr}\Sigma^2}}.
\end{aligned}$$

We know that

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

where

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}S_n^2] =& \frac{1}{(n-2)^2}\sum_{i,j=1}^{n-2}\mathrm{E}[z_i^T\Sigma z_i z_j^T\Sigma z_j] \\
=& \frac{1}{(n-2)^2}\left[\sum_{i=1}^{n-2}\mathrm{E}(z_i^T\Sigma z_i)^2 + \sum_{i\neq j}\mathrm{E}[z_i^T\Sigma z_i z_j^T\Sigma z_j]\right] \\
=& \frac{1}{(n-2)^2}\left\{(n-2)[2\mathrm{tr}\Sigma^2 + (\mathrm{tr}\Sigma)^2] + [(n-2)^2 - (n-2)](\mathrm{tr}\Sigma)^2\right\} \\
=& \frac{2}{n-2}\mathrm{tr}\Sigma^2 + (\mathrm{tr}\Sigma)^2.
\end{aligned}$$

Therefore,

$$E\mathrm{tr}(S_n-\Sigma)^2=\frac{2\mathrm{tr}\Sigma^2}{n-2}.$$

Hence

$$\frac{\sqrt{\mathrm{tr}(S_n-\Sigma)^2}}{\sqrt{\mathrm{tr}\Sigma^2}}=\frac{1}{\sqrt{n-2}}O_p(1).$$

$$0\leq\frac{|\mathrm{tr}(\mathrm{I}_p-C)(S_n-\Sigma)|}{\sqrt{\mathrm{tr}\Sigma^2}}\leq 3\sqrt{p}\sqrt{n-2}\frac{\varepsilon_n}{np}O_p(1).$$

We conclude

$$\frac{|\mathrm{tr}(\mathrm{I}_p-C)(S_n-\Sigma)|}{\sqrt{\mathrm{tr}\Sigma^2}}\xrightarrow{P}0.$$

The Lemma is proved. □

Proof of Theorem 1. From (18),

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\left[2\ln(\mathrm{PB})+p\ln 2-\widehat{\mathrm{tr}(B\Sigma)}\right]-\frac{\sum_{i=1}^p a_i(\xi_i^2-1)}{\sqrt{2\sum_{i=1}^p a_i^2}}\xrightarrow{P}0.$$

Under the null hypothesis, we know that

$$\xi\sim N_p(\mathbf{0}_p,\mathbf{I}_p).$$

Hence,

$$\mathrm{E}\xi_i^2=1\quad \mathit{and}\quad \mathrm{Var}(\xi_i^2)=2.$$

We know that

$$\max_{1\leq i\leq p}a_i=\lambda_1(A)=\lambda_1(B\Sigma),$$

and

$$\sum_{i=1}^p a_i^2=\mathrm{tr}A^2=\mathrm{tr}(B\Sigma)^2.$$

As

$$B\leq(m+2n_1+2n_2-\frac{m+n_1+n_2}{2})(V^{-1}+S_1+S_2)^{-1}\leq(\frac{m+3n_1+3n_2}{2})V,$$

and

$$\begin{aligned}B&\geq(m+2n_1+2n_2)(2V^{-1}+2S_1+2S_2)^{-1}-\frac{m+n_1+n_2}{2}(V^{-1}+S_1+S_2)^{-1}\\&=\frac{n_1+n_2}{2}(V^{-1}+S_1+S_2)^{-1}\\&=\frac{n_1+n_2}{2}(\frac{1}{k}\mathrm{I}_p+S_1+S_2)^{-1}\\&\geq\frac{n_1+n_2}{2}\frac{1}{\lambda_1(\frac{1}{k}\mathrm{I}_p+S_1+S_2)}\mathrm{I}_p,\end{aligned}$$

we have

$$\begin{aligned}\frac{\max_{1\leq i\leq p}|a_i|}{\sqrt{\sum_{i=1}^p a_i^2}}&\leq\frac{\frac{m+3n_1+3n_2}{2}k\lambda_1(\Sigma)\lambda_1(\frac{1}{k}\mathrm{I}_p+S_1+S_2)}{\frac{n_1+n_2}{2}\sqrt{\mathrm{tr}\Sigma^2}}\\&=\frac{\lambda_1(\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}}\frac{\frac{m+3n_1+3n_2}{2}k}{\frac{n_1+n_2}{2}}\lambda_1(\frac{1}{k}\mathrm{I}_p+S_1+S_2)\\&=\frac{\lambda_1(\Sigma)}{\sqrt{\mathrm{tr}\Sigma^2}}\frac{m+3n_1+3n_2}{n_1+n_2}\lambda_1(\mathrm{I}_p+kS_1+kS_2).\end{aligned}$$

Because

$$k = \frac{\varepsilon_n}{np\lambda_1(S_1 + S_2)}, \quad \varepsilon_n \to 0,$$

we have

$$\lambda_1(\mathrm{I}_p + kS_1 + kS_2) \le k\lambda_1(S_1 + S_2) + \lambda_1(\mathrm{I}_p) = \frac{\varepsilon_n}{np} + \lambda_1(\mathrm{I}_p) = O(1).$$

By conditions (12) and (13), we have

$$\frac{\max_{1\le i\le p}|a_i|}{\sqrt{\sum_{i=1}^p a_i^2}} \xrightarrow{P} 0.$$

By Lemma 1,

$$\frac{\sum_{i=1}^p a_i(\xi_i^2 - 1)}{\sqrt{2\sum_{i=1}^p a_i^2}} \to N(0,1) \quad \text{as } p \to \infty.$$

Therefore,

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)}\right] \xrightarrow{d} N(0,1).$$

By now, (1) in Theorem 1 has been proved.
Under the alternative hypothesis, we have

$$\sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}(\bar{X} - \bar{Y}) \sim N_p\left(\sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}\delta, \mathrm{I}_p\right).$$

Let

$$\eta = \sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}(\bar{X} - \bar{Y}), \quad \eta_0 = \eta - \sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}\delta,$$

$$\begin{aligned}
\sum_{i=1}^p a_i\xi_i^2 &= \left[\eta_0 + \sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}\delta\right]^T \Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}\left[\eta_0 + \sqrt{\frac{n_1n_2}{n}}\Sigma^{-\frac{1}{2}}\delta\right] \\
&= \eta_0^T\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}\eta_0 + \frac{n_1n_2}{n}\delta^T B\delta + 2\sqrt{\frac{n_1n_2}{n}}\delta^T B\Sigma^{\frac{1}{2}}\eta_0.
\end{aligned}$$

Since

$$\eta_0 \sim N_p(\mathbf{0}_p, \mathrm{I}_p),$$

$$\mathrm{E}\left[\frac{\sqrt{\frac{n_1n_2}{n}}\delta^T B\Sigma^{\frac{1}{2}}\eta_0}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\middle| S_n\right] = 0, \quad \text{and} \quad \mathrm{Var}\left[\frac{\sqrt{\frac{n_1n_2}{n}}\delta^T B\Sigma^{\frac{1}{2}}\eta_0}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\middle| S_n\right] = \frac{\frac{n_1n_2}{n}\delta^T B\Sigma B\delta}{2\mathrm{tr}(B\Sigma)^2}.$$

By (22), (A1) and (A4),

$$\mathrm{Var}\left[\frac{\sqrt{\frac{n_1n_2}{n}}\delta^T B\Sigma^{\frac{1}{2}}\eta_0}{\sqrt{2\mathrm{tr}(B\Sigma)^2}}\middle| S_n\right] - \frac{\frac{n_1n_2}{n}\delta^T C\Sigma C\delta}{2\mathrm{tr}\Sigma^2} \xrightarrow{P} 0.$$

$$\begin{aligned}
\frac{n_1n_2}{n}\frac{\delta^T C\Sigma C\delta}{\mathrm{tr}\Sigma^2} &= \frac{n_1n_2}{n}\frac{\delta^T(C - \mathrm{I}_p + \mathrm{I}_p)\Sigma(C - \mathrm{I}_p + \mathrm{I}_p)\delta}{\mathrm{tr}\Sigma^2} \\
&\le 2\frac{n_1n_2}{n}\frac{\delta^T\Sigma\delta}{\mathrm{tr}\Sigma^2} + 2\frac{n_1n_2}{n}\frac{\delta^T(\mathrm{I}_p - C)\Sigma(\mathrm{I}_p - C)\delta}{\mathrm{tr}\Sigma^2}
\end{aligned}$$

With (A3), we have

$$\frac{n_1 n_2}{n} \frac{\delta^T (\mathrm{I}_p - C)\Sigma(\mathrm{I}_p - C)\delta}{\mathrm{tr}\Sigma^2} \leq \frac{n_1 n_2}{n} \frac{9{\varepsilon_n}^2 \delta^T \Sigma \delta}{n^2 p^2 \mathrm{tr}\Sigma^2}.$$

$$0 \leq \frac{n_1 n_2}{n} \frac{\delta^T C\Sigma C\delta}{\mathrm{tr}\Sigma^2} \leq \frac{n_1 n_2}{n} \frac{\delta^T \Sigma \delta}{\mathrm{tr}\Sigma^2} + \frac{n_1 n_2}{n} \frac{9(\frac{\varepsilon_n}{np})^2 \delta^T \Sigma \delta}{\mathrm{tr}\Sigma^2}.$$

Because $n_1 n_2 \delta^T \Sigma \delta / (n \mathrm{tr}\Sigma^2) \to 0$,

$$\mathrm{Var}\left[\frac{\sqrt{\frac{n_1 n_2}{n}} \delta^T B \Sigma^{\frac{1}{2}} \eta_0}{\sqrt{\widehat{\mathrm{tr}(B\Sigma)^2}}} \middle| S_n \right] \xrightarrow{P} 0.$$

Hence, we can conclude that

$$\frac{\sqrt{\frac{n_1 n_2}{n}} \delta^T B \Sigma^{\frac{1}{2}} \eta_0}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \xrightarrow{P} 0,$$

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \left[\sum_{i=1}^{p} a_i \xi_i^2 - \frac{n_1 n_2}{n} \delta^T B \delta - \eta_0^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} \eta_0 \right] \xrightarrow{P} 0.$$

With (17) and (18), we have

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \left[2\ln(\mathrm{PB}) + p\ln 2 - \frac{n_1 n_2}{n} \delta^T B \delta - \eta_0^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} \eta_0 \right] \xrightarrow{P} 0.$$

In the proof of Theorem 1, we proved that

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \left[\psi^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} \psi - \widehat{\mathrm{tr}(B\Sigma)} \right] \xrightarrow{d} N(0,1),$$

where ψ is a random vector distributed according to $N_p(\mathbf{0}_p, \mathrm{I}_p)$. Hence, we have

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \left[\eta_0^T \Sigma^{\frac{1}{2}} B \Sigma^{\frac{1}{2}} \eta_0 - \widehat{\mathrm{tr}(B\Sigma)} \right] \xrightarrow{d} N(0,1).$$

Therefore,

$$\frac{1}{\sqrt{2\mathrm{tr}(B\Sigma)^2}} \left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)} - \frac{n_1 n_2}{n} \delta^T B \delta \right] \xrightarrow{d} N(0,1).$$

(2) in Theorem 1 has been proved. □

Proof of Theorem 2. Denote the spectral decomposition of S_n by PDP^T, where $D = \mathrm{diag}[d_1, d_2, \ldots, d_p]$. We can rewrite C as

$$\begin{aligned} C =& \frac{1}{m+3n} \left[2(m+2n)(\mathrm{I}_p + 2k(n-2)PDP^T)^{-1} - (m+n)(\mathrm{I}_p + k(n-2)PDP^T)^{-1} \right] \\ \triangleq& PHP^T, \end{aligned}$$

where $H = \mathrm{diag}[h_1, h_2, \ldots, h_p]$ and

$$h_i = \frac{1}{m+3n} \left[\frac{2(m+2n)}{1+2k(n-2)d_i} - \frac{m+n}{1+k(n-2)d_i} \right], i \in \{1, \ldots, p\}.$$

$$
\begin{aligned}
1-h_i &= \frac{k(n-2)d_i}{m+3n}\left[\frac{4(m+2n)}{1+2k(n-2)d_i}-\frac{m+n}{1+k(n-2)d_i}\right] \\
&\leq \frac{3k(n-2)d_i}{1+k(n-2)d_i}
\end{aligned}
$$

Now substitute the expression of k into the inequality,

$$
1-h_i \leq 3(n-2)\frac{\varepsilon_n}{np} \xrightarrow{p} 0. \tag{A11}
$$

$$
\begin{aligned}
&\operatorname{tr}(CS_n)^2-\frac{1}{n-2}(\operatorname{tr}(CS_n))^2 \\
=&\sum_{i=1}^{p}(d_ih_i)^2-\frac{1}{n-2}\left(\sum_{i=1}^{p}d_ih_i\right)^2 \\
=&\sum_{i=1}^{p}[d_i(1+h_i-1)]^2-\frac{1}{n-2}\left[\sum_{i=1}^{p}d_i(1+h_i-1)\right]^2 \\
=&\sum_{i=1}^{p}d_i^2-2\sum_{i=1}^{p}(1-h_i)d_i^2+\sum_{i=1}^{p}(1-h_i)^2d_i^2-\frac{1}{n-2}\left(\sum_{i=1}^{p}d_i-\sum_{i=1}^{p}(1-h_i)d_i\right)^2 \\
=&\sum_{i=1}^{p}d_i^2-2\sum_{i=1}^{p}(1-h_i)d_i^2+\sum_{i=1}^{p}(1-h_i)^2d_i^2-\frac{1}{n-2}\left(\sum_{i=1}^{p}d_i\right)^2+\frac{2}{n-2}\sum_{i=1}^{p}d_i\sum_{i=1}^{p}(1-h_i)d_i-\frac{1}{n-2}\left(\sum_{i=1}^{p}(1-h_i)d_i\right)^2 \\
=&\sum_{i=1}^{p}d_i^2-\frac{1}{n-2}\left(\sum_{i=1}^{p}d_i\right)^2-2\sum_{i=1}^{p}(1-h_i)d_i^2+\frac{2}{n-2}\sum_{i=1}^{p}d_i\sum_{i=1}^{p}(1-h_i)d_i+\sum_{i=1}^{p}(1-h_i)^2d_i^2-\frac{1}{n-2}\left(\sum_{i=1}^{p}(1-h_i)d_i\right)^2.
\end{aligned}
$$

By [1], we know that

$$
\operatorname{tr}S_n^2=\frac{n}{n-2}\operatorname{tr}\Sigma^2+\frac{n}{(n-2)^2}(\operatorname{tr}\Sigma)^2+o_p(\operatorname{tr}\Sigma^2),
$$

and

$$
\frac{1}{n-2}(\operatorname{tr}S_n)^2=\frac{n}{(n-2)^2}(\operatorname{tr}\Sigma)^2+o_p(\operatorname{tr}\Sigma^2).
$$

Noting that $p/n\in(0,\infty)$, by

$$
\operatorname{tr}\Sigma \leq \sqrt{\operatorname{tr}\mathrm{I}_p^2\operatorname{tr}\Sigma^2}=\sqrt{p}\sqrt{\operatorname{tr}\Sigma^2},
$$

we have

$$
\frac{\operatorname{tr}\Sigma}{\sqrt{n}} \leq \sqrt{\frac{p}{n}}\sqrt{\operatorname{tr}\Sigma^2}=O(\sqrt{\operatorname{tr}\Sigma^2}).
$$

Hence

$$
\frac{\operatorname{tr}S_n^2}{\operatorname{tr}\Sigma^2}=O_p(1),
$$

and

$$
\frac{1}{n-2}\frac{\operatorname{tr}S_n^2}{(\operatorname{tr}\Sigma)^2}=O_p(1).
$$

With (A11),

$$
0\leq\sum_{i=1}^{p}(1-h_i)d_i^2\leq 3(n-2)\frac{\varepsilon_n}{np}\operatorname{tr}S_n^2=3(n-2)\frac{\varepsilon_n}{np}O_p(\operatorname{tr}\Sigma^2).
$$

Similarly, we have

$$0 \leq \frac{2}{n-2}\sum_{i=1}^{p} d_i \sum_{i=1}^{p}(1-h_i)d_i \leq \frac{\varepsilon_n}{np}(\mathrm{tr}S_n)^2,$$

therefore

$$\frac{2\sum_{i=1}^{p} d_i \sum_{i=1}^{p}(1-h_i)d_i}{(n-2)\mathrm{tr}\Sigma^2} \leq (n-2)\frac{\varepsilon_n}{np}O_p(1).$$

As $\varepsilon_n \to 0$,

$$\frac{\sum_{i=1}^{p}(1-h_i)d_i^2}{\mathrm{tr}\Sigma^2} \xrightarrow{P} 0,$$

and

$$\frac{2\sum_{i=1}^{p} d_i \sum_{i=1}^{p}(1-h_i)d_i}{(n-2)\mathrm{tr}\Sigma^2} \xrightarrow{P} 0.$$

Therefore, with (A11),

$$\begin{aligned}\mathrm{tr}(CS_n)^2 - \frac{1}{n-2}(\mathrm{tr}(CS_n))^2 &= (1+o_p(1))\left[\sum_{i=1}^{p} d_i^2 - \frac{1}{n-2}\left(\sum_{i=1}^{p} d_i\right)^2\right] \\ &= (1+o_p(1))\left[\mathrm{tr}S_n^2 - \frac{1}{n-2}(\mathrm{tr}S_n)^2\right].\end{aligned}$$

Then,

$$\frac{\mathrm{tr}(CS_n)^2 - \frac{1}{n-2}(\mathrm{tr}(CS_n))^2}{\mathrm{tr}S_n^2 - \frac{1}{n-2}(\mathrm{tr}S_n)^2} \xrightarrow{P} 1. \tag{A12}$$

The authors of [1] have proved that

$$\frac{\mathrm{tr}S_n^2 - \frac{1}{n-2}(\mathrm{tr}S_n)^2}{\mathrm{tr}\Sigma^2} \xrightarrow{P} 1.$$

Thus, by the above and (A4) and (A12), we have

$$\frac{\mathrm{tr}(CS_n)^2 - \frac{1}{n-2}(\mathrm{tr}(CS_n))^2}{\mathrm{tr}(C\Sigma)^2} \xrightarrow{P} 1,$$

which implies (22) holds. □

Proof of Corollary 1. The power of the proposed test is

$$\begin{aligned}\beta(\delta) =& \Pr(T_{\mathrm{PB}} \geq z_{1-\alpha}) \\ =& \Pr\left(\frac{1}{\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}}}\left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)}\right] \geq z_{1-\alpha}\right) \\ =& \mathrm{E}_{\Sigma}\left[\Pr\left(\frac{1}{\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}}}\left[2\ln(\mathrm{PB}) + p\ln 2 - \widehat{\mathrm{tr}(B\Sigma)} - \frac{n_1 n_2}{n}\delta^T B\delta\right] \geq z_{1-\alpha} - \frac{n_1 n_2}{n}\frac{\delta^T B\delta}{\sqrt{2\widehat{\mathrm{tr}(B\Sigma)^2}}}\middle| S_n\right)\right]\end{aligned}$$

By (20) and (22), we have

$$\beta(\delta) - \mathrm{E}_{\Sigma}\left[\Phi\left(\frac{\frac{n_1 n_2}{n}\delta^T B\delta}{\sqrt{\widehat{\mathrm{tr}(B\Sigma)^2}}} - z_{1-\alpha}\right)\right] \xrightarrow{P} 0.$$

□

Appendix B. R code

```
rm(list = ls(all = TRUE))
library(MASS)
library(Matrix)
#install.packages("lava")
library(lava)
n1=70
n2=70
n=n1+n2
p=1000
M=1000
m=2*p
mu1=rep(0,p)
#Sigma 1
Sigma1=diag(1,p)
#Sigma 2
#ro=0.4
#Sigma2_0=matrix(1,p,p)
#for (i in 1:p) {
#  for (j in 1:p) {
#    k<-abs(j-i)
#    Sigma2_0[i,j]=ro^{k}
#  }
#}
#Sigma2<-Sigma2_0
#Sigma 3
#Sigma3_1=diag(0.85,25)+matrix(0.15,25,25)
#list2 <- NULL
#for (i in 1:(p/25)){
#  list2[[i]] <- Sigma3_1
#}
#Sigma3<-as.matrix(bdiag(list2))
Sigma=Sigma1
delta=0.975
t1=proc.time()
p0=delta*p
mu20=mvrnorm(1,rep(1,p),diag(rep(1, p)))
mu20_xiabiao=sort(sample(1:p,p0))
for (i in 1:p0){mu20[mu20_xiabiao[i]]=0}
#alternative 1
scal=sqrt((t(mu20)%*%solve(Sigma)%*%(mu20))/2)
#alternative 2
#scal=sqrt(t(mu20)%*%(mu20)/sqrt(tr(t(Sigma)%*%Sigma))/0.1)
mu2=mu20/rep(scal,p)

c=0
T_BF=rep(0,M)
for (q in 1:M) {
xi<-mvrnorm(n1,mu1,Sigma)
yi<-mvrnorm(n2,mu2,Sigma)
x_mean<-rep(0,p)
for(i in 1:p){
x_mean[i]=mean(xi[,i])
}
```

```
y_mean<-rep(0,p)
for(i in 1:p){
y_mean[i]=mean(yi[,i])
}
z1=matrix(0,n1,p)
for (l in 1:n1) {
z1[l,]=x_mean
}
z2=matrix(0,n2,p)
for (l in 1:n2) {
z2[l,]=y_mean
}
A<-t(xi-z1)%*%(xi-z1)+t(yi-z2)%*%(yi-z2)
k<-1/log10(n)/(eigen(A)$values[1])/(p)
V<-k*diag(rep(1, p))
B=((m+2*(n))*solve(solve(V)+2*(A))-(m+n)/2*solve(solve(V)+A))
T=n1*n2/(n)*(t(x_mean-y_mean)%*%B%*%(x_mean-y_mean))

S_n<-A/(n-2)
mu_T=tr(B%*%S_n)
sigma_T<-tr((B%*%S_n)%*%(B%*%S_n))-1/(n-2)*(tr(B%*%S_n))^2

T_BF[q]=(T-mu_T)/sqrt(2*sigma_T)
if(T_BF[q]>=qnorm(0.95)){c=c+1}
}
t2=proc.time()
t=t2-t1
cat("power =", c/M,"time",t[3][[1]],"s","\n")
```

References

1. Bai, Z.; Saranadasa, H. Effect of high dimension: By an example of a two sample problem. *Stat. Sin.* **1996**, *6*, 311–329.
2. Hotelling, H. The Generalization of Student's Ratio. *Ann. Math. Stat.* **1931**, *2*, 360–378. [CrossRef]
3. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; Technical Report; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1958.
4. Srivastava, M.S.; Du, M. A test for the mean vector with fewer observations than the dimension. *J. Multivar. Anal.* **2008**, *99*, 386–402. [CrossRef]
5. Srivastava, M.S. A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivar. Anal.* **2009**, *100*, 518–532. [CrossRef]
6. Srivastava, M.S.; Katayama, S.; Kano, Y. A two sample test in high dimensional data. *J. Multivar. Anal.* **2013**, *114*, 349–358. [CrossRef]
7. Chen, S.X.; Qin, Y.L. A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Stat.* **2010**, *38*, 808–835. [CrossRef]
8. Feng, L.; Zou, C.; Wang, Z.; Zhu, L. Two-sample Behrens-Fisher problem for high-dimensional data. *Stat. Sin.* **2015**, *25*, 1297–1312. [CrossRef]
9. Cai, T.T.; Liu, W.; Xia, Y. Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2014**, *76*, 349–372.
10. Lopes, M.; Jacob, L.; Wainwright, M.J. A more powerful two-sample test in high dimensions using random projection. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 1206–1214.
11. Srivastava, R.; Li, P.; Ruppert, D. RAPTT: An exact two-sample test in high dimensions using random projections. *J. Comput. Graph. Stat.* **2016**, *25*, 954–970. [CrossRef]
12. Zoh, R.S.; Sarkar, A.; Carroll, R.J.; Mallick, B.K. A powerful Bayesian test for equality of means in high dimensions. *J. Am. Stat. Assoc.* **2018**, *113*, 1733–1741. [CrossRef] [PubMed]
13. Wang, R.; Xu, X. On two-sample mean tests under spiked covariances. *J. Multivar. Anal.* **2018**, *167*, 225–249. [CrossRef]
14. Kuelbs, J.; Vidyashankar, A.N. Asymptotic inference for high-dimensional data. *Ann. Stat.* **2010**, *38*, 836–869. [CrossRef]

15. Thulin, M. A high-dimensional two-sample test for the mean using random subspaces. *Comput. Stat. Data Anal.* **2014**, *74*, 26–38. [CrossRef]
16. Gregory, K.B.; Carroll, R.J.; Baladandayuthapani, V.; Lahiri, S.N. A two-sample test for equality of means in high dimension. *J. Am. Stat. Assoc.* **2015**, *110*, 837–849. [CrossRef] [PubMed]
17. Yu, W.; Xu, W.; Zhu, L. A combined p-value test for the mean difference of high-dimensional data. *Sci. China Math.* **2018**, *62*, 961. [CrossRef]
18. Chen, S.X.; Li, J.; Zhong, P.S. Two-sample and ANOVA tests for high dimensional means. *Ann. Stat.* **2019**, *47*, 1443–1474. [CrossRef]
19. Zhang, J.T.; Guo, J.; Zhou, B.; Cheng, M.Y. A simple two-sample test in high dimensions based on L^2-norm. *J. Am. Stat. Assoc.* **2020**, *115*, 1011–1027. [CrossRef]
20. Zhu, Y.; Bradic, J. Significance testing in non-sparse high-dimensional linear models. *Electron. J. Stat.* **2018**, *12*, 3312–3364. [CrossRef]
21. Aitkin, M. Posterior bayes factors. *J. R. Stat. Soc. Ser. B Methodol.* **1991**, *53*, 111–128. [CrossRef]
22. Wang, R.; Xu, X. Least favorable direction test for multivariate analysis of variance in high dimension. *Stat. Sin.* **2021**, *31*, 723–747. [CrossRef]