Two-Sample Test for High Dimensional Data

Du Liheng BSc(Statistics)

UID:3035532870

Supervisor: Prof.Jeff J.F.Yao

Department of Statistics and Actuarial Science
Faculty of Science

The University of Hong Kong

u3553287@connect.hku.hk

1. Introduction

High dimensional data appears in many statistical studies and has various modern-era applications. Usually in a multivariate analysis problem, a sample $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}$ of p-dimension random observations is provided. While p is fixed and small enough i.e. smaller than 10, classical dimensional reduction methods like Principal Components Analysis (PCA) are developed. However, these methods are no fitted when p is no longer fixed. For instance, it's hard to distinguish a few principal components out of over a hundred columns and calculate the transformation. We call a data set high dimensional when its sample size n is smaller than the dimension p. Specially, in this paper we focus on the sets where the ratio $y = p/n$ tends to a fix number between 0 and 1 when p and n both go to infinity.

The two sample significance test is a classical statistical inference. The hypothesis test starts with $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$ where μ_1, μ_2 are the mean vectors of two $N_i \times p$ sample matrices $\mathbf{x_i}$, $i = 1, 2$. In this paper, we first introduce the classical Hotelling's T^2 test for the two samples mean analysis problem. However, with high dimensional data, the inference outcome becomes distorted and the T^2 statistic of Hotelling's test will be undefined. This is referred to as **the Effect of High Dimension** (EHD). Two more tests proposed by Dempster (1958) and Bai and Saranadasa (1996) to deal with high dimensional data will be introduced later and compared with Hotelling's test based on their asymptotic power function. Then in the last chapter, a simulation is conducted to confirm the mathematical proof.

2. Hotelling's T^2 Test

An important methodology in multivariate analysis is the Hotelling's T^2 test. The two sample mean test is one classical application of it. Let's start from the one sample univariate case, where we have the t-statistic:

$$t = \frac{\sqrt{n}(\overline{x} - \mu)}{s} \tag{2.1}$$

to test whether the mean of a population equals to a given value μ. The $\overline{x}$ denotes the sample mean, n the sample size and s the sample standard deviation. Under the normal assumption where the population follows $N(0, \sigma^2)$, this statistic follows a Student-t distribution with $n - 1$ degrees of freedom. In the two samples univariate case, we now assume:

1. $x \sim N(0, \sigma_1^2), y \sim N(0, \sigma_2^2)$, x and y are independent.
2. Assume the population variance $\sigma_1^2 = \sigma_2^2$ equals to an unknown constant
3. The sample size N_1 and N_2 could be different

Based on these assumptions, it's easy to derive the new t-statistic which follows a Student-t distribution with $N_1 + N_2 - 2$ degrees of freedom:

$$t = \frac{\overline{x} - \overline{y}}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{2.2}$$

where $s_p = \sqrt{\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}}$ is an estimator of the pooled standard deviation of the two samples. When we take the square of the statistic and then divide both the denominator and the numerator by the population variance σ^2, we have:

$$T^2 = \frac{\frac{\overline{x}-\overline{y}}{\sigma^2}}{\frac{s_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\sigma^2}} \tag{2.3}$$

By Central Limited Theorem and $\frac{s^2}{\sigma^2} \sim \chi^2_{N_1+N_2-2}$, the distribution of T^2 under normality assumption is known precisely as following a F distribution. Hotelling (1931) proposed this T^2 statistic for the two-sample case and expand it to the multivariate population. Suppose $\mathbf{x}_1, \cdots, \mathbf{x}_{N_1}$ and $\mathbf{y}_1, \cdots, \mathbf{y}_{N_2}$ are two samples from $N(\mu_i, \mathbf{\Sigma}), i = 1, 2$. The hypothesis test is conducted based on the null hypothesis $\mu_1 = \mu_2$.Then we know the sample mean $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ are distributed according to $N(\mu_i, n_i^{-1}\mathbf{\Sigma}), i = 1, 2$. Therefore, we derive that $\overline{\mathbf{x}} - \overline{\mathbf{y}}$ is distributed following $N(\mathbf{0}, \tau\mathbf{\Sigma})$ under H_0 where τ is the pooling coefficient $1/N_1 + 1/N_2$. Finally, we can get the analogue of the statistic (2.3):

$$T^2 = \frac{N_1 N_2}{N_1 + N_2}(\overline{\mathbf{x}} - \overline{\mathbf{y}})' \mathbf{S^{-1}} (\overline{\mathbf{x}} - \overline{\mathbf{y}}) \tag{2.4}$$

where $\mathbf{S} = \frac{1}{N_1+N_2-2}[\sum_{j=1}^{N_1}(\mathbf{x_k} - \overline{\mathbf{x}})(\mathbf{x_k} - \overline{\mathbf{x}})' + \sum_{j=1}^{N_2}(\mathbf{y_k} - \overline{\mathbf{y}})(\mathbf{y_k} - \overline{\mathbf{y}})']$ is the sample variance matrix. This T^2 (2.4) is distributed as T^2 in (2.3) with $N_1 + N_2 - 2$ degrees of freedom. We have:

Proposition 2.1 *Let* $\mathbf{x_1}, \cdots, \mathbf{x_n}$ and $\mathbf{y_1}, \cdots, \mathbf{y_n}$ *be two samples from* $N_p(\mu_i, n_i^{-1}\mathbf{\Sigma}), i = 1, 2$. *Then the distribution of*

$$\frac{(N_1 + N_2 - p - 1)T^2}{p(N_1 + N_2 - 2)}$$

is a F distribution with p *and* $N_1 + N_2 - p - 1$. *Under the null hypothesis, this F-distribution is central while for the alternative hypothesis, the distribution is non-central with a* ***NCP****(Non-centrality parameter)* $\lambda = \frac{N_1 N_2}{N_1+N_2}\mu' \mathbf{\Sigma^{-1}} \mu$.

The T^2 statistic, if viewing it from the geometric perspective, is the square of the Mahalanobis distance between the two samples. It can be regarded as a dissimilarity measure between two random vectors $\mathbf{x}$ and $\mathbf{y}$. It has many robust properties e.g. the T^2 test is **UMPI**(uniformly most powerful invariant) among all tests that are invariant under the group of nonsingular linear transformations proved by Anderson (1962). This property can be easily proved by Neyman-

Pearson lemma since the T^2 test is a likelihood ratio test. Furthermore, for all tests of $H_0 : \mu = 0$ where $\mu = |\mu_1 - \mu_2|$, when their power depends only on $\mu'\mathbf{\Sigma^{-1}}\mu$, the T^2 test is UMP and admissible according to Simaika (1941).

However, the main restriction for applying Hotelling's T^2 test is that the population dimension p can not excess its corresponding statistic's degrees of freedom N_1+N_2-2. Otherwise, the sample covariance matrix is no longer non-singular and the statistic T^2 becomes undefined. One classic remedy for this problem is Dempster's *Non-exact Test*, proposed in his papers Dempster (1960). This methodology is also later found by Bai and Saranadasa (1996) to have higher power when the population dimension is high even when T^2 can be well defined. In next two parts we will present Dempster and Bai and Saranadasa's modifications of Hotelling's T^2 for high dimensional data.

3. Dempster's Non-exact Test

The T^2 statistic for the two-sample mean test has been introduced in (2.4). However, when sample size N_1 and N_2 are too small compared with the dimension of the sample p, the inverse S^{-1} does not exist and the Hotelling's test is not applicable any more. However, there are still approaches to test the difference between the sample means as the covariance structure of the p variables is not the basic statistic itself. The non-exact test built by Dempster (1958)is one of the classic methods to deal with data in high dimensional cases. He first restructured the origin data through matrix transformation and then approximate the statistic by Chi-square distribution to make it testable.

3.1 *Introduction of Dempster's Non Exact Test*

Let $\mathbf{x_1}, \cdots, \mathbf{x_{n_1}} \overset{iid}{\sim} \mathcal{N}(\mu_\mathbf{1}, \mathbf{\Sigma})$ and $\mathbf{y_1}, \cdots, \mathbf{y_{n_2}} \overset{iid}{\sim} \mathcal{N}(\mu_\mathbf{2}, \mathbf{\Sigma})$. One may first arrange all the data in a $p \times n$ matrix

$$\mathbf{Y} = (\mathbf{x_1}, \cdots, \mathbf{x_{n_1}}, \mathbf{y_1}, \cdots, \mathbf{y_{n_2}}),$$

where $n = n_1 + n_2$. Next, we can define an orthogonal $n \times n$ matrix $\mathbf{H}$ whose first two columns are $\mathbf{1_n}/\sqrt{n}$ and $(n_2\mathbf{1}'_{\mathbf{n_1}}, -n_1\mathbf{1}'_{\mathbf{n_2}})/\sqrt{n_1 n_2/n}$. The other columns of $\mathbf{H}$ are arbitrary orthonormal vectors. Applying this transformation to the data leads to

$$\mathbf{Z} = (\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_n}) = \mathbf{YH}. \tag{3.1}$$

We have, with $\mu = \mu_\mathbf{1} - \mu_\mathbf{2}$,

$$\mathbf{z_1} \sim \mathcal{N}(\frac{1}{\sqrt{n}}(n_1\mu_\mathbf{1} + n_2\mu_\mathbf{2}), \mathbf{\Sigma}), \quad \mathbf{z_2} \sim \mathcal{N}(\sqrt{\frac{n_1 n_2}{n}}\mu, \mathbf{\Sigma}), \tag{3.2}$$

$$\mathbf{z_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), i = 3, \cdots, n$$

However, these vectors are independent since $\mathbf{H}$ is orthogonal. Moreover, $\mathbf{z_1}$ is related to the ground mean from two samples and most information on the difference of the two population means is contained in $\mathbf{z_2}$. These vectors do not provide reliable information on $\mathbf{\Sigma}$. To estimate $\mathbf{\Sigma}$, one should rely on $\mathbf{z_i}$, i = 3, . . . , n. Define

$$\mathbf{A} = \sum_{i=3}^{n} \mathbf{z_i}\mathbf{z_i}'$$

The statistic is defined as

$$\mathbf{F} = \frac{\mathbf{Q_2}}{\mathbf{Q_3} + \cdots + \mathbf{Q_n}} \tag{3.3}$$

where $Q_i = ||\mathbf{z_i}||^2$. It's easy to see that under null hypothesis, Q_2 and $Q_i, i \geqslant 3$ are i.i.d distributed. For large p, this distribution can be well approximated by a multiple of chi-square: $Q_i \sim m\chi_r^2$ for some factor m and degree of freedom r.

Notice that Q_i is distributed according to a weighted sum of chi-squares,$\sum_{1\leqslant j\leqslant p} \lambda_j \chi^{\mathbf{2}}_{\mathbf{1},\mathbf{g}}$ where λ_j are eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{g}$ equals to $\sqrt{\frac{n_1 n_2}{n}}\mu$. Co-list these two equations we have:

$$\begin{aligned}\mathbf{E}(Q_2) &= \sum_{i=1}^{n} \lambda_i = tr(\Sigma) = mr \\ \mathbf{Var}(Q_2) &= 2\sum_{i=1}^{n} \lambda_i^2 = tr(\Sigma^2) = 2m^2 r\end{aligned} \tag{3.5}$$

Hence we easily solve that: $m = \frac{tr(\Sigma^2)}{tr(\Sigma)}$ and $r = \frac{tr(\Sigma)^2}{tr(\Sigma^2)}$. This result in

$$\mathbf{F} \sim \mathbf{F_{r,(N_1+N_2-2)r}}$$

where under the null hypothesis, $F_{r,(N_1+N_2-2)r}$ represents an F-type random variable on the degree of freedom r and $(N_1+N_2-2)r$. Since the statistic F is in the form of fraction, only the single parameter r is unknown and requires estimation. From the geometric perspective, r could be regarded as a reduced dimensionality from the ideal p dimensions which would hold when the variance of $\mathbf{Y}$ is an unit matrix. To estimate $\hat{r}$ of r, two methods are suggested by Dempster:

The first one is a statistic which is sufficient for m and r and only depends on r. It uses $Q_3, Q_4, \cdots, Q_n$:

$$t = (n-2)[\ln(\frac{1}{n-2}\sum_{3}^{n} Q_i)] - \sum_{3}^{n} \ln Q_i \tag{3.4}$$

and it can be shown that:

$$t \sim [\frac{1}{r} + \frac{1+\frac{1}{n-2}}{3r^2}]\chi^2_{n-3} \tag{3.5}$$

has very slight distortion of significance level so that solving the expectation equation:

$$t \sim [\frac{1}{r} + \frac{1+\frac{1}{n-2}}{3r^2}](n-3)$$

we can have a good approximation $\hat{r_1}$ even for small r. The other estimator could be more precise as it considers the projection and angles among vectors $\mathbf{Y_3}, \cdots, \mathbf{Y_n}$. We denote θ as the angle between two vectors , similar with the previous one, we have:

$$\begin{aligned} -\ln\sin\theta^2 &\sim (\frac{1}{r} + \frac{3}{2r^2})\chi^2_1 \\ u &\sim (\frac{1}{r} + \frac{3}{2r^2})\chi^2_{\binom{n-2}{2}} \end{aligned} \tag{3.6}$$

where negative (3.6) is actually the sum of the natural logs of the squared sines of all angles in (3.5). These $\binom{n-2}{2}$ angles are pairwise independent when p is large enough. Hence, a second estimator $\hat{r_2}$ is defined as the solution of the equation below:

$$t + u = [\frac{1}{\hat{r_2}} + \frac{1 + \frac{1}{n-2}}{3\hat{r_2^2}}](n-3) + [\frac{1}{\hat{r_2}} + \frac{3}{2\hat{r_2^2}}]\binom{n-2}{2}$$

So far a description of applying this method is completed. In the next few sectors, some complementary findings and a discussion about Dempster's and Hotelling's test will be shown.

3.2 *One way to find Dempster's* **H** *matrix using Gram–Schmidt process*

To demonstrate Dempster's non-exact test, one main complex problem is to find a group of appropriate matrix $\mathbf{H}$ to conduct the matrix transformation in (3.1). In this part, a related Gram-Schmidt process is introduced to solve this problem.

Based on the settings from the beginning of this chapter, we write the first 2 rows of a new orthogonal $n \times n$ matrix $\mathbf{A}$ as $\mathbf{1_n}/\sqrt{n}$ and $(n_2\mathbf{1}'_{\mathbf{n_1}}, -n_1\mathbf{1}'_{\mathbf{n_2}})/\sqrt{n_1 n_2/n}$. Hence, z_1 will be the weighted group mean and z_2 will be the pooled mean difference. Starting from the third row to the n-th row of $\mathbf{A}$, for the row $k, k = 3, \cdots, n$, we fill it with identity matrix's k-th row. Now we have:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n_1}}\tau & \frac{1}{\sqrt{n_1}}\tau & \frac{1}{\sqrt{n_1}}\tau & \cdots & -\frac{1}{\sqrt{n_2}}\tau & \cdots & -\frac{1}{\sqrt{n_2}}\tau \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

where $\tau = \frac{1}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$. Then we write the first 2 rows of the orthogonal $n \times n$ matrix $\mathbf{B}$ as $\mathbf{1_n}/\sqrt{n}$ and $(n_2\mathbf{1}'_{\mathbf{n_1}}, -n_1\mathbf{1}'_{\mathbf{n_2}})/\sqrt{n_1 n_2/n}$, the first two rows are the same as $\mathbf{A}$ but all other entries are 0. Denote the k-th row of $\mathbf{B}$ and j-th column of $\mathbf{A}$ as $\mathbf{b_k}$ and $\mathbf{a_j}, j, k = 1, 2, 3, \cdots, n$. The matrix $\mathbf{H}$ is the same as $\mathbf{B}$ at this stage. Then for $k \geqslant 3$, the k-th row $\mathbf{h_k}$ of $\mathbf{H}$ is solved by:

$$\begin{aligned} \mathbf{v_k} &= \mathbf{a'_j} - \sum_{k=1}^{j-1} \mathbf{b_k a_j b_k} \\ \mathbf{h_k} &= \frac{\mathbf{v_k}}{||\mathbf{v_k}||} \end{aligned} \tag{3.7}$$

Now we have a matrix which is originated from $\mathbf{Y}$ and satisfy the orthogonality of $\mathbf{H}$. The intermediate matrix $\mathbf{A}$ and $\mathbf{B}$ are intentionally chosen as above so that this method works well no matter if p is larger than n or not which means this process works well even under high dimensional situation.

The arbitrariness of the last (n-2) rows of matrix $\mathbf{H}$ is regarded as a practical disadvantage of Dempster's method.Although it won't directly affect the value of F statistic, it does have distortion effect for $\hat{r_1}$ and $\hat{r_2}$ which are important for the critical value. In practice, different matrix $\mathbf{H}$ will tend to produce some outliners as elaborated in Dempster's practice (1960).

3.3 *A Complementary Finding for Dempster's NCP*

Dempster's statistic follows a F-distribution with the numerator presenting the distance between two means. Therefore, under alternative hypothesis, the distribution is non-central. In Dempster (1958)'s original paper , no definition of the non-centrality parameter is given. Bai and Saranadasa (1996) also only provides an approximation through normal distribution to imitate the power function. In this part, we are to show that the non-centrality parameter for Dempster's test exists and shares exactly the same form as the Hotelling's non central parameter.

We start with defining a new constant $\delta = \Sigma^{-\frac{1}{2}}|\mu_1 - \mu_2|$ where μ_1 and μ_2 are two sample means in our hypothesis test and $\delta = o(1)$. For each k-th factor in vector μ_1 and μ_2, we can define $\delta_k = |\mu_{1k} - \mu_{2k}|$ where $k = 1, \cdots, p$ so that $\delta = (\delta_1, \delta_2, \cdots, \delta_p)'$, then since $Q_i, i = 2, \cdots, p$ can be regarded as the length of one of the $n-1$ vectors in $\mathbb{R}^p, p > n_1 + n_2 - 2$, we can use eigenvalue decomposition to project Q_2 into the p directions with the scalar $\lambda_i, i = 1, \cdots, p$ on the i-th direction.

Thanks for the orthogonal property of eigenvalue and eigenvectors, now we can write the Dempster's statistic in the form of:

$$F = (n-2)\frac{\sum_{i=1}^{p}\lambda_i(y_i^2 + 2\tau^{-\frac{1}{2}}\delta_i y_i + \tau^{-1}\delta_i^2)}{\sum_{j=1}^{n-2}\sum_{i=1}^{p} z_{ij}^2\lambda_i} \tag{3.8}$$

where $y_i, z_{ij}, i = 1, \cdots, p, j = 1, \cdots, n$ are following i.i.d. $\mathbb{N}(0,1)$ distribution variables and $\lambda_1, \lambda_2, \cdots, \lambda_p$ are eigenvalues of Σ. To find the non-centrality parameter, we only needs to focus on the numerator which can be re-written into:

$$\sum_{i=1}^{p}\lambda_i(y_i + \tau^{-\frac{1}{2}}\delta_i)^2$$

It's obvious to see under the alternative hypothesis, Q_2 the numerator is distributed according to a non-central weighted sum of chi-squares,$\sum_{1\leqslant j\leqslant p}\lambda_j\chi^{\mathbf{2}}_{\mathbf{1},\mathbf{g_j}}$ where $g_j = \tau^{-\frac{1}{2}}\delta_j$ is the corresponding non-central Chi-square parameter and 1 is the degree of freedom.

Hence the statistic F follows a non-central F distribution with non-central parameter $\Lambda = \sum_{1\leqslant j\leqslant p} g_j^2 = \frac{n_1 n_2}{n}\mu^{'}\mathbf{\Sigma^{-1}}\mu, \mu = \mu_1 - \mu_2$ which is exactly the same as the Hotelling's non centrality parameter.

3.4 *Comparison with Hotelling's T^2 Test*

It is worth notice that even when T^2 can be well-defined, i.e. $p \leqslant n_1 + n_2 - 2$, as long as the dimension is high enough, the power function of Hotelling's T^2 test grows much less than Dempster's. This out-performing result does not conflict with Hotelling's UMP test property since Dempster's test does not obey invariance. This problem is due to the non-affineness property since Dempster used the notions of length and angle in k-dimensional space. If k linear combinations of the given k variables are used as input, then vectors' lengths and angles will be different, so does the significance level. However, the Hotelling's test also only gives unique test when the transformations are restricted to linear ones. The reason that Dempster's test performs better when dimension is high enough is mainly due to the restriction of parameters in the power function when p goes larger, as shown in Bai and Saranadasa (1996). For this part, it will be discussed later in (4.2) together with Bai and Saranadasa's test.

Although Dempster provides a good solution to the high dimensional test, the construction process for it is still too complex. In the next chapter, Bai and Saranadasa's test will be introduced, which shares the same level power with Dempster's and can be applied without the normality assumption. Furthermore, mathematically speaking it's simpler compared with Dempster's since an Edgeworth expansion can be easily found by the methodology from Bai and Rao (1991).

4. Bai-Saranadasa's Test

As discussed in 3.4, Dempster's test is restricted to data under normality assumption which is the same on this point with Hotelling's T^2 test. Except from this, the arbitrary choice of the transformation matrix $\mathbf{H}$ causes distortion of the estimator $\hat{r}$. Targeting these two problems, Bai and Saranadasa (1996) came up with a new test. For high dimensional data, it is approximately equivalent to Dempster's test and mathematically simpler. Its power performance is also superior to Hotelling's T^2 test when it can be well defined.

4.1 *Introduction of Bai-Saranadasa's Test*

To introduce this test, again, we construct the two-sample mean hypothesis test with $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$. The test requires following assumptions:

1. $\mathbf{x_{ij}} = \Gamma \mathbf{z_{ij}} + \mu_j, i = 1, \cdots, N_j, j = 1, 2$, where Γ is a $p \times m$ matrix with $m \leqslant \infty$ and $\Gamma\Gamma' = \Sigma$. $\mathbf{z_{ij}}$ are i.i.d. random vector with m independent variables satisfying $\mathbb{E}(\mathbf{z_{ij}}) = \mathbf{0}$, $\mathbf{Var}(\mathbf{z_{ij}}) = I_m$;

2. For m non-negative constants $v_k, k = 1, \cdots, m$, whenever $v_1 + \cdots + v_m = 4$, if one of $v_k = 1$ then $\mathbb{E} \prod_{k=1}^{m} z_{ijk}^{v_k} = 0$ and if two of them equal 2, $\mathbb{E} \prod_{k=1}^{m} z_{ijk}^{v_k} = 1$;

3. $\frac{p}{N_1+N_2} \to y > 0$ and $\frac{N_1}{N_1+N_2} \to \kappa \in (0, 1)$;

4. $\mu'\Sigma\mu = o(\tau tr\Sigma^2)$ and $\lambda_{max} = o(\sqrt{tr\Sigma^2})$ are true where $\mu = \mu_1 - \mu_2, \tau = \frac{N_1+N_2}{N_1 N_2}$ and λ_{max} is the largest eigenvalue of Σ, the covariance matrix.

Assumption 1 is similar to the factor model in multivariate analysis. However, in the factor model the number of factors m should be limited to be smaller than p. For our test, although m is largely arbitrary, we tend to prefer $m \geqslant p$ so that Σ is regular and allow its fundamental properties. Still, the flexibility given by the free choice of m provides richer collection of dependence structure.

Assumption 2 is applied to help all components of $\mathbf{z_{ij}}$ reach a certain kind of pseudo-independence. However, if $\mathbf{z_{ij}}$ itself is already independent then this assumption is trivial.

Assumption 3 and 4 are aiming to restrict the order of p and n under the strong assumption that $\Sigma_1 = \Sigma_2$, i.e. the two samples' covariance matrix are same with each other. However, it is hard to verify especially for high dimensional data. Ledoit and Wolf (2002) have provided the generalization of Assumption 4 to unequal variance where they assume p and n are still of the same order:

$$\mu'\Sigma_i\mu = o[(N_1 + N_2)^{-1} tr(\Sigma_1 + \Sigma_2)^2], i = 1, 2 \tag{4.1}$$

All the variables in the 4 assumptions could depend on $n = N_1 + N_2 - 2$. Next we consider the statistic:

$$M_n = ||\overline{\mathbf{x}}_{\mathbf{1}} - \overline{\mathbf{x}}_{\mathbf{2}}||^2 - \tau tr(\mathbf{S_n}) \tag{4.2}$$

$$\mathbf{S_n} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{x_{ij}} - \overline{\mathbf{x}}_{\mathbf{i}})(\mathbf{x_{ij}} - \overline{\mathbf{x}}_{\mathbf{i}})', \qquad \overline{\mathbf{x}}_{\mathbf{i}} = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x_{ij}}, i = 1, 2.$$

As we know from the hypothesis, under null hypothesis $\mathbb{E}\mathbf{M}_n = 0$. Then following Assumption 1-4, under H_0, when $n \to \infty$, by adopting CLT and law of large number,

$$Z_n = \frac{M_n}{\sqrt{Var(M_n)}} \xrightarrow{D} \mathbb{N}(0, 1). \tag{4.3}$$

Then we denote σ_M^2 as the variance of M_n. Under null hypothesis and normality assumption by simple numeric calculation we have:

$$\sigma_M^2 = 2\tau^2(1 + \frac{1}{n})tr\Sigma^2 \tag{4.4}$$

It is also shown by Bai and Saranadasa that if normality doesn't hold, the difference between σ_M^2 and $Var(M_n)$ is a multiple of $(1 + o(1))$. Hence, (4.3) still holds asymptotically even if the population doesn't follow a normal distribution. Now to complete the test, we only need to find a ratio-consistent estimator (elaborated later in Theorem 4.4) of $tr(\Sigma^2)$ to replace the denominator of the statistic. The first one comes to mind could be $tr(S_n^2)$. However, as p is quite flexible, $tr(S_n^2)$ won't have to be unbiased or ratio-consistent even under normality assumption. The way to find a fitted estimator is through a Wishart distribution.

Lemma 4.1 *Given a p-variable population which follows multivariate normal distribution:* $\mathbf{X}\sim\mathbb{N}(\mu, \Sigma)$*, the sample deviation matrix* $\mathbf{S_n} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{x_{ij}} - \overline{\mathbf{x}}_{\mathbf{i}})(\mathbf{x_{ij}} - \overline{\mathbf{x}}_{\mathbf{i}})'$ *follows a Wishart distribution:*

$$\mathbf{S_n}\sim\frac{1}{n}\mathbb{W}_p(n, \Sigma)$$

Then, we may verify that:

$$B_n^2 = \frac{n^2}{(n+2)(n-1)}(tr(S_n^2) - \frac{1}{n}(trS_n)^2),$$

is an unbiased and ratio-consistent estimator of $tr(\Sigma^2)$. Please note that this B_n^2 has two nice properties:

1. By Cauchy-Schwarz inequality, $tr(S_n^2) - \frac{1}{n}(trS_n)^2 \leqslant 0$ always holds;

2. Under Assumption 1-4, B_n^2 remains to be a ratio-consistent estimator of $tr(\Sigma^2)$ without the normal assumption.

Finally, substituting B_n^2 into (4.3) we have the completed Bai-Saranadasa's test statistic:

$$\begin{aligned} Z &= \frac{||\mathbf{\bar{x}_1} - \mathbf{\bar{x}_2}||^2 - \tau tr(\mathbf{S_n})}{\tau\sqrt{\frac{2n(n+1)}{(n+2)(n-1)}(tr(S_n^2) - \frac{1}{n}(trS_n)^2)}} \\ &= \frac{\frac{N_1N_2}{N_1+N_2}(\mathbf{\bar{x}_1} - \mathbf{\bar{x}_2})'(\mathbf{\bar{x}_1} - \mathbf{\bar{x}_2}) - tr(S_n)}{\sqrt{\frac{2(n+1)}{n}}B_n} \to \mathbb{N}(0,1) \end{aligned} \tag{4.5}$$

According to (4.5), the test rejects H_0 if $Z > \xi_\alpha$.

4.2 *Power Analysis for All Three Tests and Marchenko–Pastur distribution*

In this sector, the asymptotic power function of all three tests presented above will be analysed and compared to show if one could dominate another under certain domain.

We start with the classic Hotelling's T^2 test. First we define $\delta = \Sigma^{-\frac{1}{2}}|\mu_1 - \mu_2|$ which is the same as in sector 3.3, where μ_1 and μ_2 are two sample means in our hypothesis test. Then, applying Assumption 3 from Sector 4.1, Bai and Saranadasa has shown that:

Theorem 4.1 *if* $\frac{p}{N_1+N_2} \to y > 0$, $\frac{N_1}{N_1+N_2} \to \kappa \in (0,1)$ *and* $||\delta||^2 = o(1)$, *then the power function of Hotelling's test*

$$\beta_H(\delta) - \Phi(-\xi_\alpha + \sqrt{\frac{n(1-y)}{2y}}\kappa(1-\kappa)||\delta||^2) \to 0$$

Proof: The proof for Hotelling's asymptotic power function is more complex compared with Dempster's and Bai-Saranadasa's but they share many common procedures. Hence, a few lemmas are introduced here first for the convenience of later proof.

Lemma 4.2 *Under the null hypothesis, when $n \to \infty$, by CLT we may easily build:*

$$\sqrt{\frac{(1-y)^3 n}{2y}}\left(\frac{T^2}{n} - \frac{y}{1-y}\right) \to \mathbb{N}(0,1)$$

so that

$$\frac{p}{n-p+1}F_\alpha(p, n-p+1) = \sqrt{\frac{2y}{(1-y)^3 n}}\xi_\alpha + \frac{y_n}{1-y_n} + o(\frac{1}{\sqrt{n}}). \qquad y_n = \frac{p}{n}$$

Then under H_1, we conduct a similar separation for the non-central distributions of $\frac{T^2}{n}$ like what have been done in (3.8). Decompose it into:

$$(\mathbf{w} + \tau^{-\frac{1}{2}}\delta)' U^{-1} (\mathbf{w} + \tau^{-\frac{1}{2}}\delta),$$

where $U = \sum_{i=1}^n \mathbf{u_i}\mathbf{u_i}'$, $\mathbf{u_i}$ and $\mathbf{w} = (w_1, \cdots, w_p)', i = 1, \cdots, n$ are i.i.d. random vectors following $\mathbb{N}(\mathbf{0}, I_p)$ so that $\mathbf{w}'\mathbf{w}$ follows central χ_p^2. It is worth noticing that $U \sim \mathbb{W}(n, I_p)$ so that we can do a further eigenvalue decomposition (EVD) $U = O' diag[d_1, \cdots, d_p] O$ where $d_1, \cdots, d_p$ are eigenvalues of U in the order from the largest to the smallest. Then using the notation above, we suggest

Lemma 4.3 *First we can re-write $\frac{T^2}{n}$ into*

$$\Omega_n = \frac{T^2}{n} = \sum_{i=1}^p (w_i^2 + 2w_i O\delta\tau^{-\frac{1}{2}} + \tau^{-1}(O\delta)^2)d_i,$$

so that applying Lemma 4.2 we have $\sqrt{n}(\sum_{i=1}^p d_i - \frac{y_n}{1-y_n}) \xrightarrow{p} 0$ and $n\sum_{i=1}^p d_i^2 \xrightarrow{p} \frac{y}{(1-y)^3}$

Now, starting from Ω_n, together with Lemma 4.2 and 4.3, Theorem 4.2 can be proved through

$$\beta_H(\delta) = P(\sum_{i=1}^p w_i^2 d_i + \frac{y_n ||\delta||^2}{\tau p(1-y_n)} \geqslant \sqrt{\frac{2y}{(1-y)^3 p}}\xi_\alpha + \frac{y_n}{1-y_n} + o(\frac{1}{\sqrt{n}}))$$

For the Dempster's asymptotic power function, similarly we have:

Theorem 4.2 *If except from Assumption 3, Assumption 4 $\mu'\Sigma\mu = o(\tau tr\Sigma^2)$ and $\lambda_{max} = o(\sqrt{tr\Sigma^2})$ is satisfied as well, we have*

$$\beta_H(\delta) - \Phi(-\xi_\alpha + \sqrt{\frac{n\kappa(1-\kappa)||\mu||^2}{\sqrt{2tr(\Sigma^2)}}}) \to 0$$

where $\mu = \mu_1 - \mu_2$.

Proof: To prove this, we use (3.8) to write the power function as:

$$\beta_D(\delta) = P(n\frac{\sum_{i=1}^p \lambda_i(y_i^2 + 2\tau^{-\frac{1}{2}}\delta_i y_i + \tau^{-1}\delta_i^2)}{\sum_{j=1}^n \sum_{i=1}^p z_{ij}^2\lambda_i} > F_\alpha(r, nr)) \tag{4.7}$$

where $y_i, z_{ij}, i = 1, \cdots, p, j = 1, \cdots, n$ are following i.i.d. $\mathbb{N}(0,1)$ distribution variables and $\lambda_1, \lambda_2, \cdots, \lambda_p$ are eigenvalues of Σ. Since $||\delta||^2 = o(1)$, combining with Assumption 4, CLT and the laws of large numbers, we can show from (4.7):

$$\frac{\sum_{i=1}^p \lambda_i(y_i^2 + 2\tau^{-\frac{1}{2}}\delta_i y_i - 1)}{\sqrt{2tr(\Sigma^2)}} \approx \frac{\sum_{i=1}^p \lambda_i(y_i^2 + 2\tau^{-\frac{1}{2}}\delta_i y_i - 1)}{\sqrt{2tr(\Sigma^2) + 4\tau^{-1}\mu'\Sigma\mu}} \xrightarrow{D} \mathbb{N}(0,1). \tag{4.8}$$

and the denominator:

$$\sum_{j=1}^n \sum_{i=1}^p z_{ij}^2\lambda_i = n(tr\Sigma)(1 + \sqrt{\frac{2}{nr}}\mathbb{N}(0,1) + o_p(\sqrt{\frac{1}{nr}})). \tag{4.9}$$

From (3.5) we know that $r = \frac{tr(\Sigma)^2}{tr(\Sigma^2)}$ and by eigenvalues property we have $\sum_{i=1}^p \delta_i^2\lambda_i = ||\mu||^2$. Hence Theorem 4.3 follows immediately with:

Lemma 4.4 *When $n, r \to \infty$, similar with Lemma 4.2, we have:*

$$F_\alpha(r, nr) = 1 + \sqrt{\frac{2}{r}}\xi_\alpha + o(\frac{1}{\sqrt{r}})$$

When it comes to Bai-Saranadasa's test, we are surprising to find that its power function is asymptotically the same with Dempster's test.

Theorem 4.3 *When Assumption 1-4 are all satisfied, we have:*

$$\beta_H(\delta) - \Phi(-\xi_\alpha + \sqrt{\frac{n\kappa(1-\kappa)||\mu||^2}{\sqrt{2tr(\Sigma^2)}}}) \to 0$$

As Bai-Saranadasa's statistic Z follows $\mathbb{N}(0,1)$, it is much easier to prove Theorem 4.4.

Proof: Let $\mathbf{\bar{z}_i}, i = 1, 2$ be the sample means of $\mathbf{z_{ij}}, j = 1, \cdots, n_i$ in Assumption 1. Define

$$M_n^0 = (\mathbf{\bar{z}_1} - \mathbf{\bar{z}_2})'\Gamma'\Gamma(\mathbf{\bar{z}_1} - \mathbf{\bar{z}_2}) - \tau tr(\mathbf{S_n}) \tag{4.10}$$

Since under null hypothesis M_n^0 asymptotically equivalent with M_n, then $Var(M_n^0)^{-\frac{1}{2}} M_n^0 \xrightarrow{D} \mathbb{N}(0,1)$ and

$$Var(M_n^0) = Var(M_n) = (1+o(1))\sigma_M^2 = 2(1+o(1))\tau^2(1+\frac{1}{n})tr(\Sigma^2) \tag{4.11}$$

Under the alternative hypothesis, $M_n = M_n^0 - \mu'\Gamma(\overline{\mathbf{z}}_\mathbf{1} - \overline{\mathbf{z}}_\mathbf{2}) + ||\mu||^2$, by Assumption 4, we have

$$Var(\mu'\Gamma(\overline{\mathbf{z}}_\mathbf{1} - \overline{\mathbf{z}}_\mathbf{2})) = \tau\mu'\Sigma\mu = o(\tau^2\Sigma^2). \tag{4.12}$$

Based on this we can derive three important convergence:

$$\begin{gathered} \frac{Var(M_n^0)}{Var(M_n)} \to 1 \\ Var(M_n^0)^{-\frac{1}{2}} M_n - ||\mu||^2 \xrightarrow{D} \mathbb{N}(0,1) \\ \frac{2(n+1)}{n}\frac{B_n^2}{Var(M_n^0)} \to 1 \end{gathered} \tag{4.13}$$

Eventually, combining these three we know that

$$Z - \frac{n\kappa(1-\kappa)||\mu||^2}{\sqrt{2tr(\Sigma^2)}} \to \mathbb{N}(0,1). \tag{4.14}$$

which complete the proof.

Now, comparing $\beta_D(\delta)$ and $\beta_B S(\delta)$ with $\beta_H(\delta)$, it is easy to see that the main difference is an extra factor $\sqrt{1-y} \in (0,1)$ which restricts the increase speed of $\beta_H(\delta)$. More intrinsically speaking, the real reason that Hotelling's test performs much worse when dimension is high (still can be defined) is due to the skewing of the sample covariance matrix S_n from Σ. By the law of large number, when the population's covariance matrix is I_p, the error for S_n is of order $o_p(\frac{1}{\sqrt{n}})$ for a fixed p. However, when dimension is high and p is only restricted to the ratio $y = \frac{p}{n}$, the ratio of the largest and the smallest eigenvalue of S_n tends to $\frac{1+\sqrt{y}}{1-\sqrt{y}}$ for $y \in (0,1)$ so that if y is close to 1 then the gap between the maximum and the minimum eigenvalue could be extremely high which makes the error serious. This result is further elaborated in Yao et al. (2015) where the Marchenko–Pastur distribution (M-P Law) is proved and well constructed to show the distribution of the eigenvalues of S_n:

The Marcenko-Pastur distribution F_{y,σ^2} with index y and scale parameter σ has the density function

$$F_{y,\sigma^2}(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2}\sqrt{(b-x)(x-a)}, & a \leqslant x \leqslant b, \\ 0, & otherwise, \end{cases} \tag{4.15}$$

where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$ with an additional point mass at the origin if $y > 1$ of the value $1 - \frac{1}{y}$. If $\sigma^2 = 1$, then it is the standard M-P Law which is the case for the three tests in this paper.

4.3 *Improvements on Bai and Saranadasa's Test*

As discussed in 4.2, Bai and Saranadasa's test has the same level power with Dempster's without the normality assumption. However, it requires a finite 4-th moment of the observation and the ratio of dimension to sample size must be asymptotically a constant $\frac{p}{N_1+N_2} \to y > 0$(4.1 Assumption 2 and 3). Chen and Qin (2010) provides a main improvement for this test which not only can be applied in the situation when both p and n tend to infinity i.e. $\frac{p}{N_1+N_2} \to \infty$, but also the situation where the sample covariance matrices $\Sigma_1 \neq \Sigma_2$.

To make the test applicable when both p and n tend to infinity, Chen and Qin introduced an important theorm about the ratio consistency:

Theorem 4.4 *when Assumption 1,2,3 and (4.1) are satisfied, then if $n, p \to \infty$, we have:*

$$\frac{tr(\hat{\Sigma_i^2})}{tr(\Sigma_i^2)} \xrightarrow{p} 1, \qquad \frac{tr(\hat{\Sigma_1\Sigma_2})}{tr(\Sigma_1\Sigma_2)} \xrightarrow{p} 1, i = 1, 2$$

This ratio consistent estimator is helpful as it makes sure that the estimator will be well behaved in the limitation of $n, p \to \infty$. Normal consistent estimator can only satisfy its well behaved property when $n \to \infty$. Based on this theorem, they construct improved test statistic:

$$Q_n = \frac{T_n}{\hat{\sigma}_{n_1}} \xrightarrow{D} \mathbb{N}(0,1) \tag{4.16}$$

where

$$\begin{aligned}
T_n &= \frac{\sum_{i\neq j}^{n_1} \mathbf{x}_{1i}'\mathbf{x}_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} \mathbf{x}_{2i}'\mathbf{x}_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \mathbf{x}_{1i}'\mathbf{x}_{2j}}{n_1 n_2} \\
\hat{\sigma}_{n_1} &= \frac{2}{n_1(n_1-1)}\widehat{tr(\Sigma_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{tr(\Sigma_2^2)} + \frac{4}{n_1 n_2}\widehat{tr(\Sigma_1\Sigma_2)} \\
\widehat{tr(\Sigma_i^2)} &= \frac{1}{n_i(n_i-1)} tr \sum_{j\neq k}^{n_i} (\mathbf{x}_{ij} - \mathbf{x}_{i(jk)})\mathbf{x}_{ij}'(\mathbf{x}_{ik} - \mathbf{x}_{i(jk)})\mathbf{x}_{ik}' \\
\widehat{tr(\Sigma_1\Sigma_2)} &= \frac{1}{n_1 n_2} tr \sum_{l=1}^{n_1}\sum_{k=1}^{n_2} (\mathbf{x}_{1l} - \mathbf{x}_{1(l)})\mathbf{x}_{1l}'(\mathbf{x}_{2k} - \mathbf{x}_{2(k)})\mathbf{x}_{2k}'
\end{aligned}$$

$\mathbf{x}_{i(jk)}$ is the mean from sample i with observations $\mathbf{x}_{ij}$ and $\mathbf{x}_{ik}$ excluded. $\mathbf{x}_{i(l)}$ is the mean from sample i with observation $\mathbf{x}_{il}$ exclude.

5. Simulations

We employ our simulation study to compare the power of the three tests. Both normal and non-normal, low dimensional and high dimensional cases are cross-considered in order to verify the difference of statistical power and the previous asymptotics.

5.1 *Simulation Strategy*

We generate our observations through three settings, which corresponds to the classical low dimension case, a case when the number of row and column are close generated by Bai and Saranadasa (1996) and a high dimension case. Please notice that all tests were made under for size $\alpha = 0.05$ with 1000 repetitions. With limited repetitions, an error of $1000^{-1/2}$ for each non central parameter is expected by Central Limited Theorem. This can be used to explain the different of data between our second test and the original simulation.

For the first setting, let $N_1 = 25$, $N_2 = 20$ and $p = 4$. The co-variance matrices under multivariate normal distribution were $\Sigma = (1-\rho)I_p + \rho J_p$, where $\rho = 0$ and 0.5, I_p is the standard identity matrix and J_p is a $p \times p$ matrix with all entries one.

For the second setting, let $N_1 = 25$, $N_2 = 20$ and $p = 40$ so that Hotelling's T^2 test should be outperformed by Dempster's and Bai's according to our previous discussion. For the normal case, $\Sigma = (1-\rho)I_p + \rho J_p$ is not changed. For the non-normal case, we use a moving average model based on Gamma distribution: Let $U_{ijk} \sim \Gamma(4,1)$ be a set of independent variables. We define X_{ijk} accordingly:

$$X_{ijk} = U_{ijk} + \rho U_{i,j+1,k} + \mu_{j,k}, (j = 1, \cdots, p; i = 1, \cdots, N_k; k = 1, 2)$$

where ρ are constants chosen to be $\rho = 0.3, 0.6$ and 0.9. μ are the two mean vectors for the observation U_{ijk}.

For the third setting, let $N_1 = 25$, $N_2 = 20$ and p continuously increases from 20 to 200. Since in some cases $p > N_1 + N_2 - 2$, the Hotelling's T^2 test is no longer applicable. We only compare the power function between Dempster's and Bai's to check their performance under null hypothesis. The observation is generated through the same moving average model in the second set.

5.2 *Results*

Notation: H: Hotelling's T^2 test (Dotted line), D: Dempster's non exact test (Dashed line), BS: Bai and Saranadasa's normal test (Solid line). $\eta = \frac{||\mu_1 - \mu_2||^2}{\sqrt{tr\Sigma^2}}$ is the standard parameter for the difference of two sample means which is 0 under null hypothesis.

The first setting is a classical low dimensional case with $p = 4$. Under this case, it's easy to observe that the power of Hotelling's T^2 test hasn't been much lower than the other two tests, though still over-performed by them. Meanwhile, there still exists a certain amount of gap between Dempster's non exact test and Bai and Saranadasa's test when both n and p are not large enough to show the asymptotic convergence of their power function.

Fig. 1. Simulated power functions under the low dimensional case.

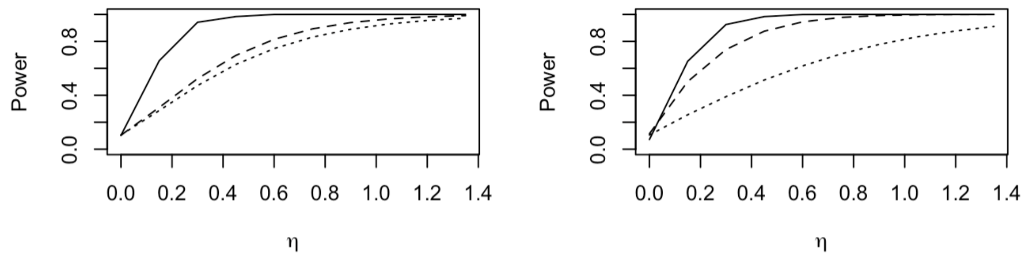

Table 5.2.1 Simulated power figures under the low dimensional case.

	$N = 45, p = 4, \alpha = 0.05$					
	$\rho = 0$			$\rho = 0.5$		
$\boldsymbol{\eta}$	**H**	**D**	**BS**	**H**	**D**	**BS**
0.15	0.2870	0.3102	0.6566	0.2554	0.5019	0.6525
0.30	0.4720	0.5234	0.9416	0.3900	0.7396	0.9251
0.45	0.6289	0.6959	0.9839	0.5133	0.8755	0.9836
0.60	0.7466	0.8151	0.9996	0.6177	0.9447	0.9991
0.75	0.8315	0.8927	1.0000	0.7066	0.9775	1.0000
0.90	0.8899	0.9394	1.0000	0.7771	0.9914	1.0000
1.05	0.9291	0.9667	1.0000	0.8338	0.9969	1.0000
1.20	0.9552	0.9823	1.0000	0.8773	0.9990	1.0000
1.35	0.9718	0.9907	1.0000	0.9104	0.9997	1.0000

When it comes to the second setting where the dimension $p = 40$ is much closer to the sample size $N = 45$, the power curve of Hotelling's increases much slower than the other two tests. For both normal and non-normal cases, as dimension goes higher, Bai and Saranadasa's test has almost the same significance level with Dempster's, which proves the theoretical asymptotic property.

Last, for the set three, as shown in Fig.4, both tests stay around 0.05 which is the set size $\alpha = 0.05$. However, it's worth noticing that when p is not high enough, Dempster's test has higher chance of having type I error. Furthermore, Bai and Saranadasa's test power is always slightly higher than Dempster's as shown in Table 5.2.3 and Fig 2, 3. This difference can be explained since Dempster's estimation relies on higher dimension to provide accuracy of its estimators of degree of freedom. This part is elaborated in 3.4.

Fig. 2. Simulated power functions with multivariate normal distribution.

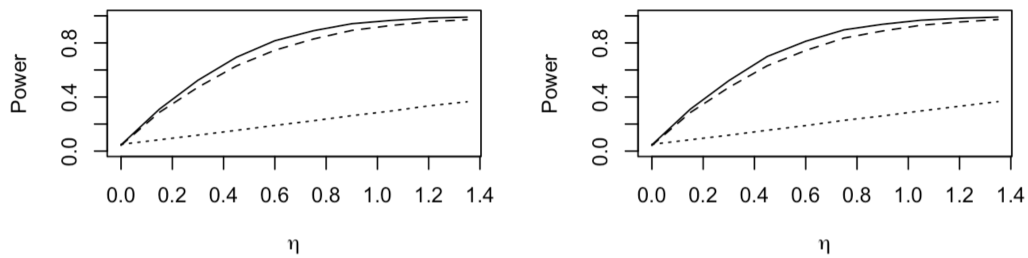

Table 5.2.2 Simulated power figures with multivariate normal distribution.

	$N = 45, p = 40, \alpha = 0.05$					
	$\rho = 0$			$\rho = 0.5$		
$\boldsymbol{\eta}$	**H**	**D**	**BS**	**H**	**D**	**BS**
0.15	0.0839	0.2870	0.3108	0.0840	0.2851	0.3102
0.30	0.1181	0.4719	0.5238	0.1176	0.4705	0.5206
0.45	0.1535	0.6297	0.6939	0.1541	0.6308	0.6995
0.60	0.1898	0.7456	0.8160	0.1885	0.7441	0.8120
0.75	0.2244	0.8285	0.8898	0.2276	0.8358	0.8966
0.90	0.2621	0.8922	0.9418	0.2601	0.8882	0.9377
1.05	0.2955	0.9278	0.9656	0.2971	0.9304	0.9677
1.20	0.3344	0.9567	0.9832	0.3316	0.9545	0.9819
1.35	0.3647	0.9711	0.9903	0.3659	0.9724	0.9909

Fig. 3. Simulated power functions with multivariate Gamma distribution.

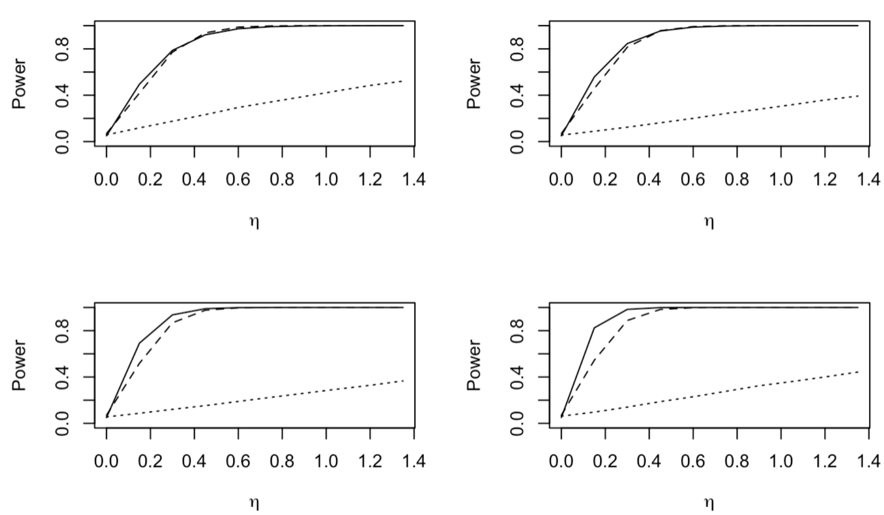

Table 5.2.3 Simulated power figures with multivariate Gamma distribution.

	$N=45, p=40, \alpha=0.05$						$N=45, p=40, \alpha=0.05$					
	$\rho=0$			$\rho=0.3$			$\rho=0.6$			$\rho=0.9$		
η	H	D	BS	H	D	BS	H	D	BS	H	D	BS
0.15	0.1181	0.4185	0.4953	0.0889	0.4577	0.5577	0.0869	0.5183	0.6922	0.0970	0.5446	0.8243
0.30	0.1753	0.7717	0.7867	0.1240	0.8140	0.8454	0.1208	0.8679	0.9360	0.1400	0.8874	0.9834
0.45	0.2326	0.9385	0.9213	0.1621	0.9564	0.9541	0.1530	0.9767	0.9894	0.1874	0.9827	0.9989
0.60	0.2941	0.9875	0.9736	0.2002	0.9925	0.9878	0.1891	0.9971	0.9985	0.2294	0.9980	0.9999
0.75	0.3421	0.9980	0.9917	0.2427	0.9990	0.9970	0.2249	0.9997	0.9998	0.2767	0.9998	1.0000
0.90	0.3890	0.9997	0.9975	0.2790	0.9999	0.9993	0.2597	1.0000	1.0000	0.3244	1.0000	1.0000
1.05	0.4372	1.0000	0.9993	0.3181	1.0000	0.9998	0.2951	1.0000	1.0000	0.3609	1.0000	1.0000
1.20	0.4843	1.0000	0.9998	0.3585	1.0000	1.0000	0.3281	1.0000	1.0000	0.4002	1.0000	1.0000
1.35	0.5212	1.0000	0.9999	0.3925	1.0000	1.0000	0.3662	1.0000	1.0000	0.4429	1.0000	1.0000

Fig. 4. Simulated power functions regarding dimension under null hypothesis.

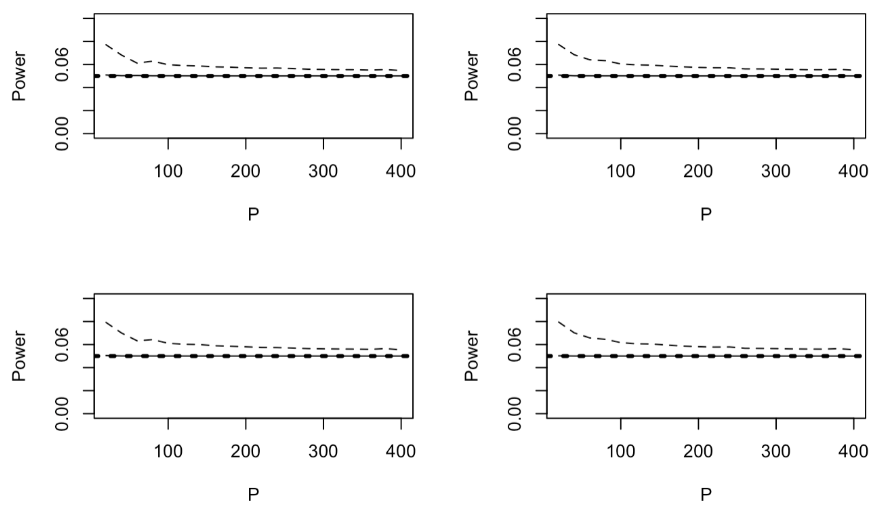

Table 5.2.4 Simulated power figures regarding dimension under null hypothesis

$N=45, \eta=0, \alpha=0.05$

	$\rho=0$		$\rho=0.3$		$\rho=0.6$		$\rho=0.9$	
P	D	BS	D	BS	D	BS	D	BS
20	0.0771	0.0509	0.0773	0.0507	0.0792	0.0504	0.0795	0.0502
40	0.0681	0.0502	0.0684	0.0502	0.0700	0.0501	0.0702	0.0501
60	0.0612	0.0504	0.0641	0.0501	0.0633	0.0501	0.0658	0.0500
80	0.0628	0.0501	0.0633	0.0501	0.0643	0.0501	0.0645	0.0500
100	0.0598	0.0502	0.0604	0.0501	0.0611	0.0501	0.0617	0.0500
120	0.0590	0.0501	0.0596	0.0501	0.0603	0.0500	0.0607	0.0500
140	0.0587	0.0501	0.0594	0.0501	0.0600	0.0500	0.0605	0.0500
160	0.0579	0.0501	0.0586	0.0501	0.0589	0.0500	0.0595	0.0500
180	0.0576	0.0501	0.0579	0.0500	0.0585	0.0500	0.0587	0.0500
200	0.0571	0.0500	0.0574	0.0500	0.0580	0.0500	0.0582	0.0500

References

Anderson, T. W. (1962). An introduction to multivariate statistical analysis. Technical report, Wiley New York.

Bai, Z. and Rao, C. R. (1991). Edgeworth expansion of a function of sample means. *The Annals of Statistics*, pages 1295–1315.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, pages 311–329.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, pages 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16(1):41–50.

Hotelling, H. (1931). The economics of exhaustible resources. *Journal of political Economy*, 39(2):137–175.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Simaika, J. (1941). On an optimum property of two important statistical tests. *Biometrika*, 32(1):70–80.

Yao, J., Zheng, S., and Bai, Z. (2015). *Sample covariance matrices and high-dimensional data analysis*. Cambridge University Press Cambridge.